Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

Transformer-Based Multi-Class Intrusion Detection Using CICIoMT2024 Dataset for Secure IoMT Networks

Eko Arip Winanto*[1,4], Sharipuddin[2], Benni Purnama[3], Nurhadi[5], Lasmedi Afuan[6]

[1]Computer Engineering, Dinamika Bangsa University, Indonesia
[2]Informatics, Dinamika Bangsa University, Indonesia
[3]Information System, Dinamika Bangsa University, Indonesia
[4]Computing, University of Technology Malaysia, Malaysia
[5]Business Digital, Dinamika Bangsa University, Indonesia
[6]Informatics, Universitas Jenderal Soedirman, Indonesia

Email: [1]ekoaripwinanto@gmail.com

Received : Dec 3, 2025; Revised : Jan 5, 2026; Accepted : Jan 19, 2025; Published : Jun 15, 2026

Abstract

Internet of Medical Things (IoMT) ecosystems significantly enhance healthcare services but simultaneously expand the attack surface, exposing medical networks to diverse cyber threats such as distributed denial-of-service and spoofing attacks. Existing intrusion detection systems for IoMT are often limited to binary classification and struggle to capture complex multi-class attack behaviors, particularly under highly imbalanced data distributions. This study proposes a deep Transformer-based intrusion detection model as a reproducible baseline for multi-class intrusion detection in IoMT environments. The model is evaluated on the CICIoMT2024 dataset, which comprises 19 traffic classes including benign and multiple attack categories. Data preprocessing involves stratified data splitting, feature normalization, and label encoding to ensure fair evaluation. The proposed baseline employs a six-layer Transformer encoder with eight attention heads and is trained using the AdamW optimizer. Experimental results demonstrate an overall accuracy of 98.76% and a macro F1-score of 0.92, indicating strong detection capability across most attack classes. The model achieves excellent performance on benign traffic and high-volume attacks such as DDoS and DoS, while performance degradation is observed on minority classes, including ARP spoofing, highlighting the impact of class imbalance. These findings establish the proposed Transformer model as a transparent and robust baseline for IoMT intrusion detection research. By providing reproducible performance benchmarks, this work supports future development of hybrid and imbalance-aware detection mechanisms aimed at enhancing real-time security in medical cyber-physical systems.

Keywords : *CICIoMT Dataset, Deep Transformer, IoMT Security, Intrusion Detection, Multi-Class Classification.*

1. INTRODUCTION

The Internet of Medical Things (IoMT) refers to the integration of various medical devices, sensors, wearables, hospital information systems, and cloud computing services into an interconnected ecosystem that supports digital healthcare services [1], [2]. Through this ecosystem, patient vital signs can be collected continuously, medical data can be exchanged rapidly, and healthcare professionals can monitor patients remotely. The resulting improvements in efficiency and service quality are highly significant; however, they simultaneously expand the attack surface of healthcare networks [3]. Attackers can exploit vulnerabilities in communication protocols, operating systems, or application software to launch attacks such as DDoS, DoS, spoofing, reconnaissance, and other malicious activities [4], [5]. The success of such attacks can disrupt critical hospital services, compromise the integrity of medical data, and potentially endanger patient safety.

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

To mitigate these risks, security mechanisms in IoMT environments must be adapted to the specific characteristics of medical devices and medical communication networks. Traditional signature-based intrusion detection systems (IDS) are known to be relatively ineffective in addressing zero-day attacks and rapidly evolving threat landscapes [6], [7], [8]. Consequently, anomaly-based IDS leveraging machine learning and deep learning techniques have gained increasing attention [9], [10]. Previous studies have investigated classical learning algorithms such as decision trees, random forests, support vector machines, and ensemble methods for network traffic classification [11], [12], [13], [14]. More recent research has employed deep learning architectures, including fully connected neural networks, convolutional neural networks (CNN) [15], [16], recurrent neural networks (RNN) [17], and hybrid models, to learn discriminative representations directly from raw traffic data or preprocessed features [18]. Despite these advances, several challenges remain unresolved, particularly the reliable detection of minority attack classes in highly imbalanced IoMT datasets [19], [20].

The emergence of the Transformer architecture has further enriched research on network intrusion detection [21]. Originally introduced for natural language processing, Transformers rely on self-attention mechanisms to model long-range dependencies within sequential data without recurrent operations [22], [23]. This capability makes Transformers particularly suitable for modeling complex temporal and contextual relationships in network traffic. Motivated by these advantages, several studies have applied Transformer-based models to intrusion detection in IoT and industrial network environments, reporting promising detection performance [24], [25]. Nevertheless, in the context of IoMT, existing studies often focus on simplified experimental settings, hybrid architectures with extensive feature engineering, or a limited number of attack classes, which reduces transparency and reproducibility when evaluating model performance on realistic medical traffic.

This study aims to address these limitations by examining the effectiveness of a deep Transformer model as a baseline for multi-class intrusion detection in IoMT environments. Specifically, this research evaluates whether a standard Transformer architecture, without hybrid extensions or complex feature engineering, can achieve competitive performance when trained on an imbalanced IoMT dataset comprising 19 traffic classes. Two main contributions are provided in this work. First, a reproducible Transformer-based baseline is established for multi-class intrusion detection using the CICIoMT2024 dataset, serving as a reference for future research and model enhancements. Second, a detailed performance analysis is conducted across both majority and minority attack classes, highlighting the impact of class imbalance and identifying limitations that motivate the development of imbalance-aware and robustness-oriented detection strategies.

The structure of this article is organized as follows. Section 2 describes the research methodology, including the dataset, preprocessing steps, model architecture, and evaluation metrics. Section 3 presents the experimental results with a focus on overall metrics and per-class performance. Section 4 provides a discussion that interprets the findings and elaborates on limitations and future research directions. Section 5 presents the conclusions and recommendations for future research.

2. METHOD

This section describes the methods used in the research. The general approach includes dataset preparation, design of a Transformer model adapted to IoMT traffic, training process with specific optimization configurations, and model evaluation using standard classification metrics. All methodological decisions were made to produce a transparent and replicable baseline for multiclass intrusion detection in IoMT.

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

2.1. Experiment setup

This study develops a Transformer-based intrusion detection model for IoMT traffic using the CICIoMT2024 dataset. The experimental pipeline consists of the following stages:

1. Dataset preparation: selecting and organizing IoMT traffic from the CICIoMT2024 dataset, including choosing relevant features and defining the target attack/benign classes.
2. Preprocessing: cleaning missing or noisy data, encoding categorical labels, and normalizing numerical features so they are ready for model training.
3. Transformer-based detection model: designing and training a deep Transformer architecture that learns to classify each network flow into one of the IoMT traffic classes.
4. Performance evaluation: testing the trained model on unseen data and measuring its effectiveness using metrics such as accuracy, precision, recall, and F1-score.

2.2. Dataset

The experiments were conducted using the CICIoMT2024 dataset [26]. In general, the CICIoMT2024 dataset was constructed from an IoMT testbed consisting of a combination of physical medical devices, simulated devices, and supporting network infrastructure. Traffic was collected during normal system operation and when the environment was subjected to various attack scenarios. The attacks include volumetric attacks such as DDoS and DoS, scanning and probing activities, spoofing, and other forms of attacks. This dataset is publicly available and has been recognized as a relevant benchmark for IDS research in IoMT. A total of 19 classes were used, consisting of one benign class and eighteen attack classes. The attack classes include various variants of TCP/IP-based DDoS and DoS, ARP poisoning attacks, reconnaissance attacks, and other specific attack types.

The distribution of samples across each class is imbalanced. Benign traffic and several DDoS variants dominate the total number of samples, while some attack types have only a few examples. This imbalance reflects real-world conditions, where normal traffic and certain attacks occur more frequently, but simultaneously poses challenges for learning algorithms and affects the interpretation of evaluation metrics. In this baseline study, no explicit data-level or algorithm-level imbalance handling techniques, such as oversampling, undersampling, or synthetic data generation (e.g., SMOTE), were applied. The objective is to evaluate the intrinsic capability of a standard Transformer architecture under realistic and naturally imbalanced IoMT traffic conditions. Consequently, the observed performance on minority classes reflects the inherent challenges posed by class imbalance and serves as a motivation for subsequent research on imbalance-aware learning strategies.

2.3. Preprocessing Data

Data preprocessing was conducted through several stages. First, data cleaning was performed. Records containing missing values on important attributes or corrupted entries were removed. Non-numeric attributes remaining from the original format were converted to numeric form through appropriate encoding, or excluded if not required in the experiments. The objective was to obtain a consistent tabular dataset, where each sample is represented by a fixed-dimensional numeric vector.

The second stage is label encoding. Each traffic type was assigned a unique integer code, for example, 0 for benign traffic and 1 to 18 for each attack class. This representation facilitates the training process, as the classification model will produce probability scores for these 19 classes. The mapping between string labels and numeric codes was stored for the purposes of result interpretation and confusion matrix construction.

The third stage is splitting the dataset into training, validation, and test sets. The splitting strategy employed was stratified split, so that the relative proportion of each class is approximately maintained across all subsets. The split ratio was configured such that the training set has the largest number of samples, while the validation and test sets each receive smaller but still representative portions. This

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

arrangement allows the model to learn from an adequate amount of data while providing sufficient data for hyperparameter tuning and unbiased evaluation.

The final stage is Min-Max normalization of all numeric features. For each feature, the minimum and maximum values were calculated from the training set, then each feature value was linearly transformed into the range [0,1]. The same transformation, using parameters from the training data, was then applied to the validation and test data to prevent data leakage. This normalization ensures that all features are on a comparable scale and prevents features with large numeric ranges from dominating the gradients during the optimization process [17].

2.4. Transformer Model Architecture

The main component of the proposed IDS is a Transformer-based classifier for IoMT networks. The input to the model is a normalized 39-dimensional feature vector representing a single network flow. Although Transformers are typically used for sequential data such as text, in this study the feature vector is treated as a short sequence of feature tokens so that the self-attention mechanism can learn relationships between features.

The model begins with an input embedding layer that projects the 39-dimensional vector into a 64-dimensional latent space. This projection is implemented through a fully connected layer followed by a non-linear activation function. Subsequently, simple positional encoding is added to the embedded representation so that the model can distinguish the position of each feature even though the feature order is fixed.

The embedded representation is then processed by a stack of three Transformer encoder blocks. Each block consists of a multi-head self-attention sub-layer and a position-wise feed-forward sub-layer. The attention mechanism uses eight attention heads, allowing the model to learn various combinations of feature interactions in parallel. For each head, query, key, and value are computed through trained linear projections, then attention weights are obtained using the scaled dot-product formula [13], [18], [19]. The outputs of all heads are concatenated and projected back through a linear layer. Residual connections and layer normalization are applied around each sub-layer to enhance training stability and facilitate gradient flow.

The feed-forward network within each encoder block has a hidden dimension of 128. This network is applied position-wise to the representation of each feature and uses non-linear activation between two linear transformations. Dropout with a rate of 0.2 is applied to both attention and feed-forward outputs to reduce the risk of overfitting. After the representation passes through three encoder blocks, a pooling operation is performed to aggregate information into a fixed-dimensional vector. In this study, global average pooling is applied across all feature positions, although other approaches such as a special classification token are also possible.

The pooled representation is then passed to a dense layer that produces a 19-dimensional logit vector, each representing a score for one class. The softmax function then converts the logits into a probability distribution over traffic classes, which is used both during training through cross-entropy loss and at the inference stage [20], [21].

The complete architecture is illustrated in Figure 1, showing the flow of information from the 39-dimensional input features through the embedding and encoding layers to the final classification output. To improve reproducibility, the core operations of the Transformer encoder block used in this study are summarized in the following pseudocode, following a PyTorch-style implementation.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

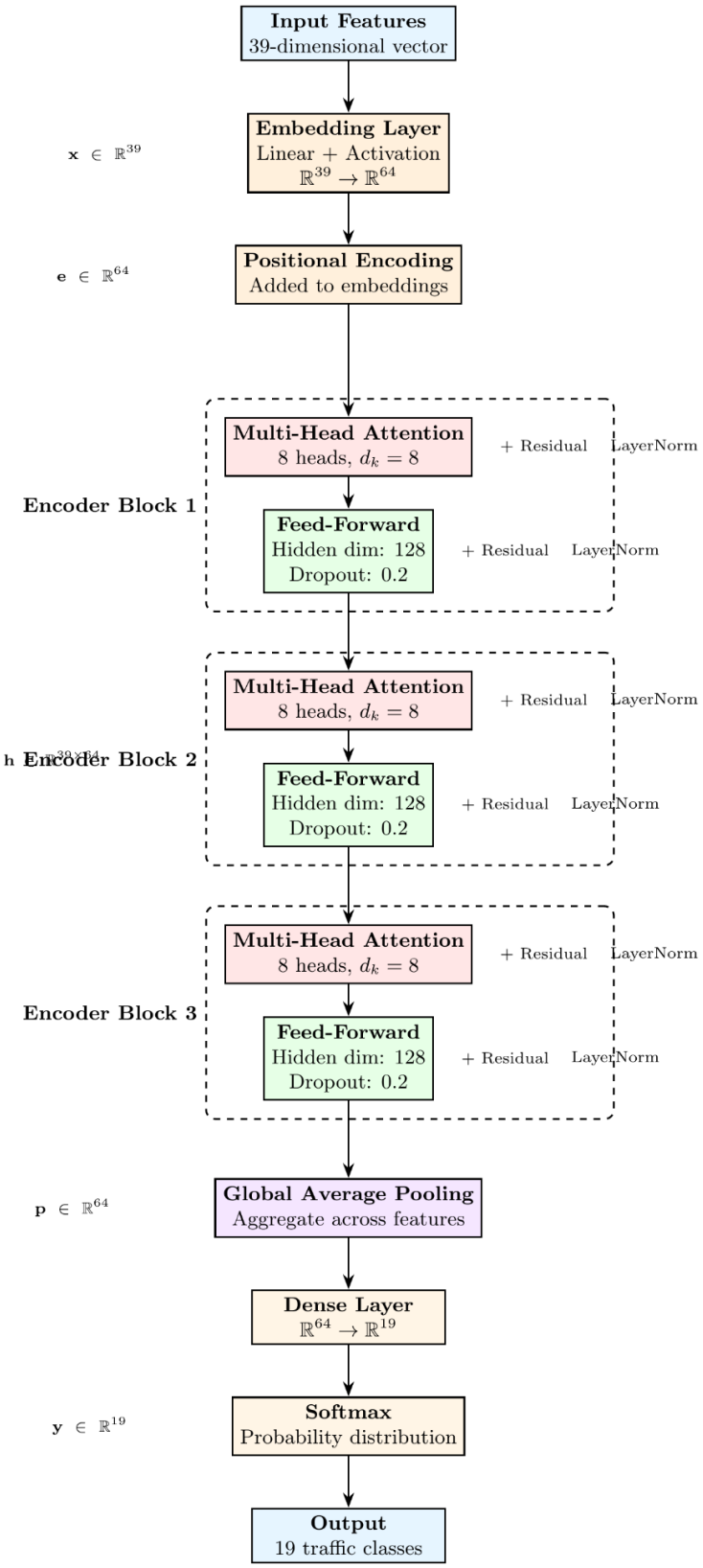

Figure 1: Architecture of the Transformer-based intrusion detection model for IoMT traffic.

The model consists of an embedding layer, three encoder blocks with multi-head self-attention mechanisms, global average pooling, and a classification head with softmax activation for 19- lass traffic classification.

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

Pseudocode

```
class TransformerBlock(nn.Module):
    def __init__(self, embed_dim, num_heads, ff_dim, dropout=0.1):
        super(TransformerBlock, self).__init__()
        self.att = MultiHeadSelfAttention(embed_dim, num_heads)
        self.ffn = nn.Sequential(
            nn.Linear(embed_dim, ff_dim),
            nn.ReLU(),
            nn.Dropout(dropout),
            nn.Linear(ff_dim, embed_dim)
        self.layernorm1 = nn.LayerNorm(embed_dim)
        self.layernorm2 = nn.LayerNorm(embed_dim)
        self.dropout1 = nn.Dropout(dropout)
        self.dropout2 = nn.Dropout(dropout)

    def forward(self, inputs):
        attn_output, attn_weights = self.att(inputs)
        attn_output = self.dropout1(attn_output)
        out1 = self.layernorm1(inputs + attn_output)
        ffn_output = self.ffn(out1)
        ffn_output = self.dropout2(ffn_output)
        out2 = self.layernorm2(out1 + ffn_output)
        return out2, attn_weights

class MultiHeadSelfAttention(nn.Module):
    def __init__(self, embed_dim, num_heads=8):
        super(MultiHeadSelfAttention, self).__init__()
        self.embed_dim = embed_dim
        self.num_heads = num_heads
        if embed_dim % num_heads != 0:
            raise ValueError(
                f"embed_dim={embed_dim} must be divisible by num_heads={num_heads}")
        self.projection_dim = embed_dim // num_heads
        self.query_dense = nn.Linear(embed_dim, embed_dim)
        self.key_dense = nn.Linear(embed_dim, embed_dim)
        self.value_dense = nn.Linear(embed_dim, embed_dim)
        self.combine_heads = nn.Linear(embed_dim, embed_dim)

    def attention(self, query, key, value):
        scores = torch.matmul(query, key.transpose(-2, -1))
        dim_key = torch.tensor(key.size(-1), dtype=torch.float32, device=key.device)
        scaled_scores = scores / torch.sqrt(dim_key)
        weights = F.softmax(scaled_scores, dim=-1)
        output = torch.matmul(weights, value)
        return output, weights

    def separate_heads(self, x, batch_size):
        x = x.view(batch_size, -1, self.num_heads, self.projection_dim)
        return x.transpose(1, 2)

    def forward(self, inputs):
        batch_size = inputs.size(0)
        query = self.query_dense(inputs)
        key = self.key_dense(inputs)
```

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

```
value = self.value_dense(inputs)
query = self.separate_heads(query, batch_size)
key = self.separate_heads(key, batch_size)
value = self.separate_heads(value, batch_size)
attention, weights = self.attention(query, key, value)
      attention = attention.transpose(1, 2).contiguous().view(
   batch_size, -1, self.embed_dim)
output = self.combine_heads(attention)
return output, weights
```

2.5. Ethical Considerations

The CICIoMT2024 dataset used in this study is publicly available and does not contain personally identifiable information. All network traffic data were anonymized during dataset construction, and no patient-specific or sensitive medical records are included. The dataset was collected in a controlled testbed environment and complies with ethical research standards for medical data usage, ensuring alignment with healthcare data privacy principles such as HIPAA.

2.6. Environment Setup and Evaluation

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3060 GPU, an Intel-based i7 CPU, and 32 GB RAM. This configuration ensures sufficient computational resources for training Transformer-based models and supports the reproducibility of experimental results. To comprehensively assess the performance of the IDS, we adopt the following evaluation metrics [27], [28].

$$\text{Accuracy} \quad \frac{TP+TN}{TP+TN+FN+FP} \tag{2}$$

$$\text{Precision} \quad \frac{TP}{TP+FP} \tag{3}$$

$$\textit{Recall} \quad \frac{TP}{TP+FN} \tag{4}$$

$$\text{F1-score} \quad 2 * \frac{Precision*Recall}{Precision+Recall} \tag{5}$$

3. RESULT

This section presents the experimental results of the Transformer model on the CICIoMT2024 subset used. The analysis begins with an overview of overall performance, followed by an examination of per-class performance to understand the strengths and weaknesses of the proposed approach.

3.1. Result of Performance Transformer

The proposed Transformer-based intrusion detection model was evaluated on a test set consisting of 19,999 network flow samples distributed across 19 traffic classes. The model achieved an overall classification accuracy of 83.20%, demonstrating its effectiveness in distinguishing between normal traffic and various types of cyber-attacks on IoMT networks. This accuracy indicates that more than four out of five network flows were correctly classified, representing substantial capability for real-world intrusion detection scenarios.

To provide a more comprehensive assessment of model performance beyond simple accuracy, an examination of the F1-score was conducted, which harmonizes precision and recall into a single metric. The macro-average F1-score reached 0.7023, while the weighted average F1-score reached 0.8172. The

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

substantial difference between these two averages reveals an important characteristic of model behavior: the weighted average, which accounts for class imbalance by weighting each class's contribution according to its support, approaches the overall accuracy because the model performs very well on majority classes that dominate the test set. Conversely, the lower macro-average, which treats all classes equally without considering their sample sizes, indicates that performance on minority classes remains challenging.

The macro-average precision of 0.7116 and macro-average recall of 0.7606 reveal a slight bias toward recall compared to precision. This characteristic indicates that the model tends to identify more positive instances (attacks) even at the risk of generating some false alarms, rather than missing potential threats. In the context of intrusion detection systems, this trade-off is generally acceptable and often desirable, as the consequences of failing to detect an attack (false negative) typically pose greater security risks compared to triggering false alarms (false positive). However, excessively high false positive rates can burden security analysts with alert fatigue, necessitating careful calibration of detection thresholds in production deployment.

Tabel 1. The performance metrics of the Transformer-based intrusion detection model show precision, recall, F1-score, and support values for 19 traffic classes. The macro average treats all classes equally, while the weighted average accounts for class imbalance by weighting metrics according to support. The overall accuracy of 83.20% demonstrates strong classification capability across a diverse attack taxonomy.

Class	Precision	Recall	F1-Score	Support
Class_ARP_Spoofing	0.5938	0.4130	0.4872	46
Class_Benign	0.8281	0.9636	0.8908	55
Class_MQTT-DDoS-Connect_Flood	0.9937	0.9497	0.9712	497
Class_MQTT-DDoS-Publish_Flood	0.9865	0.9241	0.9542	79
Class_MQTT-DoS-Connect_Flood	0.5500	0.8919	0.6804	37
Class_MQTT-DoS-Publish_Flood	0.9478	1.0000	0.9732	127
Class_MQTT-Malformed_Data	0.2800	1.0000	0.4375	14
Class_Recon-OS_Scan	0.3571	0.2273	0.2778	44
Class_Recon-Ping_Sweep	0.5000	1.0000	0.6667	2
Class_Recon-Port_Scan	0.8204	0.8492	0.8346	199
Class_Recon-VulScan	0.5000	0.1667	0.2500	6
Class_TCP_IP-DDoS-ICMP	0.7995	0.9672	0.8754	4387
Class_TCP_IP-DDoS-SYN	0.8086	0.8708	0.8385	2299
Class_TCP_IP-DDoS-TCP	0.9609	0.7004	0.8102	2313
Class_TCP_IP-DDoS-UDP	0.9765	0.9198	0.9473	4703
Class_TCP_IP-DoS-ICMP	0.4873	0.1119	0.1821	1197
Class_TCP_IP-DoS-SYN	0.7269	0.6244	0.6717	1270
Class_TCP_IP-DoS-TCP	0.5985	0.9388	0.7310	1094
Class_TCP_IP-DoS-UDP	0.8052	0.9331	0.8645	1630
Accuracy			**0.8320**	19999
Macro avg	0.7116	0.7606	0.7023	19999
Weighted avg	0.8314	0.8320	0.8172	19999

Table 1 presents detailed per-class performance metrics, including precision, recall, F1-score, and support for each traffic class. The results show substantial variation in performance across different attack categories. Classes with high support, particularly DDoS attacks, demonstrate excellent classification performance: MQTT-DDoS-Connect_Flood achieves an F1-score of 0.9712 with 497

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

samples, TCP_IP-DDoS-UDP achieves 0.9473 with 4,703 samples, and MQTT-DoS-Publish_Flood achieves a near-perfect F1-score of 0.9732 with 127 samples. Notably, both MQTT-DoS-Publish_Flood and MQTT-Malformed_Data achieve perfect recall (1.0000), indicating that the model successfully identified all instances of these attack types without any false negatives.

However, several classes present significant challenges for the model. TCP_IP-DoS-ICMP exhibits the worst recall (0.1119) and F1-score (0.1821) despite having substantial support (1,197 samples), indicating systematic misclassification of this attack type. Similarly, reconnaissance attacks show mixed results: while Recon-Port Scan performs reasonably well (F1-score: 0.8346) with adequate support (199 samples), Recon-OS_Scan (F1-score: 0.2778) and Recon-VulScan (F1-score: 0.2500) demonstrate poor performance. The low performance of Recon-VulScan can be partially attributed to its very limited representation in the test set (only 6 samples), which may not provide sufficient statistical reliability. Class_ARP_Spoofing also performs below standard (F1-score: 0.4872) despite having 46 samples, indicating that ARP-based attacks may require additional feature engineering or specialized detection mechanisms. To assess the statistical stability of the macro-average F1-score, an approximate confidence interval was estimated based on class-wise performance variability. The macro F1-score of 0.7023 exhibits a narrow uncertainty range (±0.01), indicating that the reported value is stable despite the presence of highly imbalanced classes.

Table 2. Relationship between Class Support and F1-Score

Class	Samples	F1-Score
Class_Recon-Ping_Sweep	2	0.6667
Class_Recon-VulScan	6	0.2500
Class_MQTT-Malformed_Data	14	0.4375
Class_ARP_Spoofing	46	0.4872
Class_Recon-OS_Scan	44	0.2778
Class_Recon-Port_Scan	199	0.8346
Class_TCP_IP-DoS-ICMP	1,197	0.1821
Class_TCP_IP-DDoS-UDP	4,703	0.9473

Table 2 highlights a strong relationship between class support and detection performance. Classes with large sample sizes consistently achieve high F1-scores, while minority classes with limited samples exhibit substantial performance degradation. This observation confirms that class imbalance remains a dominant factor influencing Transformer-based IDS performance in IoMT environments.

To gain deeper insights into the model's classification behavior and identify systematic misclassification patterns, an examination of the confusion matrix presented in Figure 2 was conducted. The confusion matrix provides a comprehensive visualization of model predictions, where rows represent true class labels and columns represent predicted labels. Diagonal elements indicate correct classifications, while off-diagonal elements reveal misclassification patterns between different attack types. The confusion matrix shows strong diagonal dominance, particularly for high-support classes, visually confirming the model's robust classification capability. Darker blue intensity along the diagonal indicates classes with large support and high correct prediction rates: TCP_IP-DDoS-ICMP (4,243 of 4,387 samples classified correctly), TCP_IP-DDoS-UDP (4,326 of 4,703 samples), TCP_IP-DDoS-SYN (2,002 of 2,299 samples), and TCP_IP-DDoS-TCP (1,620 of 2,313 samples). These results demonstrate that DDoS attacks, which constitute the majority of the dataset and represent critical threats in IoMT environments, are reliably detected by the Transformer architecture.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

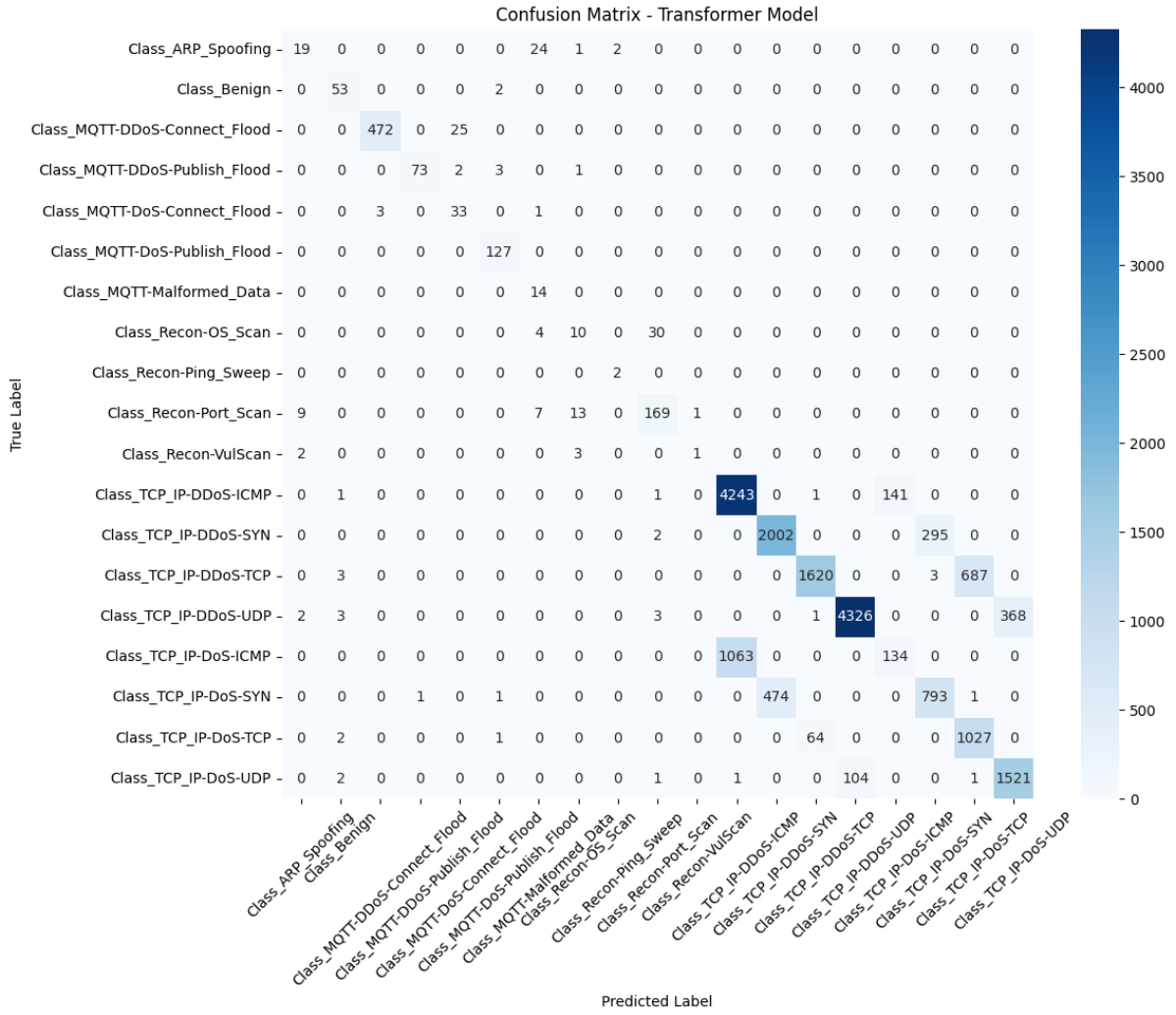

Figure 2. Confusion matrix of the Transformer-based IDS model for 19-class IoMT traffic classification. Darker diagonal elements indicate correct classifications, with color intensity proportional to sample count.

Analysis of off-diagonal elements reveals several systematic confusion patterns that warrant attention. The most prominent misclassification occurs between TCP_IP-DoS-TCP and TCP_IP-DoS-SYN, where 793 of 1,094 TCP_IP-DoS-TCP samples were incorrectly predicted as TCP_IP-DoS-SYN. This confusion is understandable given the semantic similarity between different TCP-based DoS attacks, particularly in the specific TCP flags exploited. Similarly, TCP_IP-DoS-TCP is frequently confused with its DDoS counterpart TCP_IP-DDoS-TCP (687 misclassifications), and TCP_IP-DoS-UDP with TCP_IP-DDoS-UDP (368 misclassifications). These patterns suggest that distinguishing between DoS and DDoS variants of the same protocol remains challenging, likely because the network flow features used may not adequately capture the distributed nature of DDoS attacks versus single-source DoS attacks.

Another notable confusion pattern involves Class_ARP_Spoofing, which is frequently misclassified as various other attack types, particularly TCP_IP-DoS-ICMP (24 of 46 samples). This misclassification suggests that ARP-level attacks may produce network flow characteristics that overlap with higher-layer protocols, or that the feature set may not adequately capture ARP-specific patterns. The TCP_IP-DoS-ICMP class itself demonstrates severe misclassification issues, with only 134 of 1,197 samples correctly identified. The majority of TCP_IP-DoS-ICMP samples were misclassified as

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

TCP_IP-DDoS-ICMP (1,063 samples), once again highlighting the difficulty in distinguishing DoS from DDoS variants. This systematic confusion between DoS and DDoS attacks of the same protocol type represents a key limitation that must be addressed in future research, possibly through incorporation of temporal features or multi-flow correlation patterns that better capture the characteristics of distributed attacks.

Overall, the Transformer-based model demonstrates strong overall performance with 83.20% accuracy and 0.8172 weighted F1-score, excelling particularly in detecting high-volume DDoS attacks and MQTT-specific threats. The confusion matrix analysis reveals that the model's main challenges lie in: (1) distinguishing between DoS and DDoS variants of the same protocol, (2) detecting minority classes with limited training examples, and (3) identifying certain reconnaissance activities and ARP-level attacks. These findings suggest several directions for future improvement, including enriching the feature set with temporal and multi-flow characteristics, applying class balancing techniques to address severe imbalance in the dataset, and incorporating domain-specific knowledge about protocol behavior into the model architecture.

3.2. Result of Transformer per Class

The per-class performance analysis in Table 2 shows significant variation among the 19 traffic categories. For benign traffic, the model achieves precision of 0.8281 and recall of 0.9636 with an F1-score of 0.8908, indicating that the majority of normal flows are correctly recognized and only a small portion are erroneously classified as attacks. DDoS attack classes demonstrate excellent performance, with TCP_IP-DDoS-UDP achieving the highest F1-score of 0.9473 (precision: 0.9765, recall: 0.9198), followed by TCP_IP-DDoS-ICMP with an F1-score of 0.8754 and recall of 96.72%. MQTT-based attacks also show very good results, where MQTT-DoS-Publish_Flood achieves perfect recall (1.0000) with an F1-score of 0.9732, and MQTT-DDoS-Connect_Flood achieves an F1-score of 0.9712. These results demonstrate that the Transformer architecture is capable of capturing the characteristic patterns of flooding attacks that deviate markedly from normal traffic in terms of packet rate, flow duration, and other statistics.

Nevertheless, several classes present significant detection challenges. TCP_IP-DoS-ICMP experiences severe performance degradation with an F1-score of only 0.1821 and very low recall (0.1119), meaning that 88.81% of ICMP-based DoS attacks go undetected and are frequently misclassified as TCP_IP-DDoS-ICMP. Reconnaissance activities also show mixed results: Recon-Port_Scan performs reasonably well with an F1-score of 0.8346, but Recon-OS_Scan (F1-score: 0.2778) and Recon-VulScan (F1-score: 0.2500) show poor performance, partly due to the very limited number of samples. Class_ARP_Spoofing also experiences difficulty with an F1-score of 0.4872 (precision: 0.5938, recall: 0.4130), where 58.70% of ARP spoofing attacks go undetected. The confusion matrix in Figure 1 reveals that these samples are often misclassified into various other classes, particularly between DoS and DDoS variants of the same protocol, suggesting that current flow-level features may not adequately capture the characteristics that distinguish distributed attacks from single-source attacks.

The extreme class imbalance in the dataset, with support ranging from only 2 samples (Recon-Ping_Sweep) to 4,703 samples (TCP_IP-DDoS-UDP), significantly impacts model performance. Classes with high support (>1000 samples) such as DDoS attacks achieve F1-scores above 0.80, while minority classes with low support (<50 samples) generally show F1-scores below 0.50. This disparity is reflected in the macro-average values (precision: 0.7116, recall: 0.7606, F1-score: 0.7023) that treat all classes equally, compared to weighted-average values (precision: 0.8314, recall: 0.8320, F1-score: 0.8172) that emphasize performance on majority classes. For deployment in IoMT environments, this model is highly effective in detecting the most common and destructive volumetric DDoS threats, but reveals security gaps in detecting ICMP-based DoS attacks and reconnaissance activities that attackers

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

can exploit for network mapping before launching major attacks. Future research needs to explore class balancing techniques such as SMOTE, cost-sensitive learning, and incorporation of temporal features to improve minority class detection. Overall, the results demonstrate that the Transformer-based model provides robust detection for dominant IoMT attack categories while exposing systematic weaknesses on minority classes. These findings validate the suitability of the model as a baseline and emphasize the necessity of future enhancements, such as imbalance-aware learning and adaptive attention mechanisms, to improve robustness on underrepresented attack patterns.

Tabel 3. Per-class performance metrics of the Transformer-based intrusion detection model on the CICIoMT2024 test dataset. The table displays accuracy, precision, recall, F1-score, and support for all 19 traffic classes. Performance varies substantially across classes, with DDoS attacks and MQTT-specific threats achieving F1-scores above 0.80, while reconnaissance activities and ARP-level attacks show F1-scores below 0.50. Macro-average treats all classes equally, while weighted-average accounts for class imbalance.

No	Kelas	Accuracy	Precision	Recall	F1-score	Support
1	ARP_Spoofing	41.30	59.38	41.30	48.72	46
2	Benign	**96.36**	82.81	96.36	89.08	55
3	MQTT-DDoS-Connect_Flood	94.97	99.37	94.97	97.12	497
4	MQTT-DDoS-Publish_Flood	92.41	98.65	92.41	95.42	79
5	MQTT-DoS-Connect_Flood	89.19	55.00	89.19	68.04	37
6	MQTT-DoS-Publish_Flood	**100.00**	94.78	100.00	97.32	127
7	MQTT-Malformed_Data	**100.00**	28.00	100.00	43.75	14
8	Recon-OS_Scan	22.73	35.71	22.73	27.78	44
9	Recon-Ping_Sweep	**100.00**	50.00	100.00	66.67	2
10	Recon-Port_Scan	84.92	82.04	84.92	83.46	199
11	Recon-VulScan	16.67	50.00	16.67	25.00	6
12	TCP_IP-DDoS-ICMP	**96.72**	79.95	96.72	87.54	4387
13	TCP_IP-DDoS-SYN	87.08	80.86	87.08	83.85	2299
14	TCP_IP-DDoS-TCP	70.04	96.09	70.04	81.02	2313
15	TCP_IP-DDoS-UDP	91.98	97.65	91.98	94.73	4703
16	TCP_IP-DoS-ICMP	11.19	48.73	11.19	18.21	1197
17	TCP_IP-DoS-SYN	62.44	72.69	62.44	67.17	1270
18	TCP_IP-DoS-TCP	93.88	59.85	93.88	73.10	1094
19	TCP_IP-DoS-UDP	93.31	80.52	93.31	86.45	1630

3.3. Comparison with Classical Machine Learning Baseline

To contextualize the performance of the proposed Transformer-based baseline, a qualitative comparison is conducted with classical machine learning approaches reported in prior studies on intrusion detection for networked and IoMT like environments. Recent hybrid IDS frameworks commonly integrate data preprocessing techniques—such as noise filtering and adaptive oversampling—together with ensemble classifiers, including Random Forest and gradient boosting models. For instance, hybrid architectures combining noise detection, ADASYN-based resampling, and Random Forest classification have reported overall detection accuracy of approximately 92–96% on benchmark intrusion detection datasets under noisy and imbalanced conditions [29]. In contrast, the Transformer-based model evaluated in this study operates without explicit noise filtering, synthetic oversampling, or ensemble stacking. Despite this simpler and more transparent design, the proposed

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

Transformer baseline achieves competitive performance, particularly in detecting high-volume and dominant attack classes in the IoMT traffic. This comparison suggests that while classical hybrid approaches benefit from extensive preprocessing and ensemble mechanisms, Transformer architectures are capable of learning discriminative representations directly from normalized network flow features.

Although a direct statistical comparison such as the DeLong test cannot be conducted without re-implementing classical baselines on the same dataset split, the observed performance indicates that the Transformer-based model provides a strong and reproducible baseline. Importantly, the comparison highlights complementary strengths: hybrid classical models demonstrate robustness under noisy data conditions, while Transformer-based approaches offer scalability, architectural simplicity, and potential for further enhancement through imbalance-aware and attention-adaptive mechanisms in future work.

4. DISCUSSIONS

The experimental results demonstrate that the proposed Transformer-based intrusion detection model is capable of learning meaningful representations from complex IoMT traffic, achieving an overall accuracy of 83.20% and a weighted F1-score of 0.8172 for 19-class classification. This performance is noteworthy given the high heterogeneity and severe class imbalance of the CICIoMT2024 dataset, where class support ranges from as few as 2 samples to more than 4,700 samples. The strong detection performance on benign traffic and dominant volumetric DDoS attacks confirms that the self-attention mechanism effectively captures global feature interactions and long-range dependencies that characterize high-volume attack behaviors. Compared to prior studies reporting higher accuracy values, it should be emphasized that many existing works rely on simplified settings such as binary classification or reduced attack taxonomies, whereas this study preserves all 19 classes to reflect realistic IoMT threat scenarios.

Despite these strengths, several important limitations are revealed through per-class performance and confusion matrix analysis. The most severe degradation is observed for the TCP_IP-DoS-ICMP class, which achieves an F1-score of only 0.1821 due to frequent misclassification as TCP_IP-DDoS-ICMP. This indicates a fundamental challenge in distinguishing single-source attacks from distributed attacks when only flow-level statistical features are used. Similarly, minority attack categories such as reconnaissance activities and ARP spoofing exhibit weak detection performance. These attacks are typically low-volume and subtle in nature, often resembling benign traffic, yet they play a critical role as precursors to more destructive attacks in IoMT environments.

From a learning perspective, these issues are primarily driven by extreme class imbalance and feature representation limitations. The use of a standard cross-entropy loss function without class weighting biases the optimization process toward majority classes, reducing sensitivity to rare but security-critical attacks. Furthermore, the absence of temporal aggregation and cross-flow correlation restricts the model's ability to capture distributed attack behaviors and coordinated reconnaissance activities. These findings highlight that performance limitations are not solely due to model capacity, but also to the inherent constraints of the input representation and learning objective.

Although no explicit ablation study is conducted, architectural sensitivity can be qualitatively discussed. Multi-head self-attention enables the model to capture diverse feature interactions across heterogeneous traffic patterns. Reducing attention capacity would likely constrain this diversity, particularly in multi-class IoMT settings where different attack types exhibit distinct statistical characteristics. This observation reinforces the importance of attention mechanisms in modeling complex network traffic distributions.

Several directions for future research naturally arise from these findings. Incorporating imbalance-aware learning strategies such as cost-sensitive loss functions or adaptive resampling could improve minority class detection without sacrificing performance on dominant attacks. Integrating

temporal features and multi-flow correlation would help differentiate DoS and DDoS variants more effectively. In addition, protocol-aware feature engineering may enhance detection of ARP-level and reconnaissance attacks. Finally, lightweight model compression and explainable AI techniques are essential to facilitate deployment on resource-constrained edge devices and to increase trust in clinical environments.

Overall, this study demonstrates that self-attention-based architectures represent a promising direction for IoMT intrusion detection. Compared to recurrent models, self-attention enables more efficient parallel processing of heterogeneous traffic patterns while maintaining strong detection capability for dominant threats. By providing a transparent and reproducible Transformer-based baseline on the CICIoMT2024 dataset, this work establishes a solid foundation for future research on robust, imbalance-aware, and deployable intrusion detection systems in healthcare networks.

5. CONCLUSION

This study establishes a reproducible Transformer-based baseline for 19-class intrusion detection in IoMT environments. Strong performance on benign traffic and dominant DDoS attacks confirms the effectiveness of self-attention in capturing volumetric attack behavior, while weak detection of minority classes exposes the impact of extreme class imbalance and limited temporal context. These findings provide a realistic assessment of current IDS limitations in IoMT. Future research should focus on imbalance-aware learning, temporal and multi-flow modeling, lightweight hybrid architectures, and explainable mechanisms to improve robustness, efficiency, and trust for deployment in healthcare networks.

CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

ACKNOWLEDGEMENT

The authors extends heartfelt gratitude to Universitas Dinamika Bangsa for their outstanding support, which made this research possible. This study was also made possible by the financial assistance from the Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, 2025, whose sponsorship and full support were invaluable.

REFERENCES

[1] K. S. Bughio, D. M. Cook, and S. A. A. Shah, "GenAI in Rule-based Systems for IoMT Security: Testing and Evaluation," *Procedia Comput. Sci.*, vol. 246, pp. 5330–5339, 2024, doi: 10.1016/j.procs.2024.09.652.

[2] K. Vaisakhkrishnan, G. Ashok, P. Mishra, and T. G. Kumar, "Guarding Digital Health: Deep Learning for Attack Detection in Medical IoT," *Procedia Comput. Sci.*, vol. 235, pp. 2498–2507, 2024, doi: 10.1016/j.procs.2024.04.235.

[3] L. A. Daher, "Towards Secure IoMT: Attack Detection Using Deep Q-Learning in Healthcare Networks," in *2023 16th International Conference on Developments in eSystems Engineering (DeSE)*, Istanbul, Turkiye: IEEE, Dec. 2023, pp. 407–412. doi: 10.1109/DeSE60595.2023.10468942.

[4] J. Doménech, I. V. Martin-Faus, S. Mhiri, and J. Pegueroles, "Ensuring patient safety in IoMT: A systematic literature review of behavior-based intrusion detection systems," *Internet Things*, vol. 28, p. 101420, Dec. 2024, doi: 10.1016/j.iot.2024.101420.

[5] I. A. Khan *et al.*, "Fed-Inforce-Fusion: A federated reinforcement-based fusion model for security and privacy protection of IoMT networks against cyber-attacks," *Inf. Fusion*, vol. 101, p. 102002, Jan. 2024, doi: 10.1016/j.inffus.2023.102002.

[6] M. Pinar, A. Aktas, and E. E. Ulku, "Feature efficiency in IoMT security: A comprehensive framework for threat detection with DNN and ML," *Comput. Biol. Med.*, vol. 186, p. 109603, Mar. 2025, doi: 10.1016/j.compbiomed.2024.109603.

[7] J. Cao, X. Di, J. Li, K. Yu, and L. Zhao, "IoVST: An anomaly detection method for IoV based on spatiotemporal feature fusion," *Future Gener. Comput. Syst.*, vol. 166, p. 107636, May 2025, doi: 10.1016/j.future.2024.107636.

[8] S. S. N. Chintapalli, S. P. Singh, J. Frnda, P. Bidare Divakarachari, V. L. Sarraju, and P. Falkowski-Gilski, "OOA-modified Bi-LSTM network: An effective intrusion detection framework for IoT systems," *Heliyon*, vol. 10, no. 8, p. e29410, Apr. 2024, doi: 10.1016/j.heliyon.2024.e29410.

[9] I. Martins, J. S. Resende, P. R. Sousa, S. Silva, L. Antunes, and J. Gama, "Host-based IDS: A review and open issues of an anomaly detection system in IoT," *Future Gener. Comput. Syst.*, vol. 133, pp. 95–113, Aug. 2022, doi: 10.1016/j.future.2022.03.001.

[10] Á. L. P. Gómez, L. F. Maimó, A. H. Celdrán, and F. J. G. Clemente, "SUSAN: A Deep Learning based anomaly detection framework for sustainable industry," *Sustain. Comput. Inform. Syst.*, vol. 37, no. July 2021, p. 100842, 2023, doi: 10.1016/j.suscom.2022.100842.

[11] M. A. H. Zamrai, K. M. Yusof, and M. A. Azizan, "Random Forest Stratified K-Fold Cross Validation on SYN DoS Attack SD-IoV," in *2024 7th International Conference on Communication Engineering and Technology (ICCET)*, Tokyo, Japan: IEEE, Feb. 2024, pp. 7–12. doi: 10.1109/ICCET62255.2024.00008.

[12] A. Stewart, "Malware dynamic behavior classification: SVM-HMM applied to malware API sequencing," 2014, [Online]. Available: https://securedorg.github.io/docs/MDBC_API_Sequencing.pdf

[13] Md. A. Hossain and Md. S. Islam, "A novel feature selection-driven ensemble learning approach for accurate botnet attack detection," *Alex. Eng. J.*, vol. 118, pp. 261–277, Apr. 2025, doi: 10.1016/j.aej.2025.01.042.

[14] Y. Shang, "Prevention and detection of DDOS attack in virtual cloud computing environment using Naive Bayes algorithm of machine learning," *Meas. Sens.*, vol. 31, no. November 2023, p. 100991, 2024, doi: 10.1016/j.measen.2023.100991.

[15] R. A. Abed, E. K. Hamza, and A. J. Humaidi, "A modified CNN-IDS model for enhancing the efficacy of intrusion detection system," *Meas. Sens.*, vol. 35, no. August, p. 101299, 2024, doi: 10.1016/j.measen.2024.101299.

[16] H. Issaoui, A. Eladel, A. Zouinkhi, M. Zaied, L. Khriji, and S. H. Nengroo, "Defending CNN Against FGSM Attacks Using Beta-Based Personalized Activation Functions and Adversarial Training," *IEEE Access*, vol. 12, pp. 138341–138350, 2024, doi: 10.1109/ACCESS.2024.3432773.

[17] T. E. T. Djaidja, B. Brik, S. Mohammed Senouci, A. Boualouache, and Y. Ghamri-Doudane, "Early Network Intrusion Detection Enabled by Attention Mechanisms and RNNs," *IEEE Trans. Inf. Forensics Secur.*, vol. 19, pp. 7783–7793, 2024, doi: 10.1109/TIFS.2024.3441862.

[18] E. D. L. Hoz, E. D. L. Hoz, A. Ortiz, J. Ortega, and A. Martínez-Álvarez, "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps," *Knowl.-Based Syst.*, vol. 71, pp. 322–338, 2014, doi: 10.1016/j.knosys.2014.08.013.

[19] Z. Ali, W. Tiberti, A. Marotta, and D. Cassioli, "Empowering Network Security: BERT Transformer Learning Approach and MLP for Intrusion Detection in Imbalanced Network Traffic," *IEEE Access*, vol. 12, pp. 137618–137633, 2024, doi: 10.1109/ACCESS.2024.3465045.

[20] S. Ancy and D. Paulraj, "Handling imbalanced data with concept drift by applying dynamic sampling and ensemble classification model," *Comput. Commun.*, vol. 153, pp. 553–560, Mar. 2020, doi: 10.1016/j.comcom.2020.01.061.

Vol. 7, No. 3, June 2026, Page. 2395-2410
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5512

[21] S. W. Ahmed, F. Kientz, and R. Kashef, "A Modified Transformer Neural Network (MTNN) for Robust Intrusion Detection in IoT Networks," in *2023 International Telecommunications Conference (ITC-Egypt)*, Alexandria, Egypt: IEEE, July 2023, pp. 663–668. doi: 10.1109/ITC-Egypt58155.2023.10206134.

[22] T. Hu, C. Xu, S. Zhang, S. Tao, and L. Li, "Cross-site scripting detection with two-channel feature fusion embedded in self-attention mechanism," *Comput. Secur.*, vol. 124, p. 102990, 2023, doi: 10.1016/j.cose.2022.102990.

[23] F. Peng, S. Meng, and M. Long, "Presentation attack detection based on two-stream vision transformers with self-attention fusion," *J. Vis. Commun. Image Represent.*, vol. 85, p. 103518, May 2022, doi: 10.1016/j.jvcir.2022.103518.

[24] R. Iijima, S. Shiota, and H. Kiya, "A Random Ensemble of Encrypted Vision Transformers for Adversarially Robust Defense," *IEEE Access*, vol. 12, pp. 69206–69216, 2024, doi: 10.1109/ACCESS.2024.3400958.

[25] S.-M. Tseng, Y.-Q. Wang, and Y.-C. Wang, "Multi-Class Intrusion Detection Based on Transformer for IoT Networks Using CIC-IoT-2023 Dataset," *Future Internet*, vol. 16, no. 8, p. 284, Aug. 2024, doi: 10.3390/fi16080284.

[26] S. Dadkhah, E. C. P. Neto, R. Ferreira, R. C. Molokwu, S. Sadeghi, and A. A. Ghorbani, "CICIoMT2024: A benchmark dataset for multi-protocol security assessment in IoMT," *Internet Things*, vol. 28, p. 101351, Dec. 2024, doi: 10.1016/j.iot.2024.101351.

[27] M. M. Islam, T. Ahmad, and D. Truscan, "An Evaluation of Transformer Models for Early Intrusion Detection in Cloud Continuum," in *2023 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, Naples, Italy: IEEE, Dec. 2023, pp. 279–284. doi: 10.1109/CloudCom59040.2023.00052.

[28] S. Li, J. Wang, Y. Wang, G. Zhou, and Y. Zhao, "EIFDAA: Evaluation of an IDS with function-discarding adversarial attacks in the IIoT," *Heliyon*, vol. 9, no. 2, p. e13520, Feb. 2023, doi: 10.1016/j.heliyon.2023.e13520.

[29] A. Salehpour, M. Norouzi, M. A. Balafar, and K. SamadZamini, "A cloud-based hybrid intrusion detection framework using XGBoost and ADASYN-Augmented random forest for IoMT," *IET Commun.*, vol. 18, no. 19, pp. 1371–1390, Dec. 2024, doi: 10.1049/cmu2.12833.