SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 8369-8382
Open Access. Online & Print

DOI: 10.5281/zenodo.12426832

AI-AUGMENTED FEDERATED LEARNING: A PREDICTIVE FRAMEWORK FOR DETECTING AND REAL-TIME CORRELATION OF IAM, NETWORK TELEMETRY, AND ENDPOINT EVENTS FOR PREVENTING LATERAL MOVEMENT ATTACKS

Md. Mushfiqur Rahman[1*], Sazzad Hossain[2]

[1]Department of System Management and Information Security, Samarkand State University, Samarkand, Uzbekistan, mushfique98@gmail.com
[2]Department of System Management and Information Security, Samarkand State University, Samarkand, Uzbekistan, sazzad69@gmail.com

Received: 14/11/2025
Accepted: 24/03/2026

Corresponding Author: Md. Mushfiqur Rahman (mushfique98@gmail.com)

ABSTRACT

Modern enterprises depend on cloud services and federated identity systems, and they use multiple types of endpoint devices, which leads to greater security risks and makes it possible for attackers to execute advanced multi-phase cyber-attacks that include lateral movement attacks. The existing security systems, which depend on fixed rules and network boundaries, fail to provide real-time attack detection because they cannot analyze diverse security data, and they cannot keep up with changing enemy attack patterns. The research develops a security framework that uses hybrid Artificial Intelligence (AI) to identify and prevent lateral movement attacks through the analysis of identity, network and endpoint telemetry data. The proposed framework uses Transformer-based models to extract contextual and temporal features, while Autoencoder-based anomaly detection identifies zero-day behaviors, and Temporal Graph Neural Networks (TGNN) create relational attack path models for enterprise systems. The hybrid decision fusion mechanism combines outputs from three different model types, which include supervised models, unsupervised models and graph-based models to create predictions about risks that enable automated policy-based decision making. The Canadian Institute for Cybersecurity Intrusion Detection System 201 (CICIDS2017) benchmark dataset, together with a custom multi-source enterprise dataset, which includes Identity and Access Management (IAM) authentication logs, endpoint event logs and simulated lateral movement paths, was used to test the framework. The experimental results show that the Transformer Autoencoder Graph Neural Networks (TAGNN) system achieves 98.5% detection accuracy while detecting complex stealthy lateral movement patterns with reduced false positive rates. The research demonstrates that hybrid AI models used for multi-telemetry correlation create a threat detection and prevention method that can handle modern enterprise cybersecurity needs.

KEYWORDS: Anomaly Detection, Enterprise Cybersecurity, Graph Neural Networks, Hybrid AI, Lateral Movement Detection and prevention.

INTRODUCTION

The rapid expansion of enterprise networks, cloud infrastructure, and distributed digital identities has significantly increased the complexity of modern cybersecurity environments (Judijanto et al., 2023; Ghadge, 2024). In current scenarios, organizations face massive levels of identity log events, network traffic, and endpoints, which are sources for valuable indicators of malicious activity (Yen et al., 2013; Tucker et al., 2025; Little Lion Scientific, 2024). But due to the massive levels, varieties, and speeds, traditional cybersecurity approaches such as rule-based systems, signature-based detection, and Security Information and Event Management (SIEM)-based solutions lack sufficiency in recognizing advanced threats (Aslan et al., 2023; Li & Liu, 2021 Mallick & Nath, 2024). Among various advanced threats, lateral movement is one of the most harmful forms of enemy tactics, in which the enemy travels undetected through the system networks once breached (Smiliotopoulos et al., 2024; Alsharabi et al., 2025; Bi et al., 2022). Figure 1 illustrates the six key stages of the cyber-attack lifecycle, progressing from reconnaissance to final attacker objectives.

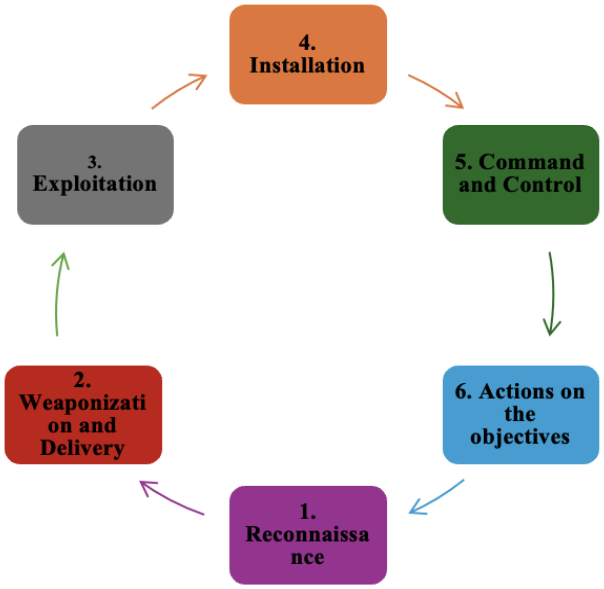

Figure 1: Cyber Attack life cycle stages (Cybersecurity for Me, 2025)

Identity-driven attacks have further risen to the forefront due to the increased use of cloud computing, two-factor authentication, and federated identity infrastructures (Sharma, 2025a; Sharma, 2025b). Use of compromised credentials, privilege escalation, and odd patterns of log-in attempts very often act as the initial step for further intrusion (McGonigal, 2024; He et al., 2025). At the same time, endpoints and network activities also give insight into other variables like process anomalies in the host process space, misuse of remote command channels, and scan traffic (Benova & Hudec, 2024; Spiekermann, 2025; Kuchar & Fujdiak, 2025). While each of these sources of telemetry is individually inadequate for the task of monitoring the whole process flow of an attack, the full strength of such analysis lies in the corollary construction of patterns of identity-driven, host-driven, and network-driven patterns (Feldman et al., 2022; Fabry, 2021; Yang et al., 2023).

AI has recently shown promise in helping to overcome these challenges, providing the capability to learn complex behavioral patterns that are very difficult to detect using traditional solutions. Solutions such as the Transformer model, Autoencoder (AE), and Graph Neural Networks (GNNs) are able to determine high-level context information, detect anomalies, and establish relationships between users, systems, and events (Berente et al., 2021; Injadat et al., 2021; Rudroff, 2024; Soliman et al., 2024). However, current solutions using AI are only able to convict using a single type of information or a single model type. Furthermore, current solutions are barely able to establish a relationship between identity information telemetry and events such as user endpoints or networks for detecting multi-stage lateral movement-type attacks (Sarker, 2022; Bagaric et al., 2021; Park et al., 2024; Bagaric et al., 2022).

Although prior studies have shown notable progress, several key limitations remain unresolved: many existing approaches depend solely on supervised learning, which cannot detect zero-day attacks, while others rely exclusively on anomaly detection, resulting in high false-positive rates (Zoppi et al., 2021; Guo, 2023; Deldar & Abadi, 2023). Some models perform well on network datasets such as the CICIDS2017 but lack representation of real IAM logs or endpoint activity (Yusof et al., 2022). Others attain good accuracy but are unable to capture relational attack paths, given that they do not use graph-based reasoning (Rosay et al., 2021). To address these gaps, this paper would put forward a hybrid AI-driven detection framework that integrates the use of Transformers for feature extraction, Feature Tokenizer (FT)-Transformers for supervised classification, Autoencoders for zero-day anomaly detection, and Temporal GNN for modeling lateral movement paths along with an automated prevention layer that applies Multi-Factor Authentication (MFA) enforcement, endpoint

isolation, credential restriction, and network blocking to stop detected attacks from propagating further. It thus offers a single solution to detect a wide variety of attack behaviors. The novelty of the research work lies in the use of a specifically tailored multi-source dataset that includes both the prime IAM log data, endpoint events, simulated lateral paths, as well as the use of the CICIDS2017 dataset, together with a hybrid AI approach that uses the Transformers, Autoencoders, as well as Temporal GNN models to recognize known, unknown, as well as multi-hop attacks. This study contributes to knowledge through the following elements.

- A combined primary–secondary data framework incorporating IAM logs, endpoint events, lateral movement paths, and CICIDS2017 for comprehensive threat representation.
- A transformer-based technique for feature extraction, which focuses on capturing context patterns as well as temporal patterns that can be derived from diverse security telemetry sources.
- Hybrid detection model employing FT-Transformer, Autoencoder, and Temporal GNN for multi-dimensional threat detection.
- A decision fusion mechanism that integrates outputs from supervised, unsupervised, and graph-based models to produce accurate and low–false-positive predictions.

The rest of the paper is divided as follows: Section 2 presents an exhaustive assessment and analysis of the pertinent studies conducted by different authors on the domain. Section 3 defines the proposed approach and consists of its structure and the methods.

LITERATURE REVIEW

This section critically reviews recent literature on AI–based approaches for detecting lateral movement and advanced cyber threats in enterprise environments.

Recent advancements that have emerged within the last few years, from 2021 to 2025, include rapid development in AI-based cybersecurity tools, digital forensics, and Zero Trust security. Ndibe et al. (2025) [36] developed AI-based forensic tools for real-time anomaly identification through their use of Machine Learning (ML) algorithms, Deep Learning (DL) algorithms and log correlation methods. The result showed fast detection of anomalies and lower rates of false identification and successful operation of automated forensic processes. In a related direction, Hoque (2025) [37] developed a system to detect lateral movement across hybrid cloud environments by constructing a temporal heterogeneous GNN which processes identity data, host information, network details and cloud system logs. The model outperformed baselines with F1 = 0.86 and Area Under Curve (AUC) = 0.80, compared to XGBoost (F1 ≈ = 0.72, AUC = 0.69).

Other studies advanced AI-driven threat detection in broader Zero Trust and enterprise environments. Cate (2025) [38] developed an AI-based cyber-defence system for Zero Trust Architectures, which anticipates security threats through its use of supervised learning, unsupervised learning, and reinforcement learning techniques. The results showed improved detection accuracy, reduced false positives, and faster response times compared to traditional tools. Complementing this, Chowdhury et al. (2025) [39] conducted a systematic review that evaluated AI-based DL models that assess cybersecurity threats in real-time for enterprise IT systems. The researchers used Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-based methods to screen 2347 documents and selected 142 studies for their analysis. The research testing real-time systems with sub-second response times showed traditional models achieved 10 to 25 percent accuracy gains, which researchers found to be less effective than Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) systems, Transformers and GNNs. The research about digital forensics demonstrated that AI systems that used improved detection methods created major advantages for the field. Similarly, Ogochukwu et al. (2025) [40] conducted research to use ML for digital forensics through supervised and unsupervised and deep learning methods on two datasets, CICIDS2017 and University of New South Wales – Network-Based 2015 (UNSW-NB15). The models achieved F1 > 0.92 and AUC > 0.95, which exceeded rule-based baselines that produced F1 scores of approximately 0.71 and AUC scores of approximately 0.69 while generating significantly fewer false positives. AI also played a central role in the evolution of Zero Trust and Identity Threat Detection and Response (ITDR) systems. Satyam et al. (2025) [41] developed an AI-driven Identity Threat Detection and Response (ITDR) system using sequence models, graph learning, and self-supervised techniques on cloud identity telemetry. The system achieved major improvements with an accuracy rate of 84.7% and, an Area Under the Receiver Operating Characteristic Curve (AUROC) value of 0.91 and a recall at one percent False Positive Rate (FPR) of 0.62 and a decrease in false positive results by 51%. The system

achieved a reduction in Mean Time to Detect (MTTD) from 45 minutes to 18 minutes and a decrease in Mean Time to Respond (MTTR) from 120 minutes to 45 minutes. Complementary Zero Trust architectures were explored by Shonubi et al. (2025) [42], who created a multi-layered Zero Trust Architecture that secures cross-domain data in federated enterprise networks through identity federation, micro-segmentation, policy enforcement, and AI-driven behavioral analytics. The proposed framework demonstrated strengthened cross-domain protection, improved access control, and enhanced resilience for high-risk operational environments.

Additional advancements were made toward AI-enhanced identity verification. Aramide et al. (2024) [43] developed an AI-driven Zero Trust identity verification model using behavioral analytics, contextual signals, and ML-based trust scoring, evaluated on a simulated dataset of 10,000 user sessions. The framework achieved strong performance with 96.3% accuracy, 1.8% False Acceptance Rate (FAR), 1.9% False Rejection Rate (FRR), and 24.5 ms trust-evaluation latency, outperforming the autoencoder baseline (92.7% accuracy, 2.4% FAR). Nangi et al. (2023) [44] created a multi-layered Zero-Trust security framework which uses AI-based identity and access controls to protect systems through its implementation of Kubernetes and service-mesh controls, policy-as-code and ML-based risk assessment. The system achieved 94.7% detection accuracy along with 3.2% false positive rate and a 1.8-second access delay, which surpassed conventional IAM systems that achieved 85% accuracy and 12% false positive rate and basic Zero-Trust systems, which achieved 90% accuracy and 7% false positive rate in simulated cloud-native environments. In addition, Datla et al. (2021) [45] used behavioral analytics and ML-based anomaly monitoring to identify cloud identity threats. The system achieved a 12% increase in detection accuracy while decreasing false positives by 18%, which allowed for faster detection of credential misuse.

Existing research continues to examine Mean Squared Error (MSE) together with network and endpoint telemetry as three separate entities, which hinders their ability to identify multi-stage cyber-attacks [36,40]. Most models lack the essential temporal graph reasoning feature, which enables tracking of multi-hop lateral movement patterns [37]. Research also remains dependent on either supervised or unsupervised learning alone, which results in either poor zero-day detection or high false positives [38,41]. The widely used benchmark datasets fail to provide an accurate representation of actual IAM and endpoint system operations, which leads to inadequate model performance in real business environments [45].

RESEARCH METHODOLOGY

This section briefly explains how the data were prepared, processed, and modeled using Transformers, Autoencoders, and GNNs.

Data Collection

The research process starts with researchers gathering secondary data from publicly available datasets, which they access through Kaggle. The study uses the CICIDS 2017 benchmark dataset as its main secondary dataset, which contains detailed network traffic data that displays both regular network behavior and malicious network activities [46]. The dataset functions as a standardized resource for cybersecurity research because it allows researchers to evaluate their testing and research on intrusion detection systems. The study uses both secondary data sources and primary data, which researchers gathered through controlled experiments that simulated actual business attack scenarios. The primary data is used to assess and validate the benchmark dataset. The study uses three primary datasets, which researchers collected from multiple sources, including IAM authentication logs that document user identity activities, endpoint event logs that track host system activity, and simulated lateral movement paths that show multi-stage attack movement across connected networks.

Data Pre-processing

The collected information is pre-processed in terms of missing value imputation, noise reduction, log standardization, and timestamp synchronization. The missing information is imputed to prevent any bias, whereas noisy and redundant data are eliminated to promote clear signal perception. Log standardization enables uniform conversion from diverse security logs to have synchronized timestamps that provide clarity to interrelated timelines.

Technique used

This section examines the key techniques used in this study, including Transformers, Autoencoders, and GNNs.

Recursive Feature Elimination (RFE)

In this study, RFE is used to select the most discriminative Transformer-generated features by

repeatedly ranking and removing features with minimal importance.

RFE Equation:

$$S_{t+1} = S_t - \arg\min(I_t) \quad (1)$$

At each iteration t the feature set S_t is reduced by removing the feature with the lowest importance score I_t.

Transformer

A Transformer is a DL model utilizing attention for capturing relationships in sequential data. For this paper, the model selected for this experiment is employed only for feature extraction [47]. In this study, it is used for extracting meaningful high-level features from raw logs on IAM, network, and events. Essentially, a Transformer is based on self-attention. Self-attention is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q is the query vector, representing the context related to the present token, K is the key vector, representing the relevance of other tokens, V is the value vector, representing the information to be computed, and $d_k\, It$ It is a key dimension for attention score scaling.

FT-Transformer

The FT-Transformer is a variant of the Transformer adapted for classification tasks in structured or tabular data in a supervised learning manner [48]. This study provides a classifier of known attack patterns, which takes as input the extracted Transformer features and outputs accurate supervised predictions of malicious versus regular events.

The prediction output is computed as:

$$\hat{y} = \text{softmax}(Wh + b) \quad (3)$$

Where yˆ represents the class probability outputs, h represents the input feature representations produced by the Transformer, and W and b are the weight matrix and bias added in the final classification stage.

Autoencoder

An autoencoder is an unsupervised DL approach for a neural network model. This model has the capacity to learn for reconstructing the inputs and identify the anomalies based on the error of reconstruction [49]. An autoencoder is used in this research for determining the zero-day attacks and anomalies, as well as learning the expected behavior.

Key Equations

Encoding:

$$z = f(W_e x + b_e \quad (4)$$

Decoding:

$$\hat{x} = g(W_d z + b_d \quad (5)$$

Reconstruction Loss:

$$L = \| x - \hat{x} \|^2 \quad (6)$$

Where x is the input feature vector, z is the encoded representation, $\hat{x}$ is the reconstructed signal, W_e and W_d are the parameter matrices for the encoder and decoder layers, b_e and b_d are the corresponding bias terms, and L is the loss.

Graph Neural Network (GNN)

GNN helps in learning through aggregation based on graph-structured data [50]. In this proposed system, GNN is used for lateral movement attack paths to establish connections between users, systems, and network links to identify malicious paths.

The node representation update is defined as:

$$h_v^{(k)} = \sigma\left(W \sum_{u \in \mathcal{N}(v)} h_u^{(k-1)}\right) \quad (7)$$

Where e $h_v^{(k)}$ is the updated embedding of node $v$at layer k, $\mathcal{N}(v)$ represents the set of neighboring nodes, $h_u^{(k-1)}$ is the previous layer embedding of each neighbour, W is the trainable aggregation weight matrix, and σ is the activation function introducing non-linearity.

Proposed Methodology

The proposed methodology establishes a complete hybrid system that detects and prevents lateral movement attacks by linking identity information with network data and endpoint security information. The researchers collected their validation research data from primary datasets, which include IAM authentication logs, endpoint event logs and simulated lateral movement paths, and from the secondary benchmark dataset known as CICIDS 2017. The collected data undergoes pre-processing through multiple procedures, which include handling missing values, standardizing log formats, removing noise and normalizing labels to achieve dataset consistency and quality. The process starts with transformer-based feature extraction that gathers contextual and sequential data from security logs, and it ends with RFE, which detects vital and unique attributes. The researchers divided the improved feature set into two parts, which they used to train a hybrid model that consists of an FT-Transformer for supervised attack classification, a fine-tuning autoencoder for unsupervised anomaly detection and zero-day detection and a temporal GNN for multi-hop lateral movement path modeling and identification. The outputs of these models are combined using a hybrid decision fusion mechanism to compute a final risk-aware prediction. Based on

this prediction, a hybrid automated response framework combining rule-based and behaviour-based prevention techniques is executed, where predefined decision rules enforce appropriate security controls such as multi-factor authentication, privilege restriction, endpoint isolation, and network path blocking, thereby ensuring real-time containment and prevention of lateral movement propagation. Figure 2 below shows the proposed architecture of this methodology.

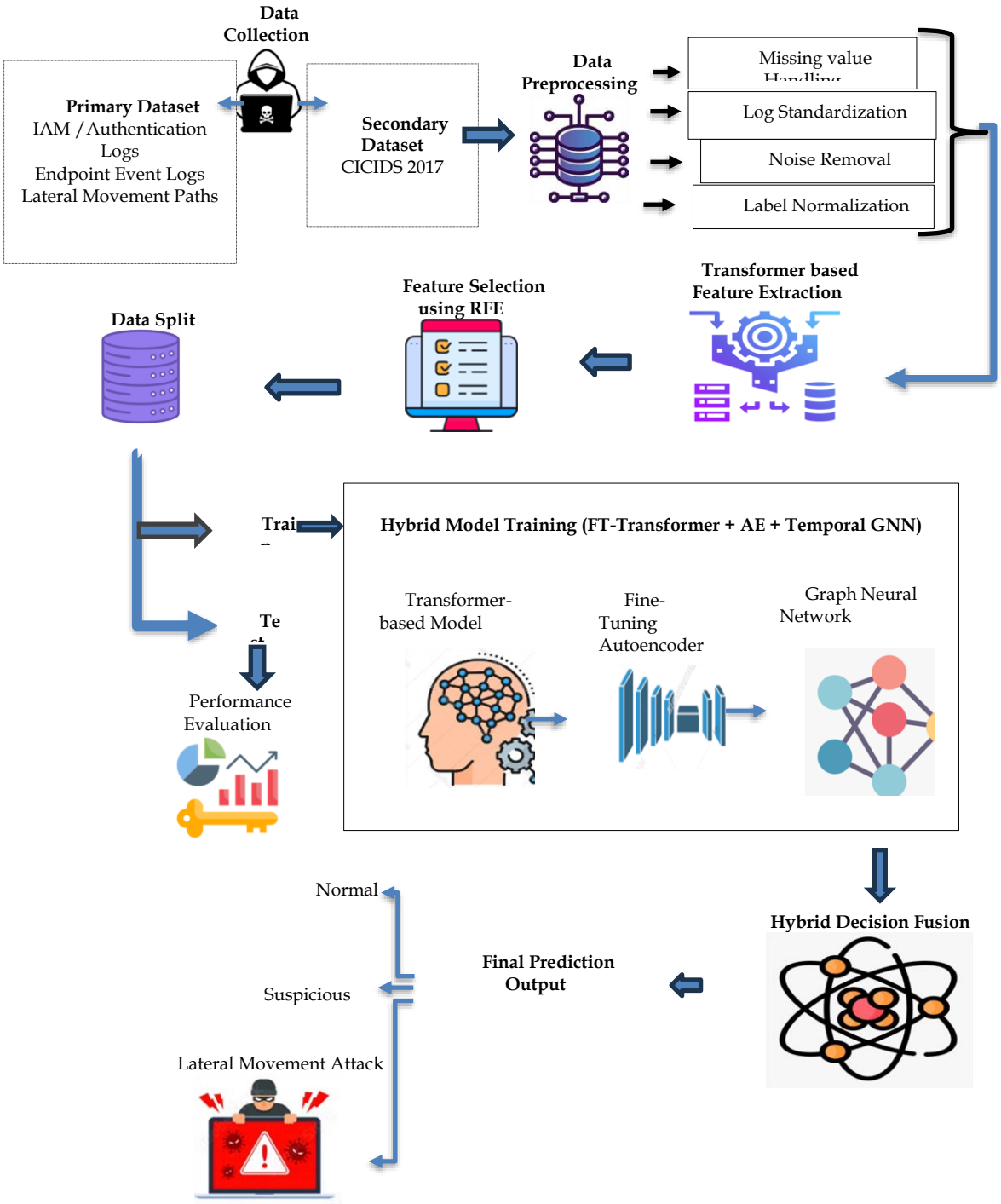

Figure 2: Proposed Architecture

Proposed Algorithm

This section describes the lateral movement detection and prevention algorithm.

Algorithm: Transformer-Assisted Lateral Movement Detection and Prediction

Step 1: Data Collection

Collect raw security data from:

- Primary dataset D_p = {IAM logs, Endpoint events, LM paths}
- Secondary dataset D_s = CICIDS2017

$$D = D_p \cup D_s$$

Step 2: Data Pre-processing

Apply four cleaning operations:

- Missing value handling

$$x_i' = \begin{cases} x_i, & x_i \neq \text{NaN} \\ \mu_i, & x_i = \text{NaN} \end{cases}$$

- Log format standardization

$$x_i'' = f_{\text{std}}(x_i')$$

- Noise removal

$$x_i''' = f_{\text{denoise}}(x_i'')$$

- Label normalization

$$y' = f_{\text{norm}}(y)$$

Final cleaned dataset:

$$X_c = \{x_i'''\}, Y_c = \{y_i'\}$$

Step 3: Transformer-based Feature Extraction

Pass cleaned data into the Transformer model:

$$H = \text{Transformer}(X_c)$$

Self-attention operation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Feature embeddings:

$$F = f_{\text{embed}}(H)$$

Step 4: Feature Selection using RFE

Initialize feature set:

$$S_0 = F$$

At each iteration t:

$$I_t = \text{Importance}(S_t)$$

Remove the least essential feature:

$$S_{t+1} = S_t - \arg\min(I_t)$$

Stop when:

$$| S_t | = k$$

Final selected features:

$$S = S_t$$

Step 5: Data Split

$$(S_{\text{train}}, S_{\text{test}}, Y_{\text{train}}, Y_{\text{test}}) = \text{Split}(S, Y_c)$$

Step 6: Hybrid Model Training

- Supervised FT-Transformer (Known Attacks)

$$\hat{y}_{\text{sup}} = \text{softmax}(Wh + b)$$

Loss:

$$L_{\text{sup}} = -\sum y \log(\hat{y}_{\text{sup}})$$

- Autoencoder (Zero-Day / Anomaly)

Encoding:

$$z = f(W_e S_{\text{train}} + b_e)$$

Decoding:

$$\hat{S} = g(W_d z + b_d)$$

Reconstruction Loss:

$$L_{\text{ae}} = \| S - \hat{S} \|^2$$

Anomaly score:

$$A = L_{\text{ae}}$$

- TGNN (Lateral Movement)

Graph $G = (V, E)$.
Node update:

$$h_v^{(k)} = \sigma\left(W \sum_{u \in N(v)} h_u^{(k-1)}\right)$$

Output:

$$\hat{y}_{\text{gnn}} = f_{\text{gnn}}(h_v^{(K)})$$

Step 7: Hybrid Decision Fusion
Collect outputs:

$$O = \{\hat{y}_{\text{sup}}, A, \hat{y}_{\text{gnn}}\}$$

Weighted fusion:

$$\hat{Y} = \alpha\hat{y}_{\text{sup}} + \beta A + \gamma\hat{y}_{\text{gnn}}$$

Where:

$$\alpha + \beta + \gamma = 1$$

Final label:

$$\text{Pred} = \begin{cases} \text{Normal}, & \hat{Y} < \tau_1 \\ \text{Suspicious}, & \tau_1 \le \hat{Y} < \tau_2 \\ \text{Lateral Movement Attack}, & \hat{Y} \ge \tau_2 \end{cases}$$

Step 8: Rule-Based Automated Prevention
Define Prevention Function:

$$P(y) = \begin{cases} \text{No Action} & y = 0 \\ \text{MFA + Privilege Restriction} & y = 1 \\ \text{Isolation + Credential Disable + network Block} & y = 2 \end{cases}$$

Operationally:
If Normal (0):

$$P(0) = \emptyset$$

If Suspicious (1):

$$P(1) = \{\text{MFA, Limited Access}\}$$

If Lateral Movement Attack (2):

$$P(2) = \{\text{Endpoint Isolation, Disable Account, Block Attack Path}\}$$

Step 9: Behaviour-Based Automated Prevention
A behaviour-based prevention mechanism is applied to dynamically respond to deviations from normal user, endpoint, and network behaviour.
Compute behaviour-based risk score:

$$R_t = \alpha A_t + \beta\hat{y}_{gnn} + \gamma C_t$$

Where:
A_t= Autoencoder anomaly score
$\hat{y}_{gnn}$= Temporal GNN behaviour output
C_t= Contextual behaviour features

$$\alpha + \beta + \gamma = 1$$

Final prevention decision:

$$P_{beh}(R_t) = \begin{cases} \emptyset, & R_t < \tau_3 \\ \{\text{MFA, Limited Access}\}, & \tau_3 \le R_t < \tau_4 \\ \{\text{Endpoint Isolation, Account Restriction, Network Segmentation}\}, & R_t \ge \tau_4 \end{cases}$$

This approach enables early, adaptive, and intelligent prevention of stealthy and zero-day lateral movement attacks by complementing rule-based controls with behaviour-driven decision making.

Step 10: Performance Evaluation

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

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

RESULTS & DISCUSSION

The section presents experimental findings for the proposed hybrid framework, which combines three components of the system. The study demonstrates that enterprise environments can achieve dependable lateral movement detection through the combination of multi-source correlation and hybrid decision fusion methods.

Figure 3 shows the number of traffic flow records per CICIDS2017 CSV file, which contains between 1,70,000 and 7,00000 records, while the Wednesday-WorkingHours file holds the highest record count. The different attack scenarios and time periods show distinct patterns of data density.

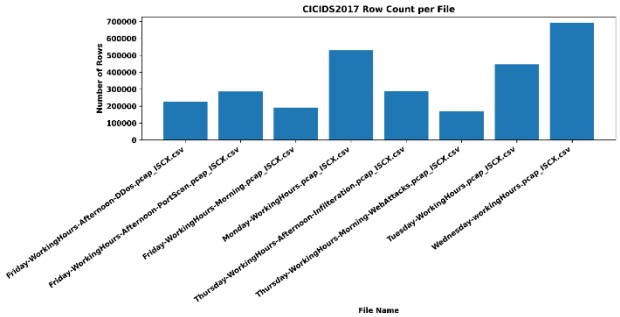

Figure 1: CICIDS2017 Row Count per File

The class-wise distribution of CICIDS traffic flow records shows that the dataset contains approximately 2,05,000 normal instances and 40,000 suspicious instances. The researchers need to create new evaluation metrics because class imbalance forces them to use weighted values together with hybrid learning methods.

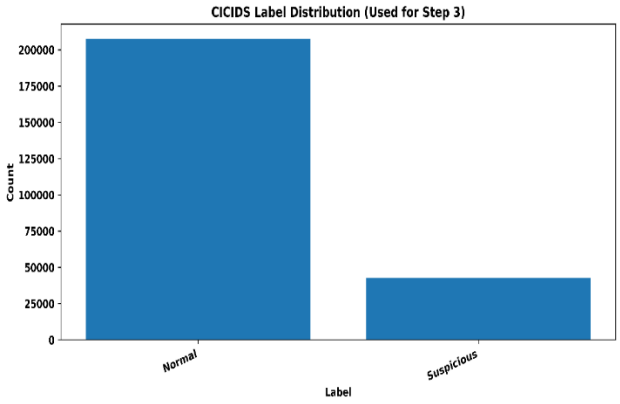

Figure 2: CICIDS Traffic Class Distribution

The CICIDS traffic records show a normalized and sampled label distribution which contains approximately 1,000,00 normal records, 20,000 suspicious records, and negligible unknown instances as shown in figure 5. The distribution maintains class imbalance yet allows for model training, which requires both stable performance and scalable capacity.

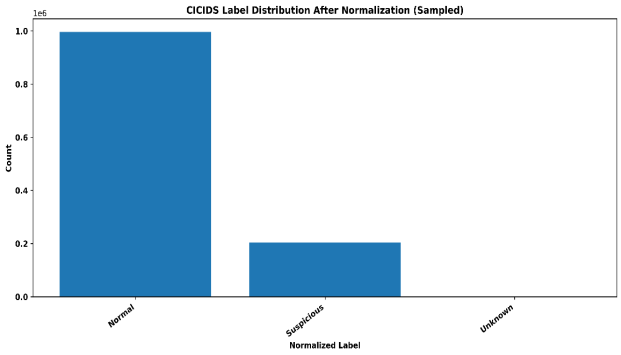

Figure 3: Normalized CICIDS Label Distribution

The L2-norm distribution of Transformer-generated CICIDS embeddings shows that most embedding norms are between 7.0 and 7.9, with a peak at 7.8, as shown in Figure 6. The system demonstrates stable feature representations that maintain their magnitude through normal operating conditions.

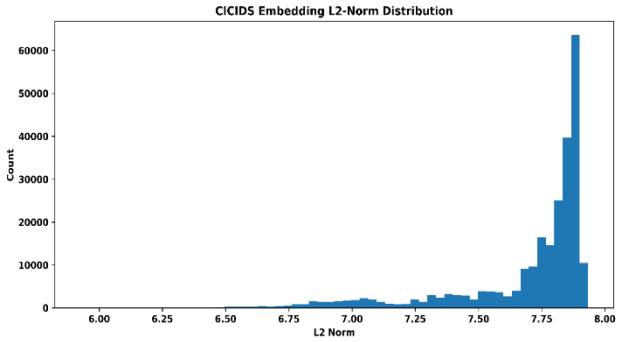

Figure 4: CICIDS Embedding L2-Norm Distribution

Figure 7 shows the most important features that RFE with Logistic Regression selected as top features, and these chosen features include f40, f9, and f18, which show maximum discriminative ability through their high absolute coefficients that reach 24 to 27. The features that received lower selection rankings still maintain relevance because their coefficients remain above 8. The RFE process accomplishes two tasks by reducing feature dimensionality and maintaining important features that enhance model performance and ability to generalize.

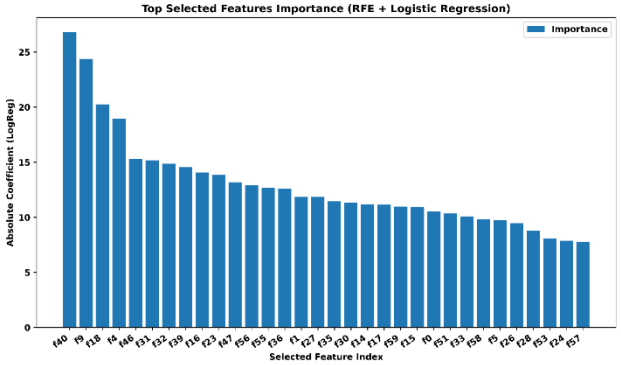

Figure 5: RFE-Selected Feature Importance

The training dataset shows autoencoder reconstruction errors through Figure 8, which displays normal samples with extremely low errors between 0 and 0.00005 MSE, while suspicious samples produce higher but scattered errors that reach about 0.0016. The autoencoder successfully understands normal behavior patterns according to this evidence.

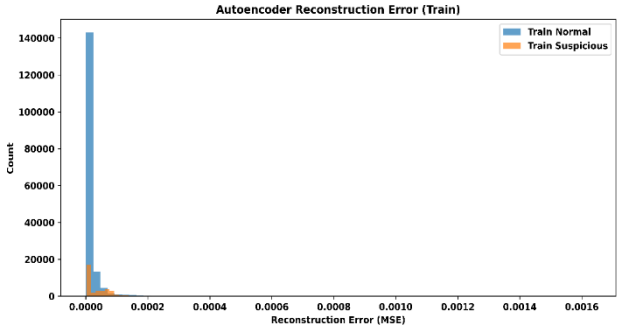

Figure 6: Autoencoder Train Reconstruction Error

Figure 9 shows the test dataset autoencoder reconstruction error, which demonstrates that normal samples' MSE values remain between 0 and 0.00005 while suspicious samples show MSE values that reach approximately 0.0007. The system demonstrates successful anomaly detection because it detects actual anomalies and operates effectively with unknown data.

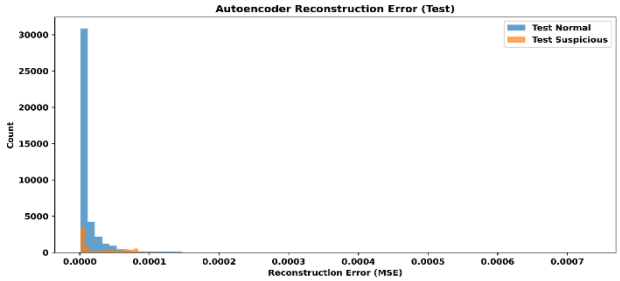

Figure 7: Autoencoder Test Reconstruction Error

Figure 10 shows autoencoder training and validation loss curves in which training loss decreases from 0.11 to 0.00, and validation loss shows the same pattern throughout 20 epochs. The system shows quick convergence while maintaining consistent learning progress without developing overfitting issues.

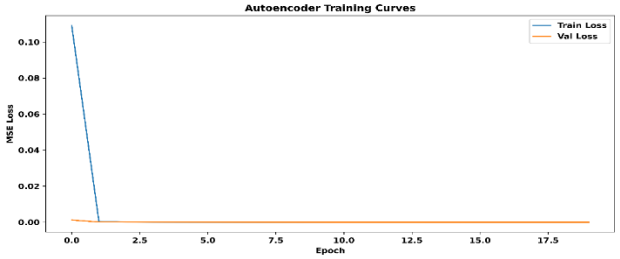

Figure 8: Autoencoder Training and Validation Loss

The FT-Transformer training and validation loss curves in Figure 11 demonstrate that training loss decreases from 0.43 to 0.26 and validation loss drops from 0.37 to 0.24 throughout the 12 epochs. The two curves maintain proximity because they show both stable convergence and effective generalization abilities.

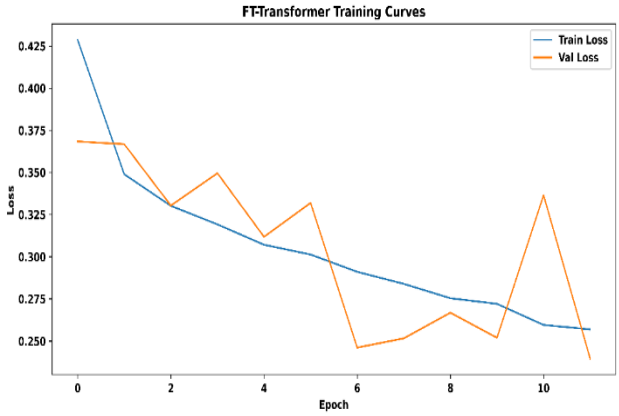

Figure 9: FT-Transformer Training and Validation Loss

The risk score distribution for identity nodes shows the results from the Temporal GNN assessment of identity nodes. The identity assessment shows that most identities display risk scores that range from approximately 0.3 to 0.6. A smaller subset of identities shows risk scores that exceed 0.7 and reach 1.0, as shown in Figure 12. The GNN demonstrates its capacity to detect compromised identities through its ability to create this long-tailed distribution.

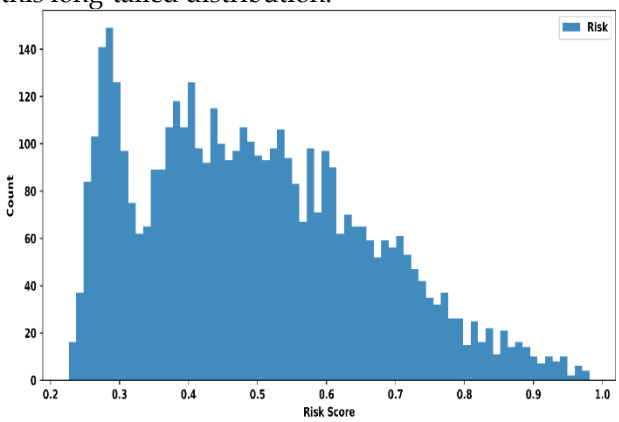

Figure 10: Identity Node Risk Distribution (Temporal GNN)

The confusion matrix for the identity-based attack detection hybrid fusion model is displayed in Figure 13. The model correctly identifies 39655 normal instances and 5549 attack instances. The results showed 1842 false positives and 2954 false negative results, which demonstrated strong discrimination ability but resulted in controlled misclassification.

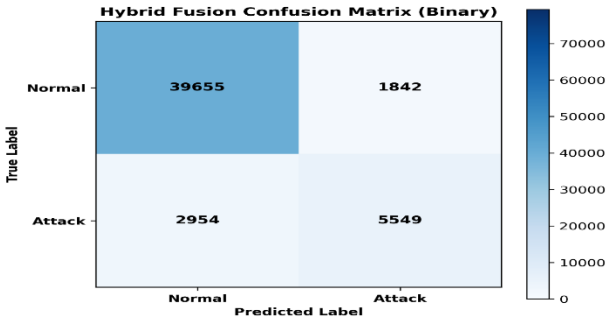

Figure 11: Hybrid Fusion Confusion Matrix

The hybrid model produced fused risk scores for identity nodes, which resulted in two distinct risk score distributions. The instances show their main distribution at low risk values, which range from 0.1 to 0.2, while their smaller tail extends to higher risk scores that exceed 0.6, as shown in Figure 14. The system demonstrates successful risk aggregation because it can distinguish between normal users and high-risk individuals.

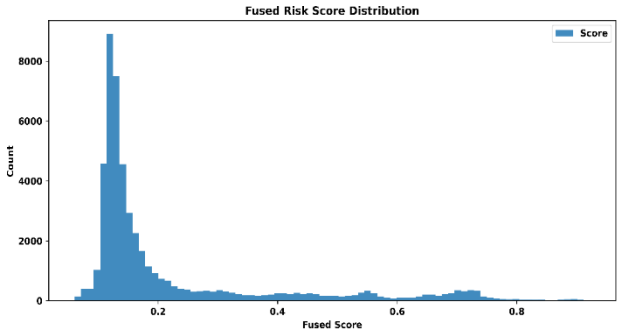

Figure 12: Fused Risk Score

Figure 15 shows the performance of the proposed model through Accuracy, Precision, Recall, and F1-Score measurements for Normal and Attack Classes. The proposed model achieved a high accuracy of 98.5% in relation to both Normal and Attack Classes. The precision was also balanced for Normal and Attack Classes with a precision of 97%. The Normal Class showed higher recall results 98% while the Attack Class reached 96%, but the Attack Class achieved an F1-score of 98%, which surpassed the Normal Class, showing both classes performed equally well.

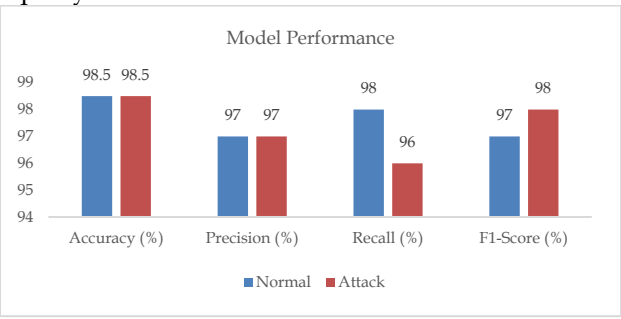

Figure 13: Hybrid Model Performance Metrics

Figure 16 demonstrates the activation of various preventive actions that are based on established rules through the operation of the proposed system. Most events (approximately 42000 instances) required no action. The system enforcement of MFA restrictions applied to 5000 cases, while 2000 high-risk events required endpoint isolation together with account disabling and attack-path block implementation. The distribution shows that most activities operate normally while organizations implement stricter controls to address higher-risk detection scenarios.

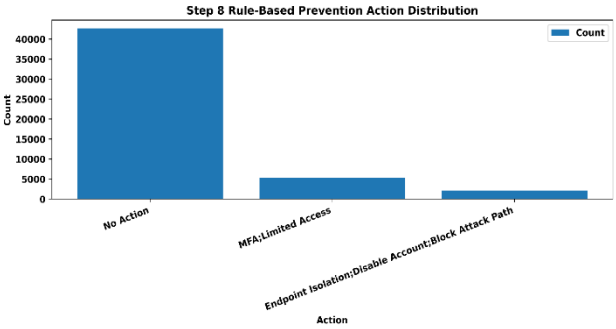

Figure 14: Rule-Based Prevention Action Distribution

Figure 17 shows the behavior-based risk score R_t, which the proposed model produces. The system categorizes most instances into low-risk values, which range from 0.15 to 0.25, because they demonstrate non-threatening behavior. The system shows a small number of cases that present a higher risk score, reaching approximately 0.9, because they exhibit behavior that may indicate malicious activity related to lateral movement.

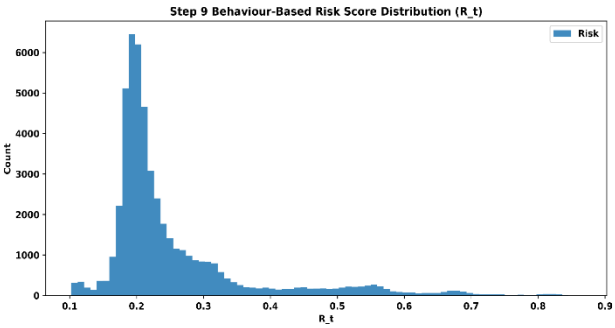

Figure 15: Behaviour-Based Risk Score Distribution

Comparative Analysis

The comparative analysis of existing studies demonstrates the progressive improvement in detection accuracy achieved through different ML and DL techniques. The Serverless AI Shield (SAS) framework developed by Pathade et al. 2026 Pathade et al., 2026) demonstrates efficient protection for serverless environments through its 94.2 percent accurate operational performance. The study conducted by Satyam et al. 2025 (Satyam et al., 2025)

used LSTM, Transformer and GNN and gradient-boosted risk scoring methods, which produced an 84.7 percent accuracy result that showed the system needed improvement to achieve better detection capabilities despite its complex design. The hybrid Random Forest (RF) and Autoencoder method developed by Aramide et al. 2024 (Aramide, 2024) achieved 96.3 percent accuracy, while Nangi et al. 2023 (Nangi et al., 2023) obtained 94.7 percent accuracy through their supervised ML classification system. The study conducted by Amouri et al. 2020 (Amouri et al., 2020) demonstrated that RF and linear regression models achieved 98 percent accuracy, which proved that ensemble-style learning effectively detects intrusions. The proposed research method outperforms all current methods through its use of a hybrid TAGNN model, which achieves the best accuracy of 98.5 percent. The system demonstrates its superiority through the combination of three distinct techniques, which include contextual feature extraction, anomaly detection, and graph-based relational learning. The proposed study presents a complete accuracy comparison with existing methods in Table 1.

Table 1: Comparative Analysis

Author(s)	Year	Technique used	Accuracy (%)
(Pathade et al., 2026)	2026	SAS	94.2
(Satyam et al., 2025)	2025	LSTM + Transformer + GNN + Gradient-Boosted Risk Scoring	84.7
(Aramide, 2024)	2024	RF + AE	96.3
(Nangi et al., 2023)	2023	Supervised ML–based classification	94.7
(Amouri et al., 2020)	2020	RF + Linear Regression	98
Proposed Study	2026	TAGNN	98.5

CONCLUSION & FUTURE SCOPE

Modern digital infrastructures have become increasingly distributed, dynamic, and interconnected, such that traditional security controls are becoming ineffective to address contemporary, sophisticated cyber threats like lateral movement. To solve the existing problem, the study developed a unified AI detection and prevention system that uses Transformer technology for feature extraction, Autoencoder technology for anomaly detection and TGNN technology to connect identity data with network and endpoint telemetry data. The TAGNN framework demonstrated outstanding performance because it reached 98.5 percent detection accuracy, which successfully identified both known and unknown security threats while delivering automated risk-based protection. Future work would expand the existing framework by using federated learning to share threat intelligence between different organizations and by using explainable AI to support better analysts, through which the system would be tested across extensive real-world enterprise and cloud-native environments

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

ACKNOWLEDGEMENTS

Sample text: We thank the anonymous reviewers for their constructive comments. This work was partially supported by the 7th Framework Programme "Project Name" funded by the EU within the Reflective Societies Work Programme 2014-2010. The authors would especially like to thank the personnel of the Research Centre for their support and technical cooperation.

REFERENCES

Albertyn, C. (2025). A comparative study of gradient boosting algorithms and transformer neural networks (Doctoral dissertation, Stellenbosch University).

Alsharabi, N., Bhardwaj, A., Alshammari, T., Alotibi, S., Alshammari, D., & Jadi, A. (2025). Threat hunting the shadows: Detecting adversary lateral movement with Elasticsearch. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3556184

Amouri, A., Alaparthy, V. T., & Morgera, S. D. (2020). A machine learning-based intrusion detection system for mobile IoT. Sensors, 20(2), 461.

Aramide, O. (2024). Zero-trust identity principles in next-gen networks. World Journal of Advanced Research and Reviews, 23(3), 3304–3316.

Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. Electronics, 12(6), 1333. https://doi.org/10.3390/electronics12061333

Bagaric, M., Svilar, J., Bull, M., Hunter, D., & Stobbs, N. (2021). The solution to pervasive bias and discrimination in criminal justice: Transparent artificial intelligence. American Criminal Law Review, 59(1).

Bagaric, M., Svilar, J., Bull, M., Hunter, D., & Stobbs, N. (2022). Transparent and fair artificial intelligence in criminal justice. American Criminal Law Review, 59, 95.

Benova, L., & Hudec, L. (2024). Comprehensive analysis and evaluation of anomalous user activity in web server logs. Sensors, 24(3), 746. https://doi.org/10.3390/s24030746

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. MIS Quarterly, 45(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Bi, J., He, S., Luo, F., Meng, W., Ji, L., & Huang, D.-W. (2022). Defense of advanced persistent threat on industrial internet of things with lateral movement modeling. IEEE Transactions on Industrial Informatics, 19(9), 9619–9630. https://doi.org/10.1109/TII.2022.3231406

Canadian Institute for Cybersecurity. (2017). CICIDS2017 dataset.

Cate, M. (2025). Building a proactive cyber defense model: Leveraging AI for threat hunting

Choi, S. R., & Lee, M. (2023). Transformer architecture and attention mechanisms in genome data analysis. Biology, 12(7), 1033.

Chowdhury, T. K. (2025). AI-powered deep learning models for real-time cybersecurity risk assessment. ASRC Procedia, 1(1), 675–704.

Cybersecurity for Me. (2025). Cybersecurity facts, figures & infographics (2025).

Datla, L. S. (2021). Identity threat detection: Techniques for preventing credential abuse. International Journal of Emerging Trends in Computer Science and Information Technology, 2(4), 95–104.

Deldar, F., & Abadi, M. (2023). Deep learning for zero-day malware detection and classification: A survey. ACM Computing Surveys, 56(2), 1–37.

Esmaeili, F., Cassie, E., Nguyen, H. P. T., Plank, N. O. V., Unsworth, C. P., & Wang, A. (2023). Anomaly detection for sensor signals utilizing deep learning autoencoder networks. Bioengineering, 10(4), 405.

Fabry, R. E. (2021). Limiting the explanatory scope of extended active inference. Biology & Philosophy, 36(1), 6. https://doi.org/10.1007/s10539-021-09782-6

Feldman, V., McMillan, A., & Talwar, K. (2022). Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In Proceedings of the IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS) (pp. 954–964). https://doi.org/10.1109/FOCS52979.2021.00096

Ghadge, N. (2024). Digital identity in the age of cybersecurity: Challenges and solutions. London Journal of Research in Computer Science & Technology, 24(1), 1–10. https://journalspress.uk/index.php/LJRCST/article/view/1444

Guo, Y. (2023). Machine learning-based zero-day attack detection: Challenges and future directions. Computer Communications, 198, 175–185.

He, S., Lei, Y., Zhang, Z., Sun, Y., Li, S., Zhang, C., & Ye, J. (2025). Identity deepfake threats to biometric authentication systems: Public and expert perspectives. arXiv. https://doi.org/10.48550/arXiv.2506.06825

Hoque, M. N. (2025). Graph neural networks for detecting lateral movement in hybrid cloud environments. Academica Global: Journal of Computer Science and Technology Studies, 1(1), 38–53

Injadat, M., Moubayed, A., Bou Nassif, A., & Shami, A. (2021). Machine learning towards intelligent systems: Applications, challenges, and opportunities. Artificial Intelligence Review, 54(5), 3299–3348.

Judijanto, L., Hindarto, D., Wahjono, S. I., & Djunarto, A. (2023). Edge of enterprise architecture in addressing cyber security threats and business risks. International Journal of Software Engineering and Computer Science, 3(3), 386–396. https://doi.org/10.35870/ijsecs.v3i3.1816

Kuchar, K., & Fujdiak, R. (2025). Analyzing anomalies in industrial networks: A data-driven approach to enhance security in manufacturing processes. Computers & Security, 153, 104395. https://doi.org/10.1016/j.cose.2025.104395

Li, Y., & Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security: Emerging trends and recent developments. Energy Reports, 7, 8176–8186. https://doi.org/10.1016/j.egyr.2021.08.126

Little Lion Scientific. (2024). Enhancing malware detection efficacy: A comparative analysis of endpoint security

and application allowlisting. Journal of Theoretical and Applied Information Technology, 102(6). https://www.jatit.org/volumes/Vol102No6/18Vol102No6.pdf

Mallick, M. A. I., & Nath, R. (2024). Navigating the cybersecurity landscape: A comprehensive review of cyber-attacks, emerging trends, and recent developments. World Scientific News, 190(1), 1–69.

McGonigal, P. T. (2024). Identity-driven targeted violence: Attending to identity, emotion, and personality-related predictors of attitudinal and behavioral prejudice (Doctoral dissertation, University of Nebraska–Lincoln).

Nangi, P. R., Obannagari, C. K. R. N., & Settipi, S. (2023). A multi-layered zero-trust security framework. International Journal of Emerging Trends in Computer Science and Information Technology, 4(3), 144–153.

Ndibe, O. S. (2025). AI-driven forensic systems for real-time anomaly detection and threat mitigation. International Journal of Research Publication and Reviews, 6(5), 389–411.

Ndibe, O. S. (2025). Integrating machine learning with digital forensics to enhance anomaly detection.

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. Patterns, 5(5).

Pasa, L., Navarin, N., & Sperduti, A. (2022). Deep learning for graph-structured data. In Handbook on Computer Learning and Intelligence (pp. 585–617).

Pathade, C., Dhimam, V., Ahmad, S., & Lareb, I. (2026). Serverless AI security: Attack surface analysis and runtime protection mechanisms. arXiv.

Rosay, A., Carlier, F., Cheval, E., & Leroux, P. (2021). From CIC-IDS2017 to LYCOS-IDS2017: A corrected dataset for better performance. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (pp. 570–575).

Rudroff, T. (2024). Revealing the complexity of fatigue: Persistent challenges and promises of artificial intelligence. Brain Sciences, 14(2), 186.

Sarker, I. H. (2022). AI-based modeling: Techniques, applications and research issues. SN Computer Science, 3(2), 158.

Satyam, V. N., Mishra, D., & Mahapatra, B. G. (2025). AI-driven identity threat detection and response systems. International Journal of Emerging Research in Engineering and Technology, 6(4), 92–101.

Sharma, Y. (2025a). Identity threat detection and response (ITDR): The next big thing in cybersecurity. International Journal of Computer Science and Information Security, 23(3). https://doi.org/10.5281/zenodo.15381861

Sharma, Y. (2025b). Identity threat detection and response (ITDR): The next big thing in cybersecurity. International Journal of Computer Science and Information Security, 23(3). https://doi.org/10.5281/zenodo.15381862

Shonubi, J. A. (2025). Multi-layered zero trust architectures for cross-domain data protection. International Journal of Research Publication and Reviews, 6(7), 146–169.

Smiliotopoulos, C., Kambourakis, G., & Kolias, C. (2024). Detecting lateral movement: A systematic survey. Heliyon, 10(4), e26317. https://doi.org/10.1016/j.heliyon.2024.e26317

Soliman, M. M., Ahmed, E., Darwish, A., & Hassanien, A. E. (2024). Artificial intelligence powered metaverse: Analysis, challenges and future perspectives. Artificial Intelligence Review, 57(2), 36.

Spiekermann, D. (2025). Positional packet capture for anomaly detection in multitenant virtual networks. International Journal of Network Management, 35(2), e2326. https://doi.org/10.1002/nem.2326

Tucker, T., Bennett, N., Kotuliak, M., Erni, S., Capkun, S., Butler, K., & Traynor, P. (2025). Detecting IMSI-catchers by characterizing identity exposing messages in cellular traffic. In Proceedings of the ISOC Network and Distributed System Security Symposium (NDSS). https://doi.org/10.14722/ndss.2025.241115

Yang, F., Lin, C. D., Zhou, Y., & He, Y. (2023). Doubly coupled designs for computer experiments with both qualitative and quantitative factors. Statistica Sinica, 33(3), 1923–1942.

Yen, T.-F., Oprea, A., Onarlioglu, K., Leetham, T., Robertson, W., Juels, A., & Kirda, E. (2013). Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks. In Proceedings of the 29th Annual Computer Security Applications Conference (pp. 199–208). https://doi.org/10.1145/2523649.252367

Yusof, M. H. M., Almohammedi, A. A., Shepelev, V., & Ahmed, O. (2022). Visualizing realistic benchmarked IDS dataset: CIRA-CIC-DoHBrw-2020. IEEE Access, 10, 94624–94642.

Zoppi, T., Ceccarelli, A., & Bondavalli, A. (2021). Unsupervised algorithms to detect zero-day attacks. IEEE Access, 9, 90603–90615.