Anomaly Detection in Cybersecurity: A Comparative Study of Deep Learning Approaches

Sahil Adwani
School of Engineering and Computer Science
Victoria University of Wellington
Wellington, New Zealand
Email: adwanisahi@myvuw.ac.nz

***Abstract*—This paper presents a comprehensive comparative study of three deep learning approaches for anomaly detection in cybersecurity: Variational Autoencoder (VAE), GANomaly, and Deep Support Vector Data Description (SVDD). Using the realistic CIC-IDS2017 dataset containing diverse attack types, we evaluate detection performance, training stability, and practical deployment considerations. Our results demonstrate that VAE achieves the most balanced performance with F1-score of 0.629, while Deep SVDD exhibits superior discriminative power (ROC-AUC: 0.862). GANomaly shows the most significant improvement through threshold optimization, with recall increasing by 163% at the cost of precision. The study provides reproducible benchmarks and practical guidelines for security practitioners implementing deep learning-based anomaly detection systems.**

***Index Terms*—Anomaly Detection, Cybersecurity, Deep Learning, Variational Autoencoder (VAE), GANomaly, Deep Support Vector Data Description (SVDD), CIC-IDS2017**

I. Introduction

The escalating sophistication of cyber threats demands advanced detection systems that can identify novel and evolving attacks beyond the capabilities of traditional signature-based methods. Deep learning-based anomaly detection has emerged as a promising approach, capable of learning normal network behaviour and identifying deviations that may indicate security breaches [1]. However, the relative performance and practical suitability of different deep learning architectures for cybersecurity applications remains underexplored, particularly for structured tabular network data.

This study addresses this gap by conducting a systematic comparison of three prominent deep learning approaches: Variational Autoencoder (VAE), GANomaly, and Deep Support Vector Data Description (SVDD). These models represent distinct philosophical approaches to anomaly detection—probabilistic reconstruction, adversarial learning, and discriminative boundary learning respectively.

Our contributions include: a consistent evaluation framework for comparing deep anomaly detection methods, comprehensive performance analysis on the realistic CIC-IDS2017 dataset, practical insights into training stability and deployment considerations and reproducible benchmarks for cybersecurity practitioners.

II. Motivation

In cybersecurity, detecting unusual patterns such as sudden spikes in traffic or unauthorized access is critical for preventing intrusions. Traditional detection systems typically rely on predefined signatures or manually crafted rules, which often fail to identify novel or subtle attacks, especially in high-dimensional and complex datasets [2].

According to recent reports, the global economic impact of cyber-crime is projected to reach $10.5 trillion annually by 2025, highlighting the urgent need for robust and adaptive threat detection systems.

Deep learning methods offer a promising alternative. They automatically learn patterns from raw data, scale well with large and noisy inputs, and are more effective at identifying previously unseen or evolving threats.

This study is motivated by the need to evaluate and compare these deep learning-based approaches systematically in the context of modern cybersecurity challenges.

III. Background and Literature

Deep learning methods have shown significant promise in detecting anomalies across various complex domains due to their ability to automatically learn complex feature representations from high-dimensional data. This study focuses on three representative models:

A. Variational Autoencoders (VAE)

Variational Autoencoders combine probabilistic modeling with autoencoder architecture to learn compressed latent representations of normal data. Unlike standard autoencoders, VAEs learn a probabilistic encoding where each input is mapped to a distribution in latent space rather than a fixed point. This probabilistic approach enables better generalization and more robust anomaly detection. The model consists of an encoder that maps inputs to parameters of a variational distribution, and a decoder that reconstructs data from latent samples. Anomalies are detected based on either high reconstruction error (indicating the model cannot properly reconstruct the input) or low likelihood under the learned latent distribution [3].

B. GAN-based Approaches

GAN-based approaches like GANomaly leverage adversarial training through a unique encoder-decoder-encoder architecture. The model employs a generator that encodes input data

to latent space, decodes it to reconstruct the input, and then re-encodes the reconstruction. This dual-encoding process allows the model to compare the latent representations of original inputs and their reconstructions. During training, the generator learns to produce realistic reconstructions of normal data while a discriminator distinguishes between actual and generated latent representations. Anomaly detection occurs through the combination of reconstruction error and latent space discrepancy - anomalous inputs produce larger differences between their original and reconstructed latent codes [4].

C. Deep Support Vector Data Description (SVDD)

Deep Support Vector Data Description learns a compact hypersphere that encompasses normal data representations in a transformed feature space. The model trains a neural network to map input data into a latent space where normal instances cluster tightly around a learned center point. The optimization objective minimizes the volume of this hypersphere while ensuring most normal data points remain within its boundaries. During inference, the distance from the hypersphere center serves as the anomaly score, with samples falling outside the boundary classified as anomalies. This discriminative approach focuses on learning the decision boundary for normal instances rather than modeling their full distribution [5], [11].

Most existing studies have evaluated these models on image-based datasets, with limited evaluation on structured cybersecurity data [6], [7]. This project contributes to bridging this gap by evaluating and comparing these methods on modern network intrusion detection data (CIC-IDS2017).

These three models were selected because they represent distinct and complementary approaches to anomaly detection. All three models are also well-suited for structured, tabular data, which is the predominant format in cybersecurity anomaly detection including the datasets used in this project.

IV. Methodology

A. Dataset

We used the CIC-IDS2017 dataset, a modern benchmark containing realistic network traffic captured over a simulated enterprise environment. The dataset includes diverse attack types: DDoS, PortScan, Infiltration, Web Attacks, and others [9].

The original dataset has a combined total of 2,830,743 network flows (rows) and 79 features (columns). Each row represents a complete network flow, which is a bidirectional communication session between two endpoints. These flows are characterized by aggregated statistical features such as packet counts, byte rates, and timing information, providing a higher-level view of network activity compared to raw packet-level data.

B. Data Preprocessing

- **Data Cleaning**: Removed 308,381 duplicate records and handled missing values
- **Feature Engineering**: Dropped 8 zero-variance features and applied one-hot encoding
- **Label Processing**: Created binary classification (Normal=0, Attack=1) with 16.88% anomaly ratio
- **One-Class Training**: Used only normal samples for training (1,467,538 samples)
- **Train-Test Split**: 70% normal traffic for training, 30% mixed traffic for testing
- **Feature Scaling**: MinMaxScaler fitted exclusively on normal training data

C. Model Architectures

All models were implemented with consistent architecture and training parameters to ensure fair comparison.

Architecture Parameters:

- Latent dimension: 64
- Hidden layers: [256, 128] neurons
- Random seed: 42

Training configurations:

- Samples: 50,000 training, 25,000 test
- Epochs: 50, Batch size: 256
- Optimizer: Adam (lr=0.001)
- Early stopping: Patience=10, ReduceLROnPlateau
- Validation split: 20%

This balanced approach ensures that performance differences directly reflect each model's fundamental algorithmic approach rather than implementation variations hence providing reliable insights into their inherent strengths and limitations for cybersecurity applications. This also prevents any single model from gaining unfair advantages through specialized tuning and overall maintains computational efficiency.

All models are trained solely on normal data. In real-world cybersecurity scenarios, large amounts of benign (normal) traffic are available, whereas labeled attack data is often scarce, unlabeled, or constantly evolving. Training only on normal data allows the models to learn a baseline of expected behavior, making them capable of detecting novel or unseen attacks that deviate from this norm.

1) Variational Autoencoder: The VAE employs a probabilistic encoder-decoder architecture to learn the distribution of normal network traffic. The encoder compresses 70 input features through sequential layers ($256 \rightarrow 128$ neurons) into a 64-dimensional latent space, capturing essential patterns of normal behavior. The decoder then reconstructs the original input through a symmetric architecture ($128 \rightarrow 256$ neurons). The model optimizes a combined loss function of reconstruction error and KL divergence. During inference, anomalies are detected based on reconstruction error—unfamiliar attack patterns produce significantly higher errors than well-learned normal traffic.

2) GANomaly: GANomaly enhances reconstruction-based detection through adversarial training. The generator follows the same encoder-decoder structure as the VAE, while a separate discriminator network ($64 \rightarrow 256 \rightarrow 128 \rightarrow 1$) learns to distinguish between latent representations of real

and generated data. The adversarial training improves reconstruction quality and enables detection of subtle anomalies. The final anomaly score combines reconstruction error (50%) and latent space discrepancy (50%), where attacks manifest as inconsistencies between original and reconstructed encodings.

3) Deep SVDD: Deep SVDD takes a discriminative approach by learning a compact hypersphere that encloses normal data representations. The network architecture maps input features to a latent space where normal samples cluster around a learned center point. The optimization minimizes the mean squared distance of normal samples from this center. Anomaly detection directly uses the Euclidean distance from the center—network flows that deviate significantly from the established "normal" region are flagged as potential attacks.

D. Threshold Optimization

Following model training, we performed systematic threshold optimization to determine the optimal classification boundary. A detection threshold is a critical parameter that determines the classification boundary between normal and anomalous network flows. Each model computes an anomaly score, and this threshold defines the cutoff point above which a flow is flagged as an attack.

We evaluated five evenly distributed thresholds to identify the optimal value that balances the fundamental recall-precision trade-off inherent to anomaly detection. A higher threshold increases precision (fewer false alarms) but reduces recall (more missed attacks), while a lower threshold achieves the opposite. Our optimization targeted the optimal balance that maximizes the F1-score, providing the most practical operational point for cybersecurity deployment.

E. Evaluation Metrics

Models were evaluated using:

- **Accuracy**: Overall percentage of correct predictions (both normal and attack traffic classified correctly).
- **ROC-AUC**: Measures the model's overall ability to distinguish between normal traffic and attacks across all possible threshold settings. Values range from 0.5 (random guessing) to 1.0 (perfect separation).
- **Precision**: True Positives / (True Positives + False Positives). Answers: "When the model flags an attack, how often is it correct?" High precision indicates low false alarm rates.
- **Recall**: True Positives / (True Positives + False Negatives). Answers: "What percentage of actual attacks did the model successfully detect?" High recall indicates comprehensive threat detection.
- **F1-Score**: Harmonic mean of precision and recall. Provides a balanced measure that considers both false alarms and missed detections, ideal for imbalanced cybersecurity datasets.
- **PR-AUC**: Evaluates model performance specifically on the attack class, particularly important given the low anomaly ratio (16.88%) in our dataset.

TABLE I
COMPREHENSIVE PERFORMANCE COMPARISON ON CIC-IDS2017

Model	ROC-AUC	F1-Score	Precision	Recall	Accuracy
VAE	0.853	0.629	0.655	0.605	0.880
GANomaly	0.790	0.466	0.389	0.582	0.778
Deep SVDD	0.862	0.591	0.639	0.549	0.867

V. EXPERIMENTAL RESULTS

Table I presents the comprehensive performance comparison after threshold optimization.

A. VAE Performance Analysis

The VAE demonstrated the most balanced performance across all metrics. After threshold optimization, it achieved precision of 0.655 while maintaining the highest recall of 0.605, resulting in the best F1-score (0.629). The model showed excellent calibration between false positives and detection rate, making it suitable for operational environments where both are critical.

B. GANomaly Performance Analysis

GANomaly showed the most dramatic improvement through threshold optimization. Recall increased by 163% (from 0.221 to 0.582), though this came at the cost of precision dropping from 0.804 to 0.389. This significant trade-off highlights the sensitivity of GAN-based approaches to threshold selection. The adversarial training instability common in GAN architectures resulted in excessive false positives, rendering it less suitable for practical cybersecurity applications.

C. Deep SVDD Performance Analysis

Deep SVDD achieved the highest ROC-AUC (0.862), demonstrating superior discriminative power between normal and anomalous patterns. It maintained competitive precision (0.639) while achieving reasonable recall (0.549). The hypersphere learning approach proved effective for capturing normal behaviour boundaries.

TABLE II
THRESHOLD OPTIMIZATION EFFECTS

Model	Default Recall	Optimized Recall	Improvement
VAE	0.372	0.605	+63%
GANomaly	0.221	0.582	+163%
Deep SVDD	0.338	0.549	+62%

Table II shows the threshold optimization impact across models. GANomaly exhibited the highest sensitivity to threshold selection, with recall improving by 163% but at the cost of significant precision reduction. VAE maintained the best balance, achieving substantial recall improvement (+63%) while preserving reasonable precision. Deep SVDD demonstrated stable performance with balanced improvements.

Table III summarizes the training characteristics. VAE exhibited stable convergence with smooth training progression. GANomaly demonstrated challenging training dynamics, with

TABLE III
TRAINING CHARACTERISTICS AND EFFICIENCY

Model	Training Stability	Convergence	Best Epoch
VAE	High	Smooth	50
GANomaly	Medium	Volatile	38
Deep SVDD	Very High	Fast	50

early stopping triggered at epoch 38. Deep SVDD showed the most efficient convergence with very high training stability.

VI. DISCUSSION

A. Practical Implications for Cybersecurity

The comparative results provide clear guidance for security practitioners:

For Enterprise Security Operations: The Variational Autoencoder (VAE) emerges as the recommended choice due to its balanced performance and stable threshold characteristics. With precision of 0.655 and recall of 0.605, VAE minimizes both false alarms and missed detections, making it suitable for environments where security teams must efficiently triage alerts without being overwhelmed.

For High-Security Scenarios: GANomaly demonstrates the largest recall improvement potential (163% increase through threshold optimization), reaching 0.582 recall. This makes it appropriate for environments where detecting the maximum number of attacks is prioritized, even at the cost of higher false positive rates (precision: 0.389).

For Resource-Constrained Deployments: Deep SVDD offers compelling advantages with its efficient training convergence and strong discriminative power (ROC-AUC: 0.862). The model's computational efficiency and stable performance make it well-suited for large-scale monitoring applications with limited resources.

B. Threshold Optimization Criticality

Our analysis reveals that threshold selection is not merely a post-processing step but a critical determinant of operational effectiveness:

GAN-based Sensitivity: GANomaly exhibited the highest sensitivity to threshold settings, with dramatic trade-offs between recall and precision. This necessitates careful tuning and continuous monitoring in production environments.

VAE Robustness: The VAE demonstrated more consistent performance across threshold variations, maintaining reasonable precision even when optimizing for higher recall. This robustness makes it preferable for environments where frequent retuning is impractical.

Conservative Defaults: The commonly used statistical threshold (mean + 2 standard deviations) proved overly conservative across all models, yielding high precision (0.989) but poor recall (0.221-0.372). This highlights the inadequacy of default thresholds for cybersecurity applications where detection rate is paramount.

C. Architectural Considerations for Cybersecurity Data

A key consideration in this study is the direct transferability of methods from image processing. While the encoder-decoder paradigm is inspired by computer vision, the nature of the data is fundamentally different. Image data exhibits high spatial correlations and local continuity with very high dimensionality (e.g., 150,528 dimensions for 224×224×3 images), whereas cybersecurity flow data is tabular with features that are often independent or loosely correlated and significantly lower dimensionality (70 features). Therefore, we cannot directly carry over architectures designed for spatial correlations without significant adaptation. Our study uses dense fully-connected networks, which are agnostic to feature order and better suited for the tabular, heterogeneous nature of network flow data.

D. Towards Continuous Operational Deployment

The ultimate objective of this research is to develop systems that run continuously and autonomously in production environments. Our offline evaluation represents the essential first step to validate detection efficacy. A real-world deployment would involve initial training on historical normal traffic, continuous real-time inference on incoming flows, automated alerting for security analyst review, and periodic retraining to adapt to concept drift. While this study presents static evaluation results, the models are designed with the architectural simplicity and computational efficiency necessary for the goal of continuous, real-time network monitoring in operational cybersecurity environments.

E. Limitations and Future Work

While this study provides comprehensive insights into deep learning approaches for cybersecurity anomaly detection, several limitations highlight productive directions for future research:

Dataset Generalization: Our evaluation focused exclusively on the CIC-IDS2017 dataset. Future work should validate these findings across diverse network environments, organizational contexts, and emerging attack patterns to ensure broader applicability.

Operational Considerations: The computational overhead, real-time inference performance, and scalability in production environments warrant deeper investigation to assess practical deployment feasibility across different organizational scales.

Architectural Scope: This study focused on three specific architectures under controlled conditions. Promising research directions include: Ensemble Methods (combining VAE's balanced performance with Deep SVDD's computational efficiency), Temporal Modelling (incorporating time-series analysis for sequential network behaviour patterns), Transfer Learning (adapting pre-trained models across different network environments and organizations), Real-time Adaptation (developing mechanisms for continuous learning against evolving threats), and Expanded Evaluation (comprehensive testing on additional real-world datasets with varied attack profiles).

Advanced Methodologies: Future work should also explore adaptive thresholding mechanisms for dynamic network

environments, enhanced explainability frameworks to support security analyst decision-making, and automated model selection techniques based on organizational security policies and resource constraints.

VII. Ethical Considerations

This research addresses several important ethical considerations in cybersecurity anomaly detection:

Data Privacy and Anonymization: The study used exclusively public, pre-anonymized datasets (CIC-IDS2017). No private user data or sensitive organizational information was utilized, ensuring compliance with privacy regulations and ethical research standards.

Potential for Misuse: While anomaly detection technology serves legitimate security purposes, it could potentially be misused for mass surveillance or privacy invasion. We emphasize that these methods should be deployed with appropriate oversight and in compliance with legal frameworks.

Algorithmic Bias and Fairness: The models learn from historical network data, which may contain biases reflecting existing network usage patterns. Careful monitoring is required to ensure that detection systems do not disproportionately flag legitimate but unusual behavior from specific user groups or network segments.

Transparency and Accountability: The significant impact of threshold optimization on model performance, particularly for GANomaly, underscores the importance of transparent parameter selection and documentation in operational deployments.

Dual-Use Considerations: The techniques developed could potentially be adapted for both defensive and offensive cybersecurity applications. We strongly advocate for responsible research and deployment focused exclusively on defensive security applications.

VIII. Conclusion

This comprehensive comparative study demonstrates that deep learning approaches offer significant promise for cybersecurity anomaly detection, with each architecture exhibiting distinct characteristics regarding performance stability and threshold sensitivity.

The Variational Autoencoder (VAE) provides the most practical balance for operational environments, achieving the highest F1-score (0.629) and recall (0.605) while maintaining robust performance across different threshold settings.

Deep Support Vector Data Description (Deep SVDD) shows exceptional discriminative power (ROC-AUC: 0.862) and training efficiency, making it particularly suitable for resource-constrained environments.

GANomaly demonstrates the most significant performance improvement potential through threshold optimization, with recall increasing by 163%. However, this comes with substantial precision trade-offs, requiring careful consideration in deployment decisions.

These findings provide valuable guidance for security practitioners implementing deep learning-based detection systems in real-world environments.

Statement

All work in this project is solely my own. The tools and libraries used include:

- Programming Language: Python
- Framework and Libraries: PyTorch, Scikit-learn, Pandas, NumPy
- Datasets: CIC-IDS2017

The complete project source code will be made publicly available at: https://github.com/adwanis/AIML339-anomaly-detection

References

[1] M. L. Ali, K. Thakur, S. Schmeelk, J. Debello, and D. Dragos, "Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study," *Appl. Sci.*, vol. 15, no. 4, p. 1903, Feb. 2025, doi: 10.3390/app15041903.

[2] N. Mohamed, "Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms," *Knowl. Inf. Syst.*, vol. 67, pp. 6969–7055, 2025, doi: 10.1007/s10115-025-02429-y.

[3] A. S. Choudhary, "An Overview of Variational Autoencoders," *Analytics Vidhya*, Mar. 31, 2025. [Online]. Available: https://www.analyticsvidhya.com/blog/2023/07/an-overview-of-variational-autoencoders/

[4] W. Lim, K. S. C. Yong, B. T. Lau, and C. C. L. Tan, "Future of generative adversarial networks (GAN) for anomaly detection in network security: A review," *Computers & Security*, vol. 139, p. 103733, 2024. [Online]. Available: https://doi.org/10.1016/j.cose.2024.103733

[5] G. Yang, D. Xu, M. Wan, X. Liu, M. Gao, and H. Chen, "Deep support vector data description of anomaly detection with positive class edge outlier exposure and maximum double hypersphere interval," in *Proc. Int. Conf. Pattern Recognit. Image Anal. (PRIA)*, 2024, pp. 2. [Online]. Available: https://ieeexplore.ieee.org/document/10287379

[6] G. Nanos, "VAE vs. GAN for image generation," *Baeldung on Computer Science*, Mar. 18, 2024. [Online]. Available: https://www.baeldung.com/cs/vae-vs-gan-image-generation

[7] X. Zhao, X. Leng, L. Wang et al., "Efficient anomaly detection in tabular cybersecurity data using large language models," *Scientific Reports*, vol. 15, no. 3344, 2025. [Online]. Available: https://doi.org/10.1038/s41598-025-88050-z

[8] "What datasets are commonly used for anomaly detection research?" *Milvus*, Accessed: Jul. 30, 2025. [Online]. Available: https://milvus.io/ai-quick-reference/what-datasets-are-commonly-used-for-anomaly-detection-research

[9] "Intrusion Detection Evaluation Dataset (CIC-IDS2017)," *Canadian Institute for Cybersecurity*, University of New Brunswick. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html

[10] A. Mahboubi, K. Luong, H. Aboutorab, H. T. Bui, G. Jarrad, M. Bahutair, S. Camtepe, G. Pogrebna, E. Ahmed, B. Barry, and H. Gately, "Evolving techniques in cyber threat hunting: A systematic review," *J. Netw. Comput. Appl.*, vol. 232, p. 104004, 2024. [Online]. Available: https://doi.org/10.1016/j.jnca.2024.104004

[11] Z. Zhang and X. Deng, "Anomaly detection using improved deep SVDD model with data structure preservation," *Pattern Recognit. Lett.*, vol. 148, pp. 1–6, 2021. [Online]. Available: https://doi.org/10.1016/j.patrec.2021.04.020