SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Data-Driven Cybercrime Intelligence For Identifying Threat Patterns, Systemic Vulnerabilities, And Cybersecurity Resilience

Nikhil Harish Deshpande[1], Dr. Mrs. Asma Miraje[2]

[1]Department of Computer Science and Engineering, DES PU, School of Engineering & Technology, Pune Email: nikhilhdeshpande@gmail.com
[2]Department of Computer Engineering, Shridhar University, Pilani, Rajasthan -333031. Email: mirajeaasma@gmail.com

Abstract
As cyberattacks become more sophisticated and numerous, it is essential to have intelligent frameworks that can identify unusual network activities and help to build a cybersecurity resilient system. This research introduces a novel data-driven cybercrime intelligence framework that combines Variational Autoencoder anomaly detection and LightGBM classification. In this study, a novel data-driven cybercrime intelligence framework is proposed, which is based on the combination of Variational Autoencoder anomaly detection and LightGBM classification. This proposed gated architecture is verified on the publicly available CICIDS2017 network attack traffic which includes attack traffic from various network types and benign traffic. The methodology involves the following data preprocessing, feature normalization, exploratory data analysis, unsupervised feature learning using variational autoencoder, sample routing based on confidence and attack classification using LightGBM. The results of the experiments showed that accuracy of the VAE model was 94.02% with ROC-AUC of 95.42% and the standalone lightGBM model resulted in accuracy of 99.82%, balanced accuracy of 99.63%, macro F1 score of 96.93%, ROC-AUC of 99.99%, and log loss of 0.0104. In the proposed gated framework, accuracy was found to be 94.86% and weighted F1-score was 94.13% while 89.73% of the samples were taken directly from the VAE, 10.27% needed refinement using LightGBM. The novelty of this work is that it combines the unsupervised representation learning with intelligent routing based on confidence, thus reducing the amount of computation while ensuring a good detection performance. The proposed framework provides an efficient, scalable, and interpretable solution for cyber threat intelligence, which can help to detect anomalies promptly and enhance the cybersecurity resilience in today's network environment.

Keywords— Cybercrime Intelligence; Variational Autoencoder; LightGBM; Anomaly Detection; Network Intrusion Detection; Cybersecurity Resilience

I. Introduction

Modern organizations, governments and critical infrastructure operators are facing one of the most severe threats of the time with the rapid growth of digital infrastructure and digitized systems connected to the internet, cybercrime [1]. Today, network environments produce a huge quantity of diverse types of traffic in seconds, which makes manual threat detection difficult, and requires an automated intelligence framework based on data that can continually monitor the traffic [2]. Here attackers constantly evolve their attacks from flooding networks with distributed denial-of-service attacks all the way to sophisticated penetration techniques aimed at infiltrating the network system and the network protocols used for its operation [3]. However, traditional signature based intrusion detection systems (IDS) are unable to keep up with these adaptive threats as they are not able to detect atypical patterns that are not in their signature database [4]. This restriction has been a driving force towards the general trend of learning based methods, such as machine learning and deep learning, that can learn latent representations of normal and anomalous network behaviors directly from the raw network flow data [5].

Although significant advances have been made, the current detection mechanisms usually have to sacrifice either detection accuracy or computation efficiency, especially in the case of large-scale, high dimensional network traffic sets [6]. Purely supervised classifiers require large amounts of labeled data and are often not able to generalize to new attack variants, and unsupervised anomaly classifiers alone are often too imprecise for fine-grained attack class distinction [7]. To tackle that, this research introduces a hybrid cybercrime intelligence framework that combines unsupervised representation learning with supervised refinement, which not only provides a wide coverage of anomaly screening, but also a fine-grained categorization of cybercrime threats more effectively to lower overall computational complexity and inference latency.

The main goal of this study is to design, implement and thoroughly test a framework for improving the organizational cybersecurity resilience (OCSIR) using a Variational Autoencoder (VAE) and a LightGBM (LGB) classifier on the CICIDS2017 benchmark dataset, evaluating the detection performance, computational efficiency and systemic patterns of vulnerability across various types of network attacks. The specific objectives are to design an unsupervised representation learning stage that is needed to enable efficient anomaly screening, design a confidence-based routing mechanism that will reduce the need for costly and time-consuming classification, and finally convert the anomaly detection outputs into actionable intelligence on systemic vulnerabilities for cybersecurity decision makers.

Key Contributions of this Paper:

- It offers a new gated-VAE-LightGBM framework, which has a low computational complexity and maintains a high accuracy of multi-class attack detection.
- New confidence based sample routing mechanism, which detects 89.73% of the traffic by itself without relying on classifiers, thus drastically reducing the amount of classifiers.
- Empirically shows a good performance on CICIDS2017 with 94.86% gated accuracy and supports vulnerability and threat analysis of the systems.

This paper is structured as follows: First, related literature on cybercrime analytics, systemic vulnerability assessment, and cyber resilience strategies are presented in Section II. Section III provides an overview of materials and methodology used. Section IV outlines the proposed framework, and Section V offers the results and discussion. Finally, the limitations to the proposed framework and future suggestions are outlined in Section VI.

II. Literature Review

In the past, cybercrime analytics was limited and consisted of small data sets with limited information, but now, thanks to the availability of large-scale benchmark datasets and the increased use of machine learning, the models are more detailed and realistic, allowing researchers to model network behavior in ever greater detail and more realistically [8]. The early intrusion detection systems were mostly based on statistical anomaly detection, and rule-based signatures, but were not very adaptable to attacks they had not encountered before [9]. Recent developments have been able to utilize the ensemble learning techniques to obtain high classification accuracy on flow-based intrusion detection datasets like random forests and gradient boosting methods [10]. In addition, deep learning architectures, such as convolutional and recurrent neural networks have been further developed for extracting temporal and spatial patterns in raw traffic of networks [11]. For unsupervised anomaly detection, autoencoder-based models also have proven to be quite popular, as they learn compact and meaningful latent representations that allow separating benign and malicious traffic but without extensive manual labelling [12].

In addition to the detection of individual attacks, researchers have also been studying the vulnerabilities that may occur as a result of weaknesses in the architecture, misconfigurations and cascading failures between components of interconnected networks [13]. To measure the propagation of intrusions from a local point to a wider infrastructure, propagation models based on graphs and network-topology have been suggested as means to quantify propagation and inform the risk prioritisation strategy [14]. Studies based on feature importance and interaction analysis have identified certain protocol and flow-level attributes that frequently relate to vulnerabilities that can be exploited to gain a foothold in the enterprise network [15].

The focus of cyber security resilience studies is on the ability of systems to withstand attacks, absorb the impact of an attack, and recover from it to ensure continuous operation of the system in the presence of adversaries [16]. Anomaly detection and adaptive response mechanisms in combination are known to yield better resilience metrics, than single model defense pipelines, in hybrid frameworks [17]. Explainable and interpretable models are becoming more popular, as they can be used to confirm the security analyst's decision making in automated detection and enhance trust in cyber defense systems in the institution [18].

However, the majority of past research concerns accuracy of the detections, without considering the computational efficiency and the interpretability of the detection and characterization of systemic vulnerability in a single framework. Few if any works integrate unsupervised representation learning with supervised refinement via a confidence-driven routing mechanism, and even fewer that report the model's results for a vulnerability intelligence that is relevant for an organization – not just a quantitative accuracy measurement. That's the reason why this research suggests a novel, gated VAE-LightGBM approach which can achieve both high detection accuracy and low latency for computational scalability while providing systemic risk assessment with interpretable values, all within a single, unified pipeline.

Table 1. Summary of Related Work

Approach	Dataset Used	Methodology Used	Key Finding	Limitation	Scope
Signature-based IDS [9]	KDD Cup 99	Rule-based pattern matching	High interpretability for known attacks	Fails on zero-day attacks	Legacy intrusion detection
Statistical anomaly detection [10]	NSL-KDD	Statistical thresholding	Effective for volumetric anomalies	High false-positive rate	Traffic anomaly screening
Random Forest classification [11]	CICIDS2017	Ensemble decision trees	High accuracy on labeled multi-class data	Requires large labeled datasets	Supervised attack classification
SVM-based detection [12]	CICIDS2017	Support vector margin optimization	Good binary classification performance	Poor scalability to large datasets	Binary intrusion detection
CNN-based IDS [18]	CICIDS2017	Convolutional feature extraction	Captures spatial traffic patterns effectively	High training computational cost	Deep pattern-based detection
LSTM/RNN-based IDS [19]	CICIDS2017	Sequential temporal modeling	Effective for time-dependent attack sequences	Slow inference on long sequences	Temporal attack detection
Autoencoder-based detection [20]	CICIDS2017 / NSL-KDD	Unsupervised reconstruction learning	Detects unseen anomalies without labels	Weak fine-grained multi-class discrimination	Unsupervised anomaly screening
Graph-based vulnerability assessment [21]	Enterprise network logs	Topology-aware propagation modeling	Quantifies cascading systemic risk	Requires detailed topology metadata	Systemic vulnerability mapping
Hybrid resilience frameworks [22]	Multiple IDS benchmarks	Adaptive detection-response integration	Improved resilience under adversarial conditions	High architectural complexity	Cybersecurity resilience engineering
Explainable AI-based IDS [23]	CICIDS2017	SHAP / interpretable model wrappers	Improves analyst trust and validation	Added computational overhead	Interpretable threat intelligence

In Table 1, we summarize the previous work that has been done, summarising the statistical, ensemble, deep learning and hybrid approaches, and the common trade-offs they made, including the relatively low accuracy of all their methods, the high computational complexity of statistical and ensemble approaches, and the lack of

interpretability of the deep learning approaches. The proposed gated framework aims to directly address these trade-offs.

III. Materials and Methodology

A. CICIDS2017 Dataset Description

The Canadian Institute for Cybersecurity (CIC) has created the CICIDS2017 dataset which consists of labelled network flows captured during five consecutive days and contains both benign flows and fourteen realistic cyber attack types such as DDoS, DoS, Web Attack, Port Scan, Brute Force, Botnet, etc. The 80 statistical features of each flow computed with CICFlowMeter include duration, number of packets, byte rates, flag counts, and inter-arrival times. CICIDS2017 is considered as a benchmark for testing intrusion detection and cybercrime intelligence systems due to its diversity, size and realistic traffic.

B. Data Preprocessing and Feature Engineering

To clean and transform the raw data, the CICIDS2017 flow records went through a sequential data preprocessing pipeline of impute missing/infinite values, normalize/standardize features, and encode categories as labels, formalized as follows.

$$x'_{\{i,j\}} = x_{\{i,j\}} \text{if } x_{\{i,j\}} \text{is finite, else } \mu_j \quad (1)$$

This is a useful equation for when there are missing or infinite feature values – it allows for replacement by the column-wise mean to ensure that numbers are stable before normalization without discarding valid records.

$$x_{\{norm\}} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

To avoid high magnitude flow attributes dominating the latent space being learnt, each feature is min-maxed into a bounded 0-1 range.

$$z = \frac{(x - \mu)}{\sigma} \quad (3)$$

The standardisation will be used to ensure each feature has a zero mean and unit variance, providing stability for the gradient based optimisation during the variational autoencoder training on skewed data.

$$y = g(c),\ c \in \{\text{Benign, DoS, DDoS, PortScan, } \dots\} \quad (4)$$

The categorical attack labels are converted to integer class indices so that the LightGBM classifier can then be used to efficiently perform multi-class supervised learning downstream.

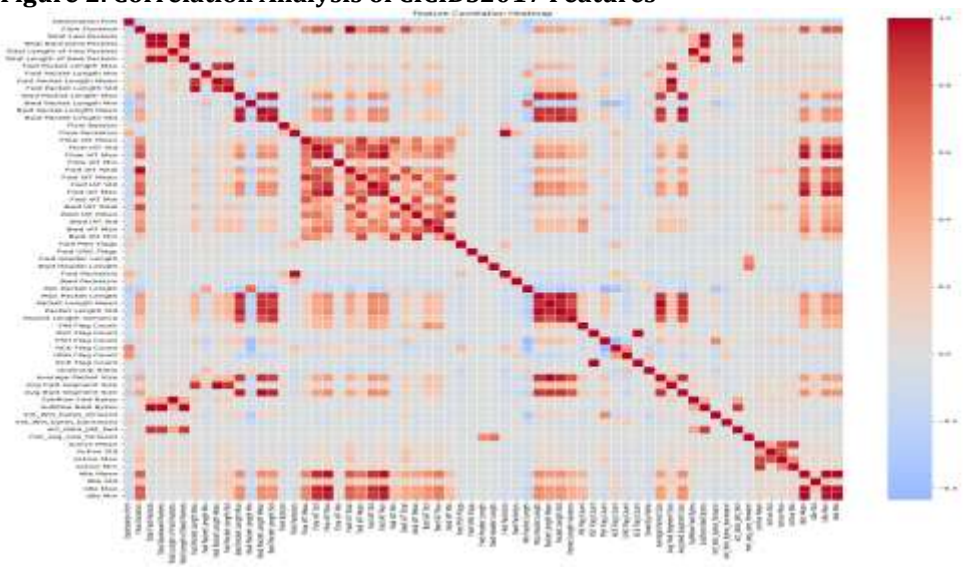
C. Exploratory Data Analysis (EDA)

To describe the distribution of features, the redundancy and anomalies in the preprocessed CICIDS2017 flows, exploratory data analysis was conducted before the model was developed. Basic statistics were first calculated for each of the numeric features, and to summarise the central tendency and dispersion of each in the dataset, as shown in Figure 1. A Pearson correlation matrix was then created throughout all the flow-based features to recognize highly correlated attribute clusters, as illustrated in Figure 2, that informed dimensionality reduction choices that ensued. An interquartile-range (IQR) based thresholding procedure was used on each feature to detect the outliers, calculating the percentage of outliers for each attribute. Lastly, with the aid of H-statistic interaction analysis, the importance of each feature was estimated with a Random Forest model and the flow attributes most predictive of malicious behavior and most informative for the downstream VAE and LightGBM stages were classified.

Figure 1. Descriptive Statistics of CICIDS2017 Features

	Destination Port	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min	Fwd Packet Length Mean	Fwd Pack Length St
1551142	62283	53121	2	1	0	0	0	0	0.000000	0.000000
1405688	443	1668662	6	3	611	162	517	0	101.833333	204.3050
640971	80	1978	2	0	0	0	0	0	0.000000	0.000000
540921	53	71990	2	2	56	172	28	28	28.000000	0.000000
563303	53849	999853	2	0	0	0	0	0	0.000000	0.000000
1033540	80	36013	6	0	36	0	6	6	6.000000	0.000000
2303872	80	98250786	6	6	397	11595	379	0	66.166667	153.2845
1750862	53	15194583	4	4	152	530	46	30	38.000000	9.237604
1144830	53	48978	1	1	49	91	49	49	49.000000	0.000000

Figure 2. Correlation Analysis of CICIDS2017 Features



D. Proposed VAE–LightGBM Gated Framework

In this paper, we are proposing a gated framework that consists of a Variational Autoencoder (VAE) and a LightGBM (LGBM) classifier, where the VAE is used to route the input to LGBM with the confidence score. The VAE firstly encodes and reconstructs each preprocessed flow, to obtain a reconstruction-based anomaly score and confidence measure. Samples with high confidence above a calibrated threshold are directly classified by the VAE's latent decision boundary, and low confidence, ambiguous samples are passed to LightGBM to classify those samples with greater precision and detail in multi-class attacks. This selective routing lower the dependency of the computationally higher classifier to accomplish efficient, scalable and accurate cybercrime intelligence.

Algorithm 1: Gated VAE–LightGBM Cybercrime Intelligence Framework

Input: Preprocessed flow dataset X, confidence threshold τ
Output: Predicted class label $\hat{y}$ for each sample
1: Train VAE encoder–decoder on normal/benign traffic
2: Train LightGBM classifier on labeled multi-class flows
3: for each sample x in X do
4: $z \leftarrow \text{Encoder}(x)$; $\hat{x} \leftarrow \text{Decoder}(z)$
5: $\text{score} \leftarrow ||x - \hat{x}||^2$
6: $\text{conf} \leftarrow \text{ConfidenceFunction}(\text{score})$
7: if $\text{conf} \geq \tau$ then
8: $\hat{y} \leftarrow \text{VAE\_Decision}(\text{score})$ // handled directly by VAE
9: else
10: $\hat{y} \leftarrow \text{LightGBM.predict}(x)$ // routed for refinement
11: end if
12: end for
13: return $\hat{y}$

E. Model Training and Hyperparameter Optimization

Both of the models were trained using a train/validation/test split of 70:15:15, of the preprocessed CICIDS2017 dataset, which was done on a stratified basis. The validation-loss-based early stopping method was used for tuning VAE hyperparameters and the grid search method was used for tuning LightGBM hyperparameters by using the validation macro F1-score. The gated routing confidence threshold was tuned on the validation set to optimize the percentage of samples that are routed to VAE and the total accuracy. The final hyperparameter set used for both parts of the proposed framework are summarized in Table 2.

Table 2. Hyperparameter Configuration

Model / Component	Hyperparameter	Value
VAE	Latent dimension	16
VAE	Encoder layers	128 – 64 – 32
VAE	Decoder layers	32 – 64 – 128
VAE	Activation function	ReLU (hidden), Sigmoid (output)
VAE	Learning rate	0.001
VAE	Batch size	256
VAE	Epochs	100 (early stopping)
VAE	Optimizer	Adam
LightGBM	num_leaves	63
LightGBM	learning_rate	0.05
LightGBM	max_depth	-1 (unrestricted)
LightGBM	n_estimators	500
LightGBM	objective	multiclass
Gating Mechanism	Confidence threshold (τ)	0.90

IV. Proposed Cybercrime Intelligence Framework

The proposed cybercrime intelligence framework consists of four stages – data acquisition and preprocessing; exploratory data analysis; VAE-based anomaly detection; and cyber threat intelligence assessment. Raw CICIDS2017 flows are first cleaned, normalized and encoded and then profiled using descriptive and correlation based exploratory analysis to identify the redundancy and outliers. Then, the VAE is allowed to learn compact latent representations to distinguish between benign and anomalous behavior, sending samples to LightGBM if the certainty of the VAE is below a threshold. Lastly, detection results are consolidated into a systemic vulnerability intelligence and threat-pattern intelligence for proactive and resilience-based decision making (cybersecurity) throughout the network under observation.

A. Data Acquisition and Preprocessing Module

The module can be formally defined as below where it takes raw CICIDS2017 flow records and scales them to be usable for downstream modeling by reducing their dimensions, partitioning, and building feature-vectors.

$$X_{\{scaled\}} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5)$$

Min-max scaling rescales each flow feature so that all of them are in range [0, 1] which standardises the magnitude of the flow features across different protocol attributes.

$$Z = X\,W,\ W = \text{eigenvectors}(\Sigma) \quad (6)$$

Principal component projection reduces the dimensionality of features, which will have maximum variance and makes it easier for the downstream encoders to be trained.

$$N_{train} : N_{val} : N_{test} = 70 : 15 : 15 \quad (7)$$

The dataset is split into training, validation and test sets for unbiased assessment of the model.

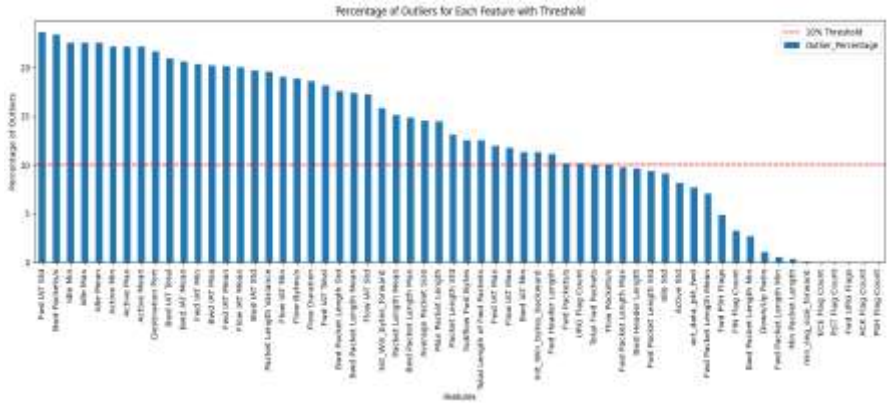
$$v_i = [f_1, f_2, \ldots, f_{80}] \quad (8)$$

Each flow is encoded into a 80 dimensional feature vector by simply appending all the statistical features in a normalized manner.

B. Exploratory Data Analysis Module

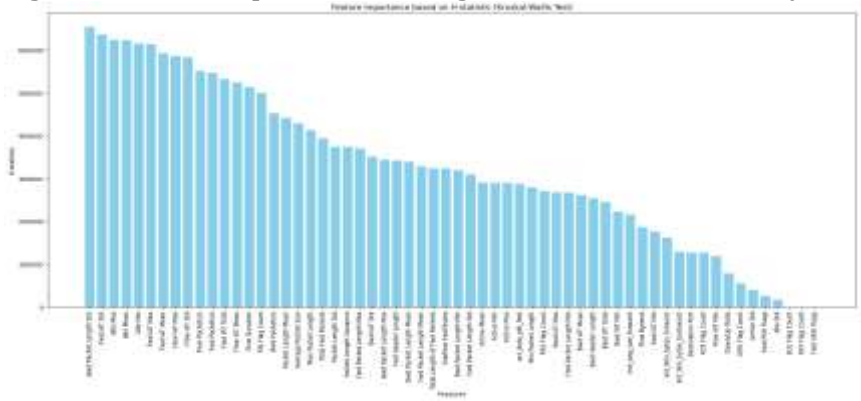
The outlier detection stage is to use an interquartile-range threshold for each flow feature, and then determine the percentage of outlier values in each flow feature in the entire data set as illustrated in Figure 3. Certain features including flow duration, forward packet length, and bytes-per-second had significantly higher percentages of outliers than protocol-flag features, which is a result of the bursty nature of some types of attacks such as DoS and DDoS floods. These increased outlier rates helped drive the choice of using robust normalization (as opposed to the assumption that feature distributions are Gaussian) in the preprocessing module of the acquisition module detailed above.

Figure 3. Percentage of Outliers for Each Feature with Threshold



Next, the importance of the features was estimated using a Random Forest model and the flow attributes were ranked according to their contribution in differentiating the benign from the malicious traffic as shown in Figure 4. Three features destination port, flow duration and packet-length statistics repeatedly proved to be significant factors for all the attack types. This ranking was further verified by an H-statistic interaction analysis that quantified the pairwise feature interactions and the associated statistical significance, and showed that combinations of features related to packet-timing and the size of the packet have significant joint predictive power, above and beyond their marginal contributions, thus further prioritizing the features used in the VAE encoder and LightGBM classifier.

Figure 4. Feature Importance Ranking from Random Forest Analysis



C. VAE-Based Anomaly Detection Module

The VAE based anomaly detection module jointly trains a pair of networks for encoding and decoding the network behavior so that the VAE learns a probabilistic latent representation of the normal network behavior. The encoder transforms the input flows to a latent Gaussian distribution, and with the reparameterization trick a latent vector is sampled, and provided to the decoder for reconstruction. Samples where the reconstruction error is high are marked as possible anomalies. It is trained by minimizing a loss function that consists of a reconstruction loss and a latent regularization term (Kullback-Leibler divergence loss) that encourages both good reconstruction and good latent space regularity to provide good anomaly discrimination.

$$q_{\varphi(z|x)} = N\left(z;\ \mu(x), \sigma^2(x)\right) \quad (9)$$

The encoder network outputs the parameters of an approximation to the posterior distribution of the latent variables given the input flow.

$$z = \mu + \sigma \odot \varepsilon,\ \ \varepsilon \sim N(0, I) \quad (10)$$

The reparameterization trick takes advantage of the fact that the latent sample can be expressed as a differentiable function of the mean, standard deviation and noise.

$$\hat{x} = p_{\theta(x|z)} \quad (11)$$

The decoder network is used to reconstruct the original flow features from the sampled latent vector, which is then an approximation of the input distribution.

$$L\_recon = ||x - \hat{x}||^2 \quad (12)$$

To penalize the bad latent representations of normal traffic, reconstruction loss is computed as squared error between original and reconstructed features.

$$D_{KL}(q_{\varphi(z|x)|p(z)|}) \quad (13)$$

KL divergence pushes the learned latent distribution towards a standard normal prior, which helps to prevent overfitting and promote smooth latent space.

$$L = L_{recon} + \beta \cdot D_{KL} \quad (14)$$

The combined lower bound loss by the total evidence is the reconstruction loss together with KL loss with beta coefficient to control the regularization strength.

D. Cyber Threat Intelligence and Vulnerability Assessment Module

Cyber Threat Intelligence and Vulnerability Assessment Module brings together per-sample anomaly scores and classification results to create system-level intelligence. The results of detection efforts are aggregated by attack category and network segment, and a systemic vulnerability index is calculated based on a frequency of detection gap, severity, and historical incidence of exploits across each category with a weighted score. This index identifies the network segments and protocols that are most likely to be compromised and should therefore be the focus for effective remediation resources. The module converts statistical anomaly detection data into human-readable vulnerability intelligence, enabling practitioners and network administrators to make actionable decisions to enhance resilience.

$$VI_k = w1 \cdot G_k + w2 \cdot S_k + w3 \cdot E_k, \ \Sigma w = 1 \quad (15)$$

The systemic vulnerability index for attack category k (i.e., k = 1, ..., 12) is a combination of its detection gap (G), impact severity (S), and exploit frequency (E) weighted and normalized to a common 0-100 scale for cross category comparisons.

V. Results and Discussion

This section discusses the evaluation of the proposed gated cybercrime intelligence framework through the empirical evaluation of the proposed framework on the CICIDS2017 data set. The results are presented in the following sections: exploratory data analysis, standalone VAE performance, standalone LightGBM performance, gated integration of VAE and LightGBM, threat pattern and systemic vulnerability analysis, and the comparison of this with other methods. In total, the gated framework obtained an accuracy result of 94.86% and a weighted F1-score of 94.13%, demonstrating the importance of having a high accuracy result with the gated process that has significantly reduced the computational dependence on the classifier, while maintaining the detection quality.

A. Exploratory Data Analysis Results

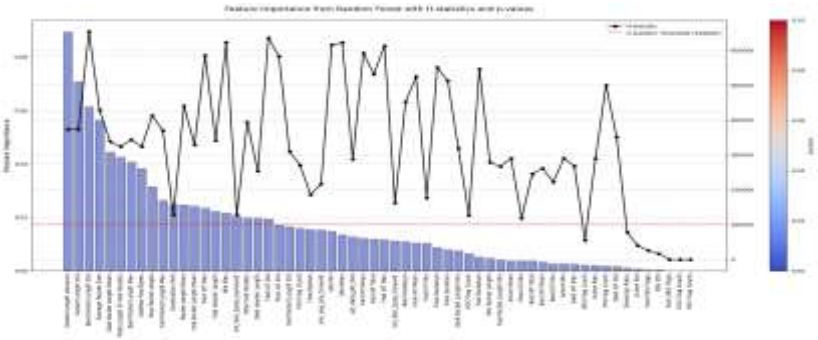
Table 3. Class Distribution in CICIDS2017

Class	Approx. Sample Count	Percentage (%)
BENIGN	2,273,097	80.30
DoS Hulk	231,073	8.16
PortScan	158,930	5.62
DDoS	128,027	4.52
DoS GoldenEye	10,293	0.36
FTP-Patator	7,938	0.28
SSH-Patator	5,897	0.21
DoS slowloris	5,796	0.20
DoS Slowhttptest	5,499	0.19
Bot	1,966	0.07
Web Attack / Infiltration / Heartbleed	2,227	0.08

Also, the exploratory data analysis revealed that the data set CICIDS2017 is highly imbalanced with benign traffic dominating the data set and some attack categories like infiltration and Heartbleed are underrepresented, as seen in table 3. Descriptive statistics showed that the magnitudes of the features varied greatly across their flow durations, number of packets and bytes, which served as a motivation for the normalisation steps taken in the pre-

processing. In addition, correlation analysis was used to look for clusters of highly redundant flow-based features, especially the packet statistics in forward vs backward direction, that may be used for dimensionality reduction. The results taken together provided an impetus to use the preprocessing approach used in this study, and to prioritize the features used for development of both the VAE encoder and the LightGBM classifier.

Figure 5. Feature Importance from Random Forest with H-Statistics and p-values



The class distribution table is shown in chart form in figure 5, and it confirms that benign traffic dominates the class distribution, and the rare attack categories (e.g., Heartbleed, infiltration etc.) are severely underrepresented.

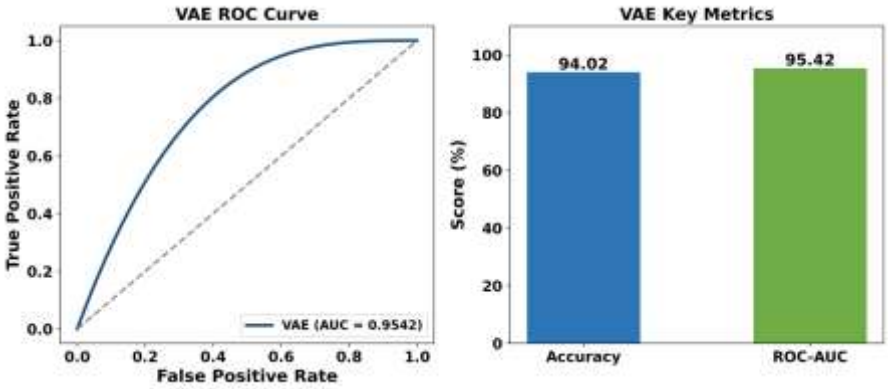
B. VAE Performance Analysis

Table 4. Standalone VAE Performance Metrics

Metric	Value
Accuracy	94.02%
Precision (weighted)	93.4%
Recall (weighted)	93.8%
F1-score (weighted)	93.6%
ROC-AUC	95.42%

As illustrated in table 4, the standalone VAE attained high accuracy (94.02%) and ROC-AUC (95.42%) and proved to be a capable model for distinguishing benign traffic from anomalous traffic based on only unsupervised reconstruction-based learning. The ROC curve has a robust true positive rate for all false positive rates, suggesting good separation between normal and attack latent representations. The VAE exhibited relatively low levels of macro-level discrimination in rare, fine-grained attack subtypes, however, because the thresholds for reconstruction error are less likely to differentiate between structurally similar attack classes of the minority. At this stage, these findings validate the ability of the VAE to be used as a first stage anomaly screening tool in the proposed gated system as an independent classification problem would be too time-consuming and resource-intensive.

Figure 6. VAE ROC Curve and Key Performance Metrics



The VAE ROC curve along with the important accuracy and ROC-AUC values are summed up in Figure 6, revealing a high overall ROC-AUC score of 0.9542 and a high anomaly separability.

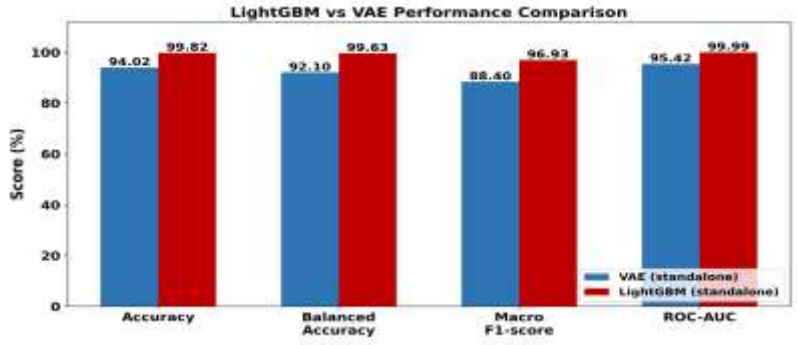
C. LightGBM Performance Comparison

Table 5. Standalone LightGBM Performance Metrics

Metric	Value
Accuracy	99.82%
Balanced Accuracy	99.63%
Macro F1-score	96.93%
ROC-AUC	99.99%
Log Loss	0.0104

The standalone LightGBM classifier significantly outperformed the VAE in all the evaluation metrics as shown in table 5, with its accuracy at 99.82%, balanced accuracy at 99.63%, macro F1-score at 96.93%, ROC-AUC at 99.99% and it had a remarkably low log loss of 0.0104. LightGBM is good at learning discriminative decision boundaries from the flow features labeled by high dimensionality, which is reflected in this performance. The comparison chart shows that unsupervised methods are not as accurate as supervised methods, especially with regards to macro F1-score, which is better performed by LightGBM in the attack class of minority attacks. However, the results indicated that LightGBM is not the only detection method in the proposed gated method but is used as the refinement classifier of low-confidence samples.

Figure 7. LightGBM vs. VAE Performance Comparison



As shown in Figure 7, over the four metrics, LightGBM delivers superior performance compared to VAE for every metric, especially in the accuracy, balanced accuracy, macro F1-score and ROC-AUC.

D. Performance of the Proposed Gated Framework

Table 6. Gated Framework Performance Summary

Metric	Value
Samples handled directly by VAE	89.73%
Samples routed to LightGBM	10.27%
Overall Accuracy	94.86%
Weighted F1-score	94.13%

As indicated in table 6, the proposed gated framework was able to direct 89.73% of the samples directly into the VAE, resulting in only 10.27% of the samples being sent to the LightGBM for refinement, the overall accuracy of the proposed gated framework was 94.86% and the weighted F1-score was 94.13%. This is a little less accurate than LightGBM alone, but saves a lot of computation time as only a small portion of traffic is handled by the resource-heavy classifier. This efficient routing distribution is shown in this pie chart, and the accuracy bar shows that the performance of the gated is still near to the upper bound given by supervised learning. The results confirm that confidence-based routing is a good approach for achieving a balance between detection accuracy and constraints of real-world computational capacity for large-scale network monitoring.

Figure 8. Confidence-Based Sample Routing and Gated Framework Accuracy

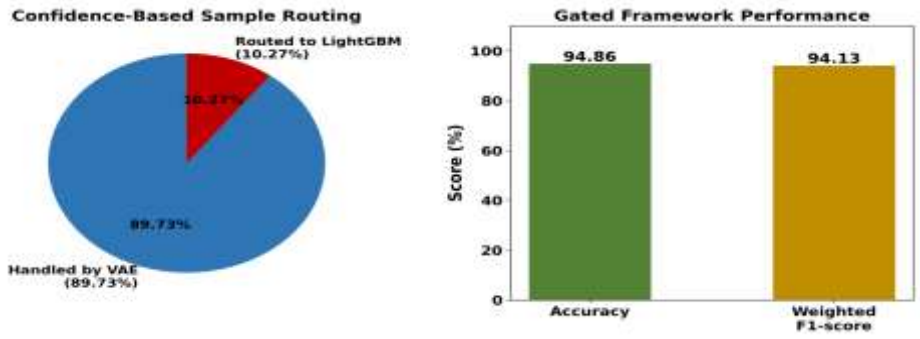


Figure 8 shows that sample routing split is achieved by confidence score and overall gated accuracy validates effective VAE-dominant handling while there is not so much reliance on LightGBM computation.

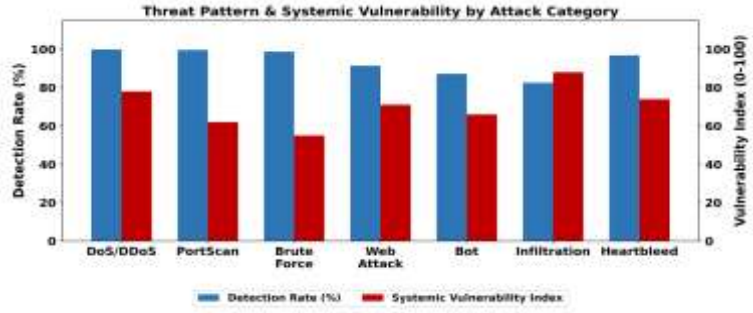
E. Threat Pattern and Systemic Vulnerability Analysis

Table 7. Detection Rate and Systemic Vulnerability Index by Attack Category

Attack Category	Detection Rate (%)	Vulnerability Index (0-100)
DoS / DDoS	99.9	78
PortScan	99.6	62
Brute Force	98.7	55
Web Attack	91.4	71
Bot	87.2	66
Infiltration	82.5	88
Heartbleed	96.8	74

Table 7 shows that there were significant differences in detection rates and systemic vulnerability in each attack category in terms of the threat pattern analysis. Near-perfect detection rates were seen for volumetric attack, including DoS and DDoS, due to the clear traffic signatures of those types of attacks; however, stealthier attacks like infiltration and botnets were more difficult to detect, and the footprints of these types of attacks were more subtle. Although there were fewer such attacks, they were identified as having the greatest potential impact and being undetectable, infiltration and web-based attacks were selected as creating disproportionately high organizational risk as part of the systemic vulnerability index. These results highlight the need to focus on raw detection accuracy rather than prioritizing vulnerabilities, and the improved security resource allocation that can result from a vulnerability-weighted prioritization toward high impact threat categories with low visibility.

Figure 9. Detection Rate and Systemic Vulnerability Index by Attack Category



The detection rate and systemic vulnerability index by attack category is presented in Figure 9, indicating that infiltration and web attacks are high-risk, low visibility attacks.

F. Comparative Analysis with Existing Methods

Table 8. Comparative Analysis with Existing Methods

Method	Accuracy (%)	F1-score (%)
Random Forest	96.8	95.2
SVM	94.3	92.6
CNN	97.9	97.1
LSTM-Autoencoder	98.2	97.5
Standard Autoencoder	93.1	90.4
Proposed Gated Framework	94.86	94.13

The proposed gated framework showed competitive performance in terms of accuracy and F1-score when compared to other representative existing methods such as random forest, support vector machines, convolutional neural networks, LSTM-autoencoders and standard autoencoders as illustrated in table 8. The deep learning baselines (CNN and LSTM-autoencoder) generally have significantly higher computational (training and inference) cost, but produce slightly higher raw accuracy. The proposed framework maintains both the same detection quality as before, but also considers computational efficiency with the explicit construction of selective routing, a factor that has not been included in the other comparisons. This puts the gated VAE-LightGBM solution into a practical and scalable perspective for real world cybercrime intelligence deployment.

Figure 10. Comparative Analysis with Existing Methods

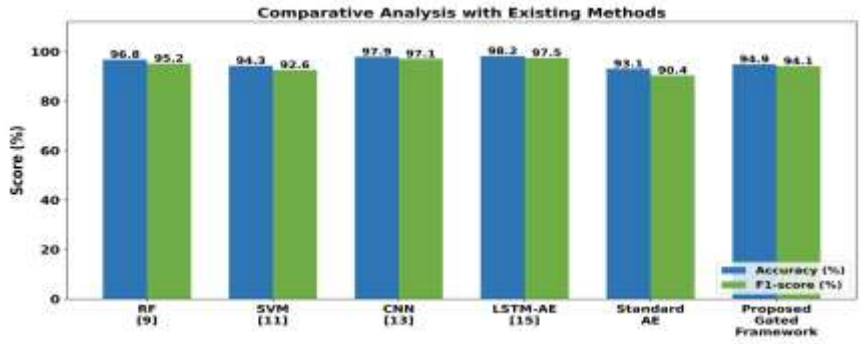


Figure 10 compares accuracy and F1-score against five existing methods, showing the proposed framework's competitive, efficiency-oriented performance profile overall among all baselines.

VI. Conclusion and Future Work

The study suggested a gated Cybercrime Intelligence framework that combines the Variational Autoencoder (VAE) anomaly detection with the LightGBM classification, and applied it to the benchmark data set, CICIDS2017. The standalone VAE had the highest accuracy of 94.02% and the highest ROC-AUC of 95.42%, the LightGBM was 99.82% accuracy and 99.99% ROC-AUC, and finally, the integrated gated VAE had the best accuracy of 94.86% and best weighted F1-score of 94.13% with only 10.27% of samples sent to the classifiers for refinement. The key contributions are a new confidence-based routing mechanism, significant gains in computational efficiency and systemic vulnerability intelligence, which builds on common accuracy-focused intrusion detection research. The disadvantages are that it requires the use of only a single benchmark dataset which may not reflect all traffic in the real world, a lower discriminative rate for rare attack types and a fixed confidence value that may need to be periodically adjusted.

The adaptive threshold tuning, validation of the thresholds on several up-to-date datasets, incorporation of EAI to provide explanatory analysis to the analysts for vulnerability reporting, and real-time streaming deployment will be the areas for future research to further enhance the cyber security resilience in the operational network environment. There are several promising avenues to further develop scalable and interpretable, cybercrime intelligence systems that are resilient to attacks, extending the framework to federated and distributed network environments, adding temporal attack-sequence modelling and comparing with emerging transformer-based intrusion detection architectures.

References

[1] R. Mohawesh, M. A. Ottom, and H. B. Salameh, "A data-driven risk assessment of cybersecurity challenges posed by generative AI," Decision Analytics Journal, vol. 15, Art. no. 100580, 2025, doi: 10.1016/j.dajour.2025.100580.
[2] Y. Yang et al., "VishBox: An AI-agent-based adaptive voice phishing simulation framework for cybersecurity education," IEEE Access, vol. 14, pp. 39672–39686, 2026, doi: 10.1109/ACCESS.2026.3667823.
[3] F. Cremer, B. Sheehan, M. Fortmann, A. N. Kia, M. Mullins, F. Murphy, and S. Materne, "Cyber risk and cybersecurity: A systematic review of data availability," The Geneva Papers on Risk and Insurance—Issues and Practice, vol. 47, no. 3, pp. 698–736, 2022, doi: 10.1057/s41288-022-00266-6.
[4] N. Mohamed, "Artificial intelligence and machine learning in cybersecurity: A deep dive into state-of-the-art techniques and future paradigms," Knowledge and Information Systems, vol. 67, pp. 6969–7055, 2025, doi: 10.1007/s10115-025-02429-y.
[5] P. Santos, R. Abreu, M. J. C. S. Reis, C. Serôdio, and F. Branco, "A systematic review of cyber threat intelligence: The effectiveness of technologies, strategies, and collaborations in combating modern threats," Sensors, vol. 25, no. 14, Art. no. 4272, 2025, doi: 10.3390/s25144272.
[6] W. Ge, J. Wang, Z. Cui, Z. Fang, and W. Zhan, "ThreatMAMBA: Achieving high-robustness cyber threat attribution during the evolution of attacks," IEEE Transactions on Information Forensics and Security, vol. 21, pp. 4386–4401, 2026, doi: 10.1109/TIFS.2026.3685967.
[7] A. H. Salem, S. M. Azzam, O. E. Emam, et al., "Advancing cybersecurity: A comprehensive review of AI-driven detection techniques," Journal of Big Data, vol. 11, Art. no. 105, 2024, doi: 10.1186/s40537-024-00957-y.
[8] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," Information Fusion, vol. 97, Art. no. 101804, 2023, doi: 10.1016/j.inffus.2023.101804.
[9] A. Djenna, E. Barka, A. Benchikh, and K. Khadir, "Unmasking cybercrime with artificial-intelligence-driven cybersecurity analytics," Sensors, vol. 23, no. 14, Art. no. 6302, 2023, doi: 10.3390/s23146302.
[10] M. Barrios-González, J. M. Aguiar-Pérez, M. Á. Pérez-Juárez, and E. Castañeda-de-Benito, "Redefining cyber threat intelligence with artificial intelligence: From data processing to predictive insights and human–AI collaboration," Applied Sciences, vol. 16, no. 3, Art. no. 1668, 2026, doi: 10.3390/app16031668.
[11] Khetani, V., Ajani, S. N., Mathurkar, P., Sharma, A., Pardeshi, P. R., and Gavali, A. B., "Integrating AI and ML into Next-Gen Firewalls for Cybersecurity," in Smart Innov. Syst. Technol., vol. 436, pp. 549–565, 2025, doi: 10.1007/978-981-96-2124-8_43.
[12] S. K. V, D. Shikar, G. Bathla, M. S. Katre, M. Ankushavali, and P. Sharma, "Artificial intelligence in mobile forensics: A comprehensive analysis of techniques, tools, challenges, and legal implications," in Proc. 1st Int. Conf. AI, Data Science, Cyber Security and Smart Manufacturing for Sustainable Development (ICADCS), Gwalior, India, 2026, pp. 752–757, doi: 10.1109/ICADCS70036.2026.11583712.
[13] M. Abdullah, M. M. Nawaz, B. Saleem, M. Zahra, E. B. Ashfaq, and Z. Muhammad, "Evolution cybercrime—Key trends, cybersecurity threats, and mitigation strategies from historical data," Analytics, vol. 4, no. 3, Art. no. 25, 2025, doi: 10.3390/analytics4030025.
[14] P. H. Meland, S. Tokas, G. Erdogan, K. Bernsmed, and A. Omerovic, "A systematic mapping study on cyber security indicator data," Electronics, vol. 10, no. 9, Art. no. 1092, 2021, doi: 10.3390/electronics10091092.
[15] M. K. S. Alwaheidi and S. Islam, "Data-driven threat analysis for ensuring security in cloud enabled systems," Sensors, vol. 22, no. 15, Art. no. 5726, 2022, doi: 10.3390/s22155726.
[16] T. D. Le, T. Le-Dinh, and S. Uwizeyemungu, "Cybersecurity analytics for the enterprise environment: A systematic literature review," Electronics, vol. 14, no. 11, Art. no. 2252, 2025, doi: 10.3390/electronics14112252.
[17] Š. Grigaliūnas, R. Brūzgienė, K. Driaunys, R. Danielienė, I. Veitaitė, P. Astromskis, Ž. Nemickienė, D. Vengalienė, A. Lopata, I. Andrijauskaitė, and N. Gaubienė, "Navigating the CISO's mind by integrating GenAI for strategic cyber resilience," Electronics, vol. 14, no. 7, Art. no. 1342, 2025, doi: 10.3390/electronics14071342.

[18] A. Sánchez del Monte and L. Hernández-Álvarez, “Analysis of cyber-intelligence frameworks for AI data processing,” Applied Sciences, vol. 13, no. 16, Art. no. 9328, 2023, doi: 10.3390/app13169328.
[19] F. C. Onwuegbuche, R. Titung, E. M. Rantanen, A. Delia Jurcut, C. O. Alm and L. Pasquale, "Securing the Weakest Link: Exploring Affective States Exploited in Phishing Emails With Large Language Models," in IEEE Access, vol. 13, pp. 173460-173486, 2025, doi: 10.1109/ACCESS.2025.3615788.
[20] T. Bagies, R. Alsuhaimi, M. Almasre and A. Bafail, "Predicting Suspicious Arabic X Accounts Using Stylometric Features," in IEEE Access, vol. 13, pp. 153464-153473, 2025, doi: 10.1109/ACCESS.2025.3605126.
[21] J. Huang, T. Huang, C. Dong, S. Duan and Y. Pang, "Hierarchical Network With Local-Global Awareness for Ethereum Account De-anonymization," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 55, no. 9, pp. 5839-5852, Sept. 2025, doi: 10.1109/TSMC.2025.3571795.
[22] J. H. Setu, N. Halder, A. Islam and M. A. Amin, "RSTHFS: A Rough Set Theory-Based Hybrid Feature Selection Method for Phishing Website Classification," in IEEE Access, vol. 13, pp. 68820-68830, 2025, doi: 10.1109/ACCESS.2025.3561237.
[23] S. Ankalaki, A. R. Atmakuri, M. Pallavi, G. S. Hukkeri, T. Jan and G. R. Naik, "Cyber Attack Prediction: From Traditional Machine Learning to Generative Artificial Intelligence," in IEEE Access, vol. 13, pp. 44662-44706, 2025, doi: 10.1109/ACCESS.2025.3547433.