Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 30, Sayı 1, 29-42, 2026

Süleyman Demirel University
Journal of Natural and Applied Sciences
Volume 30, Issue 1, 29-42, 2026

DOI: 10.19113/sdufenbed.1703191

Learning from the Normal: Anomaly-Based Intrusion Detection Using Isolation Forest, LOF, and One-Class SVM

Bashar ALHAJAHMAD[*1]

[1]Siirt Üniversitesi, Mühendislik Fakültesi, Bilgisayar Bölümü, 56000, Siirt, Türkiye

(Alınış / Received: 21.05.2025, Kabul / Accepted: 18.03.2026, Online Yayınlanma / Published Online: 24.04.2026)

Keywords
Unsupervised anomaly detection,
Intrusion detection systems,
One-Class SVM,
Local Outlier Factor,
Principal Component Analysis,
SHAP explainability.

Abstract: The increasing sophistication of cyberattacks necessitates intrusion detection systems capable of identifying previously unseen threats without relying on labeled attack data. This study presents a systematic cross-dataset evaluation of three classical unsupervised anomaly detection algorithms, namely Isolation Forest, Local Outlier Factor (LOF), and One-Class Support Vector Machine (OC-SVM), under a realistic novelty detection protocol. Models were trained exclusively on benign traffic and evaluated on mixed test sets using two widely adopted benchmark datasets, NSL-KDD and CIC-IDS2017. Performance was assessed using precision, recall, F1-score, confusion matrices, and computational runtime. Experimental results reveal dataset-dependent optimality. OC-SVM achieved the highest performance on NSL-KDD with an F1-score of 0.9948, driven primarily by error-rate and host-level aggregation features, while LOF outperformed other models on CIC-IDS2017 with an F1-score of 0.9890, effectively capturing timing irregularities and burst flow behavior characteristic of DDoS traffic. Principal Component Analysis (PCA) significantly reduced computational cost, achieving up to 65 to 70 percent reduction in training time for kernel-based models without degrading detection accuracy. SHAP-based explainability analysis demonstrated that model decisions align with statistically meaningful network behavior indicators rather than spurious correlations. The findings highlight that anomaly detection performance is strongly influenced by dataset structure and attack dynamics, and no single model universally dominates across heterogeneous intrusion scenarios. The proposed framework provides practical guidance for selecting lightweight, interpretable unsupervised detection models tailored to specific network environments.

Anomali Tabanlı Saldırı Tespiti için Normal Trafikten Öğrenme: Isolation Forest, LOF ve One-Class SVM Yaklaşımları

Anahtar Kelimeler
Gözetimsiz anomali tespiti,
Saldırı tespit sistemleri,
One-Class SVM,
Local Outlier Factor,
emel Bileşenler Analizi,
SHAP açıklanabilirliği.

Öz: Bu çalışma, ağ tabanlı saldırıların tespitine yönelik olarak yaygın biçimde kullanılan üç gözetimsiz anomali tespit algoritmasının — Isolation Forest, Yerel Aykırı Değer Faktörü (LOF) ve Tek Sınıf Destek Vektör Makinesi (One-Class SVM) — karşılaştırmalı bir analizini sunmaktadır. Çalışmada, kamuya açık bir siber güvenlik veri seti kullanılmış ve hesaplama maliyetini azaltmak ile model performansını artırmak amacıyla Temel Bileşenler Analizi (PCA) uygulanmıştır. Modeller yalnızca normal ağ trafiği verileriyle eğitilmiş, ardından hem normal hem de saldırı örneklerini içeren karışık veri üzerinde test edilmiştir. Performans değerlendirmesi, sınıflandırma başarımını ölçmek amacıyla doğruluk, geri çağırma, F1 puanı ve karışıklık matrisleri gibi temel metrikler kullanılarak gerçekleştirilmiştir. Elde edilen sonuçlar, One-Class SVM algoritmasının %99,06 geri çağırma oranı ve 0,8511 F1 puanı ile en yüksek genel performansı sağladığını, geniş yelpazedeki saldırı türlerini etkili şekilde tespit ederken kabul edilebilir düzeyde yanlış pozitif oranını koruduğunu göstermektedir. Isolation Forest algoritması yüksek doğruluk (%78,56) elde etmiş olmasına rağmen, düşük geri çağırma performansı nedeniyle yanlış pozitiflerin en aza indirilmesinin öncelikli olduğu senaryolarda daha uygun bar seçenek olarak değerlendirilmektedir. LOF algoritması ise görece yüksek yanlış alarm oranı nedeniyle daha dengeli fakat daha az sağlam bir performans sergilemiştir

*Corresponding author: bashar.aptech@gmail.com

1. Introduction

In recent years, the frequency and sophistication of cyber-attacks have increased significantly. Traditional rule-based security mechanisms that depend on predefined signatures and known attack patterns often fail to detect novel or previously unseen threats. This inherent limitation has prompted a growing shift toward Machine Learning (ML)-based anomaly detection methods, which can identify malicious behavior without relying on labeled attack data [1].

A central challenge in cybersecurity is the scarcity of the labeled datasets required for supervised learning. In real-world environments where unknown or emerging threats frequently arise, obtaining comprehensive labeled data is impractical. Consequently, unsupervised anomaly-detection methods have garnered attention. These methods operate by learning the statistical patterns of normal system behavior and subsequently identifying deviations from these patterns as potential anomalies [2].

Anomalies are data points that diverge from established behavioral norms and fail to align with the overall distribution of a dataset. Such irregularities may indicate fraudulent activities, network intrusions, or other forms of malicious conduct. Anomaly detection encompasses a wide array of methodologies, ranging from traditional statistical techniques to advanced AI-based models. The choice of detection strategy is often dictated by specific factors, such as the structure of the input data, computational efficiency, and the need for real-time detection [3].

Given the high dimensionality and complexity of network traffic data, dimensionality reduction techniques such as Principal Component Analysis (PCA) are essential to improve computational efficiency and mitigate overfitting. Among unsupervised ML approaches, three algorithms have gained particular prominence in cybersecurity owing to their effectiveness and efficiency: Isolation Forest, Local Outlier Factor (LOF), and One-Class Support Vector Machine (One-Class SVM).

Machine learning (ML) has emerged as a foundational component of modern cybersecurity, particularly in anomaly detection [24]. Unlike traditional rule-based security systems that rely on predefined signatures and struggle to detect zero-day exploits or previously unseen malware, ML approaches offer the ability to dynamically learn system behavior and identify deviations indicative of cyber threats. As highlighted in a recent review [7], ML has been applied across diverse cybersecurity domains, including malware detection, Industrial Control System (ICS) protection, and intrusion detection. The same study underscores the growing threat of adversarial attacks, in which malicious actors deliberately manipulate ML models to evade detection, further emphasizing the need for robust and adaptive anomaly detection mechanisms.

Despite the extensive body of research in intrusion detection, several practical challenges remain. First, many studies evaluate their models on a single benchmark dataset, most commonly NSL-KDD, without cross-dataset validation. Second, although deep learning and hybrid architectures often report high detection accuracy, they typically require large labeled datasets and substantial computational resources, limiting their applicability in real-time or resource-constrained environments. Third, systematic evaluation of dimensionality reduction techniques such as PCA—particularly in conjunction with lightweight unsupervised models—remains limited.

To address these gaps, this study conducts a structured comparative evaluation of three widely adopted lightweight unsupervised anomaly detection algorithms—Isolation Forest, LOF, and One-Class SVM—under a strict anomaly-detection paradigm in which models are trained exclusively on benign traffic.

Unlike many prior works, this research extends the evaluation beyond a single dataset by incorporating cross-dataset validation and computational efficiency analysis. In addition, SHAP-based explainability analysis is employed to interpret feature contributions, improving transparency and deployment relevance.

PCA is integrated as a label-free dimensionality reduction technique to enhance computational efficiency while preserving most of the statistical variance of the original feature space. The models are evaluated both with and without PCA to quantify its impact on detection performance and runtime efficiency.

By combining cross-dataset validation, dimensionality reduction analysis, runtime profiling, and explainability assessment, this study aims to provide a deployment-oriented comparison of lightweight unsupervised anomaly detection methods for modern intrusion detection systems.

2. Related Work

From a methodological and historical perspective, existing anomaly detection approaches can be broadly categorized into classical unsupervised machine learning methods, deep learning–based models, and hybrid or generative frameworks.

Among classical unsupervised machine learning approaches, tree-based, density-based, and boundary-based models have gained particular attention in cybersecurity due to their efficiency and ability to operate without labeled attack data.

Isolation Forest (iForest) isolates observations by recursively partitioning the dataset using randomly selected features and split values. Based on the assumption that anomalies are both rare and distinct, iForest efficiently identifies these outliers through shorter path lengths in the data tree. The algorithm offers linear time complexity and minimal memory usage, making it well suited for high-dimensional data and real-time intrusion detection tasks [4]. In addition, recent studies have highlighted the effectiveness of the Isolation Forest algorithm in detecting network anomalies without requiring labeled data, demonstrating high scalability and robustness across various cybersecurity datasets [23]. The LOF identifies anomalies by evaluating the local density deviation of each data point relative to its neighbors. A data point is flagged as an outlier if it resides in a region of significantly lower density than its surroundings. LOF's adaptive nature makes it particularly effective for detecting subtle anomalies in sparse or nonuniform datasets [5].

A One-Class Support Vector Machine (SVM), a variant of the traditional Support Vector Machine, is designed to model the boundary of normal data distributions in a high-dimensional space. It learns a decision function that encapsulates the majority of the data and classifies observations outside this boundary as anomalies. Because of its robustness in handling nonlinear and high-dimensional data, a One-Class SVM is widely used in various cybersecurity applications [6].
Beyond lightweight machine learning models, recent research has increasingly explored deep learning, generative, and hybrid AI-based approaches to enhance anomaly detection performance.

STEP-GAN, a GAN-based framework designed to train solely on normal data while generating synthetic anomalies, demonstrated strong detection capability on ICS and UNSW-NB15 datasets [8]. Similarly, AI-powered real-time anomaly detection systems have shown improvements in defending against zero-day vulnerabilities and insider threats [9].

Ensemble-based ML frameworks have also been proposed to strengthen detection performance while emphasizing feature engineering and privacy considerations [10]. Deep learning methods, particularly CNN-based intrusion detection systems, have demonstrated high detection accuracy but often suffer from computational complexity and latency issues that limit real-time applicability [11].

Comparative studies on IoT network anomaly detection have evaluated various ML models, including SVMs, ANNs, Decision Trees, Logistic Regression, and k-NN [12]. These studies reveal that neural networks can capture complex intrusion patterns but face scalability challenges in heterogeneous IoT environments.

Integrated anomaly detection systems for critical infrastructures such as substations have combined host-based and network-based detection strategies, successfully identifying DoS and MITM attacks [13]. Hybrid frameworks integrating ensemble and deep learning techniques have further improved detection rates but introduced increased architectural complexity [14].

Recent hybrid deep learning approaches such as AE-DTNN combine autoencoders and transformer networks, achieving near-perfect performance across multiple datasets including NSL-KDD and CSE-CIC-IDS2018 [15]. Similarly, convolutional autoencoder architectures have been deployed on embedded systems to detect previously unseen attacks while maintaining acceptable inference times [16].

Although these advanced models demonstrate strong performance, they often rely on labeled data, significant computational resources, or complex architectures. In contrast, lightweight unsupervised approaches remain attractive for practical deployment due to their interpretability, efficiency, and independence from labeled attack data.

Although numerous intrusion detection studies have been published over the past decade, a systematic quantification of dataset usage and reported performance metrics remains limited in many comparative works. To address this gap and to contextualize the present study within the broader IDS research landscape, a structured literature survey was conducted. Indexed publications from IEEE Xplore, Web of Science (WoS), and Scopus between 2014 and 2024 were examined using dataset-specific keywords (e.g., "NSL-KDD", "CIC-IDS2017", "UNSW-NB15", "CSE-CIC-IDS2018") in titles and abstracts.

Table 1 summarizes the approximate number of studies employing major benchmark datasets in intrusion detection research over the last decade. These estimates reflect dataset popularity trends and illustrate the evolution from legacy benchmarks toward more realistic modern datasets.

Table 1. Approximate adoption frequency of major IDS benchmark datasets in indexed publications (2014–2024). Counts are estimated from searches in IEEE Xplore, Web of

Science, and Scopus using dataset-specific keywords in intrusion detection context.

Dataset	Approx. No. of Studies	Typical Models Reported
NSL-KDD	~120	One-Class SVM, RF, CNN, Hybrid
CIC-IDS2017	~75	CNN, Autoencoders, SVM, RF
UNSW-NB15	~60	Deep learning, ensemble methods
CSE-CIC-IDS2018	~40	Hybrid deep & tree-based

Note: Counts are estimated from searches in IEEE Xplore, Web of Science, and Scopus (2014–2024).

Beyond dataset adoption frequency, understanding the typical performance levels achieved by different classification methods is essential for positioning new contributions. Therefore, representative intrusion detection studies published in indexed venues were examined to extract reported performance metrics. Table 2 presents a summary of selected studies, including the dataset used, the classification method applied, and the primary performance metric reported (Accuracy, F1-score, or Recall).

Table 2. Representative performance results reported in recent intrusion detection studies across major benchmark datasets. Metrics correspond to those reported by the original authors.

Reference	Dataset	Model	Reported Metric
[6]	NSL-KDD	One-Class SVM	High detection performance (reported accuracy > 98%)
[15]	NSL-KDD	AE-DTNN (Autoencoder–Dense–Transformer)	Near-perfect accuracy (> 99%)
[15]	CSE-CIC-IDS2018	AE-DTNN	Near-perfect accuracy (> 99%)
[16]	Embedded IDS traffic	Convolutional Autoencoder	High detection accuracy with improved inference time
[11]	NSL-KDD	CNN-based IDS	High classification accuracy (> 98%)
[12]	IoT intrusion datasets	SVM, ANN, DT, LR, k-NN	Competitive performance across multiple classifiers
[14]	Multiple cybersecurity datasets	Hybrid ML framework	Improved detection compared to rule-based systems
[23]	Various cybersecurity datasets	Isolation Forest	Scalable anomaly detection with robust performance
[26]	Intrusion detection datasets	SVM with PCA	Performance improvement after dimensionality reduction

This comparative overview highlights that while deep learning and hybrid architectures often report high accuracy values, lightweight unsupervised approaches remain competitive—particularly when evaluated under realistic anomaly-detection assumptions and without reliance on labeled attack data.

The survey results reveal three key observations. First, NSL-KDD remains the most widely used benchmark dataset in intrusion detection research, largely due to its historical adoption and reproducibility. However, more recent datasets such as CIC-IDS2017 and UNSW-NB15 have experienced growing adoption because they better reflect contemporary network traffic characteristics.

Second, although deep neural models—including CNNs, autoencoders, and hybrid transformer-based frameworks—often report near-perfect accuracy, these results are frequently obtained under supervised or semi-supervised settings with labeled attack data. In contrast, fully unsupervised anomaly detection settings typically report more moderate but realistic performance values.

Third, there is limited literature conducting unified side-by-side comparisons of classical unsupervised models—specifically Isolation Forest, LOF, and One-Class SVM—across multiple benchmark datasets while simultaneously evaluating dimensionality reduction effects, computational efficiency, and model interpretability. The present study addresses this gap through a structured cross-dataset experimental design incorporating both NSL-KDD and CIC-IDS2017,

evaluated with and without PCA transformation, alongside SHAP-based explainability analysis.

3. Materials and Methods

The proposed framework follows a structured and reproducible pipeline for unsupervised anomaly detection on network traffic datasets. The methodology was designed to ensure fair cross-dataset evaluation, computational transparency, and reproducibility. The overall workflow is illustrated in Figure 1.

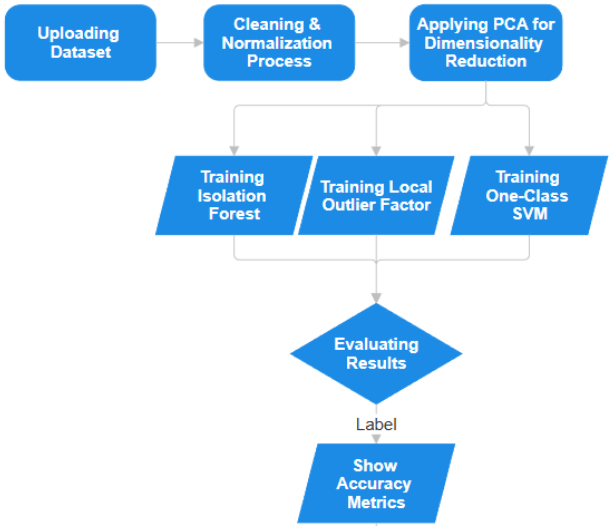

Figure 1. Overall experimental framework for cross-dataset unsupervised anomaly detection

The experimental pipeline consists of five main stages:

1. Dataset acquisition and preprocessing
2. Feature standardization
3. Optional dimensionality reduction using PCA
4. Model training on benign traffic only
5. Evaluation on mixed (benign + attack) test data

3.1. Datasets

To ensure robustness and generalizability, two publicly available benchmark datasets were used:

- NSL-KDD

The NSL-KDD dataset is a refined version of the KDD Cup 1999 benchmark and is widely used in intrusion detection research. It contains 41 numerical features describing network traffic behavior. The dataset includes multiple attack types (Neptune, Smurf, PortSweep, Back, GuessPassword, BufferOverflow, Rootkit, FTPWrite) in addition to normal traffic.

For the purpose of anomaly detection, the original multi-class labels were transformed into a binary representation. Specifically, normal traffic instances were encoded as **0**, whereas all attack categories were aggregated and encoded as **1**. This binary reformulation enables evaluation under a unified anomaly detection framework, where the task is to distinguish benign behavior from any form of malicious activity.

A total of 583,750 instances were used, with 80% allocated for training and 20% for testing.

- CIC-IDS2017

To address cross-dataset validation concerns, the CIC-IDS2017 dataset was additionally incorporated. This dataset reflects modern network traffic and includes realistic attack scenarios such as DDoS, brute force, and infiltration attacks. From the available traffic records, benign and DDoS samples were selected for binary anomaly detection evaluation.

Including CIC-IDS2017 enables evaluation of model generalization across legacy and modern intrusion datasets.

3.2. Data preprocessing

To ensure data integrity and model fairness, the following preprocessing steps were applied consistently across both datasets:

- Removal of missing or null-labeled samples
- Elimination of duplicate records
- Binary relabeling (normal vs. attack)
- Feature standardization using StandardScaler
- Optional dimensionality reduction using PCA

Standardization was performed prior to model training to ensure comparable feature scales.

3.3. Dimensionality reduction using PCA

High-dimensional network traffic data often contain redundant, correlated, and noisy attributes that increase computational burden and may introduce overfitting risks in anomaly detection models. To mitigate these issues while preserving the statistical structure of the data, Principal Component Analysis (PCA) was employed as a label-free dimensionality reduction technique [17,19].

For the NSL-KDD dataset, the original 41 numerical features were reduced to 10 principal components, preserving approximately 95.6% of the total variance. This level of variance retention indicates that the reduced feature space maintains the essential structural information required for reliable anomaly discrimination. For the CIC-IDS2017 dataset, PCA was applied optionally to evaluate performance–efficiency trade-offs and to analyze its impact on detection capability and computational scalability.

PCA was selected over label-dependent feature selection techniques such as minimum Redundancy Maximum Relevance (mRMR) or Mutual Information because the experimental framework strictly follows an unsupervised novelty detection protocol. Since class labels are excluded during training, label-driven feature ranking methods would violate the methodological assumptions of this study.

Although nonlinear dimensionality reduction techniques such as autoencoders, variational autoencoders, and transformer-based embeddings have demonstrated strong performance in intrusion detection research, their adoption introduces several methodological and operational trade-offs that conflict with the primary objective of this work, namely lightweight, interpretable, and computationally efficient unsupervised anomaly detection. Recent surveys and empirical studies have shown that deep learning–based intrusion detection systems, while powerful, often require substantial computational resources, extensive hyperparameter tuning, and complex architectural design [28, 29, 30].

Autoencoder-based feature reduction learns compressed latent representations through iterative neural network optimization. While such representations may capture nonlinear feature interactions, they require gradient-based training, architectural tuning, and careful hyperparameter configuration. In high-dimensional traffic datasets, autoencoder training may increase sensitivity to initialization, introduce additional optimization instability, and raise the risk of overfitting, particularly under fully unsupervised settings where attack labels are unavailable for validation [29].

Similarly, deep learning approaches based on convolutional neural networks, recurrent neural networks, and transformer-inspired architectures rely on hierarchical representation learning and frequently depend on GPU acceleration for efficient training. Although these methods can achieve high detection performance under supervised conditions, they significantly increase computational complexity, training time, and memory consumption, which may limit their suitability for real-time or resource-constrained deployment environments [28, 29].

In contrast, PCA offers several advantages aligned with the design principles of the proposed framework:

1. Label-Free Operation: PCA does not require class labels and is therefore fully compatible with strict unsupervised anomaly detection.
2. Deterministic and Closed-Form Nature: PCA is computed through eigenvalue decomposition of the covariance matrix and does not require iterative gradient optimization, improving reproducibility and stability.
3. Computational Efficiency: The closed-form solution of PCA introduces minimal additional training overhead compared to neural encoders.
4. Variance Preservation: Empirical results demonstrate that retaining ten principal components preserves approximately 95–96% of the variance in NSL-KDD, ensuring minimal information loss.
5. Model Compatibility: PCA improves scalability particularly for kernel-based models such as OC-SVM, where computational complexity increases with feature dimensionality.

The experimental findings confirm that PCA achieved substantial runtime reduction, up to 65–70% for kernel-based methods, without degrading detection performance across both datasets. These results are consistent with prior intrusion detection studies that emphasize the trade-off between representation complexity and computational feasibility in practical security systems [25,26,27].

Although nonlinear feature learning techniques may capture complex manifold structures that cannot be represented through linear projection, the objective of this study is not to maximize detection accuracy through deep architectural complexity. Rather, it is to evaluate lightweight, interpretable, and computationally feasible unsupervised anomaly detection mechanisms under realistic deployment constraints. Future research may explore hybrid strategies integrating nonlinear representation learning with density-based or boundary-based anomaly detection models. Nevertheless, within the scope of this study, PCA provides an effective balance between simplicity, scalability, interpretability, and detection reliability.

3.4. Anomaly detection algorithms

Three classical unsupervised algorithms were evaluated:

- Isolation Forest (iForest)

Isolation Forest isolates anomalies via random feature partitioning [4]. Anomalous samples require fewer splits to be isolated. The method has linear time complexity and low memory consumption, making it suitable for high-dimensional intrusion detection tasks [18].

- Local Outlier Factor (LOF)

LOF detects anomalies by measuring local density deviation relative to neighboring samples [5, 20]. It is particularly effective for detecting localized irregularities but sensitive to neighborhood parameter selection.

- One-Class Support Vector Machine (One-Class SVM)

One-Class SVM estimates a decision boundary enclosing normal samples [6, 21]. Samples outside this boundary are classified as anomalies. Using an RBF kernel allows nonlinear separation in high-dimensional feature spaces.

These algorithms represent tree-based, density-based, and boundary-based anomaly detection paradigms, enabling comprehensive comparative analysis.

3.5. Hyperparameter configuration

To ensure fairness and reproducibility, consistent hyperparameters were used across experiments. Final parameter settings are summarized in Table 3.

Table 3. Final hyperparameter settings

Model	Parameters
Isolation Forest	n_estimators = 100, contamination = 0.02
LOF	n_neighbors = 20, novelty = True
One-Class SVM	kernel = RBF, gamma = "scale", nu = 0.02

Hyperparameters were selected based on empirical validation and standard practice in prior literature. Automated tuning techniques such as Grid Search were not applied to preserve comparability across datasets.

3.6. Training and evaluation protocol

The anomaly detection framework follows a strict unsupervised paradigm:
Training Phase

- Models are trained exclusively on benign (normal) samples.
- Attack samples are completely excluded during training.

Testing Phase

- Evaluation is performed on mixed datasets containing both benign and attack samples.
    - Predictions are converted to binary anomaly labels.

This setup simulates realistic zero-day detection scenarios in which labeled attack data are unavailable during deployment.

3.7. Experimental pipeline

To ensure reproducibility, the complete procedure is summarized below in Figure 2.

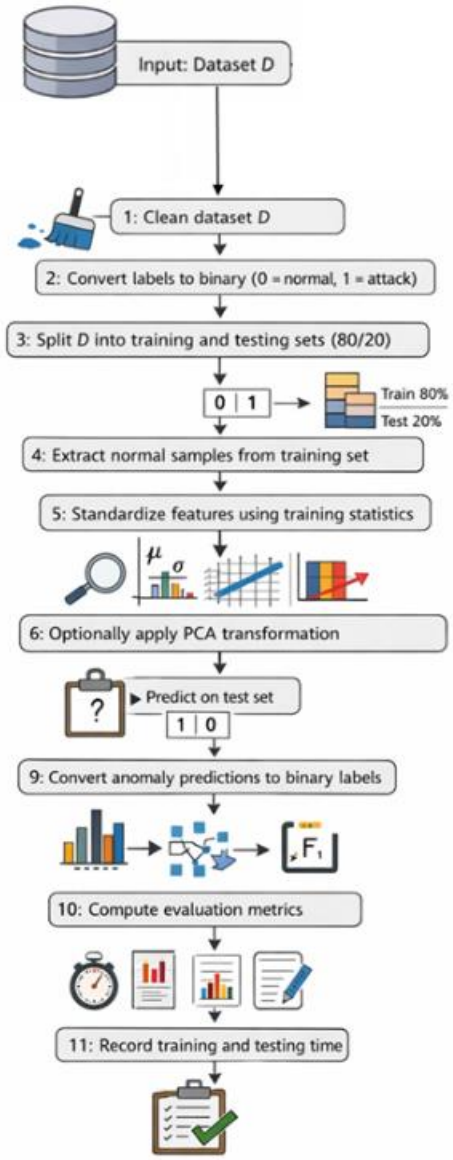

Figure 2. Detailed workflow of the proposed unsupervised anomaly detection pipeline

3.8. Evaluation metrics

The performance of the anomaly detection models was assessed using standard classification metrics appropriate for imbalanced datasets. Specifically, the following measures were computed:

- Precision, defined as the proportion of correctly identified attack instances among all instances predicted as attacks;
- Recall (Detection Rate), defined as the proportion of actual attack instances correctly identified by the model;
- F1-score, representing the harmonic mean of precision and recall, providing a balanced measure of detection effectiveness;
- Confusion Matrix, summarizing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN);
- Training Time (seconds), measuring the computational cost of model fitting;

- Testing Time (seconds), representing inference latency on unseen data.

Accuracy was not adopted as a primary performance indicator due to the class imbalance inherent in intrusion detection datasets, where normal traffic typically dominates attack samples. In such contexts, accuracy may produce misleadingly optimistic results. Therefore, greater emphasis was placed on precision, recall, and F1-score to provide a more reliable assessment of anomaly detection capability.

3.9. Computational environment

All experiments were implemented in Python 3.x using the Scikit-learn machine learning library [22]. Model training and evaluation were performed under a CPU-based execution environment without GPU acceleration. This setup reflects realistic deployment conditions in resource-constrained or production network monitoring systems, where specialized hardware may not be available.

Runtime measurements were systematically recorded for both training and testing phases in order to evaluate computational efficiency and to quantify the impact of dimensionality reduction via PCA. This analysis enables a practical comparison of performance trade-offs between detection accuracy and computational cost.

4. Results

This section presents the empirical findings obtained from applying the proposed unsupervised anomaly detection framework to two benchmark intrusion detection datasets: NSL-KDD and CIC-IDS2017. The evaluation focuses on comparative detection performance, computational efficiency, and model interpretability across algorithms and feature representations.

Three classical unsupervised models were analyzed: Isolation Forest, Local Outlier Factor, and One-Class Support Vector Machine. Each model was evaluated in both RAW and PCA-reduced feature spaces to quantify the impact of dimensionality reduction on detection accuracy and runtime efficiency.
Performance comparisons are structured dataset-wise and include quantitative metric analysis, confusion matrix interpretation, computational efficiency assessment, and SHAP-based feature attribution analysis. The F1-score is treated as the primary comparative metric due to its balanced consideration of precision and recall in imbalanced intrusion detection settings.

4.1. Results on the NSL-KDD dataset

The NSL-KDD dataset was used to evaluate the behavior of classical unsupervised anomaly detection models under a realistic novelty detection protocol. The dataset consists of 815,997 instances, including 576,710 benign samples and 239,287 attack samples. As described in Section 2, models were trained exclusively on normal traffic and evaluated on mixed test data to assess true anomaly detection capability.

4.1.1. Detection performance analysis

To examine algorithmic differences and the effect of dimensionality reduction, each model was evaluated in both RAW and PCA-reduced feature spaces. The quantitative results are summarized in Table 4.

Table 4. Performance comparison on NSL-KDD under RAW and PCA feature representations

Model	PCA	Precision	Recall	F1	Train Time (s)	Test Time (s)
Isolation Forest	No	0.9902	0.9925	0.9914	2.29	0.90
LOF	No	0.8978	0.1528	0.2611	219.13	167.81
One-Class SVM	No	0.9902	0.9994	0.9948	1482.21	319.71
Isolation Forest	Yes	0.9902	0.9926	0.9914	2.33	0.84
LOF	Yes	0.8837	0.1522	0.2597	43.40	23.48
One-Class SVM	Yes	0.9902	0.9991	0.9946	519.78	143.84

The results indicate that the One-Class SVM (RAW configuration) achieved the highest overall performance, with an F1-score of 0.9948 and recall of 0.9994. This near-complete detection of attack samples highlights the effectiveness of boundary-based novelty detection in the NSL-KDD feature space. Isolation Forest demonstrated similarly strong detection performance (F1 = 0.9914) while maintaining substantially lower computational cost. Its training time remained approximately 2.3 seconds, several orders of magnitude lower than that of the OC-SVM, confirming its suitability for large-scale or real-time environments.
In contrast, LOF exhibited poor sensitivity to attack samples, with recall values near 0.15 across both configurations. This suggests that density-based modeling is less effective for the statistical structure of NSL-KDD, where attack samples do not form clearly separable density clusters relative to benign traffic.

4.1.2. Impact of PCA on scalability

PCA did not significantly alter detection performance for Isolation Forest or OC-SVM. The F1-score of OC-SVM remained nearly unchanged (0.9948 to 0.9946),

indicating that the first ten principal components preserved sufficient variance for accurate anomaly detection.
However, PCA substantially improved computational efficiency:

- OC-SVM training time decreased by approximately 65 percent
- Testing time was reduced by more than 50 percent
- LOF training time decreased from 219.13 seconds to 43.40 seconds

These results confirm that dimensionality reduction enhances scalability without compromising detection accuracy.

4.1.3. Confusion matrix and error analysis

To further examine classification behavior, confusion matrices for the RAW configurations are presented in Table 5.

Table 5. Confusion matrices for NSL-KDD (RAW feature space)

Model	TN	FP	FN	TP
Isolation Forest	112,997	2,345	1,797	237,490
LOF	111,179	4,163	202,733	36,554
One-Class SVM	112,982	2,360	139	239,148

The OC-SVM produced only 139 false negatives, demonstrating exceptional detection reliability. From a security perspective, minimizing false negatives is critical, as missed attacks may lead to severe consequences.
Isolation Forest generated 1,797 false negatives, indicating slightly lower sensitivity but still strong detection capability.
LOF produced 202,733 false negatives, confirming its limited effectiveness for NSL-KDD.
These findings reinforce the superiority of OC-SVM for this dataset and validate the conclusions drawn from the F1-score analysis.

4.1.4. Explainability analysis using SHAP

To enhance interpretability, SHAP analysis was conducted on the best-performing configuration, namely the OC-SVM in RAW feature space (see Figure 3).

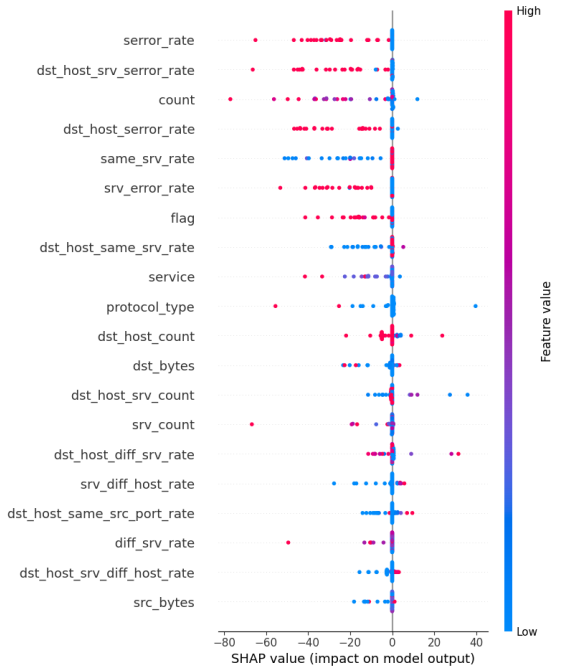

Figure 3. SHAP summary plot for NSL-KDD (OC-SVM – RAW configuration)

The SHAP summary plot visualizes global feature influence on anomaly decisions. As shown in Table 6, features are ordered by mean absolute SHAP value, indicating their overall contribution to the decision function.

Table 6. Top 10 most influential features for NSL-KDD (OC-SVM – RAW**)**

Rank	Feature	Mean Absolute SHAP
1	serror_rate	13.58
2	dst_host_srv_serror_rate	12.59
3	count	12.30
4	dst_host_serror_rate	11.44
5	same_srv_rate	11.38
6	srv_error_rate	10.11
7	flag	6.90
8	dst_host_same_srv_rate	5.14
9	service	4.32
10	protocol_type	3.95

The SHAP results demonstrate that detection is primarily driven by:

1. Error-rate features such as serror_rate and srv_error_rate
2. Host-level aggregation metrics
3. Connection frequency and service distribution attributes

This indicates that attack detection in NSL-KDD is largely governed by abnormal connection failure behavior and aggregated host traffic irregularities. These patterns are characteristic of Denial-of-Service and probing attacks within the dataset.

Importantly, the model relies on statistically meaningful traffic indicators rather than arbitrary correlations, reinforcing the transparency and operational trustworthiness of the proposed framework.

4.2. Results on the CIC-IDS2017 dataset

4.2.1. Dataset characteristics

The second experimental evaluation was conducted on the CIC-IDS2017 dataset to assess the robustness and generalizability of the proposed unsupervised anomaly detection framework under modern, high-volume traffic conditions.

The analyzed subset consists of 225,745 total samples, including 97,718 benign instances and 128,027 DDoS attack samples. As in the NSL-KDD experiments, model training was performed exclusively on benign traffic samples. During testing, models were evaluated on a mixed dataset containing both benign and DDoS traffic.

Unlike NSL-KDD, which includes heterogeneous attack types, this subset represents a large-scale volumetric DDoS scenario, enabling evaluation under burst-driven traffic behavior.

4.2.2. Detection performance analysis

The performance of each algorithm under RAW and PCA feature representations is summarized in Table 7.

Table 7. Performance comparison on CIC-IDS2017 under RAW and PCA feature representations

Model	PCA	Precision	Recall	F1	Train Time (s)	Test Time (s)
Isolation Forest	No	0.9754	0.1120	0.2009	0.64	0.45
LOF	No	0.9818	0.9962	0.9890	7.97	15.56
One-Class SVM	No	0.9954	0.6359	0.7761	176.39	27.33
Isolation Forest	Yes	0.9953	0.6169	0.7617	0.62	0.42
LOF	Yes	0.9830	0.9523	0.9674	2.01	6.69
One-Class SVM	Yes	0.9954	0.6351	0.7755	53.95	10.17

The performance trends observed in CIC-IDS2017 differ substantially from those in NSL-KDD.
The Local Outlier Factor in RAW feature space achieved the highest performance, with recall of 0.9962 and F1-score of 0.9890. This indicates that density-based modeling effectively captures the structural clustering of DDoS traffic.
Isolation Forest performed poorly in RAW configuration (recall = 0.1120), but improved substantially after PCA transformation, where recall increased to 0.6169 and F1-score rose from 0.2009 to 0.7617.

The One-Class SVM exhibited stable performance across both configurations, with F1 around 0.776, though recall remained lower than LOF.

4.2.3. Impact of PCA and computational efficiency

PCA significantly improved computational scalability in CIC-IDS2017.
For the One-Class SVM:

- Training time decreased from 176.39 seconds to 53.95 seconds
- Testing time decreased from 27.33 seconds to 10.17 seconds

LOF also benefited from dimensionality reduction, with testing time decreasing from 15.56 seconds to 6.69 seconds.
These results confirm that PCA enhances scalability in high-dimensional flow-based traffic data without substantially degrading detection performance.

4.2.4. Explainability analysis using SHAP

To interpret the decision behavior of the best-performing configuration, SHAP analysis was conducted on the LOF model in RAW feature space.

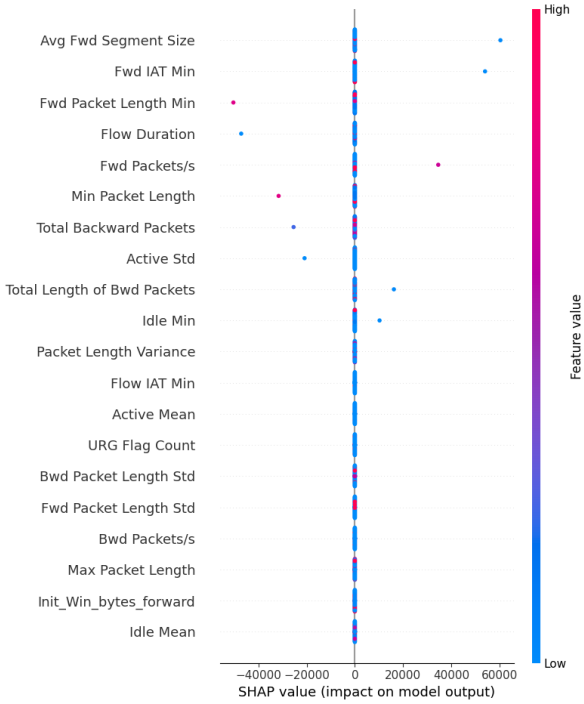

Figure 4. SHAP summary plot for CIC-IDS2017 (LOF – RAW configuration)

The SHAP summary plot ranks features according to mean absolute contribution to the anomaly score (see Table 8).

Table 8. Top 10 most influential features for CIC-IDS2017 (LOF – RAW)

Rank	Feature	Mean Absolute SHAP
1	Avg Fwd Segment Size	1210.08
2	Fwd IAT Min	1081.14
3	Fwd Packet Length Min	1009.33
4	Flow Duration	944.61
5	Fwd Packets/s	691.72
6	Min Packet Length	635.58
7	Total Backward Packets	510.17
8	Active Std	417.36
9	Total Length of Bwd Packets	325.63
10	Idle Min	206.31

The SHAP results reveal a fundamentally different feature influence structure compared to NSL-KDD. Detection in CIC-IDS2017 is dominated by:

1. Timing-related features such as Fwd IAT Min, Flow Duration, and Idle Min
2. Packet size statistics such as Avg Fwd Segment Size and Fwd Packet Length Min
3. Flow intensity metrics such as Fwd Packets/s

These features reflect the intrinsic dynamics of DDoS attacks, which generate extremely short inter-arrival times, high packet rates, burst behavior, and abnormal flow duration distributions.

In contrast to NSL-KDD, which is driven by connection error metrics, CIC-IDS2017 detection is governed by burst intensity and timing irregularities. This explains why density-based modeling is particularly effective for this dataset.

4.3. Cross-Dataset comparative analysis

To synthesize the experimental findings across both benchmark datasets, the best-performing configuration for each dataset is summarized in Table 9.

Table 9. Best-performing configurations across datasets

Dataset	Best Model	F1-score	Dominant Feature Type
NSL-KDD	One-Class SVM	0.9948	Error-rate and host-level metrics
CIC-IDS2017	LOF	0.9890	Timing and flow-based features

The results reveal a critical observation: the optimal anomaly detection algorithm is dataset-dependent. For NSL-KDD, the One-Class SVM achieved the highest F1-score (0.9948) and the lowest false negative rate.

SHAP analysis demonstrated that detection was dominated by connection error rates, host-level aggregation statistics, and service distribution irregularities. These characteristics form relatively well-defined nonlinear decision boundaries in feature space, which are effectively captured by the RBF kernel of the OC-SVM. Consequently, boundary-based novelty detection proved most suitable for this dataset.

In contrast, CIC-IDS2017 exhibited markedly different statistical behavior. The Local Outlier Factor achieved the highest F1-score (0.9890) with near-perfect recall. SHAP analysis indicated that detection was primarily influenced by timing-related features, flow duration, packet size statistics, and burst traffic dynamics. DDoS traffic in CIC-IDS2017 forms dense clusters clearly separable from benign flows, making density-based modeling more effective than boundary-based approaches.

This divergence confirms that anomaly detection performance is strongly influenced by dataset structure and attack dynamics. The framework demonstrates adaptability rather than overfitting, as different algorithms emerged as optimal under different statistical regimes. Furthermore, PCA consistently reduced runtime across both datasets without degrading detection performance, reinforcing its utility as a scalable preprocessing strategy.

4.4. Computational efficiency analysis

Beyond predictive accuracy, computational efficiency is critical for operational intrusion detection systems. Dimensionality reduction via PCA significantly improved scalability in both datasets. In NSL-KDD, OC-SVM training time decreased by approximately 65%, with testing time reduced by more than half. In CIC-IDS2017, OC-SVM training time decreased by approximately 69%, while LOF testing time was reduced by nearly 75%. These reductions demonstrate that PCA effectively mitigates the computational burden associated with high-dimensional feature spaces, particularly for kernel-based methods.

Algorithm-specific observations further highlight efficiency differences. Isolation Forest maintained minimal runtime across all configurations, consistent with its linear complexity and suitability for large-scale environments. LOF benefited substantially from PCA in CIC-IDS2017, suggesting that dimensionality reduction stabilizes local density estimation in high-dimensional spaces. OC-SVM showed the greatest sensitivity to dimensionality, with PCA dramatically improving training scalability.

These findings emphasize that performance evaluation must jointly consider detection capability and computational feasibility, particularly for real-time deployment scenarios.

4.5. Practical implications for deployment

The experimental findings provide actionable guidance for real-world cybersecurity deployment.
In high-security environments where false negatives must be minimized, the boundary-based OC-SVM (as observed in NSL-KDD) is preferable due to its extremely high recall and minimal miss rate.

In DDoS-intensive network environments characterized by burst-based flow behavior, density-based modeling through LOF (as observed in CIC-IDS2017) provides superior detection capability.

For resource-constrained or real-time monitoring systems requiring low computational overhead, Isolation Forest offers a favorable trade-off between efficiency and detection performance due to its low runtime and linear complexity. These insights reinforce the importance of context-aware model selection rather than reliance on a single universal algorithm.

5. Conclusion

This study presented a systematic and comparative evaluation of three classical unsupervised anomaly detection algorithms—Isolation Forest, Local Outlier Factor (LOF), and One-Class Support Vector Machine (OC-SVM)—for network intrusion detection across two widely used benchmark datasets: NSL-KDD and CIC-IDS2017. Unlike many prior studies relying on supervised learning and labeled attack data, the proposed framework adopted a fully unsupervised protocol, where models were trained exclusively on benign traffic to simulate realistic zero-day detection scenarios.The experimental results demonstrated that algorithm effectiveness is strongly dataset-dependent. On NSL-KDD, the boundary-based OC-SVM achieved superior detection performance (F1 = 0.9948), driven primarily by error-rate and host-level aggregation features. In contrast, on CIC-IDS2017, the density-based LOF model achieved the highest F1-score (0.9890), with detection dominated by timing-related and flow-level burst features characteristic of DDoS traffic. These findings highlight that anomaly structure, rather than algorithmic complexity alone, determines optimal model selection.

Dimensionality reduction using Principal Component Analysis (PCA) consistently reduced computational cost across both datasets—achieving up to 65–70% reduction in training time for kernel-based models—without degrading detection accuracy. This confirms that PCA offers an effective trade-off between scalability and performance in high-dimensional intrusion detection tasks.

To enhance transparency and operational trustworthiness, SHAP-based explainability analysis was conducted for the best-performing configuration on each dataset. The interpretability results revealed dataset-specific dominant feature families, confirming that the models captured meaningful statistical characteristics of attack behavior rather than relying on spurious correlations. This strengthens the reliability of the proposed framework for real-world cybersecurity deployment.

Overall, the study makes three key contributions:

1. A unified cross-dataset evaluation of classical unsupervised anomaly detection models under realistic training assumptions.
2. A systematic analysis of the impact of dimensionality reduction on both predictive performance and computational efficiency.
3. An explainability-driven investigation of feature influence patterns across heterogeneous intrusion datasets.

The findings emphasize that no single anomaly detection algorithm is universally optimal. Instead, model selection should be guided by dataset characteristics, attack dynamics, and deployment constraints. Future work will extend this framework to additional modern intrusion datasets, incorporate adversarial robustness evaluation, and explore hybrid combinations of density- and boundary-based methods to further enhance detection reliability.

Declaration of Ethical Code

In this study, we undertake that all the rules required to be followed within the scope of the "Higher Education Institutions Scientific Research and Publication Ethics Directive" are complied with, and that none of the actions stated under the heading "Actions Against Scientific Research and Publication Ethics" are not carried out.

References

[1] Tatineni, S. 2021. Machine learning approaches for anomaly detection in cybersecurity: a comparative analysis. International Journal of Computer Engineering and Technology, 12(1), 42–50.

[2] Segurola-Gil, L., Moreno-Moreno, M., Irigoien, I. ve diğerleri. 2024. Unsupervised anomaly detection approach for cyberattack identification. International Journal of Machine Learning and Cybernetics, 15, 5291–5302.

[3] Chandola, V., Banerjee, A., Kumar, V. 2009. Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58.

https://doi.org/10.1145/1541880.1541882

[4] Liu, F. T., Ting, K. M., Zhou, Z. H. 2012. Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data, 6(1), 1–39.

[5] Breunig, M. M., Kriegel, H.-P., Ng, R. T., Sander, J. 2000. LOF: Identifying density-based local outliers. ACM SIGMOD Record, 29(2), 93–104.

[6] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., Williamson, R. C. 2001. Estimating the support of a high-dimensional distribution. Neural Computation, 13(7), 1443–1471.

[7] Handa, A., Sharma, A., Shukla, S. K. 2019. Machine learning in cybersecurity: a review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1306.

[8] Adiban, M., Siniscalchi, S. M., Salvi, G. 2023. A step-by-step training method for multi-generator GANs with application to anomaly detection and cybersecurity. Neurocomputing, 537, 296–308.

[9] Goswami, M. 2024. AI-based anomaly detection for real-time cybersecurity. International Journal of Research and Review in Technology, 3(1), 45–53.

[10] Yaseen, A. 2023. The role of machine learning in network anomaly detection for cybersecurity. Sage Scientific Review of Applied Machine Learning, 6(8), 16–34.

[11] Alabadi, M., Çelik, Y. 2020. Anomaly detection for cyber-security based on convolution neural network: A survey. Uluslararası İnsan Bilgisayar Etkileşimi, Optimizasyon ve Robotik Uygulamaları Kongresi (HORA), IEEE, 1–14.

[12] Inuwa, M. M., Das, R. 2024. A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks. Internet of Things, 26, 101162.

[13] Hong, J., Liu, C. C., Govindarasu, M. 2014. Integrated anomaly detection for cyber security of the substations. IEEE Transactions on Smart Grid, 5(4), 1643–1653.

[14] Choppadandi, A., Kaur, J., Chenchala, P. K., Agarwal, A., Nakra, V., Pandian, P. K. G. 2021. Anomaly detection in cybersecurity: leveraging machine learning algorithms. ESP Journal of Engineering and Technology Advances, 1(2), 34–41.

[15] H. Kamal, M. Mashaly, "AE-DTNN: Autoencoder–Dense–Transformer Neural Network Model for Efficient Anomaly-Based Intrusion Detection Systems," Machine Learning and Knowledge Extraction, vol. 7, no. 3, p. 78, 2025.

[16] N. Borgioli, F. Aromolo, L. T. X. Phan, G. Buttazzo, "A convolutional autoencoder architecture for robust network intrusion detection in embedded systems," Journal of Systems Architecture, vol. 156, p. 103283, 2024.

[17] Jia, W., Sun, M., Lian, J. ve diğerleri. 2022. Feature dimensionality reduction: a review. Complex & Intelligent Systems, 8, 2663–2693.

[18] Liu, F. T., Ting, K. M., Zhou, Z. H. 2008. Isolation Forest. 2008 IEEE International Conference on Data Mining (ICDM), IEEE, Pisa, Italy, 15–19 Aralık 2008, 413–422.

[19] Jolliffe, I. T., Cadima, J. 2016. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.

[20] Scikit-learn developers. sklearn.neighbors.LocalOutlierFactor. https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.LocalOutlierFactor.html (access date: 20.04.2025).

[21] Tax, D. M. J., Duin, R. P. W. 2004. Support vector data description. Machine Learning, 54(1), 45–66.

[22] Scikit-learn developers. sklearn.svm.OneClassSVM. https://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html (access date: 20.04.2025).

[23] Al Farizi, W. S., Hidayah, I., & Rizal, M. N. (2021, September). Isolation forest based anomaly detection: A systematic literature review. In 2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE) (pp. 118-122). IEEE.

[24] Saadah, B. (2025). ANOMALY DETECTION IN MNIST DATASET USING ONE-CLASS SVM. Jurnal Kecerdasan Buatan dan Teknologi Informasi, 4(3), 264-270.

[25] Jolliffe, I. (2011). Principal component analysis. In International encyclopedia of statistical science (pp. 1094-1096). Springer, Berlin, Heidelberg.

[26] Almaiah, M. A., Almomani, O., Alsaaidah, A., Al-Otaibi, S., Bani-Hani, N., Hwaitat, A. K. A., ... & Aldhyani, T. H. (2022). Performance investigation of principal component analysis for intrusion detection system using different support vector machine kernels. Electronics, 11(21), 3571.

[27] Aoufi, S., Derhab, A., & Guerroumi, M. (2020). Survey of false data injection in smart power grid: Attacks, countermeasures and challenges. Journal of Information Security and Applications, 54, 102518.

[28] Ferrag, M.A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches,

datasets, and comparative study. Journal of Information Security and Applications, 50, 102419.

[29] Shone, N., Ngoc, T.N., Phai, V.D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. IEEE Transactions on Emerging Topics in Computational Intelligence, 2(1), 41–50.

[30] Vinayakumar, Vinayakumar, R., Alazab, M., Soman, K.P., Poornachandran, P., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. IEEE Access, 7, 41525–41550.