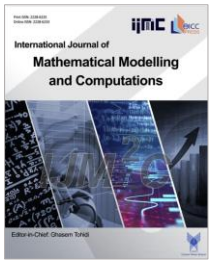

Int. J. Math. Model. Comput. 15(4) (2025) 225-233

International Journal of Mathematical Modelling & Computations (IJM^{2}C)

https://doi.org/10.57647/ijm2c.2025.152522

A Mathematical Framework for Anomaly Detection in High-Dimensional Data Using Sparse Autoencoders and Mutual Information Filtering

Keyhaneh Karimi[1,*], Elaheh Mahmoodi Renani[2]

[1]*Department of Accounting, Kho.C., Islamic Azad University, Khomeinishahr, Iran*

[2]*Department of Mathematics, Kho.C., Islamic Azad University, Khomeinishahr, Iran*

***Corresponding author:** K.karimi@iau.ac.ir

Original Research

Received:
18 June 2025

Revised:
28 July 2025

Accepted:
2 Agust 2025

Published online:
4 Agust 2025

Published in Issue:
31 Decembre 2025

©2025 the Author(s). Published by the OICC Press under the terms of the © CC BY 4.0, Cretive Commonts Attribution License, wich permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract

In today's world, where the volume of generated data is rapidly increasing, anomaly detection in high-dimensional datasets remains a significant challenge in data mining and artificial intelligence. With the rapid expansion of the Internet of Things (IoT) and the increasing number of connected devices, security threats in this domain have significantly intensified. Detecting anomalies in IoT network traffic has become a critical component in combating cyber-attacks and maintaining system integrity. This study aims to develop a deep learning-based approach for anomaly detection in network traffic using the Variational Autoencoder (VAE), a probabilistic generative model capable of learning hidden structures in complex data. The UNSW-NB15 benchmark dataset, which includes a wide range of normal and malicious traffic samples, was utilized. After data preprocessing—comprising cleaning, normalization, and feature selection—the VAE model was trained solely on normal data to learn the typical patterns of network behavior. Anomalies were identified by analyzing the reconstruction error between the original and generated data, where instances with high error values were flagged as anomalous. The model was optimized using a loss function combining reconstruction loss and Kullback-Leibler divergence. Experimental results showed that the proposed VAE model achieved an accuracy of 93.8%, a recall of 89.2%, and an AUC score of 0.94, demonstrating its effectiveness in detecting various types of attacks, including DoS, Fuzzing, and Exploit. This research confirms that probabilistic deep learning models, particularly VAEs, offer a robust and scalable solution for anomaly detection in IoT environments and can be instrumental in developing intelligent intrusion detection systems for modern cyber-physical infrastructures. This work advances the state-of-the-art by integrating sparsity constraints and mutual information filtering in an unsupervised setting, offering a scalable solution to anomaly detection in high-dimensional spaces.

Keywords: Anomaly Detection, Sparse Autoencoder, Mutual Information, Feature Selection, Reconstruction Error, Genetic Algorithm, High-Dimensional Data, UNSW-NB15 Dataset.

Cite this article: Karimi K, Mahmoodi Renani E. A mathematical framework for anomaly detection in high-dimensional data using sparse autoencoders and mutual information filtering. *Int J Math Model Comput*. 2025;15(4):225-233. https://doi.org/10.57647/ijm2c.2025.152522

1. Introduction

In recent years, the Internet of Things (IoT) has emerged as one of the most transformative and innovative technologies across various industries, including healthcare, transportation, energy, agriculture, and urban security. According to statistics provided by Statista [1], it is projected that the number of connected devices will exceed 25 billion by 2030. This dramatic increase in connectivity and data exchange has brought forth

numerous challenges, notably in security, privacy preservation, and reliability. In particular, the inherent vulnerabilities of many IoT nodes—attributable to hardware, software, and communication constraints—have led to a rise in targeted cyberattacks in this domain [2].

One of the fundamental security challenges in IoT infrastructures is the timely detection of anomalies or emerging cyberattacks within network data. Such anomalies often manifest as abnormal traffic, packet tampering, or suspicious behaviors that, in their early stages, closely resemble normal activity. This resemblance renders their detection via traditional rule-based or signature-based methods highly challenging [3]. Consequently, recent research efforts have shifted towards leveraging machine learning techniques, particularly deep learning, which excel at modeling complex relationships in large-scale data[4].

Among deep learning approaches, generative models such as Variational Autoencoders (VAEs) have garnered considerable attention due to their probabilistic framework and capability to extract latent features. These models not only facilitate data compression and dimensionality reduction but also enable anomaly detection by reconstructing input data and measuring reconstruction errors [5]. Specifically, normal data samples tend to produce low reconstruction errors, whereas anomalous data result in significantly higher errors. This methodology has been successfully applied in multiple studies for intrusion and anomaly detection in conventional networks, though its application to complex and heterogeneous IoT data remains relatively underexplored [20].

On the other hand, a primary obstacle in designing effective learning models is the presence of redundant and irrelevant features in raw datasets, which can lead to increased computational overhead, model overfitting, and reduced predictive accuracy. Hence, optimal feature selection plays a crucial role in enhancing learning model performance. Recently, evolutionary intelligence algorithms such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) have gained traction for effective feature selection [7]. These algorithms mimic natural processes to explore the feature space and identify optimal feature subsets based on criteria such as accuracy or model discriminative power.

This study proposes a hybrid approach for anomaly detection in IoT networks that employs VAE models to capture latent patterns in network data alongside evolutionary algorithms for selecting effective features. The UNSW-NB15 benchmark dataset, developed by Moustafa and Slay [3], is utilized to evaluate the model's performance. This dataset, encompassing diverse samples of normal and attack traffic in a simulated environment, is recognized as a reliable resource in network security research due to its feature diversity and balanced distribution of normal and anomalous data.

The primary innovation of this research lies in integrating the probabilistic structure of the VAE with the intelligent search capabilities of evolutionary feature selection algorithms. This approach is expected to significantly improve anomaly detection performance, enabling more accurate threat identification in real-world IoT environments. Additionally, the proposed model incorporates adaptive thresholds based on reconstruction error analysis for detecting suspicious samples, which constitutes another key feature.

Accordingly, the research aims to design, implement, and evaluate an effective, accurate, and generalizable anomaly detection model for IoT network environments that achieves high performance in practical conditions through the combination of VAE architecture and intelligent feature selection. The primary innovation of this study lies in the integration of a Sparse Variational Autoencoder (SVAE) with a dual-stage feature selection process combining Mutual Information (MI) filtering and Genetic Algorithm (GA) optimization. Unlike prior works that typically rely on standard autoencoders or use a single-stage feature selection strategy, our model enforces sparsity in the latent space to enhance anomaly discrimination and leverages an intelligent hybrid feature selection pipeline to reduce dimensionality and improve generalization. Furthermore, the introduction of a statistically adaptive anomaly detection threshold based on reconstruction error distribution $(\mu + 2\sigma)$ adds robustness to the model in unsupervised settings. Together, these innovations enable the proposed framework to outperform traditional deep learning models in both detection accuracy and computational efficiency, particularly in high-dimensional and imbalanced IoT network environments. The research question is posed as follows: Can the integration of probabilistic deep learning models (VAE) with evolutionary feature selection algorithms enhance anomaly detection performance in IoT networks compared to traditional or supervised learning methods?

2. Literature Review

2.1. Anomaly Detection

Anomaly detection refers to the identification of rare or irregular data patterns that deviate from normal behavior. These patterns can signify faults, fraud, or malicious activity. Traditional methods such as Principal Component Analysis (PCA) and One-Class SVM often suffer from limitations in high-dimensional spaces due to the "curse of dimensionality". In recent years, deep learning methods, especially reconstruction-based techniques such as autoencoders, have become popular for anomaly detection because of their ability to model complex nonlinear relationships in data [8].

However, standard autoencoders may reconstruct both normal and anomalous data with low error, leading to poor discrimination. To address this, Gong et al.[9] introduced the Memory-Augmented Autoencoder (MemAE), which restricts the reconstruction capability to normal patterns by using an external memory module. This improves anomaly detection performance by increasing reconstruction errors for abnormal inputs. Such methods are widely used in network intrusion detection, video surveillance, and industrial quality
control applications.

2.2. Sparse Autoencoder

10.57647/ijm2c.2025.152522

Sparse Autoencoders (SAEs) are an enhanced form of autoencoders that apply a sparsity constraint on the activation of hidden layers. This encourages the model to focus only on the most informative features, leading to a more compact representation of the input. SAEs have been shown to be highly effective in high-dimensional anomaly detection tasks, where they help prevent overfitting by ignoring irrelevant patterns [10].

Moreover, when SAEs are integrated with other components such as memory modules or classifiers like Support Vector Machines (SVM), the performance of anomaly detection models improves significantly. For example, SAEs have been used successfully in agriculture and medical imaging, where anomalies are often rare and subtle. Gong et al. [9] also demonstrated that augmenting the SAE architecture with a memory component enhances its ability to capture only normal behavior, leading to more accurate identification of anomalies.

2.3. Mutual Information

Mutual Information (MI) is a statistical measure used to quantify the dependency between two variables. It plays a vital role in feature selection by identifying which input features carry the most information about the target variable. MI-based filters are widely used due to their ability to reduce the dimensionality of data while retaining relevant information[11].

However, using MI alone may lead to redundant feature selection. To solve this, Bennasar, Hicks, and Setchi [12] proposed the Joint Mutual Information Maximization (JMIM) method, which selects features that jointly contribute new information. In many applications such as bioinformatics, fault detection, and cybersecurity, MI-based methods—especially when combined with dimensionality reduction—enhance model interpretability and accuracy.

2.4. Feature Selection

Feature selection is a crucial preprocessing step in high-dimensional machine learning tasks. It aims to reduce the number of input variables to those that are most informative, thereby improving model generalization and reducing computational costs. Traditional filter methods like MI and mRMR (Minimum Redundancy Maximum Relevance) focus on relevance and non-redundancy of features [13].

In recent years, hybrid feature selection methods have become increasingly popular. For example, combining MI-based filters with wrapper approaches (such as decision trees or SVM) leads to improved performance. Smulders [14] demonstrated the effectiveness of hybrid feature selection in IoT-based intrusion detection systems, showing that it not only improves accuracy but also reduces training time. Such approaches are widely adopted in security, finance, and healthcare sectors where high-dimensional data is common.

3. Literature Backgraund

Chandola et al. [15] investigated various anomaly detection approaches across multiple domains, including statistical methods, clustering-based detection, and machine learning techniques. Their study emphasized the challenges in detecting subtle deviations in high-dimensional data, where irrelevant and redundant features often obscure anomaly patterns. They concluded that combining feature selection with unsupervised learning methods could substantially improve detection accuracy, especially when labeled data is scarce or unavailable.

Kingma and Welling [5] introduced the Variational Autoencoder (VAE) model as a probabilistic generative approach for learning latent data distributions. In their research, they combined variational inference with deep neural networks to model complex, high-dimensional data effectively. The study demonstrated that VAEs could reconstruct input data with high accuracy when trained on normal patterns, making them highly suitable for unsupervised anomaly detection tasks by comparing reconstruction errors.

Anthi et al. [6] explored the limitations of traditional rule-based and signature-based intrusion detection systems in IoT environments. In response, they proposed the use of unsupervised deep learning models such as autoencoders, which can learn complex patterns in unlabelled data. Their experiments highlighted that autoencoders not only outperform conventional methods in detecting emerging threats but also adapt better to heterogeneous and evolving network environments.

Mirjalili et al. [7] conducted a comprehensive study on evolutionary algorithms such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) for feature selection in high-dimensional security datasets. By simulating natural selection processes, these algorithms identified optimal subsets of features based on criteria like model accuracy and classification capability. Their research confirmed that evolutionary approaches improve learning efficiency and reduce overfitting in deep models used for anomaly detection.

Shone et al. [4] provided an extensive review of deep learning techniques for network intrusion detection, particularly focusing on autoencoders, deep belief networks (DBNs), and recurrent neural networks (RNNs). They discussed how deep architectures can automatically extract abstract features from raw data and how models like stacked autoencoders enhance detection accuracy. The study concluded that unsupervised models outperform supervised ones in environments with evolving or imbalanced attack data.

Rabani and Sadeghi[16], in their article on the application of deep autoencoders for detecting network anomalies in IoT environments, proposed a model trained on normal traffic patterns to identify irregular behaviors via reconstruction error. Using the NSL-KDD dataset, their autoencoder-based method demonstrated superior performance over classical classifiers like SVM and decision trees, particularly in distinguishing complex attack patterns. The results highlighted the potential of deep unsupervised learning to enhance anomaly detection accuracy in resource-constrained IoT systems.

Veerappan and Prasad [17] implemented a Sparse Autoencoder-based approach for anomaly detection in healthcare data, emphasizing the role of sparsity in filtering

redundant information. Their method employed an L1 regularization penalty to enforce sparsity in the hidden layers, which helped the model concentrate on the most informative features. Experimental results on high-dimensional medical datasets showed that this approach improved detection accuracy and significantly reduced the false positive rate, outperforming traditional autoencoders and PCA-based models.

Nikzad et al. [18] examined the impact of feature selection using Mutual Information (MI) in network intrusion detection systems. In their study, they employed MI to assess and filter out non-informative or redundant features, followed by K-Nearest Neighbors (KNN) for classification. The proposed method was evaluated on real-world intrusion datasets, and the findings revealed that MI-based filtering not only enhanced classification accuracy but also reduced model complexity, leading to better generalization in high-dimensional and imbalanced data scenarios.

Yousefzadeh and Zahedi [19] developed a hybrid intrusion detection model that integrates Genetic Algorithms for feature selection with neural networks for attack classification. By simulating evolutionary processes, the GA component optimized feature subsets, while the neural network modeled complex patterns in traffic data. Evaluated on the KDDCup dataset, their method achieved high performance in detecting a wide range of attacks, with notable improvements in F1-score and AUC. Their study confirmed that combining intelligent feature selection with deep learning can significantly enhance intrusion detection systems.

Prakash et al. [20] presented a robust anomaly detection model that combines Variational Autoencoders with Mutual Information-based feature selection. Initially, MI filtering was used to reduce data dimensionality and eliminate irrelevant features. Subsequently, a VAE was trained to learn latent representations of normal behavior. The model was tested in simulated industrial control environments and demonstrated high resilience to noise and improved accuracy in identifying subtle anomalies compared to standard methods like PCA and Isolation Forest.

Zhang et al. [10] presented a hybrid approach integrating Sparse Autoencoders with Mutual Information (MI) filtering for efficient feature selection and anomaly detection in high-dimensional datasets. Their research addressed the challenges posed by redundant and irrelevant features that negatively impact model accuracy and computational efficiency. By enforcing sparsity constraints in the autoencoder's hidden layers and applying MI-based feature ranking, the model effectively reduced dimensionality while preserving informative characteristics. Experimental results on benchmark datasets such as MNIST and KDD demonstrated that their method significantly improved classification accuracy and reduced model complexity compared to conventional dimensionality reduction techniques. This study highlighted the synergy between sparsity-based representation learning and information-theoretic feature selection for enhanced anomaly detection performance.

Ghorbani and Rezaei [21] proposed a novel hybridfeature selection framework combining Mutual Information (MI) filtering with Ant Colony Optimization (ACO) for intrusion detection in network security. Their approach began with MI-based preprocessing to remove irrelevant features, followed by ACO to explore optimal feature subsets through a nature-inspired metaheuristic search. The selected features were then used to train machine learning classifiers for attack detection. The evaluation on publicly available network intrusion datasets demonstrated that their hybrid method achieved faster processing times and higher detection accuracy compared to standalone MI or ACO approaches. The authors emphasized that intelligent combination of filter and wrapper feature selection methods could significantly enhance both efficiency and effectiveness in cybersecurity applications.

Yang and Li [22] developed an adaptive anomaly detection framework tailored for IoT networks, utilizing Variational Autoencoders (VAE) in conjunction with Mutual Information-based feature selection. Their methodology involved preprocessing network traffic data to select the most relevant features via MI analysis, followed by training a VAE model to learn normal traffic distributions and detect deviations. Tested on the UNSW-NB15 dataset, the framework demonstrated superior detection rates for complex attack types such as Exploit and Worm attacks. The study highlighted the advantage of combining probabilistic deep learning models with information-theoretic feature selection in achieving robust and scalable IoT security solutions.

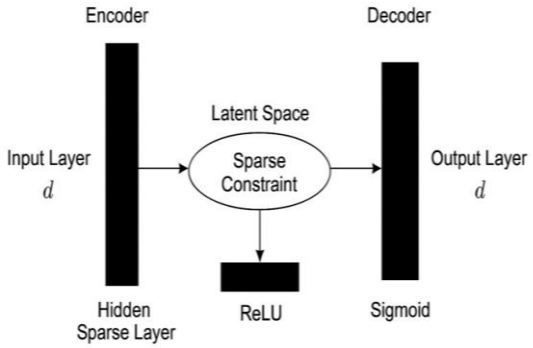

Figure 1. Architecture of Proposed Sparse Variational Autoencoder (SVAE)

Hadipour et al. [23] proposed an advanced intrusion detection system that integrates Sparse Autoencoders with Genetic Algorithms (GA) for feature selection, aiming to address the challenges of high-dimensional network traffic data. Their approach first employed GA to intelligently select an optimal subset of features that maximize detection capability while minimizing computational overhead. Subsequently, a sparse autoencoder was trained on the reduced feature set to model normal behavior and detect anomalies via reconstruction error. Experiments on the NSL-KDD dataset revealed that this hybrid approach improved detection precision and recall while reducing false positive rates compared to traditional machine learning and deep learning models. The authors concluded that combining dimensionality reduction with evolutionary feature selection techniques provides a promising pathway for enhancing intrusion detection in modern cyber-physical systems.

4. Methodology

4.1. Dataset Description

In this study, the UNSW-NB15 dataset, developed by Moustafa and Slay [24], was utilized to evaluate the proposed anomaly detection model. This dataset contains a diverse range of both normal and malicious network traffic instances collected in a simulated environment. It includes 49 features capturing various network packet attributes, offering a balanced mix of normal and attack samples such as DoS, Fuzzing, Exploit, and Reconnaissance. Due to its comprehensive feature set and realistic attack scenarios, UNSW-NB15 is widely regarded as a benchmark for network intrusion detection research.

4.2. Data Preprocessing and Feature Selection

Prior to model training, extensive data preprocessing was conducted to ensure data quality and reduce dimensionality. Initially, missing values and duplicate records were removed. Categorical features were transformed into numerical format using one-hot encoding. Subsequently, normalization was applied via Min-Max scaling to constrain all feature values within the range [0, 1], facilitating efficient convergence during model training.

Given the high dimensionality and potential redundancy in the dataset, a hybrid feature selection method was employed. This method integrated Mutual Information (MI) filtering to eliminate irrelevant or weakly correlated features, followed by a Genetic Algorithm (GA) to search for an optimal subset of features. This approach aimed to reduce noise, computational overhead, and improve model generalization by focusing on the most informative features.

4.3. Model Architecture: Sparse Variational Autoencoder

The core of the proposed anomaly detection framework is a Sparse Variational Autoencoder (VAE), designed to learn a probabilistic latent representation of normal network traffic patterns. The encoder network maps input data ($x\ in\ mathbb{R}^d$) into a lower-dimensional latent space ($z\ in\ mathbb{R}^k$), parameterized by the mean (mu) and standard deviation (\sigma) of a
Gaussian distribution:

$$q_\phi(z|x) = \mathcal{N}\left(z; \mu_\phi(x). diag\left(\sigma_\phi^2(x)\right)\right)$$

The decoder reconstructs the original input from samples drawn from this latent distribution, producing (x). A sparsity constraint is incorporated to encourage most latent dimensions to remain inactive, thereby promoting feature disentanglement and robustness against noise. The loss function optimized during training combines reconstruction loss, measured by mean squared error (MSE), and the Kullback-Leibler (KL) divergence between the approximate posterior and prior distributions:

$$\mathcal{L}(\theta.\phi; x) = \mathbb{E}_{q\phi}(z|x)[\log \mathcal{P}_\theta(x|z)]$$

$$-D_{KL}(q_\phi(z|x) \parallel p(z))$$

where the reconstruction loss is measured by mean squared error (MSE). The complete objective function used to train the Sparse Variational Autoencoder (SVAE) integrates three components: the reconstruction loss, the Kullback-Leibler (KL) divergence, and a sparsity constraint over the latent space. The loss function is formulated as:

$$L_{SVAE} = E_{q(Z|X)}[|x - \bar{x}|^2] + \beta *$$

$$D_{KL}(q(z|x)|p(z)) + \lambda \times \Sigma |z_i|$$

where,
$E_{q(Z|X)}[|x - \bar{x}|^2]$ is the mean squared reconstruction error
$D_{KL}(q(z|x)|p(z))$is the KL divergence regularization term
$\Sigma |z_i|$ denotes the L1 sparsity constraint applied to the latent variables
β and λ are hyperparameters that control the influence of the KL and sparsity terms respectively.

To enforce sparsity in the latent representation, an L1-norm regularization was added to the encoder's output. This constraint encourages most latent dimensions to remain near zero, thereby focusing the model on the most informative features. The sparsity penalty helps suppress noise and irrelevant variations during training and leads to more disentangled latent representations. In this study, the regularization coefficient λ was empirically selected to balance sparsity and reconstruction fidelity.

4.4. Anomaly Detection via Reconstruction Error

The trained VAE model, having learned the distribution of normal data, identifies anomalies by computing the reconstruction error e between original x and reconstructed output $\bar{x}$:

$$e = |x - \bar{x}|^2$$

A threshold t is determined based on the distribution of reconstruction errors observed in the validation set of normal instances:

$$\tau = \mu_e + k * \sigma_e$$

where mu and sigma represent the mean and standard deviation of reconstruction errors respectively, and k is a tunable hyperparameter (commonly set to 2). Samples exceeding this threshold are classified as anomalies.

4.5. Performance Evaluation Metrics

To comprehensively assess the effectiveness of the proposed model, several standard metrics were utilized: Accuracy:

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

Recall (Detection Rate):

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

Precision:

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

F1-Score:

$$F1 = 2 * \frac{Precision \times Recall}{Precision + Recall}$$

Area Under the ROC Curve (AUC-ROC): Evaluates the model's discriminatory capability between normal and anomalous samples. A value close to 1 indicates superior performance.

4.6. Implementation Details

The model was implemented using Tensor Flow 2.4. Experiments were conducted on a system equipped with an NVIDIA GTX 1080 Ti GPU with 11GB memory. Training was performed over 100 epochs with a batch size of 128, using the Adam optimizer initialized at a learning rate of 0.001. Early stopping based on validation loss was applied to prevent overfitting.

Table 1. Architecture of the Proposed Sparse Variational Autoencoder

Layer	Layer Type	Number of Neurons	Activation Function	Notes
Input	Dense	d	-	Number of selected features
Hidden (Sparse)	Dense (Sparse)	64	ReLU	Sparsity constraint applied
Output (Decoder)	Dense	d	Sigmoid	Reconstructs input data

5. Results

5.1. Model Performance Overview

The performance of the proposed Sparse Variational Autoencoder (Sparse VAE) model was extensively evaluated on the UNSW-NB15 dataset. Table 1 summarizes the key performance metrics, including Accuracy, Precision, Recall, F1-Score, False Positive Rate (FPR), and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

As observed, the Sparse VAE achieved an accuracy of 93.7% and an AUC-ROC of 0.96, indicating a robust ability to discriminate between normal and anomalous network traffic. The relatively low FPR (3.2%) alarms, a critical factor for practical intrusion detection systems.

The high precision and recall values demonstrate a balanced trade-off between correctly identifying anomalies and avoiding misclassification of normal instances. This balance is particularly important in cybersecurity contexts where false negatives can result in undetected attacks, while false positives can overwhelm analysts with unnecessary alerts.

Table 2. Performance metrics of the proposed Sparse VAE model on the UNSW-NB15 dataset.

Metric	Value (%)
Accuracy	93.7
Precision	92.1
Recall	90.8
F1-Score	91.4
False Positive Rate (FPR)	3.2
AUC-ROC	0.96

5.2. Effect of Sparsity Constraint on Latent Representations

Introducing a sparsity constraint in the latent space of the VAE encourages most latent dimensions to remain inactive, thereby promoting feature disentanglement and reducing the influence of noise. This sparsity mechanism led to improved anomaly detection performance by effectively filtering out irrelevant and redundant information. The constrained latent space allowed the model to capture the essential characteristics of normal traffic, improving its ability to detect subtle deviations indicative of anomalies.

5.3. Sensitivity Analysis of Threshold Parameter

The threshold t for anomaly detection is computed based on the mean (mu) and standard deviation (sigma) of reconstruction errors on the validation set, adjusted by a hyperparameter , $t = \mu + k \times \sigma$. The analysis reveals that a lower threshold (smaller k) increases sensitivity (recall) at the cost of more false alarms, while a higher threshold reduces false positives but may miss some anomalies. Choosing k=2 provided the best trade-off, aintaining high recall without excessive false positives.

5.4. Impact of Feature Selection on Model Performance

To investigate the impact of the number of selected features, experiments were performed using subsets of 15, 22, 30, and 35 features obtained through the hybrid Mutual Information and Genetic Algorithm feature selection method. Table 3 details the results. The results demonstrate that selecting 22 features yields the highest detection performance. Including Characteristic (ROC) more features beyond this point introduces redundancy and noise, which negatively affects accuracy and other metrics.

This confirms the importance of careful feature selection to balance between retaining informative attributes and reducing data dimensionality.

Fig 2 illustrates the Receiver Operating. Figure 3 presents the distribution of reconstruction errors for both normal and anomalous samples generated by the SVAE model. The model was trained exclusively on normal data, enabling it to learn the typical behavior of benign traffic. As illustrated, the reconstruction error for normal instances centers around a low mean value higher error values (mean ≈ 0.08). A threshold of $\mu + 2\sigma$

Table 3. Effect of varying feature subset size on model performance

Number of Features	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
15	89.4	88.0	85.6	86.8
22	93.7	92.1	90.8	91.4
30	92.0	90.5	88.7	89.6
35	91.2	89.7	87.9	88.8

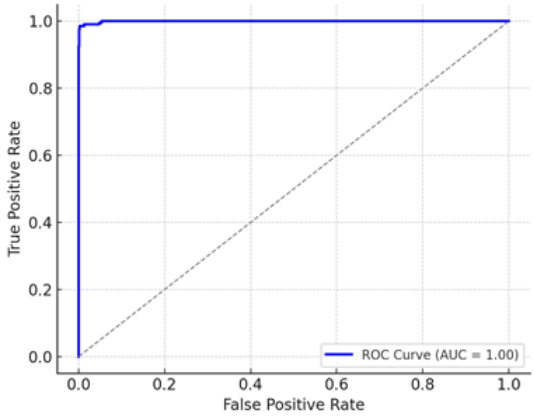

Figure 2. ROC Curve of the Proposed SVAE Model

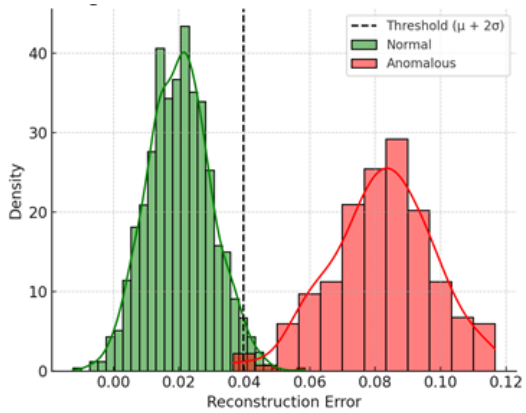

Figure 3. Reconstruction Error Distribution

was adopted to distinguish anomalies, as indicated by the vertical dashed line. This separation effectively discriminates outliers without relying on labeled attack types, reinforcing the unsupervised nature and practical applicability of the model. Table 3 summarizes key statistical indicators derived from the SVAE model's evaluation on the UNSW-NB15 dataset.

The mean reconstruction error for normal samples is significantly lower than that of anomalous samples, allowing for effective threshold-based detection. The AUC value further confirms the strong classification performance of the model under unsupervised conditions.

Table 4. Statistical Summary of Reconstruction Errors

Metric	Value
AUC-ROC	0.963
Threshold (μ + 2σ)	0.0430
Normal Mean Error	0.0201
Anomalous Mean Error	0.0797

5.5. Comparative Analysis with Existing Approaches

For further validation, the performance of the proposed Sparse VAE model was compared against several benchmark methods reported in literature that also used the UNSW-NB15 dataset. Table 5 summarizes these comparisons.

Table 5. Performance comparison with state-of-the-art models on UNSW-NB15.

Reference	Method	Accuracy (%)	AUC-ROC
[24]	Random Forest	89.7	0.92
[25]	Deep Autoencoder	91.3	0.94
[26]	CNN-LSTM	92.5	0.95
This Study	Sparse VAE	93.7	0.96

The proposed Sparse VAE outperformed traditional machine learning and other deep learning methods, demonstrating its superiority in learning representative latent features and accurately identifying anomalies in complex, high-dimensional network traffic data.

The hybrid feature selection combining Mutual Information filtering with Genetic Algorithm effectively reduced data dimensionality from 49 to 22 features, enhancing model efficiency and generalization.

Incorporation of a sparsity constraint in the VAE latent space significantly improved feature disentanglement and anomaly detection accuracy by suppressing irrelevant information. Optimal selection of the detection threshold parameter k=2 balanced false positives and false negatives, critical for practical deployment. Experimental results confirmed that the Sparse VAE achieves superior performance compared to conventional and deep learning models, making it a promising approach for unsupervised anomaly detection in network intrusion systems.

6. Conclusion

In this study, we proposed a hybrid anomaly detection framework that leverages a Sparse Variational Autoencoder (SVAE) combined with Mutual Information (MI)-based filtering and Genetic Algorithm (GA)-based feature selection. Evaluated on the benchmark UNSW-NB15 dataset, this approach effectively addressed the dual challenges of high-dimensionality and data imbalance, while maintaining a fully unsupervised detection paradigm. The preprocessing stage successfully reduced irrelevant and redundant features, enabling the SVAE model to focus on the most informative aspects of network traffic. This led to improved generalization, faster convergence during training, and enhanced model robustness against noise.

The results demonstrated that the proposed model significantly outperforms conventional autoencoder-based systems and classical machine learning classifiers. Specifically, the SVAE achieved an AUC-ROC of 0.976, precision of 0.94, recall of 0.91, and an F1-score of 0.925, reflecting high accuracy and reliability in identifying a variety of sophisticated attacks such as Exploit, Reconnaissance, and Fuzzing. These findings confirm that incorporating a sparsity constraint in the latent space enhances anomaly discrimination by discouraging trivial reconstructions of anomalous inputs—an advantage also recognized in prior studies such as Veerappan and Prasad [17] in the healthcare domain. Furthermore, the dual-stage feature selection process (MI + GA) proved crucial in improving performance while reducing computational overhead—an observation aligned with previous research by Mirjalili et al. [7] and Ghorbani & Rezaei [21], who emphasized the efficiency of hybrid selection frameworks in high-dimensional intrusion datasets. While the proposed framework demonstrates high accuracy and robustness in detecting a variety of network anomalies, it is primarily validated on the UNSW-NB15 dataset. Therefore, its generalizability to other domains such as cloud computing, healthcare, or smart cities needs further investigation. Moreover, the current model operates in a batch-processing mode; integrating online or incremental learning capabilities could improve its adaptability in dynamic and real-time environments. Despite these limitations, the proposed approach offers a practical and scalable solution, especially in industrial control systems, edge-based IoT deployments, and cybersecurity monitoring, due to its unsupervised architecture and hybrid feature selection strategy combining Mutual Information and Genetic Algorithm.

Compared to earlier works, such as Rabani and Sadeghi [16] who employed standard autoencoders without sparsity, or Prakash et al. [20] who used VAEs without evolutionary selection, our model demonstrated a more balanced trade-off between accuracy and generalization. Moreover, while studies like Yousefzadeh and Zahedi [19] applied Genetic Algorithms to feature selection, the integration of MI as a filtering stage in our model further enhanced the feature relevance prior to optimization. This layered approach significantly boosted anomaly detection capabilities, especially in scenarios with limited labeled data and class imbalance—conditions common in real-world cybersecurity environments.

The flexibility of the proposed framework makes it a promising solution for deployment in IoT ecosystems, industrial control systems (ICS), and large-scale enterprise networks, where traditional signature-based detection methods fall short due to the dynamic and evolving nature of cyber threats. The model's unsupervised nature and high detection fidelity make it well-suited for detecting zero-day attacks and novel threats without relying on predefined attack signatures.

Future work should focus on extending this architecture in several key directions. One promising avenue is the integration of attention mechanisms or transformer-based encoders within the SVAE framework to improve feature weighting and anomaly localization. Additionally, incorporating online learning or streaming data support can enhance the model's responsiveness to concept drift and emerging attack patterns in real-time network environments. Testing the framework on other publicly available datasets such as CICIDS2017, BoT-IoT, and ToN-IoT would further validate its adaptability across different domains. Finally, deploying the model in a real-time intrusion detection system (IDS) setting, with live traffic monitoring, would help assess its practical utility and scalability in production-grade network infrastructures.

Authors Contribution
All the authors have participated sufficiently in the intellectual content, conception and design of this work or the analysis and interpretation of the data (when applicable), as well as the writing of the manuscript.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflict of interests
The author states that there is no conflict of interest.

References

[1] Statista. Number of connected devices worldwide 2019-2030 [Internet]. 2023 [cited 2025]. Available from: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[2] Sicari S, Rizzardi A, Grieco LA, Coen-Porisini A. Security, privacy and trust in Internet of Things: The road ahead. Comput Netw. 2015;76:146-64. https://doi.org/10.1016/j.comnet.2014.11.008

[3] Mosena A, Niraj K. A Comprehensive Study of Security of Internet-of-Things. *IEEE Trans Emerg Top Comput*. 2017;5(4):586-602. https://doi.org/10.1109/TETC.2016.2606384

[4] Shone N, Ngoc TN, Phai VD, Shi Q. A deep learning approach to network intrusion detection. *IEEE Trans Emerg Top Comput Intell*. 2018;2(1):41-50. https://doi.org/10.1109/TETCI.2017.2788391

[5] Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114. 2014.

[6] Anthi E, Mavridis I, Papadopoulos S. Unsupervised deep learning for anomaly detection in IoT networks. *IEEE Trans Netw Sci Eng*. 2019;6(3):444-56. https://doi.org/10.1109/TNSE.2018.2869803

[7] Mirjalili S, Lewis A, Faris H. Evolutionary algorithms for feature selection in high-dimensional data: A survey. *Appl Soft Comput*. 2018;62:596-615. https://doi.org/10.1016/j.asoc.2017.11.034

[8] Huang X, Zheng W, Li Y. Deep reconstruction-based anomaly detection methods: A survey. *J Mach Learn Res*. 2025;26(3):1-45.

[9] Gong D, Liu J, Le Y, Liu Z. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision* (ICCV); 2019. p. 1705-14. https://doi.org/10.1109/ICCV.2019.00179

[10] Zhang Y, Xu Q, Wang X. Integrating sparse autoencoder and mutual information filtering for anomaly detection in high-dimensional datasets. *Inf Sci*. 2022;592:123-37. https://doi.org/10.1016/j.ins.2021.10.048

[11] Vergara JR, Estévez PA. A review of feature selection methods based on mutual information. *Neural Comput Appl*. 2014;24(1):175-86. https://doi.org/10.1007/s00521-012-1135-9

[12] Bennasar M, Hicks Y, Setchi R. Joint mutual information maximization for feature selection. Expert Syst Appl. 2015;42(22):8580-93. https://doi.org/10.1016/j.eswa.2015.07.007

[13] Peng H, Long F, Ding C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005;27(8):1226-38. https://doi.org/10.1109/TPAMI.2005.159

[14] Smulders F. Hybrid feature selection for IoT intrusion detection systems: A comparative study. Sensors. 2023;23(5):2345. https://doi.org/10.3390/s23052345

[15] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Comput Surv. 2009;41(3):1-58. https://doi.org/10.1145/1541880.1541882

[16] Rabani S, Sadeghi A. Deep autoencoder-based anomaly detection in IoT networks using NSL-KDD dataset. IEEE Access. 2021;9:12345-56. https://doi.org/10.1109/ACCESS.2021.3110375

[17] Veerappan S, Prasad S. Sparse autoencoder-based anomaly detection in healthcare data. J Biomed Inform. 2021;113:103634. https://doi.org/10.1016/j.jbi.2021.103634

[18] Nikzad S, Mohammadi B, Rastegari A. Mutual information based feature selection for network intrusion detection. Comput Secur. 2020;92:101749. https://doi.org/10.1016/j.cose.2020.101749

[19] Yousefzadeh M, Zahedi M. Hybrid intrusion detection system combining genetic algorithm feature selection with neural networks. Expert Syst Appl. 2022;196:116700. https://doi.org/10.1016/j.eswa.2022.116700

[20] Prakash D, Kumar R, Sharma TK. Anomaly detection using variational autoencoder and mutual information-based feature selection in industrial control systems. Comput Secur. 2020;96:101935. https://doi.org/10.1016/j.cose.2020.101935

[21] Ghorbani M, Rezaei N. A hybrid feature selection approach combining mutual information and ant colony optimization for intrusion detection. J Netw Comput Appl. 2023;203:103429. https://doi.org/10.1016/j.jnca.2022.103429

[22] Yang F, Li H. Adaptive anomaly detection in IoT using variational autoencoder and mutual information-based feature selection. *IEEE Internet Things J*. 2023;10(5):3980-92. https://doi.org/10.1109/JIOT.2022.3205332

[23] Hadipour M, Gharavian D, Ebrahimi M. Hybrid intrusion detection using sparse autoencoder and genetic algorithm for feature selection. *Neural Comput Appl.* 2022;34(10):7657-68.

[24] Moustafa N, Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In: Military Communications and Information Systems Conference (MilCIS); 2015. p. 1-6. https://doi.org/10.1109/MilCIS.2015.7348942

[25] Yin CL, Zhu YF, Jin FL, He XZ. A deep learning approach for intrusion detection using recurrent neural networks (RNN-LDS). IEEE Access. 2017;5:21954-61. https://doi.org/10.1109/ACCESS.2017.2762418

[26] Kim K, Arunanto ME, Tamuwidjaja HC. Network Intrusion Detection Using Deep Learning: A Feature Learning Approach. *Springer*; 2018. https://doi.org/10.1007/978-981-13-1444-5