International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 13-22 March 2026-Sept 2026 www.csjournals.com

A NIDPS System Using Robust Denoising Autoender for Zero-Day Attacks

Anu Devi[1] , Dr. Sahil Saharan [2] and Dr. Anupam Bhatia[3]
[1]MCA Student, [2]Asst. Professor, [3]Associate Professor
Deptt. of Comp. Sc. and Application, Chaudhary Ranbir Singh University, Jind, Haryana, India

Abstract: Currently in network security modern inputs, Zero-Days attack and adversarial attacks is major challenge. Evolution of information and communication technologies has significantly increased the dependency of individuals, organizations, and governments on interconnected computer networks. Conventional intrusion detection system are enabled to identify these attacks. The current research proposed robust network security monitoring and prevention system which is based on Denoising Autoencoder. For this model we utilized UNSW-NB15 dataset. Model is trained on normal traffic using semi-supervised strategy so that the system can predict new- attacks easily. During training adversarial data augmentation is used to add some noisy data. This enhances the robustness of our model. Experimental evaluation represent the accuracy of our model is 0.94. Classification report shows precision is 0.69, Recall/ Detection Rate is 0.86 and F1 score is 0.74.

Keywords: Network Security, Zero-Day Attacks, Adversarial Robustness,UNSW-NB15.

1. Introduction:

In the contemporary world, everything depends on the internet and network for example: social media, online banking, E-commerce. With the rapid growth of network utilization and digitalization, cyber attacks are also increased with it[1]. As network traffic and utilization increase attack detection and prevention become more challenging for intrusion detection system[2]. Network intrusion is also an important problem among these. Network intrusion is an attack in which unauthorized person or group tried to enter in your network (e.g. Wifi, server or application) without your permission. After obtaining unauthorized access to the network, attackers attempt to compromise, destroy system servers, controlled or damaged the overall system.

Network security and network intrusion prevention is very important in today's world. To ensure network security Network Intrusion Detection and Prevention Systems are required. To protect system network from attacker NIDPS are required so that personal important information are not leaked or misused by anyone. Due to systems vulnerabilities,. Network Intrusion Detection and Prevention System are designed to detect and prevent our system

from internal or external attacks. NIDPS system ensure integrity, privacy of sensitive data. NIDPS system also provide real-time alerts to security administrator.

In the current technological era, Machine Learning (ML) and Deep Neural Architectures (DNA)-based Intrusion Detection Systems (IDS) have introduced significant advancements in the field of cybersecurity. Various standards machine learning techniques are used for development of NIDPS system like Random Forest[2][3][5],Decision Tree[3][6][10]. Machine learning methods like Naive Bayes, SVM[10] are also widely used. But ML existing models has some limitations. First, ML models trained on labeled data such as NSL-KDD[10] and CIC-IDS2017 [3][4] and CIC-IDS2018[4]dataset. Therefore, these models cannot predict Zero day attacks (unknown attacks). Secondly, neural network / architecture are highly lack inherent robustness to adversarial attacks, where small changes in input can lead machine learning models to produce incorrect or misleading predictions.

To address these challenges this paper propose a model Network Intrusion Detection & Prevention System (NIDPS) based on Denoising Autoencoder[2][3]architecture**.** The main objective of this model is to detect Zero-Day Attack and Adversarial Robustness. This model evaluated on UNSW NB-15[6] dataset. This model is trained using **Denoising Autoencoder architecture** which learns robust representations by reconstructing clean inputs from noisy data. Some noisy data is added during training to improve model robustness. For model evaluation we use **Confusion Matrix, Fidelity Profile (Precision), Recall / Detection Rate, and F1-Score.**

A network security monitoring engine continuously analyzes live packet transmission streams. Security systems easily diagnose underlying anomalies and unauthorized operation. Intrusion Prevention System take actions to prevent network system from intruders. IPS system performs various actions like block IP address, generate alerts messages etc.

In 2007 Estonia cyber attack is considered one of the first nation-scale cyber warfare incidents. In this attack government websites, banks, media networks, parliament systems are temporarily down. In 2022, AIIMS (Delhi) experienced a major cyber attack. In this attack patient records are inaccessible, online services are down and healthcare operations also affected.

2. Literature Review:

S.No.	Authors	Dataset	Algorithms used	Results	Contributions	Limitations
1.	Alkasassbeh et al. (2025)	UGRansome	Deep Q-Network (RL)	~98% accuracy	RL effective for zero-day ransomware detection	High training cost; limited real-world validation
2.	YangGuo (2022)	Multiple IDS datasets	Autoencoders, RF,GANs, TL	Upto >98% (GAN)	ML(esp.GANs) improves zero-day detection	No true zero-day datasets; scalability issues
3.	Shamshair Ali et al.	CICIDS2017, NSL-KDD, CSE-CIC-IDS2018, SCADA, IoT	DT, RF, MLP, LSTM, CNN, Autoencoders, RL, Hybrid IDS	~90–99% accuracy (DL highest)	Comparative study of ML, DL, and hybrid IDS for zero-day detection	High false positives; dataset dependence; no standard benchmark
4.	Vikash Kumar, Ditipriya Sinha	CICIDS2018, Real-time traffic	Heavy-Hitter Analysis,Grap based Detection	~91% accuracy	Detects both high-volume andlow-volume zero-day attack	No adversarial robustness evaluation
5.	V. Heydari, K. Nyarko	UNSW-NB15	RF, Transformer NIDS, ADV_NN	ADV_NN: 86.55% clean, 84.07% (PGD), >80% (FGSM), >85 % (black-box);	ADV-NN improve robustness against adversial attacks	High training cost, limited testing on unseen attacks,weak cross validation
6.	Sydney M. Kasongo, Yanxia Sun	UNSW-NB15	SVM, LR, ANN, DT, kNN	DT best in binary (~90.9%); ANN best in multiclass (~77.5%); feature selection improved most models	Feature selection improve performance, DT strong for binary & ANN for multiclass	Poor detection of minor attacks, weak SVM/LR performance

7.	Benyamin Tafreshian & Shengzhi Zhang (2025)	UNSW-NB15, NSL-KDD	Adversarial Training, Ensemble IDS	improved robustness and attack detection accuracy under adversarial conditions.	Adversarially trained IDS models are more resilient to evasion attacks.	Increased computational complexity and training cost.
8.	Lan Goodfellow et al. (2015	General Deep Learning Datasets	FGSM, Adversarial Training	Demonstrated that small perturbations can mislead deep neural networks.	introduced adversarial attacks and robustness training concepts.	Focused on image-based DL models, not IDS traffic data
9.	sanidhya Sharma & Zesheng Chen (2024)	NSL-KDD, UNSW-based IDS setup	DNN, PGD, Boundary, HopSkipJump Attacks	Achieved high adversarial attack success rates against IDS models.	ML-based IDS are highly vulnerable to adversarial evasion attacks	Mainly evaluated supervised IDS models.
10.	Moraboena et al.	NSL-KDD, CICIDS2017	Symmetric Deep Autoencoder (SDAE)	Accuracy: **91.76%**	SDAE improves intrusion detection & reduces false alarms	Tested only on benchmark datasets; lacks real-world IoT/WSN evaluation

3. Research Framework:

This segment outlines the procedure followed for model development. It includes dataset sourcing, pre-processing, dimensionality optimization, model learning, and performance auditing to ensure model accuracy.

3.1. **Structural Framework of the Model:-** System architecture defines the whole process from data collection to Anomaly Detection. After data collection data is preprocessed and the data splits into two distinct sections. Autoencoder learns baseline structure and Adversarial Synthesis modules generate fake attack data. Then dataset is merged and processed by deep feature model to capture security patterns. Reconstruction Error Feedback loop updates model continuously and distinguish network easily. On the basis of reconstruction error decision engine evaluates normal traffic and anomaly detection.

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 13-22 March 2026-Sept 2026 www.csjournals.com

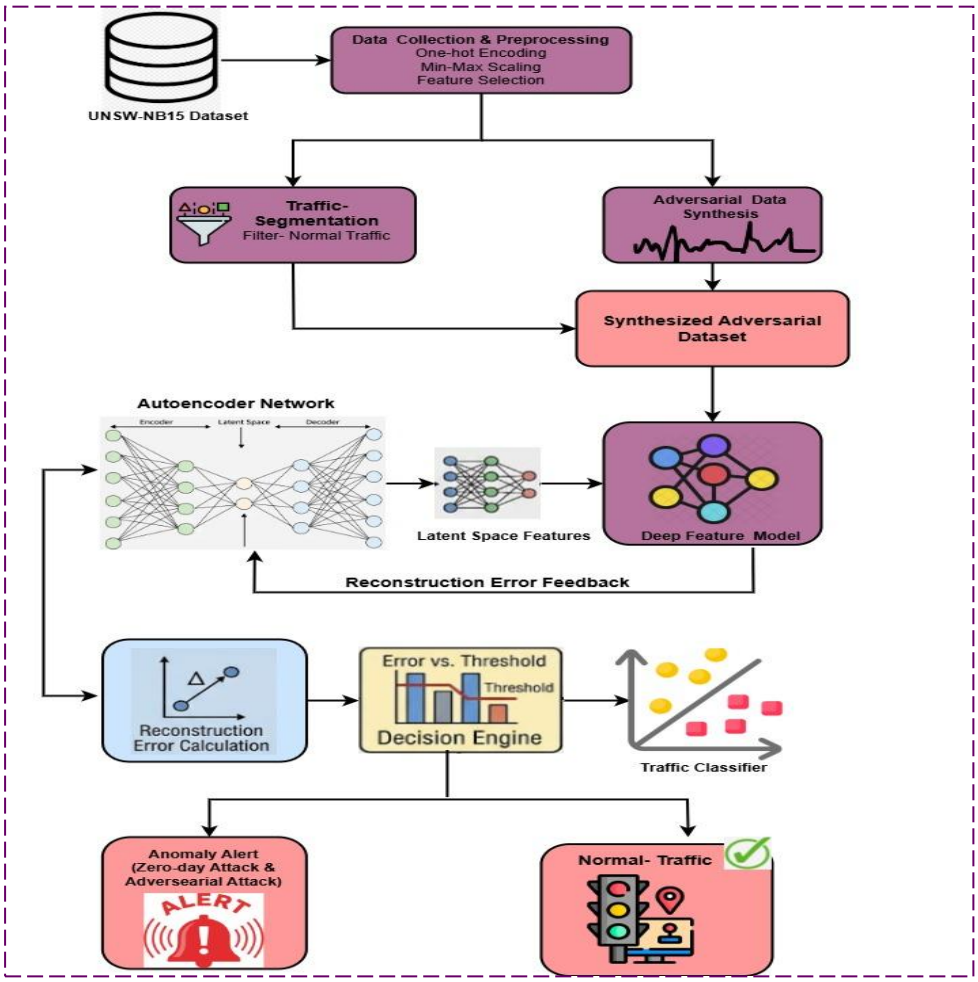


Fig 3.1 : Framework for robust and Zero day Attack System to detect Anomalies

3.2. Data Ingestion and Dataset Sourcing : The dataset used in this model is a secondary data. The dataset will be collected from Kaggle. It provides open-source dataset for the students, researcher. It provides the facility of online coding environment like Jupiter. In this model, UNSW-NB15 dataset[5][6][8] is utilized because it simulates modern traffic.

3.3. Data Preprocessing: Data preprocessing is the process in which we clean, transform and organize raw data into meaningful insights. Datasets are large, complex and noisy which negatively impact model performance. Data pre-processing steps used in this model are:-

- **One-Hot Encoding:** Categorical values such as attacks name are converted into numerical form because machine learning / Deep learning algorithms not understand categorical data[8].
- **Min- Max Scaling:** All numerical features are scaled within bounded interval of[0,1] by using min-max normalization[8] on the dataset. So that all features contributes equally during training.

Formula: $x' = \frac{x - xmin}{xmax - xmin}$

- **Adversarial Data Augmentation:** To improve the model robustness we can add some noisy data during training[5].

3.4. Feature Selection: In this section features are selected after data preprocessing for model development. It helps in improving model performance and eliminating redundant data. Heatmap are used to find correlation between features. For reducing model complexity we remove weakly related or not used features.

- **Correlation Analysis:** This technique is applied to remove less related features and it improves the model accuracy and its performance
- **Autoencoder-based latent representation:**

During training, the autoencoder[2][3] compress 49 features into a 16-dimensional latent space . It keeps the most important patterns and removing redundant information.

3.4 Model Training: After **f**eature selection model is trained using various machine learning techniques.

- **Deep Learning Technique:** At the time of model training Denoising Autoencoder (DAE) architecture is implemented in Tensorflow / Keras framework. During this phase **semi-supervised** learning strategy is adopted. The model is trained on 'Normal' traffic to learn hidden patterns of normal network behavior[2].
- During model training some noisy data is added by using Gaussian noise injection. The model parameters are optimized using the Adam optimizer**,** while the decoding disperancy is minimized using **the** Mean Squared Error (MSE) loss function[2].

3.5 Model Testing & Evaluation: This section outlines the testing and evaluation of the developed model. During testing, the autoencoder detects anomalies using reconstruction error. The system pipeline is verified through normal and unseen attack data to check its performance.

- **Threshold Determination:** A threshold value is set using the 98 percentile of reconstruction error from normal data**.** Values above this limit are treated as abnormal.
- **Anomaly Detection:** Each input passed through the model and reconstruction error is calculated. If the error is higher than threshold determined value then it is classified as an attack otherwise as normal traffic.

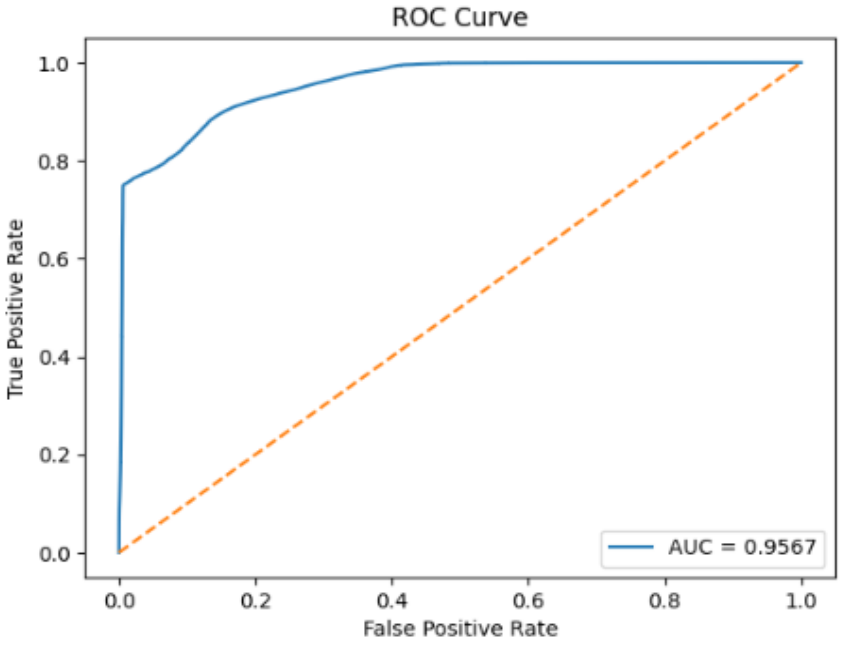
- **Model Evaluation:** Model's absolute reliability is monitored using a combination of specific evaluation metrics. Instead of depending only on accuracy we conducted a detailed analysis using Precision, Recall (Detection Rate), and F1-Score Mean to understand the model's reliability and detection capability. Confusion matrix is used to measure the false positive rate of model.

3.6 Evaluation and Quantitative Analysis: This section represents the experimental results obtained from the proposed NIDPS model using UNSW-NB15 dataset. We evaluates the operational reliability using classification report, histogram etc.

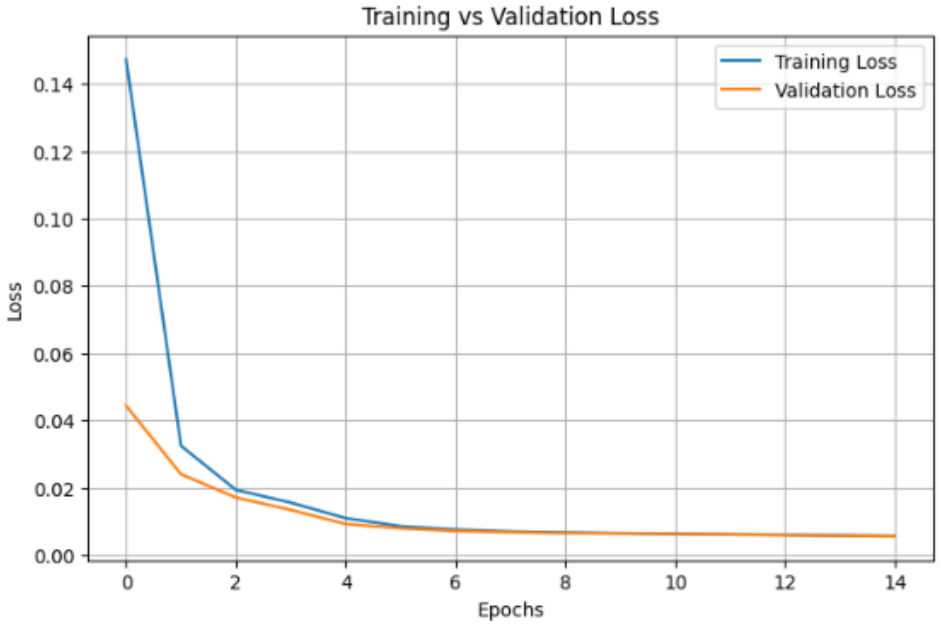
Table 1: Classification Report

Model Performance Evaluation Metrics	Normal traffic	Attack
Accuracy	0.94	0.94
Precision	0.99	0.40
Recall/ Detection Rate	0.95	0.79
F1 Score	0.97	0.53
Support	2,295,222	95,053

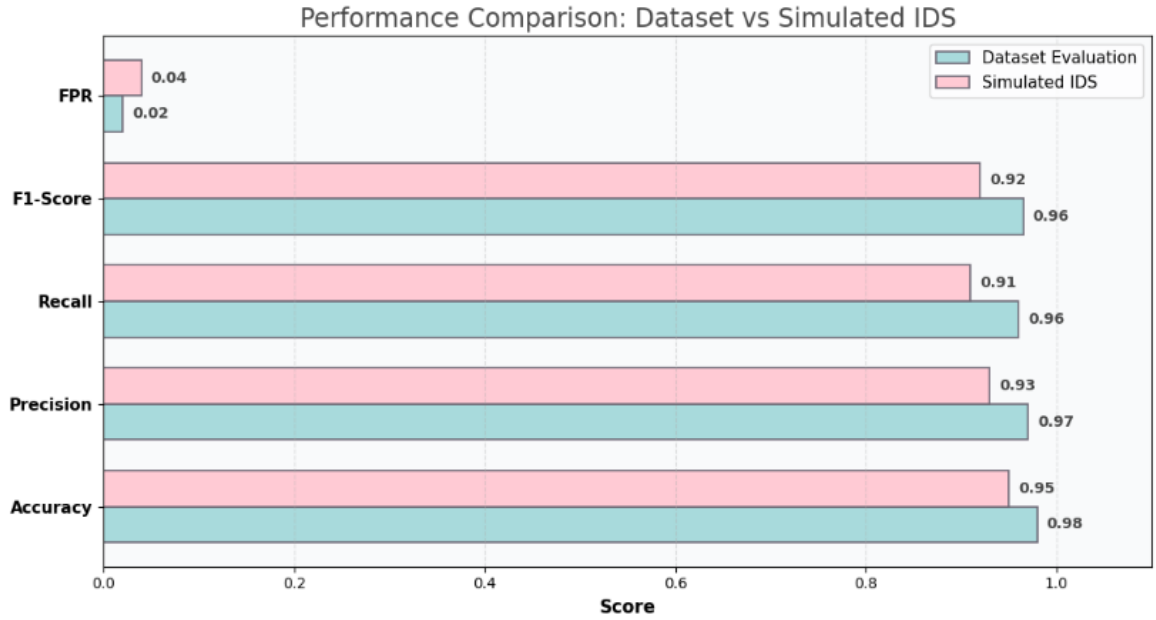
- **Receiver Operating Characterstics(ROC) Curve Analysis:** To assess a predictive framework's reliability ROC graph is used. Below chart visually demonstrates how effectively the underlying algorithm separates distinct binary classifications.



- **Training & Validation Loss Graph:** This graph represent that how much error is computed during training and the error computed at the time of validation of unseen data which is not seen during training.



- **Model Performance Comparison:**



3.7 Conclusion:

In this research work, in modern network environment to overcome from the problem of Zero-day attacks and adversarial attacks a robust NIDPS system is developed. Traditional and

signature- based supervised learning models are enabled to detect new attacks because they are trained only on labeled data. In this project we use Denoising Autoencoder architecture on UNSW-NB15 dataset.

3.8 Future Scope: In current Deep Learning and Autoencoder model treat as "Black Box " testing. These models detect attack but not determined that why this packet treat as an attack. In future SHAP and LIME AI frameworks can integrate with it so that administrator recognized that why alert triggers. Because of data privacy and cybersecurity policies many organizations cannot share their raw network traffic data. To solve this problem autoencoder model can be extended on federated learning architecture. This model train on local different nodes and no need of sharing data on central server. It continued data privacy for organizations.

4. References

[1] M. Alkasassbeh, E. H. Omoush, M. Almseidin, and A. Aldweesh, "A self-adaptive intrusion detection system for zero-day attacks using deep Q-networks," *IEEE Access*, vol. 13, pp. 174280–174295, 2025. doi: 10.1109.

[2] Y. Guo, "A review of machine learning-based zero-day attack detection: Challenges and future directions," *Computer Communications*, vol. 198, pp. 50–65, 2023. doi: 10.1016/j.comcom.2022.11.001.

[3] S. Ali, S. U. Rehman, A. Imran, G. Adeem, Z. Iqbal, and K.-I. Kim, "Comparative Evaluation of AI-Based Techniques for Zero-Day Attacks Detection," *Electronics*, vol. 11, no. 23, p. 3934, 2022. doi: 10.3390/electronics11233934.

[4] V. Kumar and D. Sinha, "A Robust Intelligent Zero-Day Cyber-Attack Detection Technique," *Complex & Intelligent Systems*, vol. 8, no. 2, pp. 1333–1354, 2022. doi: 10.1007/s40747-021-00396-9.

[5] V. Heydari and K. Nyarko, "Enhancing Adversarial Robustness in Network Intrusion Detection: A Novel Adversarially Trained Neural Network Approach," *Electronics*, vol. 14, no. 16, p. 3249, 2025. doi: 10.3390/electronics14163249.

[6] S. M. Kasongo and Y. Sun, "Performance analysis of intrusion detection systems using a feature selection method on the UNSW NB15 dataset," *Journal of Big Data*, vol. 7, p. 105, 2020. doi: 10.1186/s40537-020-00379-6.

[7] B. Tafreshian and S. Zhang, "A Defensive Framework Against Adversarial Attacks on Machine Learning-Based Network Intrusion Detection Systems," *arXiv preprint*, 2025. doi: 10.48550/arXiv.2502.15561.

[8] M. Sarhan, S. Layeghy, M. Gallagher, and M. Portmann, "From zero-shot machine learning to zero-day attack detection," *International Journal of Information Security*, vol. 22, pp. 947–959, 2023. doi: 10.1007/s10207-023-00676-0.

[9] S. Sharma and Z. Chen, "A Systematic Study of Adversarial Attacks Against Network Intrusion Detection Systems," *Electronics*, vol. 13, no. 24, p. 5030, 2024. doi: 10.3390/electronics13245030.

[10] S. Moraboena, G. Ketepalli, and P. Ragam, "A Deep Learning Approach to Network Intrusion Detection Using Deep Autoencoder," *Revue d'Intelligence Artificielle*, vol. 34, no. 4, pp. 457–463, 2020. doi: 10.18280/ria.340410.