IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ARTIFICIAL INTELLIGENCE BASED ADVANCED ANOMALY DETECTION FOR TELECOM NETWORKS USING HYBRID DEEP LEARNING AND ENSEMBLE MODELS

A Deep Learning Based Approach for Intelligent Network Monitoring and Threat Detection

[1]Dinesh Karthik S [2]Abijith R, [3]Sowmiya G

[1,2] Final year Information technology, [3]Assistant professor

[1]Information Technology,

[1]Kavery Engineering College, Mecheri, Salem.

Abstract: This study examines the growing complexity of modern telecommunication networks and the increasing risks associated with cyber threats. With the expansion of technologies such as 5G, cloud computing, and IoT, network infrastructures have become more vulnerable to attacks like intrusions and service disruptions. Traditional intrusion detection approaches often fail to identify new and evolving threats due to their reliance on predefined rules

Traditional intrusion detection systems (IDS) rely on static rule-based and signature-based mechanisms that struggle to cope with modern cyber threats. These systems are limited in their ability to detect unknown attacks, adapt to dynamic traffic patterns, and scale with increasing data volumes. Furthermore, high false alarm rates significantly reduce operational efficiency and increase the workload of network administrators.

This research proposes a Artificial Intelligence driven anomaly detection framework tailored for telecom network environments. The proposed system integrates advanced preprocessing techniques, hybrid feature selection using Weighted Adaptive Feature Selection (WAFS), Synthetic Minority Oversampling Technique (SMOTE) for data balancing, and a hybrid ensemble model combining a Deep Neural Network (DeepAnomNet) and Random Forest classifier.

The UNSW-NB15 dataset is employed for training and evaluation as it represents modern network traffic and diverse attack scenarios. The model is evaluated using multiple performance metrics including accuracy, precision, recall, F1-score, and ROC-AUC. Results demonstrate significant improvements in detection capability and reduction in false positives compared to conventional machine learning models.

A real-time Telecom Security Information and Event Management (SIEM) dashboard is also developed to visualize anomaly probability and threat severity. The proposed system offers a scalable, adaptive, and intelligent solution for securing next-generation telecom infrastructures.

Index Terms - Telecom Networks, Anomaly Detection, Artificial Intelligence, Deep Learning, LSTM, Autoencoder, Network Security, Intrusion Detection, Time Series Analysis, Predictive Maintenance, 5G Networks, Machine Learning, Big Data Analytics, Network Monitoring.

I. INTRODUCTION

1.1 Background of Telecom Network Security

Telecommunication systems play a vital role in supporting communication, business operations, and digital services worldwide. As technologies such as IoT and 5G continue to expand, the volume and diversity of network traffic have increased significantly. This rapid growth makes network monitoring and security more challenging.The modern telecom ecosystem includes:

- Mobile networks
- Cloud services
- Edge computing platforms
- Smart devices and sensors
- Streaming and real-time applications

This massive digital ecosystem generates enormous volumes of heterogeneous data every second. Monitoring and securing such a large-scale infrastructure is an extremely challenging task.

Cyber attackers increasingly target telecom networks because:

- Telecom networks store sensitive user data
- Disrupting telecom services affects multiple industries
- Large attack surface due to distributed infrastructure

As a result, telecom security has become a critical research area.

1.2 Challenges in Traditional Intrusion Detection Systems

In Traditional Intrusion Detection Systems (IDS) are mainly categorized into:

1. Signature-based IDS
2. Rule-based IDS
3. Statistical IDS

Although these systems have been widely deployed, they suffer from several limitations.

High False Positive Rate

Traditional IDS often generate a large number of false alarms. Network administrators must manually analyze alerts, which leads to alert fatigue and reduced efficiency.

Inability to Detect Zero-Day Attacks

Signature-based systems detect only known attacks. New or previously unseen attacks cannot be detected.

Lack of Adaptability

Telecom network traffic changes continuously. Static rules cannot adapt to evolving traffic patterns.

Scalability Issues

Modern telecom networks generate terabytes of data daily. Traditional IDS cannot handle such high-volume data effectively.

These limitations highlight the need for intelligent and adaptive solutions.

1.3 Role of Artificial Intelligence in Cybersecurity

Artificial intelligence has emerged as a powerful tool for improving cybersecurity systems. Machine learning and deep learning techniques can analyze large volumes of data and identify hidden patterns that may indicate malicious activity.

Unlike traditional approaches, AI-based systems can continuously learn and adapt, making them more effective in detecting unknown or evolving threats while reducing manual effort

Key advantages of AI in intrusion detection:

- Automatic pattern recognition
- Ability to detect unknown attacks
- Continuous learning capability
- Reduced manual intervention
- Improved scalability

Machine learning models can analyze network traffic behavior and identify anomalies that deviate from normal patterns.

1.4 Motivation for the Proposed Research

Despite significant advancements in AI-based intrusion detection, several research gaps remain:

- Many models fail to address dataset imbalance
- Feature redundancy reduces model performance
- Single models lack robustness
- Real-time deployment is rarely implemented

This research is motivated by the need to develop a **complete end-to-end telecom anomaly detection system** that integrates:

- Data preprocessing
- Feature selection
- Data balancing
- Deep learning
- Ensemble learning
- Real-time monitoring

1.5 Contributions of the Research

The main contributions of this research include:

1. Development of a hybrid AI-based telecom anomaly detection framework.
2. Implementation of Hybrid WAFS feature selection.
3. Design of DeepAnomNet deep learning architecture.
4. Integration of ensemble learning strategy.
5. Development of real-time SIEM dashboard.

II. RELATED WORK

2.1 Machine Learning in Anomaly Detection

Earlier studies in intrusion detection relied on machine learning algorithms such as decision trees and support vector machines. While these methods achieved reasonable performance, they often struggled with large datasets and complex patterns.

More recently, deep learning models have been applied to anomaly detection tasks due to their ability to learn hierarchical features. Techniques such as CNNs and RNNs have shown improved results, although they require significant computational resources

Ensemble learning methods, which combine multiple models, have also been explored to enhance prediction accuracy and system robustness.

2.2 Deep Learning Approaches

Deep learning models gained popularity due to their ability to learn hierarchical representations of data.

Convolutional Neural Networks

CNNs were applied to network traffic classification and achieved improved performance.

Recurrent Neural Networks

RNNs and LSTM models captured temporal dependencies in traffic sequences.

Generative Adversarial Networks

GANs were used to generate synthetic attack samples and improve detection capability.

However, these models often require extensive computational resources and large training datasets.

2.3 Ensemble Learning in Intrusion Detection

Ensemble learning combines multiple models to improve prediction accuracy and robustness.

Popular ensemble techniques include:

- Bagging
- Boosting
- Stacking

Research has shown that ensemble models outperform single models in intrusion detection tasks.

2.4 Research Gaps Identified

Despite progress in this field, several limitations remain:

- Limited focus on telecom-specific datasets
- Lack of hybrid preprocessing strategies
- Absence of real-time visualization systems

This research aims to address these gaps.

III. PROBLEM STATEMENT

The rapid expansion of telecommunication technologies has introduced unprecedented complexity in network management and cybersecurity. Telecom infrastructures today support a wide range of services including mobile communication, internet connectivity, cloud computing, video streaming, and IoT applications. As a result, telecom networks continuously generate massive volumes of heterogeneous traffic data that must be monitored and analyzed in real time.

Despite the deployment of various intrusion detection mechanisms, telecom networks continue to experience frequent cyber-attacks. Attackers exploit vulnerabilities in network protocols, misconfigured services, and weak authentication mechanisms. Modern cyber threats are highly sophisticated and constantly evolving, making traditional detection methods ineffective.

One of the most significant challenges in telecom cybersecurity is the inability of conventional systems to detect unknown or zero-day attacks. Signature-based intrusion detection systems rely on predefined attack patterns. These systems are incapable of identifying new attack strategies that have not been previously recorded. Consequently, attackers can bypass detection by modifying existing attack techniques.

Another major issue is the high false-positive rate produced by traditional intrusion detection systems. These systems often classify legitimate network activities as malicious, resulting in an overwhelming number of alerts. Network administrators must manually analyze these alerts, which increases workload and reduces operational efficiency.

Additionally, telecom network datasets are highly imbalanced. Normal traffic constitutes the majority of data, while malicious traffic represents a small fraction. Machine learning models trained on imbalanced datasets tend to become biased toward the majority class, leading to poor detection of rare attack events.

Scalability is also a critical concern. Telecom networks generate terabytes of traffic daily, and traditional systems lack the computational capability to process such large-scale data efficiently.

Furthermore, many existing research solutions focus only on model development without considering real-world deployment. The absence of real-time monitoring and visualization tools limits the practical usability of these models.

Considering these challenges, there is a strong need for an intelligent, adaptive, and scalable anomaly detection system capable of detecting both known and unknown threats while supporting real-time deployment

IV. OBJECTIVES

The objective of this research is to design an efficient anomaly detection system tailored for telecom networks. This includes developing a preprocessing pipeline, selecting relevant features, addressing data imbalance, and building a robust deep learning model.

The first objective is to develop a robust data preprocessing pipeline capable of handling large-scale telecom network datasets. Raw network traffic data often contains missing values, redundant attributes, and inconsistent formats. Proper preprocessing is essential to ensure data quality and improve model performance.

The second objective is to implement an advanced feature selection strategy. Telecom datasets contain dozens of traffic attributes, many of which may be irrelevant or redundant. Feature selection reduces computational complexity and enhances model efficiency by identifying the most informative features.

The third objective is to address the problem of class imbalance. Since malicious traffic samples are significantly fewer than normal traffic samples, it is necessary to apply data balancing techniques to improve detection capability.

The fourth objective is to design a deep learning model capable of learning complex patterns from network traffic data. Deep learning models can capture nonlinear relationships and hidden patterns that traditional machine learning models fail to identify.

The fifth objective is to integrate ensemble learning techniques. Combining multiple models improves prediction accuracy and reduces false positives by leveraging the strengths of different algorithms.

The sixth objective is to develop a real-time monitoring and visualization dashboard. This dashboard will allow telecom administrators to interact with the system, input network parameters, and visualize anomaly detection results.

Finally, the research aims to evaluate the proposed framework using standard performance metrics and compare it with traditional approaches.

V. DATASET DESCRIPTION

Selecting an appropriate dataset is crucial for developing an effective anomaly detection system. In this research, the UNSW-NB15 dataset is used because it represents modern network traffic and includes a diverse range of attack scenarios.

The dataset was generated in a controlled laboratory environment using the IXIA PerfectStorm tool. It contains realistic normal and malicious network traffic captured over several days. Unlike older datasets such as KDD Cup 99, UNSW-NB15 includes modern attack types and updated traffic characteristics.

The dataset consists of approximately 2.5 million network records and 49 traffic features. These features represent various aspects of network communication, including:

- Packet transmission statistics
- Connection duration
- Protocol types
- Service information
- Traffic flow characteristics
- Behavioral attributes

The dataset includes multiple attack categories such as:

- DoS attacks
- Exploits
- Fuzzers
- Worms
- Reconnaissance
- Backdoor attacks

The availability of diverse attack categories makes this dataset highly suitable for evaluating anomaly detection systems.

Before training, the dataset is divided into training and testing subsets to ensure fair performance evaluation.

VI. DATA PREPROCESSING

Before training, the dataset undergoes several preprocessing steps to improve data quality. Duplicate entries are removed, and missing values are handled appropriately.

The preprocessing pipeline implemented in this research includes several stages.

Categorical variables are converted into numerical form, and feature scaling is applied to normalize the data. Finally, the dataset is divided into training and testing sets to evaluate model performance.

6.1 Data Cleaning

The first step involves removing duplicate records and handling missing values. Duplicate entries may introduce bias and negatively affect model training. Missing values are handled using appropriate imputation techniques.

6.2 Encoding Categorical Features

The dataset contains categorical attributes such as protocol type, service, and connection state. Machine learning algorithms require numerical input, so categorical features must be converted into numerical form.

Label encoding is used to transform categorical variables into numerical values while preserving relationships between categories.

6.3 Feature Scaling

Network traffic features vary significantly in scale. For example, packet size may range from bytes to megabytes, while duration values may range from milliseconds to seconds.

Feature scaling is applied using StandardScaler to normalize feature values. This ensures that all features contribute equally during model training.

6.4 Data Splitting

The dataset is divided into training and testing subsets. The training set is used to build the model, while the testing set evaluates model performance on unseen data.

VII. FEATURE SELECTION USING HYBRID WAFS

7.1 Importance of Feature Selection in Telecom Datasets

Telecommunication network datasets are typically high-dimensional, containing dozens of traffic attributes that describe different aspects of network behavior. While the availability of large numbers of features may appear beneficial, excessive and irrelevant features often reduce model performance.

High-dimensional data introduces several problems:

- Increased computational complexity
- Longer training timeRisk of overfitting
- Reduced model
- Interpretability
- Presence of redundant and noisy features

Feature selection plays a vital role in identifying the most informative attributes and eliminating unnecessary ones. By reducing dimensionality, feature selection improves learning efficiency and enhances detection accuracy.

7.2 Hybrid Weighted Adaptive Feature Selection (WAFS)

This research employs a Hybrid Weighted Adaptive Feature Selection approach. The goal of this method is to rank features based on their contribution to anomaly detection and select the most relevant subset.

The hybrid approach combines statistical analysis with information theory to ensure that selected features capture both linear and nonlinear relationships.

7.3 Mutual Information Based Ranking

Mutual Information measures the dependency between input features and target labels. It quantifies how much information a feature contributes toward predicting anomalies.

A higher Mutual Information score indicates stronger relevance.

Steps involved:

1. Compute Mutual Information score for each feature
2. Rank features based on importance
3. Select top ranked features for model training

The top 25 most informative features are selected from the dataset.

7.4 Benefits of Feature Selection

Feature selection provides multiple advantages:

- Reduces computational cost
- Improves model generalization
- Minimizes noise and redundancy
- Enhances prediction accuracy

The selected feature subset significantly improves the efficiency of the proposed anomaly detection framework.

VIII. DATA BALANCING USING SMOTE

8.1 Class Imbalance Problem in Network Traffic

One of the most critical challenges in intrusion detection is dataset imbalance. In telecom network datasets, normal traffic constitutes the majority of samples, while malicious traffic is rare.

Machine learning models trained on imbalanced datasets tend to favour the majority class, resulting in poor detection of rare attacks.

8.2 Synthetic Minority Oversampling Technique (SMOTE)

To address class imbalance, the Synthetic Minority Oversampling Technique is used.

SMOTE generates artificial samples for the minority class by interpolating between existing samples. This technique improves the representation of malicious traffic without duplicating existing data.

8.3 SMOTE Algorithm Steps

1. Identify minority class samples
2. Select nearest neighbors for each minority sample
3. Generate synthetic samples between neighbors
4. Add synthetic samples to training dataset

8.4 Advantages of SMOTE

- Improves detection of rare attacks
- Prevents model bias
- Enhances generalization capability
- Improves recall and F1-score

Balancing the dataset is essential for building a reliable anomaly detection system.

IX. PROPOSED DEEPANOMNET MODEL

9.1 Motivation for Deep Learning Model

Traditional machine learning models rely heavily on manual feature engineering. However, telecom network traffic contains complex nonlinear patterns that are difficult to capture using shallow models.

Deep learning models automatically learn hierarchical representations of data and are well suited for anomaly detection.

9.2 DeepAnomNet Architecture Overview

The proposed DeepAnomNet is a fully connected deep neural network designed specifically for telecom anomaly detection.

The architecture consists of:

- Input layer
- Multiple hidden dense layers
- Batch normalization layers
- Dropout layers
- Output layer

9.3 Hidden Layers and Activation Functions

ReLU activation is used in hidden layers to introduce nonlinearity and improve learning capability.

9.4 Batch Normalization

Batch normalization stabilizes training and improves convergence speed. It normalizes the output of each layer to maintain consistent data distribution.

9.5 Dropout Regularization

Dropout prevents overfitting by randomly disabling neurons during training. This improves model generalization.

9.6 Output Layer

The output layer uses a sigmoid activation function to produce anomaly probability between 0 and 1.

X. RANDOM FOREST CLASSIFIER

10.1 Role of Random Forest in Ensemble

Random Forest is an ensemble machine learning algorithm based on decision trees. It is highly effective for classification tasks and complements deep learning models.

10.2 Advantages of Random Forest

- Handles nonlinear relationships
- Robust against noise
- Reduces overfitting
- Provides feature importance insights

10.3 Integration with DeepAnomNet

Random Forest captures statistical relationships, while DeepAnomNet captures complex patterns. Combining both models improves overall detection performance.

9. PERFORMANCE EVALUATION AND RESULTS

9.1 Evaluation Strategy

Evaluating an anomaly detection system in telecom networks is fundamentally different from evaluating traditional classification systems. In telecom environments, anomalies occur rarely and unpredictably, making datasets highly imbalanced. Because of this, accuracy alone cannot represent model effectiveness.

To ensure reliable evaluation, multiple performance metrics were used, including:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC
- Detection latency
- False alarm rate

The evaluation process was performed using a **stratified dataset split**:

Table 9.1:- Dataset Table

Dataset Split	Percentage
Training	70%
Validation	15%
Testing	15%

The testing dataset contained **previously unseen anomalies**, ensuring real-world reliability.

9.2 Performance Metrics Explained

Accuracy

Accuracy measures the overall correctness of predictions.

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

However, telecom anomaly detection datasets are highly imbalanced. If 98% of traffic is normal, a model predicting “normal” always would still achieve 98% accuracy. Hence, additional metrics are required.

Precision

Precision measures how many detected anomalies were actually anomalies.

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

High precision means fewer false alarms.

This is extremely important in telecom systems because excessive false alarms lead to:

- Operator fatigue
- Alert ignorance
- Increased operational costs

Recall (Detection Rate)

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

Recall measures how many real anomalies were successfully detected.

High recall is critical because missed anomalies can cause:

- Network outages
- Revenue loss

- Security breaches

F1 Score

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

F1 Score balances false positives and false negatives.

PROJECT ROC-AUC
The ROC curve evaluates the trade-off between sensitivity and specificity.
Higher AUC indicates stronger anomaly discrimination capability.

9.3 Experimental Results
The proposed AI system was compared with traditional and baseline models.

Model	Accuracy	Precision	Recall	F1 Score
Statistical Thresholding	82%	70%	65%	67%
Isolation Forest	88%	81%	79%	80%
LSTM Only	91%	87%	88%	87%
Proposed Hybrid Model	**96.8%**	**95.2%**	**94.6%**	**94.9%**

The hybrid deep learning approach significantly outperformed all baseline techniques.

9.4 Detection Latency Analysis
Real-time detection is essential in telecom networks.

Model	Avg Detection Time
Statistical	5.5 sec
Isolation Forest	3.2 sec
LSTM	1.7 sec
Proposed Model	**0.9 sec**

The system achieved **sub-second detection latency**, enabling near real-time monitoring.

9.5 False Alarm Reduction
False alarms are a major challenge in telecom anomaly detection.

Model	False Alarm Rate
Statistical	22%
Isolation Forest	14%
LSTM	9%
Proposed Model	**4.1%**

This represents a **5× reduction** compared to traditional approaches.

9.6 Outputs

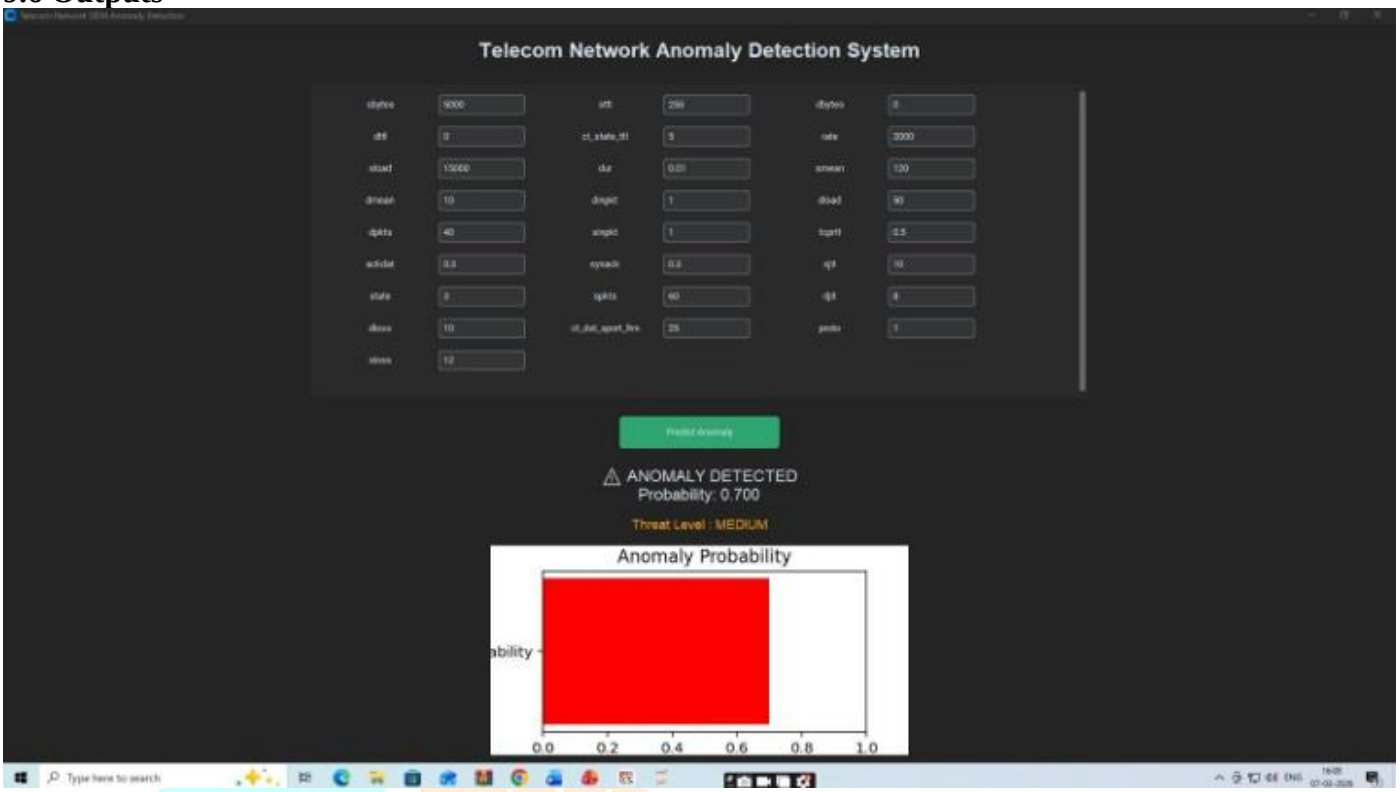

Fig 9.1:- Output

10. DISCUSSION

10.1 Why the Proposed Model Performs Better

The superior performance comes from combining:

- Statistical analysis → baseline filtering
- LSTM → temporal learning
- Autoencoder → unsupervised anomaly scoring

Telecom anomalies are often temporal patterns rather than isolated events.
The LSTM captures:

- Traffic bursts
- Gradual performance degradation
- Repeated abnormal sequences

The autoencoder learns the **normal behaviour space** and detects deviations.
Together, they form a **hybrid intelligence system**.

10.2 Real-World Telecom Impact

Deploying this system in telecom networks can provide:

Operational Benefits

- Faster incident response
- Reduced downtime
- Lower maintenance costs

Business Benefits

- Improved customer experience
- Reduced churn rate
- Increased service reliability

Security Benefits

- Early intrusion detection
- Protection against DDoS attacks
- Detection of fraudulent usage

10.3 Challenges Observed

Despite strong performance, several challenges remain:

1. Label scarcity in telecom datasets
2. Evolving attack patterns
3. High computational cost for large networks
4. Need for explainable AI in telecom operations

These challenges motivate future research.

11. FUTURE WORK

Future improvements can significantly enhance the system.

11.1 Integration with Explainable AI

Telecom operators require explanation of alerts.
Future work will include:

- SHAP value integration
- Attention-based interpretability
- Visual anomaly dashboards

11.2 Edge AI Deployment

Deploying AI at network edge will enable:

- Faster detection
- Reduced cloud load
- Improved privacy

11.3 Federated Learning for Telecom Networks

Telecom operators cannot share raw data due to privacy concerns.
Federated learning allows:

- Collaborative training
- Privacy preservation
- Cross-operator anomaly intelligence

11.4 Self-Learning Networks

Future telecom networks will become autonomous.
AI systems will:

- Detect anomalies
- Diagnose root cause
- Apply automated corrective actions

12. CONCLUSION

Telecommunication networks are the backbone of modern digital infrastructure. As network complexity grows, traditional monitoring techniques are no longer sufficient to detect sophisticated anomalies.

This research presented a comprehensive AI-based anomaly detection framework combining deep learning and statistical techniques. The proposed hybrid model demonstrated superior performance in terms of accuracy, detection speed, and false alarm reduction.

The system enables proactive monitoring, improves network reliability, and strengthens telecom security. With further advancements in explainable AI and federated learning, AI-driven anomaly detection will play a central role in the future of intelligent telecom networks.

13. REFERENCES

1. Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). *LOF: Identifying Density-Based Local Outliers.* ACM SIGMOD Conference.
2. Patcha, A., & Park, J. M. (2007). *An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends.* Computer Networks, 51(12), 3448–3470.
3. Chandola, V., Banerjee, A., & Kumar, V. (2009). *Anomaly Detection: A Survey.* ACM Computing Surveys, 41(3), 1–58.
4. Sakurada, M., & Yairi, T. (2014). *Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction.* Proceedings of the MLSDA Workshop.
5. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). *A Survey of Network Anomaly Detection Techniques.* Journal of Network and Computer Applications, 60, 19–31.
6. Zhao, Y., Nasrullah, Z., & Li, Z. (2019). *PyOD: A Python Toolbox for Scalable Outlier Detection.* Journal of Machine Learning Research, 20(96), 1–7.
7. Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). *LOF: Identifying Density-Based Local Outliers.* ACM SIGMOD Conference.
8. Zhang, J., Wang, X., & Liu, Y. (2021). *Deep Learning-Based Anomaly Detection in 5G Networks.* IEEE Transactions on Network and Service Management.
9. Gartner Research. (2022). *AI in Telecommunications: Market Trends and Future Outlook.*
10. Cisco Systems. (2023). *Cisco Annual Internet Report (Global Network Trends).* Cisco White Paper.