Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 46 (2025), No. 6, pp. 1881–1890
DOI : 10.47974/JIOS-2017

Hybrid machine learning model for network traffic anomaly detection using time-series forecasting

Pratik Kumar Swain [§]
Faculty of Engineering and Technology
Sri Sri University
Cuttack 754006
Odisha
India

Ansuman Patnaik [†]
Stellantis Bangalore
Bangalore 560001
Karnataka
India

Suneeta Satpathy *
Center for Cyber Security
SOA University
Bhubaneswar 751030
Odisha
India

Abstract

This research study employs a hybrid model capable of detecting anomalies in network infrastructure to enhance cyber security. The proposed model analyzes anomalies and classifies cyber-attacks by combining ARIMA for time-series forecasting with advanced AI (Artificial Intelligence) models like autoencoders and Isolation Forests. ARIMA generates residuals based on deviations from predictions by capturing standard traffic patterns, which is then classified by AI models. The proposed approach leverages ARIMA's temporal dependency handling and robust AI classification along with addressing limitations in

[§] *ORCID-ID:* **0009-0005-2577-0508**
* *ORCID-ID:* **0000-0002-6943-171X**

[§] *E-mail:* **pratikkumarswain.official@gmail.com**
[†] *E-mail:* **ansuman.patnaik@stellantis.com**
* *E-mail:* **suneeta1912@gmail.com (Corresponding Author)**

© Taru Publications

traditional methods. The model justifies improved intrusion detection for dynamic cyber security environments with the detection of attacks like brute force, DDoS, and SQL injections when evaluated on the CIC-IDS 2018 dataset.

Subject Classification: *68M10, 68T05, 68T09.*

Keywords: *Anomaly detection, ARIMA, Autoencoders, Cyber-attacks, Isolation forest, Intrusion detection, Time-series forecasting.*

1. Introduction

Network infrastructure security has become crucial due to the exponential growth of digital communication and interconnected systems. Cyber-attacks, including brute force, man-in-the-middle, and Distributed Denial of Service (DDoS) attacks, pose significant risks to individuals, organizations, and governments [1] [2]. Traditional detection methods like signature-based, rule-based, and statistical techniques often fail to detect evolving cyber threats due to their limited adaptability and inability to handle complex, high-dimensional data. To address this, the ARIMA model is employed to capture temporal patterns in network traffic. However, ARIMA alone struggles with complex anomalies. Thus, machine learning models such as Auto-encoders and Isolation Forests are used to analyze ARIMA residuals for better threat detection. This study proposes a hybrid model combining ARIMA, Auto-encoders, Isolation Forests, and XGBoost which is validated using the CIC-IDS 2018 dataset. Further, a comprehensive literature review of existing studies on time series forecasting and anomaly detection using Machine Learning algorithms is done in Section 2 followed by the explanation of the proposed hybrid methodology in Section 3. Section 4 describes the results achieved by the proposed model along with a discussion followed by a conclusion and future scope of the work in Section 5.

2. Literature Review

Numerous studies have explored advanced machine learning and deep learning techniques to detect and mitigate cyber-attacks, yet many struggle with capturing temporal correlations in time series data. Kumar and Shilpa [3] introduced SAND, a statistical anomaly detection method using subsequence datasets and feature weighting to identify anomalies by measuring divergence between observed and predicted patterns. Choi et al. [4] compared deep learning algorithms for anomaly classification but

overlooked deployment challenges, such as model size and stability, crucial in resource-limited settings. Lai et al. [5] proposed a time series model for generating synthetic datasets and compared nine algorithms, though the study lacked robust model evaluations, limiting its real-world application. Other studies, like those by Kim et al. [6] and Shilpa, Kumar, and Jha [7], highlighted deep learning methods but did not fully address resource constraints or conceptual limitations. Bojarajulu and Tanwar [8] developed the DCAM model, though scalability was limited. Overall, the literature reveals a shift from traditional methods to hybrid approaches for more effective anomaly detection and cyber-attack prevention. Garg et al. [9] proposed GDN models with improvements in feature extraction and temporal dependency learning, while Shilpa, Kumar, and Jha [10] introduced an MST-GAT-based model for complex feature analysis. This shift from traditional to deep and graph-based models highlights the need for hybrid approaches to enhance anomaly detection and cyber-attack prevention.

3. Methodology

The proposed hybrid model combines ARIMA for time series forecasting with ML-based anomaly detection methods like Autoencoders and Isolation Forests. Final classification is done by XGBoost, identifying anomalies through forecasting, residual generation, and machine learning-based detection, as shown in Fig. 1.

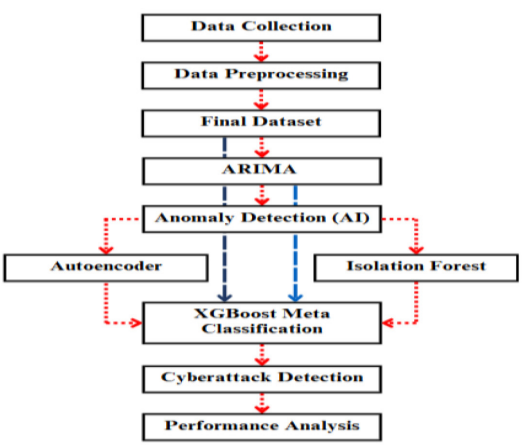

Figure 1

Proposed Hybrid Model

3.1 *Dataset Used*

The CIC-IDS 2018 dataset [11] is used to evaluate the proposed hybrid model for cyber-attack classification. It includes real-world benign and malicious traffic with 83 features and supports XGBoost classification of various attacks like DDoS, SQL Injection, and Brute force.

3.2 *Data Pre-Processing*

To evaluate the proposed model, the CIC-IDS 2018 dataset was pre-processed for time series forecasting, residual generation, anomaly detection, and final attack classification using labelled data. Sixteen features relevant to cyber-attack anomaly classification are selected for temporal deviation analysis as shown in Fig. 2, are further studied for use in time forecasting. Selected features are then normalized using Min-Max scaling to a 0–1 range to prevent dominance by larger numerical values. Thereafter Linear Interpolation estimates missing values using surrounding known data points.

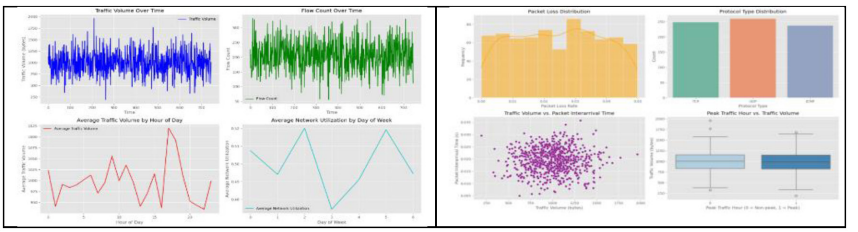

Figure 2

Key Features Analysis

The dataset is split 70:30 for training and testing; ARIMA learns temporal patterns from normal traffic, while ML detects anomalies from residuals. The hybrid model, primarily using XGBoost, classifies cyberattacks from the testing set containing both normal and attack traffic.

3.3 *Time-Series Forecasting*

ARIMA (AutoRegressive Integrated Moving Average) is a statistical model for time series forecasting, requiring stationary data confirmed via the ADF test and differencing. After ensuring stationarity, parameters (p, d, q) are selected using grid search and AIC. It forecasts future values by learning from historical patterns, with residuals representing prediction errors.

3.4 *Residual Analysis Using AI Techniques*

The residuals from the ARIMA model are analysed using Autoencoders and Isolation Forests, two unsupervised machine learning algorithms chosen for their effectiveness in anomaly detection.

3.3.1 Autoencoders

An Autoencoder is a neural network for unsupervised learning, processing ARIMA residuals to detect cyberattack-related anomalies. It includes an input layer, encoder with ReLU activation, latent space for key patterns, decoder, and output layer for reconstruction error, learning normal traffic patterns during training. In testing, high reconstruction errors indicate anomalies, aiding in identifying potential cyber threats, as illustrated in Figure 3 and 4.

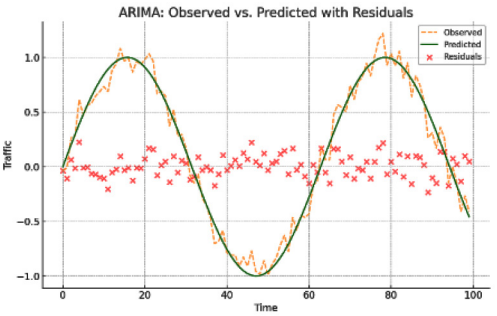

Figure 3

ARIMA Predictions

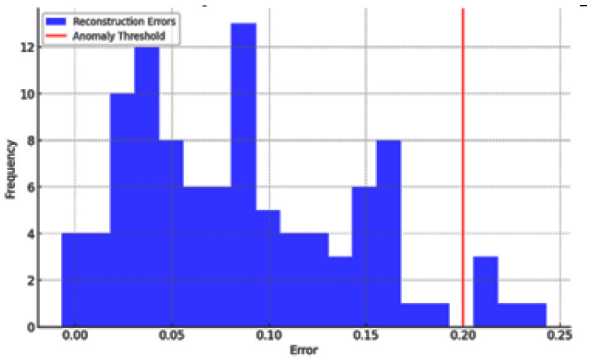

Figure 4

Autoencoder Reconstruction Errors (Residual Analysis)

3.3.2 Isolation Forests

In this approach, ARIMA-generated residuals are input into the Isolation Forest which is unsupervised anomaly detection for anomaly classification. ARIMA-generated residuals, representing network traffic deviations, are used as feature vectors. The model assigns anomaly scores based on fewer splits for anomalies. This scalable approach efficiently handles high-dimensional data, making it suitable for large datasets like network traffic. Higher scores suggest greater anomaly likelihood, and a decision boundary classifies each residual as normal or anomalous (Fig. 5).

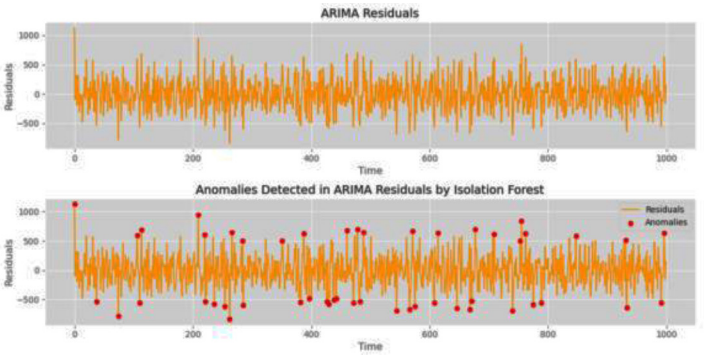

Figure 5

Anomaly Detection using Isolation Forest

3.4 *Meta-Classification with XGBoost*

XGBoost serves as a meta-classifier, combining ARIMA residuals, Autoencoder errors, and Isolation Forest scores to improve anomaly detection. Trained on labeled traffic data with optimized hyperparameters, it captures both temporal and complex patterns, enhancing network traffic anomaly detection.

3.5 *Evaluation Metrics*

The proposed hybrid model, combining ARIMA, autoencoder, Isolation Forest, and XGBoost, is evaluated on the CIC-IDS 2018 dataset using metrics like Accuracy, Precision, Recall, F1-score, and AUC to ensure effective and robust network anomaly detection performance.

4. Results & Discussions

The integration of ARIMA with AI techniques enhances network traffic anomaly detection, validated through metrics like accuracy, precision, recall, F1-score, and AUC. Analysis of specific attack types further illustrates the model's robustness and reliability, outperforming standalone methods.

4.1 *Time-Series Forecasting Results*

The ARIMA model effectively forecasts normal network traffic, capturing periodic patterns as shown in fig 6 with low prediction error (MAPE: 1.9%, RMSE: 0.024). Its precision enhances anomaly detection by minimizing noise. Optimized via AIC and grid search, ARIMA adapts well to dynamic environments, making it suitable for real-time traffic modelling and AI-based classification.

4.2 *Cyber Attack Detection Results*

The hybrid anomaly detection framework was evaluated for accuracy in detecting network traffic deviations. ARIMA + Autoencoder achieved an F1-score of 96.7%, while ARIMA + Isolation Forest improved recall to 97.8%. The combined hybrid approach with XGBoost outperformed all, reaching 98.4% accuracy and 98.8% precision, showcasing superior performance (as in fig 7).

4.3 *Attack-Wise Detection Accuracy*

Table 1 shows detection performance on the CIC-IDS 2018 dataset. DDoS attacks had 99.1% accuracy, while infiltration attacks reached 97.8%,

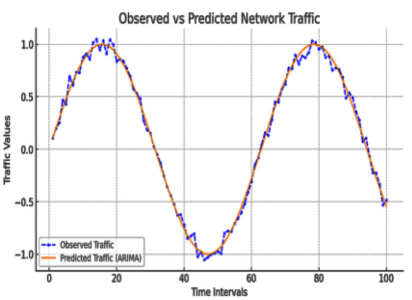

Figure 6

Time Forecasting Graph

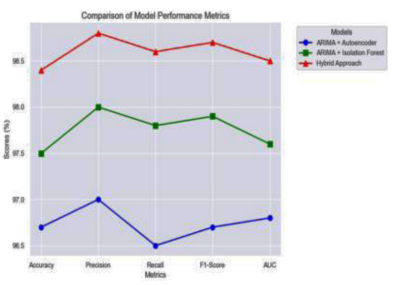

Figure 7

Cyber Attack Detection Comparison

Table 1
Anomaly Detection Comparison

Attack Type	Accuracy (%)	Precision (%)	Recall (%)	F-1 Score (%)
DDoS	99.1	98.9	99.3	99.1
Brute Force	98.7	98.4	98.9	98.6
Infiltration	97.8	97.6	98.0	97.8
SQLi	98.0	97.9	98.1	98.0
Benign	99.5	99.6	99.4	99.5

reflecting challenges in detecting stealthy threats. Benign traffic was classified with 99.5% accuracy.

4.4 Comparative Analysis with Existing Models

The proposed hybrid model's performance in cyber-attack detection is compared with existing models using ARIMA, Isolation Forest, and XGBoost.as shown in Table 2.

5. Conclusion & Future Scope

The proposed hybrid framework combines ARIMA for time-series forecasting with AI techniques like autoencoders, Isolation Forests, and

Table 2
Comparative Analysis in Classifying Cyber Attacks

Reference	Algorithms	Dataset	Accuracy
Keerthana et al. [12]	AARMA	Own Custom Dataset	89.7%
Kumar [13]	Graph-Based Machine Learning	CTU-13 and IoT-23	97.7%
Fouladi, Ermiş, and Anarim [14]	ARIMA + Chaos Theory	MAWILab	98.8%
Sharma, Patel, and Menaria [15]	ANN + CNN + LSTM + RNN	UNSW-NB15	96.9%
Garg et al. [9]	Customized Convolutional Neural Network (CCNN)	N-Balot	97.0%
Own Proposed Model	ARIMA + Autoencoder + Isolation Forest + XGBoost	CICIDS-2018	99.1%

XGBoost for anomaly detection. Evaluated on CIC-IDS 2018 DDoS attacks were detected with 99.1% accuracy and 99% F1-score, while infiltration attacks reached 97.8% for both. The model showed 99.5% accuracy in identifying benign traffic. Future improvements include GPU-accelerated forecasting and AI augmentation to enhance scalability and reduce latency.

References

[1] A. K. Saini, M. L. Saini, D. K. Srivastava, S. Singh, P. Vats, and T. K. Tak, "Secure communication framework using lattice-based encryption and image steganography," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 28, no. 5-B, pp. 1855–1864 (2025).

[2] S. Satpathy, P. K. Swain, S. N. Mohanty, and S. S. Basa, "Enhancing security: Federated learning against man-in-the-middle threats with gradient boosting machines and LSTM," in 2024 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–8 (2024).

[3] P. R. Kumar and B. Shilpa, "An IoT-based smart healthcare system with edge intelligence computing," in Reconnoitering the Landscape of Edge Intelligence in Healthcare, Apple Academic Press, pp. 31–46 (2024).

[4] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines," *IEEE Access*, vol. 9, pp. 120043–120065 (2021).

[5] K. H. Lai, D. Zha, J. Xu, Y. Zhao, G. Wang, and X. Hu, "Revisiting time series outlier detection: Definitions and benchmarks," in Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track (Round 1) (2021).

[6] S. Kim, K. Choi, H.-S. Choi, B. Lee, and S. Yoon, "Towards a rigorous evaluation of time-series anomaly detection," in Proc. AAAI Conf. Artificial Intelligence, vol. 36, no. 7, pp. 7194–7201 (2022).

[7] B. Shilpa, P. R. Kumar, and R. K. Jha, "LoRa DL: A deep learning model for enhancing the data transmission over LoRa using autoencoder," *J. Supercomput.*, vol. 79, no. 15, pp. 17079–17097 (2023).

[8] B. Bojarajulu and S. Tanwar, "Customized convolutional neural network model for IoT botnet attack detection," *Signal, Image Video Process.,* pp. 1–13 (2024).

[9] H. Garg, S. Jhunthra, R. Goel, S. Sharma, V. Gupta, R. Sharma, and P. Dass, "Understanding user polarisation regarding COVID-19 vaccines through social network analysis," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 27, no. 8, pp. 2301–2336 (2024).

[10] B. Shilpa, P. R. Kumar, and R. K. Jha, "Spreading factor optimization for interference mitigation in dense indoor LoRa networks," in 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), pp. 1–5 (2023).

[11] G. B. Mohammad, Shitharth, and P. R. Kumar, "Integrated machine learning model for URL phishing detection," *International Journal of Grid and Distributed Computing*, vol. 14, no. 1, pp. 513–529 (2021).

[12] N. Keerthana, H. M. A. Ghanimi, P. V. V. S. Srinivas, N. R. Pendli, F. T. Ayasrah, M. A. Bennet, S. Sengan, and P. Dadheech, "Blockchain-based cybersecurity: A predictive privacy and security of cyber-physical systems," *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 28, no. 5-B, pp. 1911–1922 (2025).

[13] P. R. Kumar, "Wireless mobile charger using inductive coupling," *J. Emerg. Technol. Innov. Res.* (JETIR), vol. 5, no. 10, pp. 40–44 (2018).

[14] R. F. Fouladi, O. Ermiş, and E. Anarim, "A DDoS attack detection and defense scheme using time-series analysis for SDN," *J. Inf. Secur. Appl.*, vol. 54, no. 102587 (2020).

[15] S. Sharma, A. K. Patel, and N. Menaria, "Integral transform of product of generalized k-Bessel function and generalized Hurwitz-Lerch Zeta function," *Journal of Interdisciplinary Mathematics*, vol. 27, no. 8, pp. 1757–1764 (2024).

Received December, 2024