

Assignment 1 Report

ML and DL Anomaly Detection for Enhancing 802.1X and Network Access Control

By Mohammed Baatiyyah - 261002493

Master of AI - 2026

Assignment content:

- Summary research papers.
- Critical Analysis.
- Connections.
- Conflicts between the findings.
- Synthesis Matrix for Literature Review:.
- References.

Summary research papers

Paper 1: Attack and Anomaly Detection in IoT Networks Using Supervised Machine Learning Approaches

In this paper the author talks about attracting cyber criminals to expose the IoT system to failure using IDS and Dataset of 70 GB contains instances belonging to majorly four categories of attack. Instead of using available feature reduction techniques like PCA, he develops attack-independent flow-based features. Multiple supervised classifiers were evaluated, including KNN, SVM, Logistic Regression, ANN, Decision Tree, and Random Forest to increase accuracy of the system to detect anomaly and normal traffic. Results show that Random Forest achieved the highest performance with 99.9% accuracy and AUC=1 for multiclass classification for detecting the anomaly and normal traffic.

Paper 2: Machine Learning in Network Anomaly Detection: A Survey

This paper discusses how to secure the large number of connected devices to the internet against anomaly detection that requires processing big data over a complex network structure with a prompt reaction. The paper presents a comprehensive review on the ML in network anomaly detection on IoT networks, Cloud/Fog/Edge networks and computer networks. From SL to RL, each category processes data in a different style, which leads to a large gap in the outcome. The techniques used for ML to test the prediction are Supervised, unSupervised, semi-Supervised and Reinforcement Learning. The study concludes that while ML significantly enhances anomaly detection capability, selecting the appropriate ML type depends on network structure, dataset availability, and attack characteristics.

Paper 3: Intrusion Detection Systems using Supervised Machine Learning Techniques: A survey

In this paper the author doing survey of intrusion detection systems (IDS) using supervised ML techniques.it study analyzes the application of supervised learning algorithms such as Random Forest, SVM, ANN, J48, and others across four widely used datasets: KDD'99, NSL-KDD, CICIDS2017, and UNSW-NB15.this paper also present a taxonomy the can guide what is suitable for algorithms across different types of IDS datasets, and it guides whether feature selection is effective or not in improving classification performance. Results show that supervised learning achieves high and promising classification accuracy, particularly Random Forest and Deep Neural Networks. Feature selection is often essential for performance improvement, data imbalance is a major concern, and large datasets benefit from deep learning techniques.

Paper 4: Combining Unsupervised Approaches for Near Real-Time Network Traffic Anomaly Detection

The authors in this paper use unsupervised machine learning techniques to improve anomaly detection accuracy for zero-day cyberattacks threats, while maintaining fast inference speed. Three hybrid models are introduced: DAGMM-EIF, DA-EIF, and MemAE-EIF with using datasets KDDCUP99, NSL-KDD, and CIC-IDS2017 to evaluate metrics used on this paper.Results show that the MemAE-EIF model achieved the best overall performance. Also, it is unnecessary to have, and often not possible to create, a training dataset that contains all the cyber attacks one wants to detect for network anomaly detection systems. The study also found the Unsupervised techniques to detect the anomaly threats are taking time from seconds to minutes.

Paper 5: Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques

This study investigates the application of various machine learning models for detecting anomalies in network traffic, specifically focusing on their effectiveness in addressing challenges such as class imbalance and feature complexity. The major purpose is to analyze the various intricacies of their performance and evaluate their usefulness in the context of network security. The study evaluates several models including Isolation Forest, Naïve Bayes, XGBoost, LightGBM, and Support Vector Machine (SVM) using the KDDCup'99 dataset. Results show that ensemble boosting models like LightGBM its the best performance, while Isolation Forest performed poorly. The study also mentioned that advanced ML models enhance network intrusion and anomaly detection compared to traditional ML.

Paper 6: Anomaly Detection in Encrypted Network Traffic Using Self-Supervised Learning

This paper proposes an Encrypted Traffic Anomaly Detection using Self_supervised Contrastive Learning (ET_SSL), a novel framework that detects anomalies in the aimed encrypted traffic without either the prior availability of labeled datasets or payload analysis. using three publicly available datasets: CIC-Darknet2020, ISCX VPN-nonVPN and UNSW-NB15. The evaluation measures that author used are detection rate, precision, recall, F1 measure, false positive rate (FPR) and throughput.the result is ET-SSL model achieved about 96.8% accuracy with a very low false positive rate (~1.2%) and can detect anomalies in encrypted traffic in with 15 ms to 25 ms latency.

Critical Analysis:

Previous studies have shown that artificial intelligence (AI) can be used to identify cyberattacks, including zero-day attacks, and to improve security systems compared to traditional methods. Supervised learning approaches provide high accuracy but rely heavily on labeled datasets and struggle to detect unknown attacks. Unsupervised and hybrid models offer better capabilities for identifying zero-day threats but may produce higher false positive rates.

Many studies have relied on datasets rather than a real-world machine learning environment, which may not be suitable for large-scale, diverse work environments. Future research should focus on real-time deployment, dataset quality, and reducing computational complexity while maintaining high detection accuracy.

Connection:

1- Use an AI for Anomaly Detection:

All papers focus on applying machine learning techniques to detect anomalies or cyber-attacks in network traffic, highlighting ML as a key solution for modern network security.

2- Focus on Intrusion Detection Systems (IDS):

Each study addresses the improvement of intrusion detection systems by analyzing network traffic patterns to identify malicious behavior especially on supervised ML.

3- Uses Datasets and Evaluation Metrics:

Most papers evaluate their models using standard cybersecurity datasets (e.g., KDD99, NSL-KDD, CICIDS) and metrics such as accuracy, precision, recall, and F1-score.

4- Goal of Improving Detection of Modern Cyber Threats:

All studies aim to enhance the ability to detect advanced threats such as unknown attacks, encrypted traffic anomalies, or large-scale network intrusions.

Conflicts between the findings:

1- Supervised vs. Unsupervised Learning Effectiveness:

Some papers report that supervised learning models achieve higher accuracy in intrusion detection, while others argue that unsupervised or self-supervised approaches are more effective for detecting unknown or zero-day attacks.

2- Need for Labeled Data:

Several studies emphasize the importance of large labeled datasets to train accurate models, whereas other papers propose approaches that avoid labeled data because it is difficult and expensive to obtain in real network environments.

3- Model Complexity vs. Real-Time Performance:

Some research suggests that deep learning and hybrid models improve detection accuracy, while other studies highlight that complex models increase computational cost and may not be suitable for real-time network monitoring.

4- Hardware and computing:

Some studies relied on previous research, including datasets, and these yielded good comparative results. Other studies focused on implementing the research in a standard operating environment, which helped achieve more positive results after training the machine learning.

Synthesis Matrix for Literature Review

Study	Authors	Year	Research Design	Key Findings/Results	Methodology	Limitations
Attack and Anomaly Detection in IoT Networks Using Supervised Machine Learning Approaches	Himani Tyagi	2020	Experimental	Flow-based Random Forest Performed Best with 99% accuracy with Very low training and testing time. Support Vector Machine (SVM) is not familiar with multiclassification as it takes longer as compared to other classifiers on training	Dataset Collection used BoT-IoT that contained DoS, DDoS, Reconnaissance, and Information Theft traffic. Use platform consists of Virtual Machines are managed through the vSphere platform. Data cleaning is performed throughout the process. The dataset used on the classifier and ratio of the train test set is taken as 70:30 respectively. feature set to Packet size, Protocols, Connection state, PRV, BRV, IPG and NR. <i>ML classifiers used KNN, RF, DT, ANN, LR.</i>	Dataset Dependency Results may not generalize to other real-world environments or different datasets. Limited Attack Types DOS, DDOS, Reconnaissance, Information Theft. The study uses only supervised machine learning.

Machine Learning in Network Anomaly Detection: A Survey	SONG WANG, JUAN FERNANO,	2021	Systematic Survey	No Single ML Model can be used for all network types. Reinforcement Learning Shows Adaptive Potential.	Its Survey Methodology based. Collects and reviews the existing old research. Based on results, the author analyzes and compares different machine learning techniques (SL,UL,SSL,RL).	There is no real-world data used, only dependent on what researchers had, and Limited Real-World Deployment Analysis.
Intrusion Detection Systems using Supervised Machine Learning Techniques: A survey	Emad E. Abdallah, Wafa' Eleisah, Ahmed Fawzi Otoom	2022	Survey	Machine learning is highly effective for IDS to get best result of accurate.feature selection is critical and required for performance improvement.For good performance, large intrusion detection data sets necessitate a deep learning technique.	Its survey methodology based. using four popular datasets: KDD'99, NSL-KDD, CICIDS2017, and UNSW-NB15 for IDS. Run the dataset on different ML models to get the best accuracy result.	No new model proposed for IDS implementation. Dependence on Existing Datasets. The study uses only supervised machine learning.
Combining Unsupervised Approaches for Near Real-Time Network Traffic Anomaly Detection	Francesco Carrera, Vincenzo Dentamaro, Stefano Galantucci, Andrea Iannacone, Donato Impedovo	2022	Experimental	MemAE-EIF got the best performance detecting anomaly with 0.9172 ms score. Hybrid models improved performance for detecting zero-day attacks.	This paper used Unsupervised anomaly detection on Hybrid architecture(Deep Learning + Extended Isolation Forest), with Three hybrid models were developed:DAGMM-EI, DA-EIF and MemAE-EIF.	Not Real-Time Production Data that can limitate Validation as it does not deploy the model in a real streaming system. Threshold tuning depends on knowing anomaly percentage.

					Datasets Used are:KDDCUP99, NSL-KDD, CIC-IDS2017.	
Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques	STEPHANIE NESS, VISHWANATH ESWARAKRISHNAN , HARISH SRIDHARA, VARUN SHINDE, NAGA VENKATA PRASAD JANAPAREDDY, AND VINEET DHANAWAT	2024	Experimental	LightGBM gets the best performance with 100% accuracy. SVM performed about 99% accuracy.	Autore uses KDDCup'99 dataset for anomaly detection, and did cleaning for missing value. Several ML models used on this study(Isolation Forest,Naïve Bayes,XGBoost,LightGBM). Data was split into a training set and a test set using an 80:20 split. The used evaluated performance metrics:Accuracy,Precision,RecallF1-score.	Not Real-Time Production Data that can limitate Validation as it does not deploy the model in a real streaming system. Threshold tuning depends on knowing anomaly percentage.
Anomaly Detection in Encrypted Network Traffic Using Self-Supervised Learning	Sadaf Sattar, Shumaila Khan, Muhammad Ismail Khan, Ainur Akhmediyarova,	2025	Experimental	ET-SSL model achieved high detection accuracy (up to 96.8%) for identifying anomalies in encrypted network traffic.	The authors use dataset encrypted (ISCX VPN-nonVPN,UNSW-NB15, CIC-Darknet2020). generate different traffic loads and evaluate the throughput, CPU usage and memory usage to	

	Orken Mamyrbaye, Dinara Kassymova, Dina Oralbekova & Janna Alimkulova				determine the scalability of ET-SSL.	
--	--	--	--	--	---	--

References:

Tyagi, H., & Kumar, R. (2020). Attack and anomaly detection in IoT networks using supervised machine learning approaches. *Revue d'Intelligence Artificielle*. [Link](#)

Wang, S., Balarezo, J. F., Kandeepan, S., Al-Hourani, A., Gomez Chavez, K., & Rubinstein, B. (2021). Machine learning in network anomaly detection: A survey. *IEEE Access*. [Link](#)

Abdallah, E. E., Eleisah, W., & Otoom, A. F. (2022). Intrusion Detection Systems using Supervised Machine Learning Techniques: A survey. *Procedia Computer Science*. [Link](#)

Carrera, F., Dentamaro, V., Galantucci, S., Iannacone, A., Impedovo, D., & Pirlo, G. (2022). Combining unsupervised approaches for near real-time network traffic anomaly detection. *Applied Sciences*. [Link](#)

Ness, S., Eswarakrishnan, V., Sridharan, H., Shinde, V., Janapareddy, N. V. P., & Dhanawat, V. (2025). Anomaly detection in network traffic using advanced machine learning techniques. *IEEE Access*. [Link](#)

Sattar, S., Khan, S., Khan, M. I., Akhmediyarova, A., Mamyrbayev, O., Kassymova, D., Oralbekova, D., & Alimkulova, J. (2025). Anomaly detection in encrypted network traffic using self-supervised learning. *Scientific Reports*. [Link](#)