iJRASET
International Journal For Research in
Applied Science and Engineering Technology

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68216

Machine Learning in Network Traffic Analysis: Classification, Optimization, and Security

Mr.Yash Gujarathi[1], Mr. Yash Potekar[2]
PVG College of Science and Commerce

***Abstract**: With the rapid expansion of digital networks, ensuring efficient and secure network traffic management has become a significant challenge. Traditional rule-based approaches struggle to handle evolving traffic patterns, particularly with the increasing use of encryption. Machine learning (ML) has emerged as a powerful alternative, providing enhanced capabilities for traffic classification, anomaly detection, and optimization. This paper presents a comprehensive review of ML-based techniques, including supervised learning, unsupervised learning, deep learning, and graph-based learning. Key challenges such as data imbalance, real-time processing, and computational overhead are explored. The study consolidates findings from multiple research papers, emphasizing the role of AI-driven models in improving cybersecurity, traffic prediction, and quality of service (QoS). Future research directions include hybrid models, federated learning, and the integration of ML with emerging networking paradigms such as Software-Defined Networking (SDN) and 5G.*

Keywords: Machine Learning, Network Traffic Analysis, Classification, Optimization, Security, Intrusion Detection

I. INTRODUCTION

The exponential growth of internet usage has led to increased demands for efficient network traffic classification, security, and optimization[1]. The ever-expanding digital infrastructure, driven by cloud computing, the Internet of Things (IoT), and 5G technology, has transformed how data is transmitted and processed across networks.[2] As organizations and consumers rely more on internet-based services, the need for intelligent network traffic analysis has become more critical than ever.

Traditional approaches, such as deep packet inspection and rule-based filtering, struggle to adapt to encrypted and dynamically evolving traffic patterns[3]. Encryption protocols such as Transport Layer Security (TLS) and Secure Sockets Layer (SSL) make it increasingly difficult for conventional methods to classify and monitor network behaviour without violating privacy regulations. Additionally, the rise of cyber threats and sophisticated malware attacks requires advanced analytical approaches to detect anomalies and mitigate potential security breaches.[4]

Machine learning (ML) techniques offer a robust alternative by leveraging statistical learning to generalize patterns and detect anomalies[5][6][7]. ML algorithms have demonstrated the ability to analyse large-scale network data efficiently, providing accurate classifications, intrusion detection, and predictive analytics. Supervised learning models such as decision trees and support vector machines (SVMs) have been widely used for traffic classification, while deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have significantly improved predictive capabilities.[3]

The importance of ML in network traffic research extends beyond security. Effective traffic classification contributes to better bandwidth management, quality of service (QoS) optimization, and congestion control.[8] With the integration of artificial intelligence (AI), automated traffic management systems can dynamically adjust network parameters, ensuring optimal performance across different environments. AI-powered traffic monitoring tools enhance cybersecurity frameworks, offering real-time threat detection and response mechanisms.[1] Furthermore, emerging paradigms such as Software-Defined Networking (SDN) and Network Function Virtualization (NFV) benefit from ML-driven traffic analysis, enabling more flexible and scalable network architectures.

Despite its advantages, ML-based network traffic analysis faces several challenges. The availability of labelled datasets, computational complexity, and the need for real-time processing pose significant hurdles to implementation. Additionally, privacy concerns related to data collection and analysis must be addressed, necessitating privacy-preserving techniques such as federated learning and differential privacy. These approaches ensure secure and decentralized data processing, reducing risks associated with centralized data storage.

Given the increasing complexity of modern network infrastructures, the role of ML in traffic analysis continues to evolve.[9] Future research will likely focus on hybrid ML models that combine deep learning with traditional approaches to enhance classification accuracy.

The adoption of edge computing and federated learning will enable decentralized and low-latency network monitoring solutions, making ML-driven traffic analysis more efficient and scalable. As digital transformation progresses, innovative ML methodologies will play a pivotal role in shaping the future of network security, optimization, and intelligence. The exponential growth of internet usage has led to increased demands for efficient network traffic classification, security, and optimization. Traditional approaches, such as deep packet inspection and rule-based filtering, struggle to adapt to encrypted and dynamically evolving traffic patterns. ML techniques offer a robust alternative by leveraging statistical learning to generalize patterns and detect anomalies [5][10][6].

II. MACHINE LEARNING APPROACHES IN NETWORK TRAFFIC ANALYSIS

A. Supervised Learning

Supervised ML models such as neural networks, decision trees, and support vector machines (SVMs) have been widely applied for network traffic classification. These models are trained on labelled datasets to predict network behaviours accurately [5][11][12]. Supervised learning provides highly accurate results when sufficient labelled data is available, making it a preferred method for intrusion detection and network anomaly detection. However, it requires a significant amount of labelled data, which can be costly and time-consuming to collect.

B. Unsupervised Learning

Clustering techniques such as K-means and DBSCAN are utilized for anomaly detection and traffic segmentation. These methods help classify traffic without requiring labelled datasets [13]. Unsupervised learning methods are particularly useful in identifying new types of cyber threats that have not been previously labelled. They can automatically detect unusual traffic patterns, making them highly effective for zero-day attack detection. However, they may produce false positives due to the absence of explicit labels, requiring additional refinement techniques.

C. Deep Learning

Deep learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), enable more complex network traffic classification. Graph neural networks (GNNs) further enhance classification performance by leveraging network topology structures [14][15]. Deep learning models, such as autoencoders and transformers, can extract complex features from network data, improving accuracy. Moreover, attention mechanisms in transformer-based models enhance the interpretability of network traffic patterns, making them increasingly valuable for real-time traffic monitoring.

D. AI-Based Optimization

Reinforcement learning and genetic algorithms have been explored to improve network performance and optimize traffic flow. Such approaches enhance quality of service (QoS) and mitigate congestion [16][17].Reinforcement learning (RL) models dynamically adapt to changing network conditions, improving efficiency. Genetic algorithms (GAs), inspired by evolutionary principles, optimize network traffic management by selecting the most efficient routing paths. These AI-driven optimization techniques are particularly useful in SDN-based architectures, where intelligent traffic routing significantly enhances performance.

III. GAPS AND CHALLENGES IN ML-BASED NETWORK TRAFFIC ANALYSIS

Despite the advancements in AI-driven traffic optimization, several challenges remain unresolved [18]. Scalability issues hinder the deployment of AI models in large-scale networks due to high computational costs and data processing constraints [19].Data privacy and security risks also pose concerns, particularly in federated learning-based models, which are vulnerable to adversarial attacks and data leakage [17]. Real-time adaptability remains a major challenge, as existing ML models often struggle to adjust dynamically to sudden traffic fluctuations and network failures [20]. Additionally, lack of standardized datasets and feature extraction methodologies affect the generalizability of AI-based network classification models [21].

A. Data Imbalance

Network traffic datasets often contain an uneven distribution of normal and anomalous instances, leading to biased ML models [12]. This imbalance can cause models to favour the majority class, resulting in poor anomaly detection and higher false negative rates. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE), cost-sensitive learning, and anomaly detection methods can help mitigate the impact of class imbalance by improving the model’s ability to recognize rare but significant events.

B. Encrypted Traffic Handling

The increasing adoption of encryption techniques complicates ML-based classification without violating privacy [9][13]. Traditional deep packet inspection (DPI) methods become ineffective, necessitating alternative approaches such as statistical flow-based features, federated learning, and privacy-preserving machine learning (PPML). However, these techniques often come with trade-offs in terms of accuracy, computational complexity, and scalability, making the effective classification of encrypted traffic a persistent challenge.

C. Computational Overhead

High-performance ML models require extensive computational resources, making real-time deployment challenging [14] [15] [16]. Deep learning models, particularly those involving recurrent neural networks (RNNs) or transformer-based architectures, can be resource-intensive, limiting their feasibility for edge or IoT devices. Optimizations such as model pruning, quantization, knowledge distillation, and the use of lightweight architectures (e.g., MobileNet or TinyML) can help reduce the computational burden while maintaining classification performance.

IV. KEY FINDINGS

ML-based classification achieves up to 97.6% accuracy using neural networks, demonstrating their effectiveness in network traffic analysis [22]. However, accuracy can vary depending on dataset quality, feature selection, and model tuning, emphasizing the need for continuous refinement.

Graph neural networks (GNNs) outperform traditional ML models in handling complex network structures, enabling improved traffic classification and anomaly detection [14]. GNNs excel at capturing relationships between network nodes, making them well-suited for identifying advanced threats and encrypted traffic patterns.

Random forest algorithms offer a balance between accuracy and efficiency, making them ideal for real-time applications *[23]*. Their ability to handle imbalanced datasets and provide interpretability makes them a preferred choice for practical deployments.

Edge AI models reduce latency by 30-50%, significantly enhancing real-time traffic analysis capabilities[24]. By processing data closer to its source, these models minimize dependence on centralized cloud resources while maintaining classification performance, making them well-suited for IoT and 5G networks.

Federated learning improves network security by decentralizing data training, reducing the risks associated with centralized data storage *[25]*. This approach ensures that sensitive network data remains on local devices, mitigating privacy concerns while enabling collaborative learning across distributed nodes.

V. FUTURE RESEARCH DIRECTIONS

Hybrid ML Models Combining Deep Learning and Traditional ML Approaches to Improve Accuracy and Efficiency:

Integrating traditional machine learning algorithms with deep learning techniques can leverage the strengths of both methods, enhancing model performance and interpretability. Hybrid approaches aim to combine the feature extraction capabilities of deep learning with the simplicity and speed of traditional algorithms, resulting in more robust and efficient models. For instance, a study discusses the integration of these techniques to improve clustering and classification tasks.[26]

Federated Learning for Privacy-Preserving Network Traffic Classification:

Federated Learning (FL) enables collaborative model training across multiple devices without sharing raw data, thus preserving privacy. Applying FL to network traffic classification allows for the development of models that can detect anomalies and classify traffic patterns while ensuring user data remains localized. Research has demonstrated the effectiveness of FL in creating accurate traffic classifiers without exposing sensitive information. [25]

Integration of ML Techniques with SDN and 5G for Enhanced Network Management:

Combining machine learning with Software-Defined Networking (SDN) and 5G technologies can lead to more intelligent and adaptive network management solutions. ML algorithms can analyse vast amounts of network data to optimize routing, predict traffic congestion, and enhance security protocols. Studies have explored the integration of ML with SDN in 5G networks to improve performance and quality of service. [27]

Hybrid Machine Learning Models Combining Deep Learning and Traditional Approaches to Improve Accuracy and Efficiency

Integrating deep learning techniques with traditional machine learning algorithms can leverage the strengths of both methodologies, enhancing classification performance and computational efficiency.

For instance, a study proposes a hybrid approach that first trains a deep network on the training data, extracts feature from the deep network, and then uses these features to re-express the data for input to a non-deep learning method for final classification[28]

VI. CONCLUSION

The integration of AI and ML techniques in network traffic optimization has shown promising results, particularly in enhancing security, improving resource allocation, and reducing latency [29].

Machine Learning (ML) has significantly advanced network traffic analysis by enhancing classification, anomaly detection, and optimization capabilities [4].By leveraging various ML techniques, including deep learning and traditional algorithms, models have achieved high accuracy in identifying complex traffic patterns and security threats. These advancements facilitate real-time decision-making, proactive threat mitigation, and adaptive network management strategies.

Despite these improvements, several challenges persist. Analysing encrypted traffic remains a significant obstacle, as traditional inspection methods are ineffective in privacy-preserving environments [30].

ML has revolutionized network traffic analysis by offering enhanced classification, anomaly detection, and optimization capabilities. Despite these advancements, challenges such as encrypted traffic analysis, real-time processing, and computational efficiency persist. Future research should focus on hybrid models, federated learning, and integration with next-generation networking technologies to further improve traffic classification and security [4][31].The findings of this study contribute to the evolution of intelligent, self-adaptive network traffic management frameworks capable of handling the growing demands of modern digital infrastructure [23].

REFERENCES

[1] D'Alconzo, I. Drago, A. Morichetta, M. Mellia, and P. Casas, 'A Survey on Big Data for Network Traffic Monitoring and Analysis', Mar. 2020, doi: 10.1109/TNSM.2019.2933358.

[2] M. Abbasi, A. Shahraki, and A. Taherkordi, 'Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey', Mar. 15, 2021, Elsevier B.V.doi: 10.1016/j.comcom.2021.01.021.

[3] M. R. Joshi and T. H. Hadi, 'A Review of Network Traffic Analysis and Prediction Techniques'.

[4] N. Alqudah and Q. Yaseen, 'Machine Learning for Traffic Analysis: A Review', in Procedia Computer Science, Elsevier B.V., 2020, pp. 911–916. doi: 10.1016/j.procs.2020.03.111.

[5] M. R. Parsaei, M. J. Sobouti, S. Raouf, and R. Javidan, 'Network Traffic Classification using Machine Learning Techniques over Software Defined Networks', 2017. [Online]. Available: www.ijacsa.thesai.org

[6] V. Sowmya, 'OPTIMIZING IOT NETWORK TRAFFIC USING DEEP Q-LEARNING ALGORITHMS: A CASE STUDY ON SMART CITY DATA FLOW'.

[7] V. Ibiyemi, K. Bamidele Moses, A. Kolawole Gabriel, and A. Junior, 'A review of Network Traffic Prediction using Deep Learning Models', 2024. [Online]. Available: www.ijres.org

[8] S. Troia, R. Alvizu, Y. Zhou, G. Maier, and A. Pattavina, 'Deep Learning-based Traffic Prediction for Network Optimization'. [Online]. Available: https://www.internet2.edu/

[9] R. Boutabaet al., 'A comprehensive survey on machine learning for networking: evolution, applications and research opportunities', Journal of Internet Services and Applications, vol. 9, no. 1, Dec. 2018, doi: 10.1186/s13174-018-0087-2.

[10] G. Bernardez et al., 'MAGNNETO: A Graph Neural Network-Based Multi-Agent System for Traffic Engineering', IEEE Trans Cogn Commun Netw, vol. 9, no. 2, pp. 494–506, Apr. 2023, doi: 10.1109/TCCN.2023.3235719.

[11] P. Kumar, D. Pandey, R. K. Srivastav, and P. K. Pandey, 'Network Traffic Analysis and Prediction Using Machine Learning', 2071. [Online]. Available: www.ijrpr.com

[12] J. Cao, D. Wang, Z. Qu, H. Sun, B. Li, and C. L. Chen, 'An improved network traffic classification model based on a support vector machine', Symmetry (Basel), vol. 12, no. 2, Feb. 2020, doi: 10.3390/sym12020301.

[13] A. Azab, M. Khasawneh, S. Alrabaee, K. K. R. Choo, and M. Sarsour, 'Network traffic classification: Techniques, datasets, and challenges', Jun. 01, 2024, KeAi Communications Co.doi: 10.1016/j.dcan.2022.09.009.

[14] G. Bernardez et al., 'MAGNNETO: A Graph Neural Network-Based Multi-Agent System for Traffic Engineering', IEEE Trans Cogn Commun Netw, vol. 9, no. 2, pp. 494–506, Apr. 2023, doi: 10.1109/TCCN.2023.3235719.

[15] S. Izadi, M. Ahmadi, and A. Rajabzadeh, 'Network Traffic Classification Using Deep Learning Networks and Bayesian Data Fusion', Journal of Network and Systems Management, vol. 30, no. 2, Apr. 2022, doi: 10.1007/s10922-021-09639-z.

[16] S. Mamdoohi, 'OPTIMIZATION AND MACHINE LEARNING METHODS TOWARD IMPROVED TRAFFIC NETWORK PERFORMANCE IN DISRUPTED ENVIRONMENTS', 2023.

[17] L. Doris, 'OPTIMIZING NETWORK TRAFFIC IN EDGE COMPUTING SYSTEMS USING REAL-TIME AI AND ML ALGORITHMS'. [Online]. Available: https://www.researchgate.net/publication/387958420

[18] A. Lakhinaet al., 'Structural Analysis of Network Traffic Flows', 2004.

[19] F. Bronzino, P. Schmitt, S. Ayoubi, H. Kim, R. Teixeira, and N. Feamster, 'Traffic Refinery: Cost-Aware Data Representation for Machine Learning on Network Traffic', Oct. 2020, doi: 10.1145/3491052.

[20] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, 'Traffic Flow Prediction with Big Data: A Deep Learning Approach', IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865–873, Apr. 2015, doi: 10.1109/TITS.2014.2345663.

[21] A. Azab, M. Khasawneh, S. Alrabaee, K. K. R. Choo, and M. Sarsour, 'Network traffic classification: Techniques, datasets, and challenges', Jun. 01, 2024, KeAi Communications Co.doi: 10.1016/j.dcan.2022.09.009.

[22] A. O. Salau and M. M. Beyene, 'Software defined networking based network traffic classification using machine learning techniques', Sci Rep, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-70983-6.

[23] S. Liu et al., 'ServeFlow: A Fast-Slow Model Architecture for Network Traffic Analysis', Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.03694

[24] P. Wang, Z. Li, M. Fu, Z. Wang, Z. Zhang, and M. Liu, 'FedEdge AI-TC: A Semi-supervised Traffic Classification Method based on Trusted Federated Deep Learning for Mobile Edge Computing', Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.06924

[25] Z. Jin, K. Duan, C. Chen, M. He, S. Jiang, and H. Xue, 'FedETC: Encrypted traffic classification based on federated learning', Heliyon, vol. 10, no. 16, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35962.

[26] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, 'Hybrid approaches to optimization and machine learning methods: a systematic literature review', Mach Learn, vol. 113, no. 7, pp. 4055–4097, Jul. 2024, doi: 10.1007/s10994-023-06467-x.

[27] R. H. Serag, M. S. Abdalzaher, H. A. E. A. Elsayed, M. Sobh, M. Krichen, and M. M. Salim, 'Machine-Learning-Based Traffic Classification in Software-Defined Networks', Mar. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/electronics13061108.

[28] P. Mavaie, L. Holder, and M. K. Skinner, 'Hybrid deep learning approach to improve classification of low-volume high-dimensional data', BMC Bioinformatics, vol. 24, no. 1, Dec. 2023, doi: 10.1186/s12859-023-05557-w.

[29] O. K. Oladele, 'The Role of Computer Science in Optimizing Telecommunication Network Performance'. [Online]. Available: https://www.researchgate.net/publication/383909817

[30] I. A. Alwhbi, C. C. Zou, and R. N. Alharbi, 'Encrypted Network Traffic Analysis and Classification Utilizing Machine Learning', Jun. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/s24113509.

[31] B. Panjavarnam, P. Poornima, and K. Subathra, 'Network Traffic Prediction in 4G- LTE Using Machine Learning Techniques', in 2024 International Conference on Communication, Computing and Internet of Things, IC3IoT 2024 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/IC3IoT60841.2024.10550364.