INNOVATIONS IN INFORMATION TECHNOLOGY
https://www.worldlinkpub.com/journals/innovations-in-information-technology/
ISSN: 3141-6004

A HYBRIDIZED MACHINE LEARNING MODEL FOR CYBER THREAT DETECTION AND CLASSIFICATION

[*1] Emmanuel Ogala, [2] Rose O. Akor, [3] Victor O. Ojodumi, and [4]Richard A. Nyinongu
[1]Department of Computer Science, Joseph Sarwuan Tarka University, Makurdi
[2]Department of Computer Science Kogi State University, Anyigba.
[3]Department of Computer Science Federal University Lokoja.
[4]Department of Computer Science Joseph Sarwuan Tarka University, Makurdi
Corresponding Author: ogalaemmanuel4edu@gmail.com; ogala.e@ksu.edu.ng

Received: 26th May 2026
Accepted for publication: 15 June 2026
Published: 01 July 2026

ABSTRACT

Cybersecurity has become increasingly critical as digital systems expand and cyber threats grow more sophisticated. Traditional methods such as signature-based and behavior-based detection struggle to adapt to evolving attack patterns. To address these limitations, this study proposes a weighted voting hybrid machine learning framework that combines multiple algorithms to improve detection accuracy and robustness. The model integrates KNN, Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and LightGBM (LGB), leveraging their complementary strengths. Empirical evaluation shows that the hybrid model outperforms individual classifiers in both binary and multiclass classification tasks. In binary classification, the hybrid approach achieved 94.58% accuracy, surpassing all baseline models. For multiclass classification, it attained 82.57% accuracy, again demonstrating superior performance. To further enhance reliability, the framework incorporates stacking with a Logistic Regression meta-learner and probability calibration, resulting in approximately 95% accuracy and a PR-AUC of 0.993 on the UNSW-NB15 dataset. The model's generalization capability was validated using the CICIDS2017 dataset through cross-dataset evaluation, where it maintained strong performance (PR-AUC ≈ 1.000) despite domain differences. Additional robustness tests, including noise injection and feature perturbation, confirmed the model's stability, with minimal data drift observed. Overall, the proposed hybrid framework provides an effective and adaptable solution for modern cyber threat detection and classification.

Keywords: Hybrid Machine Learning, Cyber Threat Detection, Intrusion Detection Systems (IDS), Ensemble Learning, Classification, Feature Selection, Network Security

1.0 INTRODUCTION

Cybersecurity has become critical due to the rapid evolution of the digital landscape, with vast amounts of sensitive data and services now residing online (Subba *et al*., 2016). Threats such as malware, Denial of Service (DoS) attacks, and unauthorized access attempts are widespread and diverse, originating from both internal and external actors (Mukherjee and Sharma, 2012). Consequently, effective and accurate threat detection and classification are essential (Tian *et al*., 2020). Various cybersecurity frameworks have evolved over time. Signature-based detection is effective against known threats but fails against advanced ones (Tang *et al*., 2016). Behaviour-based techniques can identify unknown malware using machine intelligence but are complex (Alshamrani, 2022). Consequently, machine learning (ML) has emerged as a promising approach, using statistical properties to detect threats, including metamorphic malware (Makhlouf *et al*., 2018). ML models employ static analysis (examining code without execution) or dynamic analysis (observing behaviour during execution) (Tian *et al*., 2020). Common ML models include k-Nearest Neighbors (k-NN), Logistic Regression, Random Forest, Naïve Bayes, Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA). However, standalone ML models face challenges such as high false-positive rates and low detection of novel attacks (Khan and Gumaei, 2022). Hybrid intelligent systems, which combine multiple algorithms, offer improved performance and adaptability (Maimo *et al*., 2018; Stolfo *et al*., 2000; Khan and Gumaei, 2022).

A taxonomy of cybersecurity techniques categorizes systems into Signature-based Intrusion Detection Systems (SIDS) and Anomaly-based Intrusion Detection Systems (AIDS) (Namjoshi and Narlikar, 2010). AIDS models "normal" behaviour and flags deviations, enabling detection of zero-day exploits, but suffers from high false positives due to benign anomalies (Niyaz *et al*., 2017). Hybrid systems integrate SIDS and AIDS to improve accuracy and reduce false alarms (LeCun *et al*., 2015). ML and deep learning (DL) methods—including SVM, Decision Trees, Random Forest, k-NN, CNN, and RNN—are widely applied, though challenges remain regarding data quality, feature selection, and model adaptability (Tang *et al*., 2016; Panda and Patra, 2007). A comparative analysis of ML techniques for malware

classification found that deep learning achieved 96% accuracy, outperforming RF, DT, SVM, KNN, SGD, LR, and NB. The study highlighted the value of hybrid approaches combining static and dynamic analysis (Al-Janabi *et al*., 2020). Another review covered ML applications in intrusion detection, malware analysis, and biometric authentication, also addressing adversarial threats like data poisoning and evasion tactics (Dong and Wang, 2016). A proposed hybrid deep learning model combining Convolutional Neural Networks (CNN) and Quasi-Recurrent Neural Networks (QRNN) achieved 99.99% accuracy and low false-positive rates (0.3% on BoT-IoT, 1% on TON_IoT), also showing a 23% speed improvement over LSTM (Al-Taleb and Saqib, 2022).

ML techniques—including k-NN, SVM, and decision trees—are essential across intrusion detection methodologies, with models updatable to address evolving threats (Liu and Lang, 2019). Finally, reinforcement learning (RL) agents offer adaptability beyond rule-based models, learning optimal security policies through real-time interaction with network environments and adjusting actions based on feedback (Liu *et al*., 2020). Existing hybrid intrusion detection systems rely on single ensemble classifiers without probability calibration, meta-learning optimization, or cross-dataset generalization analysis. Consequently, their reliability, robustness against domain shift, and decision confidence in security-critical environments remain unaddressed. This study bridges these gaps by proposing a novel hybrid framework that integrates meta-learning optimization with calibrated probability estimates, systematic feature alignment, and comprehensive cross-dataset robustness evaluation, thereby enhancing detection accuracy, reliability, and deployability in evolving network environments.

2.0 MATERIALS AND METHODS

2.1 Source of Data

The UNSW-NB15 dataset used for this study was collected from the publicly available GitHub repository at the following address: https://github.com/abhinav-bhardwaj/IoT-Network-Intrusion-Detection-System-UNSW-NB15/tree/master/datasets. The CICIDS2017 dataset, on the other hand was obtained from through the Official dataset page from the Canadian Institute for Cybersecurity (CIC) address: https://www.unb.ca/cic/datasets/ids-2017.

2.2 Method of Data Collection

This UNSW-NB15 dataset, originally developed by the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS), contains realistic network traffic and a wide range of modern attack scenarios. It is well-suited for machine learning applications in cybersecurity, supporting both binary classification (normal vs. malicious) and multiclass classification (differentiating among specific attack types). The dataset includes 49 features extracted using the Argus and Bro-IDS tools, providing comprehensive information for building and evaluating threat detection systems.

2.3 Model Design Architecture

The architecture defines the structural and strategic foundation for the proposed cyber threat detection framework. This phase establishes the system architecture, identifies the learning paradigms to be employed, specifies dataset utilization, and outlines the high-level data flow that governs model development and evaluation. It functions as a conceptual blueprint that ensures coherence between the research objectives, methodological choices, and expected performance outcomes. The preliminary design is represented through the architectural workflow presented in Figure 1. As illustrated in Figure 1, the architecture is organized into interconnected stages that reflect a progressive hybrid-learning pipeline. The design begins with dataset preparation using the UNSW-NB15 dataset as the primary training and model-development source. The dataset undergoes preprocessing operations, including cleaning, encoding of categorical attributes, normalization, and feature preparation to ensure algorithm compatibility.

The processed data are then partitioned into training and testing subsets using a stratified 80:20 split to preserve class distributions for both binary and multiclass tasks. To evaluate generalization capability, the architecture also incorporates cross-dataset validation using the CICIDS2017 dataset. After applying consistent preprocessing and feature-alignment protocols, the trained hybrid framework is evaluated on CICIDS2017 without full retraining. This stage assesses transferability and robustness under dataset shift, which is critical for operational deployment in dynamic network environment.

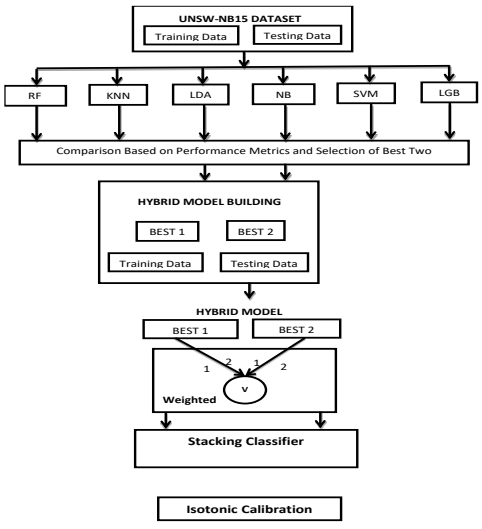


Figure 1: Architecture for the Hybrid Machine Learning Model for Cyber Threat Detection and Classification

2.4 Algorithm for the Hybrid Model

The algorithm for the implementation of the model shows the training, testing of the six base algorithms and evaluations of

their performances to determine the best two, and integrating the two into a single hybrid classifier. It illustrates the step-by-step procedural design adopted in developing a hybrid classifier using the UNSW-NB15 and CICIDS2017 datasets. Each step in the algorithm represents a structured and modular phase of the research implementation, aligned with the defined objectives. The UNSW-NB15 dataset serves as the primary dataset for model training and internal evaluation, while the CICIDS2017 dataset is used for cross-dataset validation to assess generalization under dataset shift. Once the environment is ready, the datasets are loaded using appropriate Python commands. Both datasets contain labelled network traffic records representing normal behaviour and multiple attack categories. Initial data exploration is conducted to examine dataset structure, detect missing or inconsistent values, and analyse class distributions. This step ensures early identification of data quality issues that could influence model reliability.

Input: UNSW-NB15 dataset, CICIDS2017 dataset, selected machine learning models X_i, where i=0 to 5, hyperparameter search space, binary and multiclass classification labels, and evaluation metrics.

Output: Trained weighted voting hybrid classifier, performance metrics on both tasks (binary *and* multiclass), cross-dataset validation results. Steps are:

1: Start.
2: Import the required libraries.
3: Load the UNSW-NB15 and CICIDS2017 datasets.
4: Perform data preprocessing.
5: Split the UNSW-NB15 dataset into training and testing sets.
6: Define six machine learning models, represented as Xi , where i=0 to 5.
7: For each model Xi, tune hyperparameters for both binary and multiclass classification tasks using UNSW-NB15.
8: Perform full cross-validation.
If No, repeat Step 7.
If Yes, proceed.
9: Evaluate all tuned models and select the two best-performing models for both tasks.
10: Build a weighted voting hybrid classifier using the two selected models.
11: Tune the hyperparameters of the hybrid classifier for both binary and multiclass tasks.
12: Evaluate the weighted voting hybrid classifier.
13: Perform full cross-validation for the hybrid classifier.
If No, repeat Steps 10–12.
If Yes, proceed.
14: Apply a stacking classifier using Logistic Regression as the meta-learner and Isotonic Calibration.
15: Undertake cross-dataset validation using CICIDS2017.
16: Compare all models based on evaluation results.
17: Exit

3.0 RESULTS

Tests were conducted for binary classification and multiclass classification. 80% of data in the dataset was used for training, and the remaining 20% is engaged in test. This was done in two stages. The first stage was the training and testing of six standalone algorithms while the second is the hybridization of the two best performing algorithms into a single model. The results of the first stage were on the various performance metrics in order to identify the best performing models to employ in the hybridization stage.

3.1 Six Models' Performances in Binary Classification

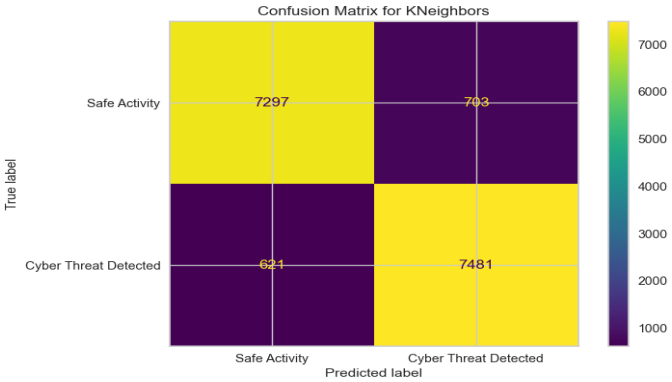
The models' performance result for binary classification for the UNSW-NB15 dataset is shown in Figure 2.

	Models	Accuracy	Precision	Recall	F1_Score	Kappa_Score	AUC_Score	average_score
0	KNeighbors	0.917774	0.917812	0.917774	0.917769	0.835531	0.917739	0.904067
1	RandomForest	0.942864	0.942876	0.942864	0.942865	0.885727	0.942877	0.933346
2	Naive Bayes	0.821761	0.830905	0.821761	0.820639	0.643880	0.822276	0.793537
3	Support Vector	0.907030	0.915185	0.907030	0.906525	0.813887	0.906581	0.892706
4	Linear Discriminant Analysis	0.873929	0.886741	0.873929	0.872791	0.747555	0.873345	0.854715
5	Light Gradient Boost	0.942367	0.942617	0.942367	0.942364	0.884747	0.942439	0.932817

Figure 2: Binary Classification Performances of the Six Models

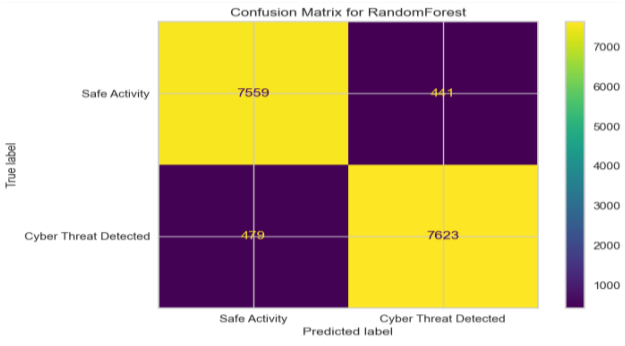
3.2 Confusion Matrix for the Six Algorithms

Further analysis can be seen in the confusion matrix in figures 3-8



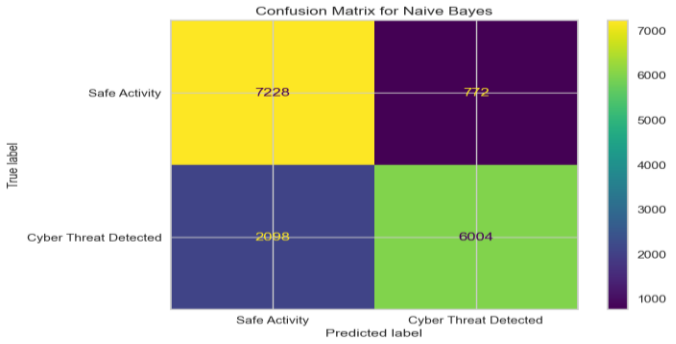
TP = 46.46%, TN = 45.32%, FP = 3.39%, FN = 4.37%

Figure 3: Binary Confusion Matrix for KNN



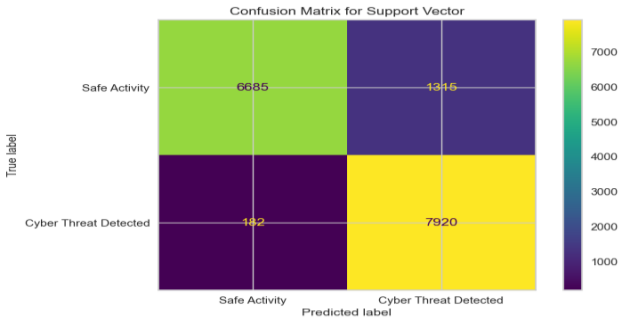
TP = 47.34%, TN = 46.94%, FP = 2.74%, FN = 2.97%

Figure: 4 Binary Confusion Matrix for Random Forest



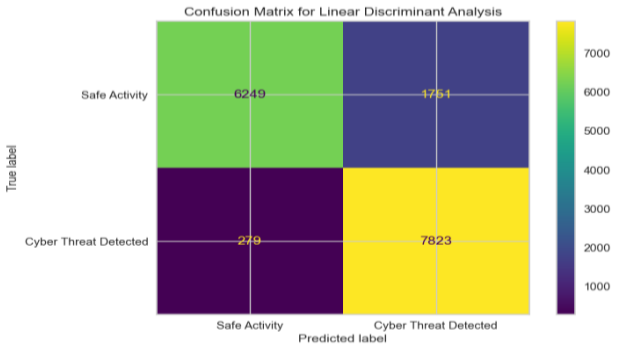
TP = 37.29%, TN = 44.88%, FP = 4.79%, FN = 13.03%

Figure 5: Binary Confusion Matrix for Naïve Bayes



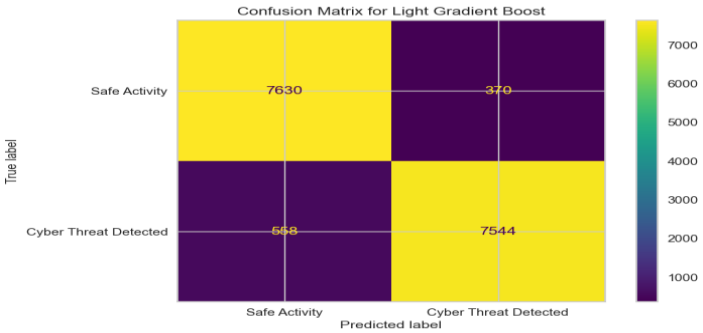
TP = 49.19%, TN = 41.52%, FP = 13.03%, FN = 8.17%

Figure 6: Binary Confusion Matrix for Support Vector Machine



TP = 48.58%, TN = 38.81%, FP =13.03%, FN = 1.73%

Figure 7: Binary Confusion Matrix for Linear Discriminat Analysis



TP = 46.85%, TN = 47.39%, FP =2.3%, FN = 3.47%

Figure 8: Binary Confusion Matrix for Light Gradient Boost

4.0 DISCUSSION

The evaluation of six baseline models revealed that Random Forest (94.29%) and Light Gradient Boosting (94.24%) outperformed KNN, SVM, LDA, and Naïve Bayes in binary intrusion detection, while Light Gradient Boosting (82.56%) led multiclass classification, with all models experiencing performance declines due to class imbalance and overlapping attack signatures. The weighted voting hybrid combining Random Forest and Light Gradient Boosting achieved 94.58% binary and 82.57% multiclass accuracy, marginally surpassing individual models with improved stability. Critically, optimization through stacking with a Logistic Regression meta-learner and isotonic probability calibration further increased binary accuracy to approximately 95%, with PR-AUC values of 0.993 on UNSW-NB15 and 1.000 on CICIDS2017, confirming enhanced ranking quality and confidence estimation. Near-zero Population Stability Index values indicated minimal distributional drift, while noise robustness and feature masking experiments demonstrated graceful degradation rather than catastrophic failure.

Cross-dataset evaluation on CICIDS2017 revealed a recall decline from 0.974 to 0.782 (approximately 19.8% relative gap), though precision remained exceptionally high at 1.000 and PR-AUC remained strong, suggesting that dataset heterogeneity—not model inadequacy—accounts for the observed performance gap. These findings collectively confirm that while hybridization alone yields modest gains, stacking with calibration produces robust, reliable, and deployment-ready intrusion detection capable of maintaining performance under heterogeneous and evolving threat landscapes, with practical benefits including higher decision trustworthiness, improved cross-dataset resilience, and reduced false positives and negatives.

5.0 CONCLUSION

Based on the findings, this study concludes that hybrid machine learning models consistently outperform individual classifiers in cyber threat detection, particularly on complex and imbalanced datasets. Random Forest and Light Gradient Boosting emerged as the most effective base learners for both binary and multiclass intrusion detection tasks. Weighted

voting and stacking ensembles enhanced predictive stability, reduced variance, and improved generalization across datasets, while probability calibration significantly improved confidence reliability - a critical requirement for operational cybersecurity decision-making. Although cross-dataset performance gaps are unavoidable, they proved manageable through careful feature alignment, calibration, and robust ensemble design.

Overall, this study successfully achieved all stated objectives, demonstrating that hybridized and optimized machine learning architectures offer a scalable, accurate, and resilient solution for modern intrusion detection systems. This work represents a next-step evolution that synthesizes their strengths while systematically resolving their limitations, resulting in a more holistic framework that balances accuracy, reliability, generalization, and deployability positioning the study as a meaningful advancement toward production-ready intelligent intrusion detection systems.

REFERENCES

Al-Janabi, M., and Altamimi, A. M. (2020). A comparative analysis of machine learning techniques for classification and detection of malware. In 2020 21st International Arab Conference on Information Technology (ACIT) (pp. 1–9). IEEE. https://doi.org/10.1109/ACIT50332.2020.9300060

Alshamrani, S. S. (2022). Design and analysis of machine learning-based technique for malware identification and classification of portable document format files. *Security and Communication Networks*, Article 7611741, 1–10. https://doi.org/10.1155/2022/7611741

Al-Taleb, N., and Saqib, N. A. (2022). Towards a hybrid machine learning model for intelligent cyber threat identification in smart city environments. *Applied Sciences*, **12**(4), 1863. https://doi.org/10.3390/app12041863

Dong, B., and Wang, X. (2016). Comparison deep learning method to traditional methods using for network intrusion detection. In 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN) (pp. 581–585). IEEE. https://doi.org/10.1109/ICCSN.2016.7586590

Khan, S., and Gumaei, A. (2022). A novel hybrid intrusion detection system for Internet of Things. Security and *Communication Networks*. Article 1302435, 1–12. https://doi.org/10.1155/2022/1302435

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, **521**, 436–444. https://doi.org/10.1038/nature14539

Liu, H. and Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences*, 9(20), 4396. https://doi.org/10.3390/app9204396

Liu, H., Lang, B., Liu, M., and Yan, H. (2020). CNN and RNN based payload classification methods for attack detection. *Knowledge-Based Systems*, **163**, 332–341. https://doi.org/10.1016/j.knosys.2018.08.011

Maimó, L. F., Perales Gómez, Á. L., Clemente, F. J. G., Pérez, G. M., and Pérez, J. M. (2018). A self-adaptive deep learning-based system for anomaly detection in 5G networks. *IEEE Access*, **6**, 7700–7712.

Makhlouf, A. B., Derdour, M., and Ghoualmi-Zine, N. (2018). Intrusion detection using deep learning techniques. In Proceedings of the 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS) (pp. 1–5). IEEE. https://doi.org/10.1109/PAIS.2018.8598481

Mukherjee, S., and Sharma, N. (2012). Intrusion detection using naive Bayes classifier with feature reduction. *Procedia Technology*, **4**, 119–128. https://doi.org/10.1016/j.protcy.2012.05.017

Namjoshi, K., and Narlikar, G. (2010). Robust and fast pattern matching for intrusion detection. Proceedings IEEE INFOCOM 2010, 1–9. https://doi.org/10.1109/INFOCOM.2010.5462149

Niyaz, Q., Sun, W., and Javaid, A. Y. (2017). A deep learning based DDoS detection system in software-defined networking (SDN). arXiv preprint arXiv:1611.07400. https://arxiv.org/abs/1611.07400

Panda, M., and Patra, M. R. (2007). Network intrusion detection using naive Bayes. *International Journal of Computer Science and Network Security*, **7**(12), 258–263.

Stolfo, S. J., Fan, W., Lee, W., Prodromidis, A. L., and Chan, P. K. (2000). Cost-based modeling for fraud and intrusion detection: Results from the JAM project. In Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00 (Vol. **2**, pp. 130–144). IEEE.

Subba, B., Biswas, S., and Karmakar, S. (2016). A neural network based multi-class intrusion detection system. In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 2644–2650). IEEE. https://doi.org/10.1109/ICACCI.2016.7732403

Tang, T. A., Mhamdi, L., McLernon, D., Zaidi, S. A. R., and Ghogho, M. (2016). Deep learning approach for network intrusion detection in software defined networking. In 2016 International Conference on Wireless Networks and Mobile Communications (WINCOM) (pp. 258–263). IEEE.

Tian, Z., Luo, C., Qiu, J., Du, X., and Guizani, M. (2020). A distributed deep learning system for web attack detection on edge devices. *IEEE Transactions on Industrial Informatics*, 16(3), 1963–1971.