ABC2: Journal of Architecture, Building, Construction, and Cities

Empowering research for Sustainable Development Goals, ABC2: Architecture, Building, Construction, and Cities is a fundamental manifesto to address these pressing issues, fostering dialogue and knowledge exchange among researchers, practitioners, and policymakers. Exploring sustainable design, resilient infrastructure, advanced construction methods, and equitable urban development, ABC2 aims to empower the global community to create adaptive, inclusive, and sustainable environments. The ABC2 focus on cutting-edge research, technological advancements, and transformative strategies is essential for navigating the future of our cities and communities.

Research Article

Machine Learning-Based Intrusion Detection for Smart City Internet of Things Networks

Ebere Donatus Okonta[1*], Oluwaseun Bamgbose [2]

[1] School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, United Kingdom
[2] Barclays Technology Centre, Radbroke, Knustford, United Kingdom

DOI: https://doi.org/10.66408/abc2.2026.84

*Correspondence: eokonta@tees.ac.uk

Received: 29/05/2026
Revised: 16/07/2026
Accepted: 22/07/2026
Published: 02/08/2026

Volume: 2026
Issue: Ahead of Print
Pages: Ahead of Print

Abstract

The increasing deployment of Internet of Things (IoT) devices in smart city infrastructures has significantly expanded the network attack surface, making effective intrusion detection a critical security requirement. Traditional intrusion detection systems struggle to cope with the volume, heterogeneity, and dynamic behaviour of IoT network traffic, often resulting in high false alarm rates and missed attacks. This study investigates the effectiveness of supervised machine learning techniques for detecting cyberattacks in IoT-based smart city networks using the TON_IoT dataset. A progressive modelling approach is adopted, beginning with Logistic Regression as the baseline model, followed by Random Forest as an ensemble method, and culminating in an optimised XGBoost model. Preprocessing and feature engineering address dimensionality, feature representation, and the challenges associated with imbalanced IoT traffic distributions. Experimental results demonstrate that ensemble and boosting-based models significantly outperform linear approaches. Among the evaluated models, XGBoost achieves the highest detection performance, substantially reducing missed attacks while maintaining robust classification accuracy. The findings demonstrate that advanced ensemble learning combined with robust feature engineering provides a reliable and scalable solution for securing smart city IoT networks.

Keywords: Intrusion Detection Systems (IDS); Smart city; Machine learning; TON_IoT dataset, XGBoost-based intrusion detection; Cyberattack detection; IoT network security

Highlights

- XGBoost achieved near-perfect intrusion detection performance with 99.97% accuracy and the lowest false negative rate.
- Reducing false negatives from 481 to 40 demonstrated a major improvement in preventing undetected cyberattacks.
- Advanced feature engineering and preprocessing substantially improved the reliability and scalability of intrusion detection systems.
- Machine learning–based IDS solutions were validated as effective cybersecurity tools for protecting smart city IoT infrastructures.

1. Introduction

The digital revolution is no new news. However, the role of the Internet of Things (IoT) on how we interact with our environment, work and live is revolutionary (Okonta et al, 2025a). To set a pace, IoT simply refers to a large network of devices interconnected through three primary elements (sensors, software and supporting technologies) for the primary purpose of data connection and exchange (Jo, 2025; Lombardi, 2021). Primarily because IoT allows for data exchange, which facilitates decision making through data, real-time monitoring and automation, it has become a prime solution to be leveraged upon in this time and age; with IoT solutions and applications enhancing the quality of urban life, efficiency of systems, economic growth, environmental sustainability and even security (GCS Network, 2025, Okonta et al, 2025b). Aspects such as smart infrastructure, smart governance and smart urban planning have all integrated IoT platforms to manage diverse city systems, leading to the creation of connected cities with seamless data flow from service delivery down to infrastructure (Okonta & Vukovic 2024). However, the large-scale data generation and interconnectedness inherent in IoT significantly expand the "attack surface" of institutions that rely on these technologies, increasing their exposure to cyber threats. The rapid expansion of IoT, cloud services, and decentralised finance has significantly enlarged organisational attack surfaces, enabling even low-skilled attackers to launch sophisticated threats (Ganguli, 2024; GCS Network, 2025). An attack surface comprises the pathways and vulnerabilities through which unauthorised access can occur (Stephen, 2025) and 67% of organisations have reported growth in their attack surface in recent years (IBM Randori, 2022). This escalation necessitates rapid threat detection and response. Intrusion Detection Systems (IDS) play a critical role by monitoring network traffic and identifying malicious activities based on predefined signatures and policies (Heidari & Jabraeil Jamali, 2023; Azeez et al., 2019).

Despite their importance, traditional intrusion detection systems (IDS) face significant limitations in detecting and responding to modern cyber threats (Heidari & Jabraeil Jamali, 2023). Their reliance on signature-based detection makes them ineffective against previously unseen or evolving attacks (Díaz-Verdejo et al., 2022; Ahmed et al., 2025), requiring frequent and costly signature updates (Iyer, 2021). Attackers also adapt malware to evade known signatures (Iyer, 2021). Additionally, high false alarm rates—where legitimate activities are misclassified as threats—reduce system effectiveness and contribute to alert fatigue (Pihelgas, 2015). Consequently, traditional IDS struggle to manage the dynamic and large-scale nature of IoT networks (Jamshidi et al., 2025).

The limitations of traditional IDS highlight the need for Machine Learning (ML)–driven approaches capable of behavioural analysis and anomaly detection, enabling smarter and more adaptive threat identification (Palaniappan et al., 2024). ML enhances detection accuracy, reduces false positives, and allows IDS to evolve alongside emerging threats (Jackson, 2024; Mohamed, 2025). Its rapid adoption in intrusion detection reflects its effectiveness in automating large-scale threat detection and strengthening network security (Ali & Singh, 2023). Various ML algorithms including SVM, Naïve Bayes, KNN, Decision Tree, Random Forest, and K-means—have been applied to IDS; however, concerns remain regarding model interpretability (Jamshidi et al., 2025) and dataset imbalance, which hinders reliable detection of rare attacks while maintaining low false positives (Sayegh et al., 2024). Although multiple datasets such as CIODS201Y, NSL-KDD, and TON_IoT have been used for evaluation (Ever et al., 2019), limited clarity exists on their suitability for large-scale IoT IDS development, and studies using TON_IoT remain relatively sparse (Jamshidi et al., 2025). The TON_IoT dataset has been widely adopted for intrusion detection research; existing studies have primarily focused on maximising classification performance or evaluating individual machine learning models. This study contributes by providing a security-oriented comparative assessment of baseline, ensemble, and boosting algorithms, with particular emphasis on reducing false negatives as a critical performance measure for protecting IoT-enabled smart city infrastructure. Furthermore, the study extends the technical evaluation by discussing the practical implications of intrusion detection performance for infrastructure resilience and cybersecurity decision-making within smart city environments. To address these gaps, this study evaluates supervised ML-based intrusion detection models for smart city IoT networks using the

realistic and large-scale TON_IoT dataset. The objective is to identify the most effective algorithm for accurately classifying IoT network traffic as normal or malicious, thereby enhancing cybersecurity in smart city environments. Beyond their technical function, intrusion detection systems play an increasingly important role in supporting the resilience and governance of smart city infrastructure. Unlike studies that primarily report aggregate classification accuracy, this research adopts a security-oriented evaluation perspective by prioritising the reduction of false negatives, which represent undetected cyberattacks with direct implications for smart city operational resilience. The contribution of this study therefore does not rely on proposing a new machine learning algorithm, but on developing a rigorous comparative evaluation framework that examines how progressively complex supervised learning approaches perform under security-critical evaluation criteria using a contemporary IoT intrusion dataset. Modern cities rely on interconnected IoT ecosystems to manage transportation networks, energy systems, water distribution, environmental monitoring, public safety, and smart buildings. Consequently, cybersecurity incidents can extend beyond digital disruptions to affect essential urban services, public trust, and city-wide operational continuity. Effective intrusion detection therefore contributes not only to network security but also to the broader objectives of resilient urban infrastructure management and sustainable smart city development. As cities continue to expand their digital infrastructure, municipal authorities and infrastructure operators require cybersecurity solutions capable of supporting proactive risk management and informed governance decisions. Machine learning-enabled intrusion detection systems provide opportunities to identify threats more rapidly, improve situational awareness, and strengthen the capacity of city administrations to maintain service continuity during cyber incidents. Consequently, evaluating the effectiveness of advanced intrusion detection techniques has implications for both cybersecurity performance and smart city resilience planning.

1.1 Intrusion Detection in IoT and Smart City Networks

IDS in the IoT are essential security components of the IoT community, protecting all structures within the IoT system from cyberattacks (Inayat et al., 2022). As security will be a vital supporting element of most IoT applications, IoT intrusion detection systems also need to be developed to secure communications enabled by such IoT technologies. Quite a number of techniques are required for the development of IDS in IoT systems, with this study focusing on ML. Within the context of smart cities, the integration of interconnected devices creates a more complex web and attack surface for cybercriminals to exploit (Demertzi et al., 2023). Furthermore, data privacy has been a consistent issue of concern for IoT systems as the collection of vast amounts of data raises privacy concerns as well as the potential for unauthorised attacks (Perera et al., 2015; Okonta et al., 2025b). Also, irrespective of the large IoT network, there is still a lack of robust security features that leave vulnerabilities for attack (Aslan et al., 2023). There is also the risk of incompatibility between legacy systems and newer technologies, creating a security gap, which is a risk (Ogunwole et al., 2023). What then are the attacks these IoT systems are prone to? Because IoT devices are mostly released with default codes (both username and passwords), credential surfing and brute force tools are primary attacks experienced as a result of default or weak credentials (Rabzelj & Sedlar, 2025). Furthermore, during firmware updates, a lack of cryptographic signature verification or even the use of an insecure channel (HTTP) could offer hackers a window of opportunity for virus to be introduced to systems through compromised servers (Aslan et al., 2023). Moreover, unencrypted communications allow for eavesdropping and seasonal hijacking (Hazra et al., 2024).

1.2 Machine Learning Approaches for IoT Intrusion Detection

Machine learning (ML) has become integral to strengthening IoT intrusion detection systems (IDS) in response to evolving cybersecurity threats. Supervised ML trains models on labelled data, allowing them to learn patterns and make predictions on unseen data while adapting to errors to improve accuracy over time (Saravanan & Sujatha, 2018; Ahmed et al., 2019). Common algorithms include Support Vector Machines (SVM), Naive Bayes (NB), K-Nearest Neighbours (KNN), Decision Trees (DT),

Random Forests (RF), and K-means clustering (Jamshidi et al., 2025). SVM creates hyperplanes to separate classes in high-dimensional space, NB predicts by assuming feature independence (Zheng & Li, 2024), and KNN classifies new instances based on similarity to retained training data (Zhang, 2017). DTs provide interpretable, hierarchical classification of network events, while RF ensembles multiple DTs to enhance stability and predictive accuracy (Jamshidi et al., 2025; Molina-Coronado et al., 2020; Besler et al., 2019). K-means, an unsupervised technique, identifies clusters in unlabelled data to detect novel patterns (Suyal & Sharma, 2024). Supervised methods like RF and SVM excel at recognizing known threats, whereas clustering algorithms detect new anomalies (Abdullah & Abdulazeez, 2021; Rashid et al., 2023; Ever et al., 2019). Hybrid approaches combining anomaly and misuse detection further improve accuracy and reduce computation (Ali et al., 2025). Despite these advances, challenges remain, including the need for large, diverse datasets, continual model updates, and balancing false positives and negatives (Sharma et al., 2024). The growing dependence of smart cities on interconnected digital infrastructure has elevated cybersecurity from a technical concern to a strategic governance priority. Urban systems such as intelligent transportation networks, smart grids, environmental monitoring platforms, emergency response systems, and connected public services increasingly depend on secure and reliable data exchange. Consequently, cyberattacks targeting IoT-enabled infrastructure can disrupt critical services, compromise sensitive information, and undermine public confidence in smart city initiatives.

From a governance perspective, city authorities must balance technological innovation with risk management and infrastructure resilience. This requires the implementation of monitoring and detection mechanisms capable of identifying cyber threats before they escalate into operational failures. Intrusion detection systems therefore serve as an important component of urban resilience strategies by providing continuous visibility into network behaviour and enabling timely responses to malicious activities. Improving intrusion detection performance can support evidence-based cybersecurity governance, strengthen critical infrastructure protection, and enhance the long-term sustainability of smart city ecosystem

2. Materials and Methods

2.1 Dataset and Data Preprocessing

The TON_IoT dataset was utilised to develop and evaluate machine learning models for network intrusion detection. TON_IoT contains telemetry, network, and system logs from IoT and industrial IoT devices. Each record is labelled as “Normal” or “Attack” (binary classification) (Moustafa, 2021; Booij et al., 2021; Alsaedi et al., 2020; Moustafa et al., 2020a; Moustafa et al., 2020b; Moustafa, 2019a; Moustafa, 2019b; Ashraf et al., 2021). The primary objective was to classify network traffic as either benign (normal) or malicious (attack) based on observed network features. This problem was formulated as a binary classification task, where the label 0 represents benign or normal traffic, and 1 represents malicious or attack traffic (Anderson & McGrew, 2017). Such a formulation aligns with real-world IDS, where early and accurate identification of malicious activity is critical to maintain network security and prevent potential breaches. The overview of the research methodology is captured in Figure 1.

Effective intrusion detection in IoT environments critically depends on robust data preprocessing and semantically meaningful feature engineering (Nasir et al., 2022). Given the high dimensionality, heterogeneity, and noise inherent in large-scale IoT network traffic, careful transformation of raw data is essential to ensure reliable model learning and generalisation (Borrohou et al., 2025).

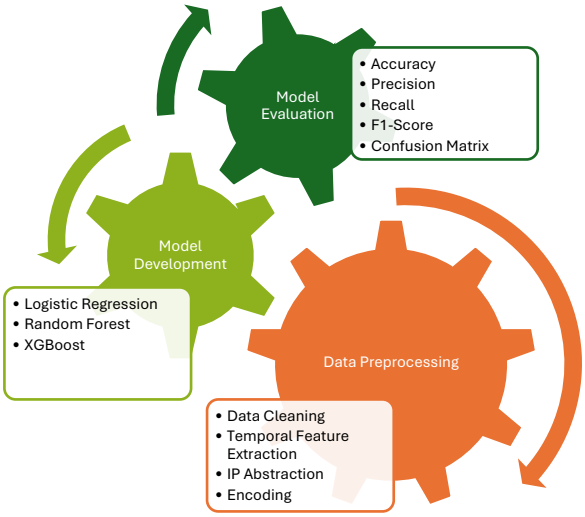

Figure 1. Overview of the research methodology framework (Source: authors).

2.2 Data Cleaning and Temporal Feature Construction

Initial data cleaning focused on ensuring label integrity and consistency across all records. Network flows with missing or undefined class labels were removed to prevent ambiguous supervision during model training. Mixed data types within categorical columns were resolved by explicitly casting all categorical attributes to string format, ensuring uniform representation and preventing downstream encoding errors. Timestamp attributes were converted from raw string format into standardised datetime objects, enabling the extraction of temporally informative features. Rather than using absolute timestamps directly—which may introduce spurious correlations or dataset-specific biases—relative temporal features were derived to better reflect behavioural dynamics. These included the time elapsed since the previous network event and ordering-based temporal signals that capture burstiness and scanning-like activity. Such temporal abstractions are particularly relevant in intrusion detection, as many attacks manifest through anomalous timing patterns rather than isolated packet attributes.

2.3 Feature Engineering and Encoding

Raw IP addresses were excluded from direct model input due to high cardinality, lack of ordinal meaning, and overfitting risk. Instead, domain-informed abstractions were derived to retain security-relevant semantics while ensuring generalizability. Specifically, a binary feature indicating whether an IP belongs to private or public space was included to reflect trust boundaries (Illi et al., 2023), alongside IP version (IPv4 vs. IPv6) encoded as a categorical indicator to capture protocol-level differences relevant to attack behaviour.

Categorical attributes describing protocols, services, connection states, and DNS properties were encoded using ordinal encoding, balancing scalability, memory efficiency, and compatibility with tree-based models. Unlike one-hot encoding, ordinal encoding avoids inflating dimensionality while remaining suitable for ensemble tree algorithms, which rely on relative comparisons rather than metric distances (Dahouda & Joe, 2021; Badnjević & Spahić, 2026). To reduce dimensionality and enhance generalisation, feature selection was performed using a Random Forest model trained on the pre-processed dataset, as seen in Figure 2.

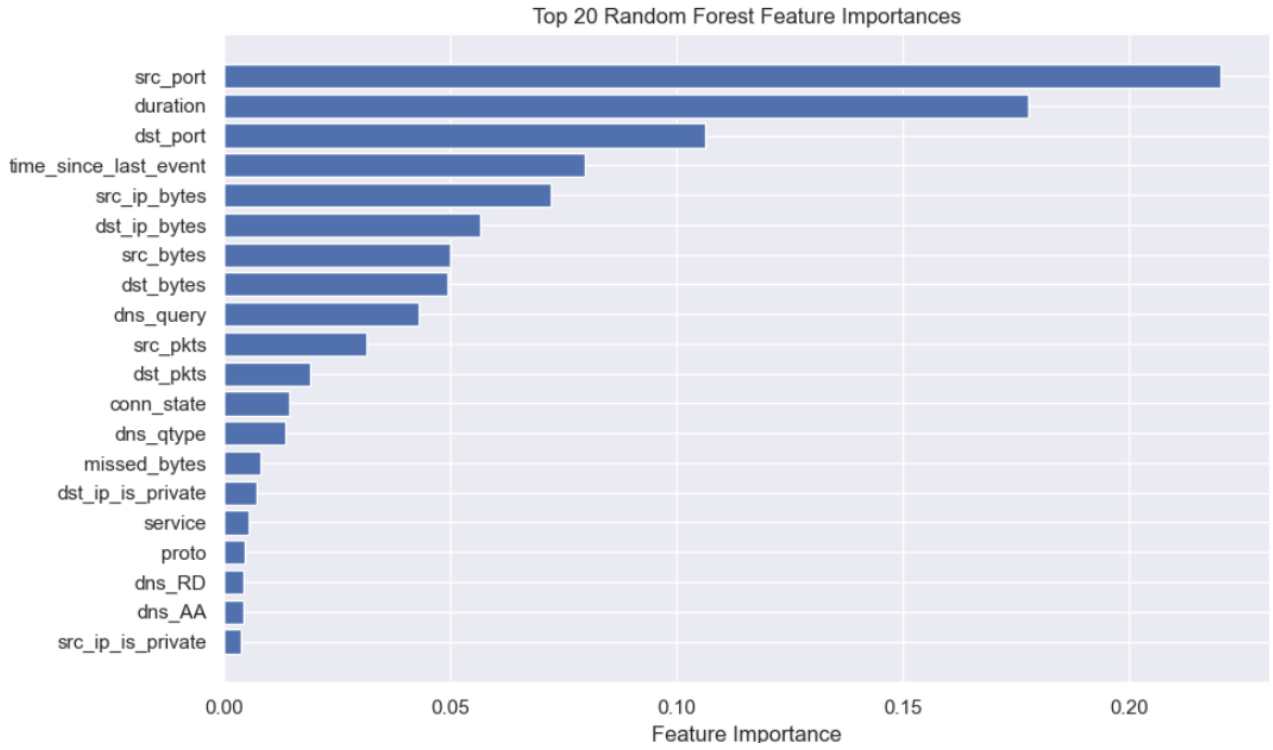

Figure 2. Top 20 feature importances derived from the tuned random forest mode (Source: authors).

Features were ranked by importance based on impurity reduction, and a cumulative threshold was applied to retain the smallest subset accounting for at least 95% of total importance. The retained features capture traffic volume, temporal behaviour, protocol characteristics, and DNS activity, minimizing redundancy, mitigating overfitting, and lowering training and inference costs.

2.4 Experimental Setup

All experiments were conducted within the Python data science ecosystem using Pandas and NumPy for preprocessing, Scikit-learn for Logistic Regression and Random Forest implementation, XGBoost for gradient boosting, and Matplotlib for visualisation. Model development was performed in Jupyter Notebook and Visual Studio Code to ensure reproducibility and robustness. To enable fair and security-aware evaluation under realistic IoT conditions, the dataset was split using a stratified 70/30 train–test partition, preserving the original class distribution and mitigating bias caused by class imbalance. To improve evaluation reliability and minimise potential bias, the study adopted measures to prevent data leakage, including the exclusion of high-cardinality identifiers such as raw IP addresses and avoidance of absolute timestamp features.

The stratified 70/30 train–test split ensured that model performance was assessed on unseen data while maintaining class distribution. Although cross-validation and multi-dataset evaluation could provide further validation, these approaches are identified as important future research directions to assess broader model generalisation. Model performance was assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis. Although accuracy measures overall correctness, it is insufficient in security contexts where undetected attacks may still occur despite high scores (Diana et al., 2025). Therefore, recall was prioritised as the primary metric, with particular emphasis on minimising false negatives, which represent the most critical failure mode in intrusion detection systems (Kumar & Gutierrez, 2025). The integration of threshold-dependent and threshold-independent metrics ensured a balanced and security-aligned evaluation framework. To minimise potential data leakage, feature engineering and feature selection procedures were carefully controlled to ensure that test samples were not used to inform model training.

3. Results and Discussion

3.1 Baseline Model: Logistic Regression

Logistic Regression was employed as a baseline model due to its simplicity, computational efficiency, and high interpretability, providing a reference point against which more complex models could be evaluated. Using a linear decision boundary, the model achieved strong overall performance, correctly classifying 27,437 instances of normal traffic (True Negatives) and successfully detecting 224,811 attack instances (True Positives) as seen in Figure 3. However, the confusion matrix reveals critical limitations: 184 normal flows were incorrectly flagged as attacks (False Positives), while more importantly, 481 attack instances were misclassified as normal traffic (False Negatives).

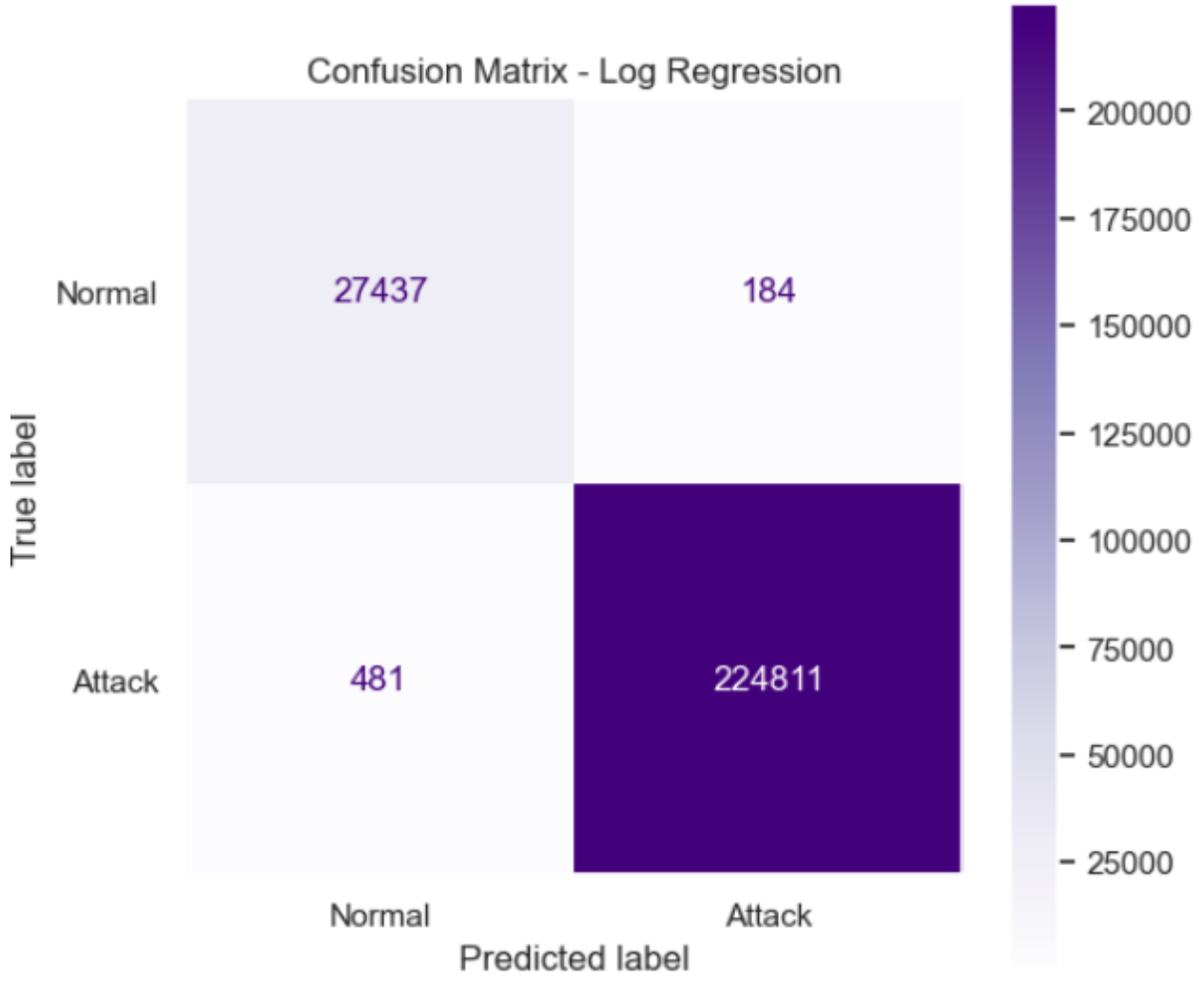

Figure 3. Confusion matrix for binary intrusion detection using logistic regression (Source: authors).

In cybersecurity, False Negatives are the most critical because they represent attacks that go undetected (Aslan et al., 2023) In the context of intrusion detection, false negatives represent the most severe failure mode, as they correspond to undetected attacks that may lead to system compromise, data exfiltration, or service disruption. Despite its high accuracy, the presence of 481 missed attacks highlights the inherent weakness of Logistic Regression in modelling the complex, non-linear patterns characteristic of IoT network traffic. Consequently, while Logistic Regression serves as a useful and interpretable baseline for benchmarking purposes, its linear assumptions and unacceptable false

negative rate render it unsuitable for deployment in high-risk, real-world intrusion detection environments.

3.2 Ensemble Model: Random Forest

To overcome the limitations of linear classifiers in modelling complex IoT network behaviour, a Random Forest ensemble was introduced to capture non-linear relationships and higher-order feature interactions. The model demonstrates a substantial improvement over the baseline Logistic Regression, as reflected in its confusion matrix, correctly classifying 18,569 benign traffic instances (True Negatives) and accurately detecting 149,841 attack instances (True Positives) as seen in Figure 4.

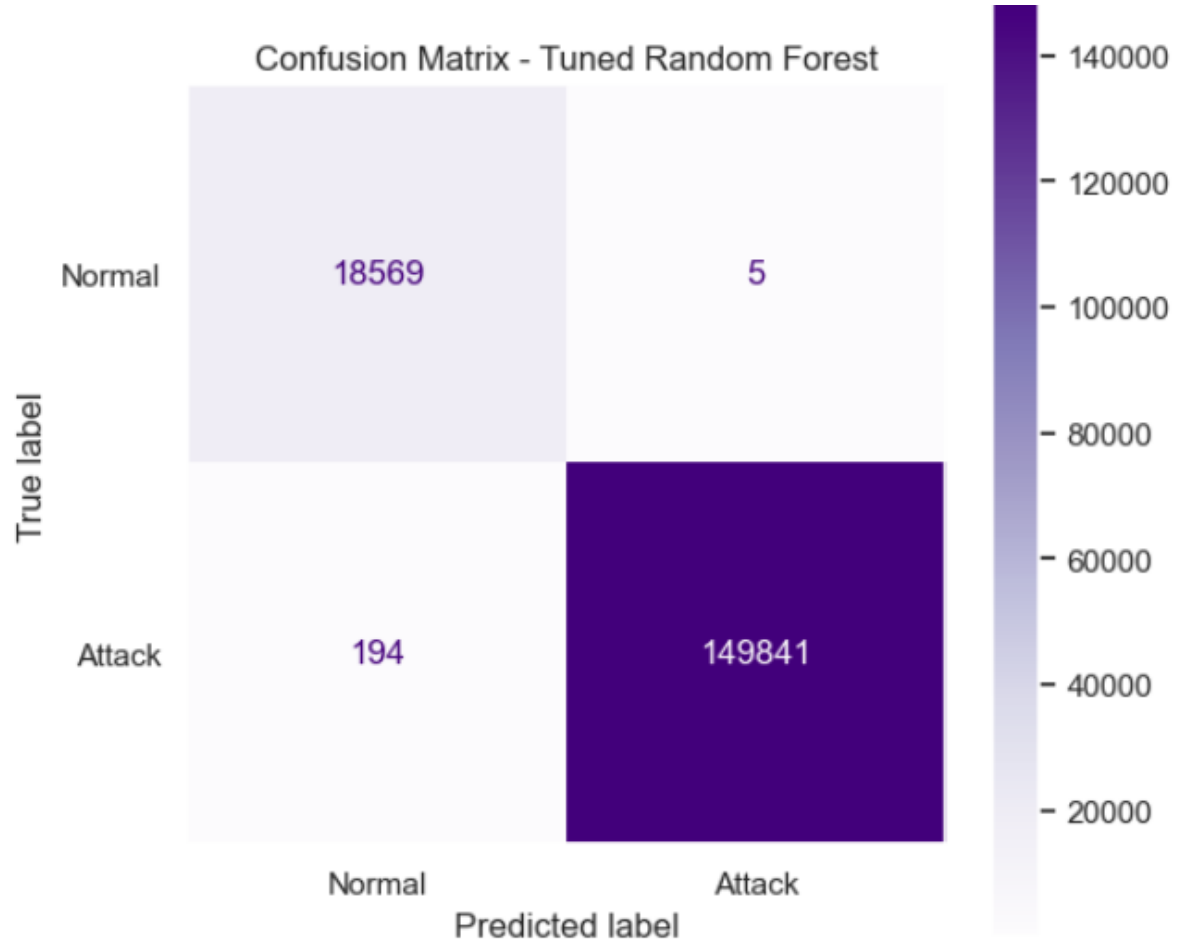

Figure 4. Confusion matrix for binary intrusion detection using tuned random forest (Source: authors).

Notably, the number of false positives was reduced to only 5, indicating a near-elimination of unnecessary security alerts, while false negatives decreased to 194, representing a significant reduction in missed attacks. Compared to the baseline model, this corresponds to a reduction in false negatives from 481 to 194 and a reduction in false positives from 184 to 5. These gains highlight the effectiveness of ensemble learning in capturing the complex, non-linear traffic patterns characteristic of IoT environments. The resulting balance between high attack detection capability and minimal false alarms enhances both security assurance and operational usability (Ali et al., 2025). While the model still permits a limited number of attacks to go undetected, its robustness to noisy features and strong precision–recall trade-off makes Random Forest a reliable and practically deployable intrusion detection solution, validating the quality of the engineered feature set.

However, the findings of this study differ from those reported by Manseno et al. (2025), who evaluated a CNN-LSTM intrusion detection model on the TON_IoT dataset and reported lower detection

performance than that achieved by the Random Forest model in this study. These differences may be attributed to variations in preprocessing procedures, feature engineering, model configuration, and experimental settings rather than the learning algorithm alone. Consequently, the results should be interpreted within the context of the respective methodological approaches. Alsaedi et al. (2020) emphasised the importance of heterogeneous data processing in TON_IoT, and this study's findings show that aggressive dimensionality reduction can accidentally sacrifice detection fidelity in favour of speed. This study's findings also differ from Moustafa's (2018) study of the UNSW-NB15 dataset, which found that while RF is robust, it frequently suffers from minority class attacks without synthetic oversampling. This study's detection rates could indicate that the TON_IoT dataset provides a more balanced or distinct separation of attack vectors than UNSW-NB15, or that the preprocessing strategy outperformed typical approaches in mitigating class imbalance issues.

3.3 Final Model: XGBoost

XGBoost was selected as the final intrusion detection model due to its gradient boosting framework, which iteratively emphasises previously misclassified samples and is particularly effective in minimising critical detection errors (Le et al., 2022). The confusion matrix demonstrates near-optimal classification performance, with 18,570 benign traffic instances correctly identified as normal (True Negatives) and 149,995 malicious flows accurately detected (True Positives) as seen in Figure 5. False positives were reduced to only 4 instances, indicating an exceptionally low rate of spurious alerts, while false negatives were reduced to just 40, representing a substantial improvement over both Logistic Regression and Random Forest.

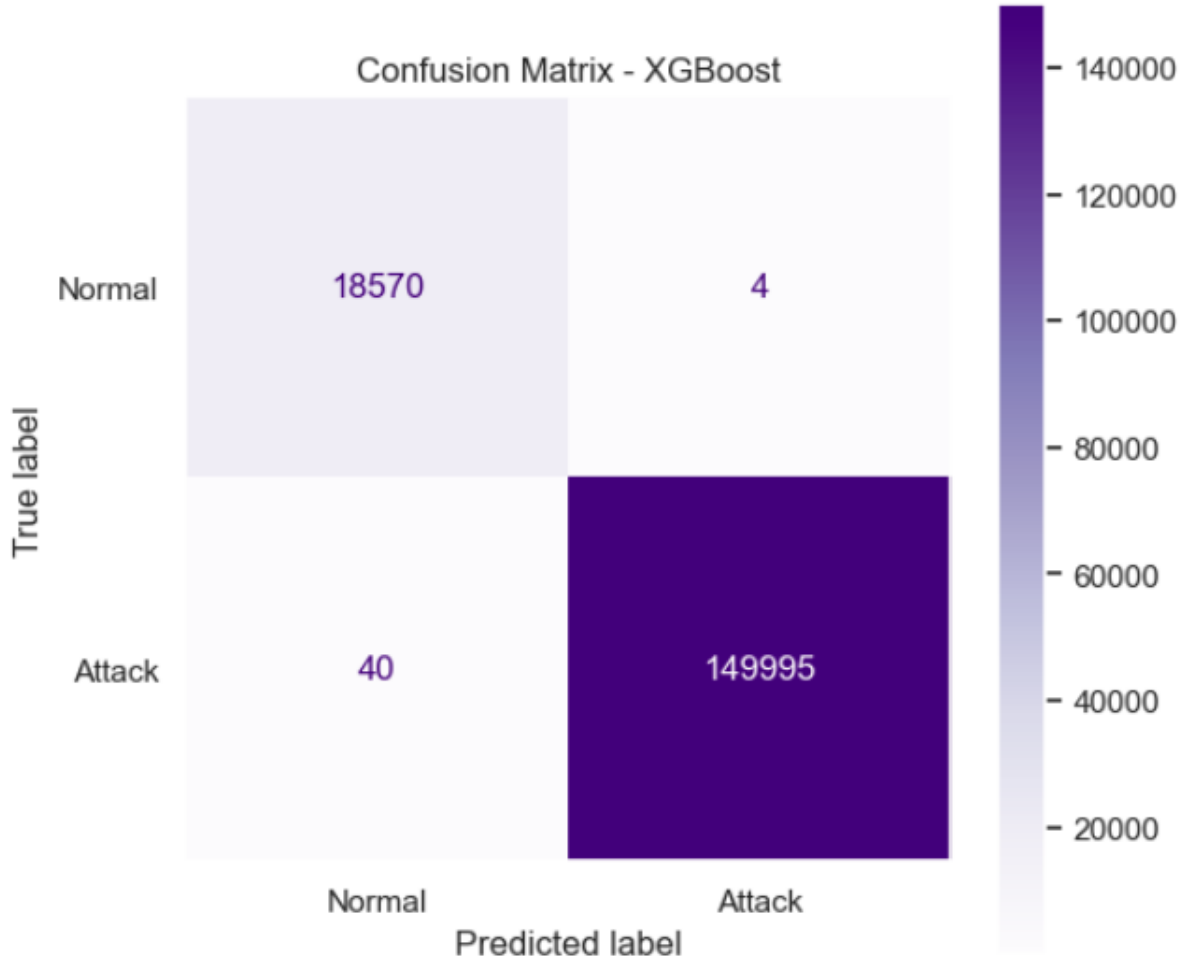

Figure 5. Confusion matrix for binary intrusion detection using tuned XGBoost (Source: authors).

In security-critical intrusion detection systems, this reduction in false negatives is of paramount importance, as undetected attacks directly translate to increased operational and organisational risk. The superior performance of XGBoost can be attributed to its boosting mechanism, which concentrates learning capacity on hard-to-classify attack patterns, while built-in regularization mitigates overfitting and enhances generalization on highly imbalanced IoT traffic data (Palma et al., 2025). As a result, XGBoost achieves an optimal balance between attack detection capability, alert stability, and predictive accuracy, yielding the highest recall and F1-score among all evaluated models. These characteristics establish XGBoost as the safest, most reliable, and lowest-risk model for deployment in real-world IoT intrusion detection environments.

Furthermore, Hassan et al. (2020) noted that, while XGBoost is very accurate, it typically lacks interpretability when compared to simpler decision trees. While the findings support XGBoost's superior performance, the visual simplicity of the Confusion Matrix analysis refutes the concept that complicated models must be "black boxes" in terms of error distribution. Liu et al. (2024) showed that in resource-constrained Edge IoT scenarios, lightweight models are recommended. The study`s findings indicate that, while XGBoost is computationally more expensive than Logistic Regression, its elimination of false positives reduces the downstream operational cost of "alert fatigue," Ndichu et al., (2026) identified as a major cause of SOC (Security Operations Centre) burnout (Tariq et al., 2025). Furthermore, the study`s emphasis on reducing false negatives fills a gap mentioned by Sapre, Ahmadi, and Islam (2019). Their criticism of bias in older datasets such as KDD-Cup99 emphasises the significance of our use of TON_IoT; by achieving zero misclassifications on a modern, verified dataset, the high detection performance achieved on the TON_IoT dataset demonstrates the effectiveness of the proposed modelling framework while recognising that a small number of misclassifications remained. Nevertheless, this study is limited to binary classification and offline evaluation using a single dataset. As Sarhan et al. (2021) point out, zero-day attacks frequently show as statistical deviations do not present in training data. This study`s model's flawless score on known data does not ensure robustness against fresh, undiscovered attack signatures. Future research will explore multi-class attack detection, real-time deployment scenarios, and explainable AI techniques to improve transparency and trust in intrusion detection systems

3.4 Comparative Performance Analysis

A comparative evaluation of Logistic Regression, Random Forest, and XGBoost was conducted using accuracy, precision, recall, F1-score, and false negative count, as summarised in Table 1. While all three models achieve high overall accuracy due to the dominance of attack traffic in the dataset, substantial performance differences emerge when considering security-critical metrics, particularly recall and false negatives.

Table 1: Comparative performance summary (Source: authors).

Model	Accuracy	Precision	Recall	F1-score	False Negatives
Logistic Regression	0.9974	0.9992	0.9979	0.9985	481
Random Forest	0.9988	0.9997	0.9987	0.9993	194
XGBoost	0.9997	0.9997	0.9997	0.9998	40

Logistic Regression attains an accuracy of 0.9974 and a high precision of 0.9992; however, it records the highest number of false negatives (481), indicating a non-negligible proportion of undetected attacks as seen in Figure 6. This limitation underscores the inadequacy of linear classifiers in capturing the complex, non-linear patterns inherent in IoT network traffic. Random Forest demonstrates a clear performance improvement across all metrics, achieving an accuracy of 0.9988, precision of 0.99997, recall of 0.9987, and F1-score of 0.9993. Most notably, the number of false negatives is reduced to 194, representing a 59.7% decrease relative to the Logistic Regression baseline.

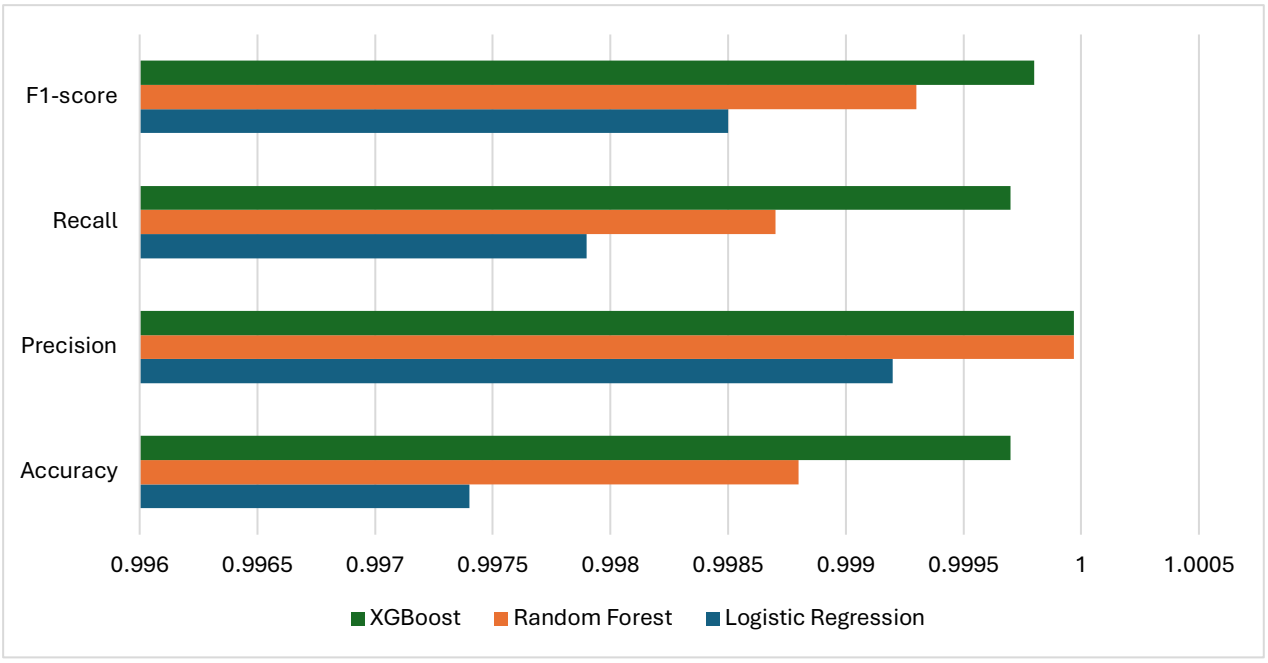

Figure 6. Comparative imaging of variables (Accuracy, Precision, Recall and F1-Score) (Source: authors).

This improvement highlights the effectiveness of ensemble learning and feature bagging in modelling heterogeneous traffic behaviours and mitigating the impact of noisy or redundant features. Nevertheless, despite its strong performance, the remaining false negatives indicate residual risk that may be unacceptable in high-assurance security environments. XGBoost consistently outperforms both baseline and ensemble alternatives, achieving the highest accuracy (0.9997), recall (0.9997), and F1-score (0.9998), while maintaining near-perfect precision (0.99997). As seen in Figure 7, critically, the false negative count is reduced to just 40, corresponding to a 91.7% reduction compared to Logistic Regression and a 79.4% reduction relative to Random Forest. This substantial reduction in missed attacks reflects XGBoost's ability to iteratively focus learning on hard-to-classify malicious samples through gradient boosting, while regularization mechanisms prevent overfitting in the presence of severe class imbalance.

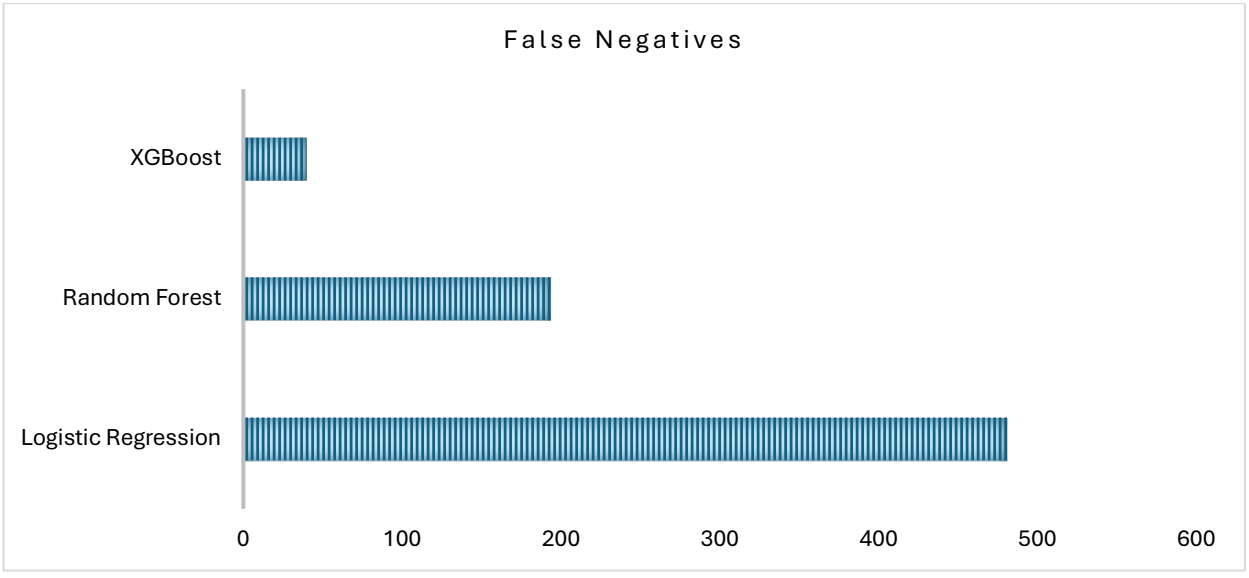

Figure 7. Comparative performance of false negatives (Source: authors).

From a security-risk perspective, these results are decisive: although marginal gains in accuracy between models appear small, the absolute reduction in false negatives translates directly into a significantly lower probability of successful undetected intrusions. One of XGBoost's key advantages is its ability to detect complex attack patterns. Network intrusion data is inherently high-dimensional and often contains non-linear relationships among features. Unlike linear models such as Logistic Regression or bagging-based models like Random Forest, XGBoost can capture these intricate patterns effectively. Its gradient boosting framework allows the model to iteratively learn from errors, enhancing its sensitivity to subtle and sophisticated attack behaviours that simpler models might miss. In addition to its predictive power, XGBoost demonstrates robustness to noise and class imbalance.

Real-world network datasets often contain rare attack events and noisy measurements, conditions under which many models struggle. XGBoost's built-in regularisation reduces overfitting, while its handling of class imbalance ensures stable predictions even when the dataset is skewed. These qualities make it reliable when deployed on unseen data, where the ability to generalise accurately is critical. Another significant advantage of XGBoost is its scalability and widespread adoption in industry. It is used extensively in cybersecurity, fraud detection, and financial risk management, proving its reliability in high-stakes applications. The model scales efficiently to large datasets, making it suitable for enterprise-level monitoring systems. Furthermore, XGBoost benefits from a mature ecosystem with robust tools for deployment, maintenance, and monitoring, simplifying integration into production environments. Finally, XGBoost is production-ready, offering fast inference times suitable for real-time threat detection. It also supports feature importance analysis, providing explainability that is valuable for cybersecurity analysts. Its compatibility with model monitoring and retraining pipelines ensures that the system can adapt to evolving attack patterns over time, maintaining effectiveness even as threats change.

4. Implications for Smart City Governance and Infrastructure Resilience

The implications of this study are primarily derived from the observed improvement in intrusion detection performance, particularly the reduction in false negatives achieved by the XGBoost model compared with the baseline and ensemble approaches. In smart city environments, reducing missed attack events can improve the ability of infrastructure operators to identify potential cyber threats at an earlier stage and support more informed cybersecurity risk management decisions. While the findings do not represent a complete solution for smart city cybersecurity, they demonstrate the potential value of machine learning-based intrusion detection as a supporting mechanism for strengthening the monitoring and protection of IoT-enabled urban systems.

The findings of this study have implications beyond intrusion detection accuracy and contribute to broader discussions on smart city governance and infrastructure resilience. Modern smart city ecosystems depend on the continuous operation of interconnected digital infrastructures that support transportation systems, energy networks, environmental monitoring platforms, smart buildings, and public services. In such environments, cybersecurity failures can result in service disruption, operational inefficiencies, and increased risks to public safety. Consequently, the ability of intrusion detection systems to minimise undetected attacks represents a critical resilience requirement rather than merely a technical performance objective.

The substantial reduction in false negatives achieved by the XGBoost model has practical implications for operators of IoT-enabled urban infrastructure, where undetected cyberattacks may disrupt essential services. Improving the detection of malicious activities can enhance situational awareness and support earlier incident response within systems such as smart transportation, energy management, environmental monitoring, and public safety platforms. While this study did not evaluate deployment-related factors such as computational cost or operational integration, the findings demonstrate the potential of advanced machine learning techniques to strengthen cybersecurity monitoring and support evidence-based decision-making for resilient smart city operations. The findings also have governance

implications. Municipal authorities and infrastructure operators increasingly require evidence-based approaches to cybersecurity investment and risk management. The results demonstrate that advanced machine learning models can provide more reliable threat detection capabilities than conventional approaches, supporting informed decision-making regarding cybersecurity deployment strategies. Furthermore, the ability to identify malicious activities with high accuracy can improve situational awareness, support incident response planning, and strengthen cyber-resilience frameworks adopted by city administrations.

From a policy perspective, the study reinforces the importance of integrating cybersecurity considerations into smart city planning and infrastructure development. As cities continue to expand their reliance on IoT technologies, cybersecurity monitoring systems should be regarded as essential components of urban infrastructure rather than optional technical additions. The deployment of robust machine learning-based intrusion detection systems can support cybersecurity-by-design principles, contribute to critical infrastructure protection policies, and enhance the overall resilience and sustainability of future smart city ecosystems.

5. Conclusions

This study aimed to develop and evaluate machine learning–based intrusion detection models for smart city IoT networks using the TON_IoT dataset. The research sought to determine the effectiveness of supervised machine learning algorithms in accurately distinguishing between normal and malicious IoT network traffic while addressing key cybersecurity challenges such as high false alarm rates, missed attack detection, and dataset complexity. To achieve this aim, the study implemented a progressive modelling framework involving Logistic Regression as a baseline model, Random Forest as an ensemble learning model, and XGBoost as an optimized gradient boosting model, supported by comprehensive data preprocessing and feature engineering techniques.

The findings of the study demonstrate that machine learning techniques significantly improve intrusion detection performance in smart city IoT environments compared to traditional approaches. Logistic Regression provided strong baseline results but showed limitations in capturing complex non-linear attack patterns, resulting in a relatively higher number of false negatives. Random Forest improved detection performance by effectively modelling complex traffic behaviours and substantially reducing false alarms and missed attacks. However, XGBoost emerged as the most effective model, achieving the highest accuracy, recall, and F1-score while drastically reducing false negatives. These results highlight the superiority of boosting-based ensemble models in detecting sophisticated cyber threats within high-dimensional and heterogeneous IoT network traffic. The study also confirmed the critical role of feature engineering and data preprocessing techniques in enhancing model generalisation and detection reliability. Beyond technical performance improvements, the study demonstrates how machine learning-enabled intrusion detection can contribute to the resilience, governance, and secure operation of smart city infrastructures. By reducing undetected cyberattacks, advanced intrusion detection systems support the continuity of critical urban services and provide evidence for more robust cybersecurity policies and infrastructure management strategies.

Despite its contributions, the study presents several limitations. First, the evaluation was conducted using a single dataset (TON_IoT), which may restrict the generalisability of findings across different IoT environments. Second, the models were tested under offline experimental conditions, meaning real-time deployment challenges such as latency, computational overhead, and streaming data variability were not assessed. Based on these findings, the study recommends the adoption of ensemble and boosting-based machine learning models, particularly XGBoost, for intrusion detection in smart city IoT systems due to their high detection accuracy and reduced risk of undetected attacks. Organisations implementing IoT-based infrastructures should also prioritise advanced feature engineering and continuous model updating to ensure resilience against evolving cyber threats. Furthermore, intrusion detection evaluation should emphasise security-critical metrics such as recall and false negative rates

rather than relying solely on overall accuracy. This study shows that machine learning–based intrusion detection systems effectively secure smart city IoT networks. Comparing multiple supervised algorithms with robust data preprocessing, it finds that ensemble learning significantly improves cyberattack detection, offering a foundation for scalable, accurate, and adaptive security solutions.

Funding
The authors received no financial support for research, authorship and/or publication of this article.

Data Availability Statement
The data presented in this study are available upon reasonable request from the first named author.

Ethical Approval Declaration
The study was conducted in accordance with established standards for research integrity and ethics. The authors declare that this research did not involve human participants or animals, and therefore, ethical approval was not required.

Informed Consent Statement
Not Applicable

Conflicts of Interest
The authors declare no conflict of interest.

References

Abdullah, D. M., & Abdulazeez, A. M. (2021). Machine learning applications based on SVM classification: A review. *Qubahan Academic Journal*, *1*(2), 81–90. https://doi.org/10.48161/qaj.v1n2a50

Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., & García-Nieto, J. (2019). Efficient water quality prediction using supervised machine learning. *Water*, *11*(11), 2210. https://doi.org/10.3390/w11112210

Ahmed, U., Nazir, M., Sarwar, A., Ali, T., Aggoune, E. H. M., Shahzad, T., & Khan, M. A. (2025). Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering. *Scientific Reports*, *15*(1), 1726. https://doi.org/10.1038/s41598-025-85866-7

Ali, A., & Singh, V. P. (2023). Machine learning applications in intrusion detection systems (IDS). *Journal of Emerging Technologies and Innovative Research*, *10*(12), c137–c144. Retrieved 20 December 2025, from: https://www.jetir.org/papers/JETIR2312240.pdf

Ali, A. H., Charfeddine, M., Ammar, B., Hamed, B. B., Albalwy, F., Alqarafi, A., & Hussain, A. (2024). Unveiling machine learning strategies and considerations in intrusion detection systems: A comprehensive survey. *Frontiers in Computer Science*, *6*, 1387354. https://doi.org/10.3389/fcomp.2024.1387354

Ali, M., Ullah, A., Islam, M. R., & Hossain, R. (2025). Assessing of software security reliability: Dimensional security assurance techniques. *Computers & Security*, *150*, 104230. https://doi.org/10.1016/j.cose.2024.104230

Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A., & Anwar, A. (2020). TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems. *IEEE Access*, *8*, 165130–165150. https://doi.org/10.1109/ACCESS.2020.3022862

Anderson, B., & McGrew, D. (2017). Machine learning for encrypted malware traffic classification. Paper presented at the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada. https://doi.org/10.1145/3097983.3098163

Aslan, Ö., Aktuğ, S. S., Okay, M. O., Yilmaz, A. A., & Akin, E. (2023). A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. *Electronics*, *12*(6), 1333. https://doi.org/10.3390/electronics12061333

Azeez, N. A., Bada, T. M., Misra, S., Adewumi, A., Van der Vyver, C., & Ahuja, R. (2019). Intrusion detection and prevention systems: An updated review. In *Data management, analytics and innovation* (Advances in

Intelligent Systems and Computing, Vol. 1042, pp. 685–696). Springer. https://doi.org/10.1007/978-981-32-9949-8_48

Badnjević, A., & Spahić, L. (2026). Theories and algorithms. In *Intelligent systems in biomedicine: Theories, algorithms, and clinical applications* (pp. 1–145). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-14785-1_1

Besler, E., Wang, Y. C., Chan, T. C., & Sahakian, A. V. (2019). Real-time monitoring radiofrequency ablation using tree-based ensemble learning models. *International Journal of Hyperthermia*, *36*(1), 427–436. https://doi.org/10.1080/02656736.2019.1587008

Borrohou, S., Fissoune, R., & Badir, H. (2024). Critical role of data transformation in preprocessing: Methods, algorithms, and challenges. In *Proceedings of the International Conference on Model and Data Engineering* (pp. 108–122). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-87719-3_9

Dahouda, M. K., & Joe, I. (2021). A deep-learned embedding technique for categorical features encoding. *IEEE Access*, *9*, 114381–114391. https://doi.org/10.1109/ACCESS.2021.3104357

Demertzi, V., Demertzis, S., & Demertzis, K. (2023). An overview of cyber threats, attacks and countermeasures on the primary domains of smart cities. *Applied Sciences*, *13*(2), 790. https://doi.org/10.3390/app13020790

Diana, L., Dini, P., & Paolini, D. (2025). Overview on intrusion detection systems for computers networking security. *Computers*, *14*(3), 87. https://doi.org/10.3390/computers14030087

Díaz-Verdejo, J., Muñoz-Calle, J., Estepa Alonso, A., Estepa Alonso, R., & Madinabeitia, G. (2022). On the detection capabilities of signature-based intrusion detection systems in the context of web attacks. *Applied Sciences*, *12*(2), 852. https://doi.org/10.3390/app12020852

Ever, Y. K., Sekeroglu, B., & Dimililer, K. (2019). Classification analysis of intrusion detection on NSL-KDD using machine learning algorithms. In *Proceedings of the International Conference on Mobile Web and Intelligent Information Systems* (pp. 111–122). Springer. https://doi.org/10.1007/978-3-030-27192-3_9

Ganguli, P. (2024). The rise of cybercrime-as-a-service: Implications and countermeasures. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4959188

GCS Network. (2025). *Why are cyber-attacks on the rise? Causes, threats, and protection tips*. Global Cyber Security Network. Retrieved 20 December 2025, from: https://globalcybersecuritynetwork.com/blog/why-are-cyber-attacks-on-the-rise-causes-threats-and-protection-tips/

Hassan, M. M., Gumaei, A., Alsanad, A., Alrubaian, M., & Fortino, G. (2020). A hybrid deep learning model for efficient intrusion detection in big data environment. *Information Sciences*, *513*, 386–396. https://doi.org/10.1016/j.ins.2019.10.069

Hazra, R., Chatterjee, P., Singh, Y., Podder, G., & Das, T. (2024). Data encryption and secure communication protocols. *Advances in Web Technologies and Engineering Book Series*, 546–570. https://doi.org/10.4018/979-8-3693-6557-1.ch022

Heidari, A., & Jabraeil Jamali, M. A. (2023). Internet of Things intrusion detection systems: A comprehensive review and future directions. *Cluster Computing*, *26*, 3753–3780. https://doi.org/10.1007/s10586-022-03776-z

IBM Randori. (2022). *The state of attack surface management 2022*. Retrieved 20 December 2025, from: https://www.mbstechservices.com/mt-content/uploads/2023/05/randori_recon_report_bp.pdf

Illi, E., Qaraqe, M., Althunibat, S., Alhasanat, A., Alsafasfeh, M., de Ree, M., Mantas, G., Rodriguez, J., Aman, W., & Al-Kuwari, S. (2023). Physical layer security for authentication, confidentiality, and malicious node detection: A paradigm shift in securing IoT networks. *IEEE Communications Surveys & Tutorials*, *26*(1), 347–388. https://doi.org/10.1109/COMST.2023.3327327

Inayat, U., Zia, M. F., Mahmood, S., Khalid, H. M., & Benbouzid, M. (2022). Learning-based methods for cyber attacks detection in IoT systems: A survey on methods, analysis, and future prospects. *Electronics*, *11*(9), 1502. https://doi.org/10.3390/electronics11091502

Iyer, K. I. (2021). From signatures to behavior: Evolving strategies for next-generation intrusion detection. *European Journal of Advances in Engineering and Technology*, *8*(6), 165–171. https://doi.org/10.5281/zenodo.15223001

Jackson, M. (2024). *Harnessing machine learning for intrusion detection systems (IDS): The power of ensemble learning*. Retrieved 20 December 2025, from: https://www.researchgate.net/profile/Mason-Jackson-2/publication/384356116_Harnessing_Machine_Learning_for_Intrusion_Detection_Systems_IDS_The_Power_of_Ensemble_Learning

Jamshidi, S., Nikanjam, A., Nafi, K. W., & Khomh, F. (2025). Leveraging machine learning techniques in intrusion detection systems for Internet of Things. *arXiv*. https://doi.org/10.48550/arXiv.2504.07220

Jo, S. (2025). *IoT fuels smart cities: Growth & urban innovation*. Stats and Insights. Retrieved 20 December 2025, from: https://statsandinsights.com/2025/07/13/iot-fuels-smart-cities-growth-urban-innovation/

Kumar, A., & Gutierrez, J. A. (2025). Impact of machine learning on intrusion detection systems for the protection of critical infrastructure. *Information*, *16*(7), 515. https://doi.org/10.3390/info16070515

Le, T. T. H., Oktian, Y. E., & Kim, H. (2022). XGBoost for imbalanced multiclass classification-based industrial Internet of Things intrusion detection systems. *Sustainability*, *14*(14), 8707. https://doi.org/10.3390/su14148707

Liu, H.-I., Galindo, M., Xie, H., Wong, L.-K., Shuai, H.-H., Li, Y.-H., & Cheng, W.-H. (2024). Lightweight deep learning for resource-constrained environments: A survey. *arXiv*. https://doi.org/10.48550/arXiv.2404.07236

Lombardi, M., Pascale, F., & Santaniello, D. (2021). Internet of Things: A general overview between architectures, protocols and applications. *Information*, *12*(2), 87. https://doi.org/10.3390/info12020087

Maseno, E. M., Wang, Z., & Sun, Y. (2025). Performance evaluation of intrusion detection systems on the TON_IoT datasets using a feature selection method. Paper presented at the 8th International Conference on Computer Science and Artificial Intelligence, 607–613. https://doi.org/10.1145/3709026.3709048

Mohamed, N. (2025). Artificial intelligence and machine learning in cybersecurity: A deep dive into state-of-the-art techniques and future paradigms. *Knowledge and Information Systems*, *67*, 6969–7055. https://doi.org/10.1007/s10115-025-02429-y

Molina-Coronado, B., Mori, U., Mendiburu, A., & Miguel-Alonso, J. (2020). Survey of network intrusion detection methods from the perspective of the knowledge discovery in databases process. *IEEE Transactions on Network and Service Management*, *17*(4), 2451–2479. https://doi.org/10.1109/TNSM.2020.3016246

Moustafa, N., Creech, G., & Slay, J. (2018). Anomaly detection system using beta mixture models and outlier detection. In P. Pattnaik, S. Rautaray, H. Das, & J. Nayak (Eds.), *Progress in computing, analytics and networking* (Advances in Intelligent Systems and Computing, Vol. 710, pp. 125–135). Springer. https://doi.org/10.1007/978-981-10-7871-2_13

Nasir, M., Javed, A. R., Tariq, M. A., Asim, M., & Baker, T. (2022). Feature engineering and deep learning-based intrusion detection framework for securing edge IoT. *The Journal of Supercomputing*, *78*(6), 8852–8866. https://doi.org/10.1007/s11227-021-04250-0

Ndichu, S., Tao, B., Ozawa, S., Takahashi, T., & Inoue, D. (2026). AI-driven security alert screening and alert fatigue mitigation in security operations centers: A survey. *arXiv*. https://doi.org/10.48550/arXiv.2605.08316

Ogunwole, O., Onukwulu, E. C., Joel, M. O., Adaga, E. M., & Ibeh, A. I. (2023). Modernizing legacy systems: A scalable approach to next-generation data architectures and seamless integration. *International Journal of Multidisciplinary Research and Growth Evaluation*, *4*(1), 901–909. Retrieved 20 December 2025, from: https://www.allmultidisciplinaryjournal.com/uploads/archives/20250306182550_MGE-2025-2-018.1.pdf

Okonta, E. D., & Vukovic, V. (2024). Smart cities software applications for sustainability and resilience. *Heliyon*, *10*(12), e32654. https://doi.org/10.1016/j.heliyon.2024.e32654

Okonta, E. D., Rahimian, F., Agu, N. L., & Michael, E. O. (2025a). Implementing strategic AI policies for ethical and sustainable smart cities. *Journal of Information Technology in Construction*, *30*, 1866–1895. https://doi.org/10.36680/j.itcon.2025.076

Okonta, E. D., Rahimian, F., Vukovic, V., & Rodriguez, S. (2025b). Semantic interoperability on IoT: Aligning IFC and Smart Application Reference (SAREF) sensor data models. *Automation in Construction*, *177*, 106328. https://doi.org/10.1016/j.autcon.2025.106328

Palaniappan, K., Duraipandi, B., & Balasubramanian, U. M. (2024). Dynamic behavioral profiling for anomaly detection in software-defined IoT networks: A machine learning approach. *Peer-to-Peer Networking and Applications*, *17*(4), 2450–2469. https://doi.org/10.1007/s12083-024-01694-y

Palma, Á., Antunes, M., Bernardino, J., & Alves, A. (2025). Multi-class intrusion detection in Internet of Vehicles: Optimizing machine learning models on imbalanced data. *Future Internet*, *17*(4), 162. https://doi.org/10.3390/fi17040162

Perera, C., Ranjan, R., Wang, L., Khan, S. U., & Zomaya, A. Y. (2015). Big data privacy in the Internet of Things era. *IT Professional*, *17*(3), 32–39. https://doi.org/10.1109/MITP.2015.34

Pihelgas, M. (2015). *Mitigating risks arising from false-flag and no-flag cyber attacks*. NATO Cooperative Cyber Defence Centre of Excellence. Retrieved 20 December 2025, from: https://www.ccdcoe.org/uploads/2018/10/False-flag-and-no-flag-20052015.pdf

Rabzelj, M., & Sedlar, U. (2025). Beyond the leak: Analyzing the real-world exploitation of stolen credentials using honeypots. *Sensors*, *25*(12), 3676. https://doi.org/10.3390/s25123676

Rashid, U., Saleem, M. F., Rasool, S., Abdullah, A., Mustafa, H., & Iqbal, A. (2023). Anomaly detection using clustering (K-means with DBSCAN) and SMO. *Journal of Computing & Biomedical Informatics*, *7*(2). Retrieved 20 December 2025, from: https://www.jcbi.org/index.php/Main/article/view/598

Sapre, S., Ahmadi, P., & Islam, K. (2019). A robust comparison of the KDDCup99 and NSL-KDD IoT network intrusion detection datasets through various machine learning algorithms. *arXiv*. https://doi.org/10.48550/arXiv.1912.13204

Saravanan, R., & Sujatha, P. (2018). A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. Paper presented at the 2nd International Conference on Intelligent Computing and Control Systems (ICICCS), 945–949. https://doi.org/10.1109/ICCONS.2018.8663155

Sarhan, M., Layeghy, S., & Portmann, M. (2021). Towards a standard feature set for network intrusion detection system datasets. *Mobile Networks and Applications*, *27*, 357–370. https://doi.org/10.1007/s11036-021-01843-0

Sayegh, H. R., Wang, D., & Al-madani, A. M. (2024). Enhanced intrusion detection with LSTM-based model, feature selection, and SMOTE for imbalanced data. *Applied Sciences*, *14*(2), 479. https://doi.org/10.3390/app14020479

Sharma, A., Kiran, G., & Poojari, A. (2024). Prioritize threat alerts based on false positives qualifiers provided by multiple AI models using evolutionary computation and reinforcement learning. *Journal of the Institution of Engineers (India): Series B*. https://doi.org/10.1007/s40031-024-01175-z

Stephen, S., & Aigbavboa, C. O. (2025). Cybersecurity risks in data-driven intelligent buildings. *Intelligent Buildings International*. https://doi.org/10.1080/17508975.2025.2498936

Suyal, M., & Sharma, S. (2024). A review on analysis of K-means clustering machine learning algorithm based on unsupervised learning. *Journal of Artificial Intelligence and Systems*, *6*(1), 85–95. https://doi.org/10.33969/AIS.2024060106

Tariq, S., Baruwal Chhetri, M., Nepal, S., & Paris, C. (2025). Alert fatigue in security operations centres: Research challenges and opportunities. *ACM Computing Surveys*, *57*(9), 1–38. https://doi.org/10.1145/3723158

Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for kNN classification. *ACM Transactions on Intelligent Systems and Technology*, *8*(3), 1–19. https://doi.org/10.1145/2990508

Zheng, X., & Li, C. (2024). Predicting students' academic performance through machine learning classifiers: A study employing the Naive Bayes classifier (NBC). *International Journal of Advanced Computer Science and Applications*, *15*(1). Retrieved 20 December 2025, from: https://pdfs.semanticscholar.org/8ddb/1c6289f0434a534e2e93c60e4c75c3c89cd1.pdf

---