Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 3||pp 30-42||June||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

Encrypted Remote Access Trojan Detection: A Machine Learning Approach with Real-World and Open Datasets

Emmanuel Sebakara[1*] & Dr. K N Jonathan[2]
[1,2]University of Lay Adventists of Kigali, Rwanda
Corresponding Email: emmanuelsaba183@gmail.com; phialn1@gmail.com

Accepted: 13 May 2025 || Published: 05 June 2025

Abstract

The increasing use of encryption by cyber attackers to conceal Remote Access Trojans (RATs) challenges traditional signature-based detection systems, which struggle with encrypted traffic and leave security gaps. In this study, we propose a privacy-preserving, machine-learning-based framework that detects encrypted RATs without decrypting traffic. Instead, it analyzes behavioral indicators and metadata, including packet timing anomalies, TLS handshake irregularities, and persistent unidirectional flows. We evaluated our approach using two datasets: a public Kaggle dataset (177,482 labeled records, 85 features) and an anonymized internal dataset from Company X (40,000 samples, 27 features). Among four tested models—Logistic Regression, Decision Tree, Random Forest, and XGBoost—Random Forest performed best, achieving 74.83% and 72.11% accuracy on the Company X and Kaggle datasets, respectively, outperforming a baseline signature-based system (53.8% accuracy). Our model also showed strong generalization, with 80% correct predictions across sample-based evaluations, demonstrating its readiness for real-world deployment. By ensuring privacy and delivering improved detection, our framework offers a scalable, adaptive alternative to traditional cybersecurity methods.

Keywords: *Remote Access Trojan (RAT), Encryption, Machine Learning, Behavioral Analysis, Cybersecurity, Privacy-Preserving Detection.*

How to Cite: Sebakara, E., & Jonathan, K. N. (2025). Encrypted Remote Access Trojan Detection: A Machine Learning Approach with Real-World and Open Datasets. *Journal of Information and Technology, 5*(3), 30-42.

1. Introduction

In today's digitally connected world, cybersecurity remains a pressing global concern for organizations, institutions, and governments. As of 2024, cyberattacks have increased by over 38% compared to the previous year, with malware-based incidents constituting nearly 22% of all breaches (Check Point Research, 2024). Among these threats, Remote Access Trojans (RATs) stand out as particularly dangerous due to their covert nature, ability to remotely control infected systems, and growing complexity through the integration of encryption.

According to Cybersecurity Ventures (2023), more than 60% of surveyed organizations experienced at least one incident involving encrypted malicious traffic. While encryption is vital for safeguarding legitimate communications, it is increasingly misused by cybercriminals to conceal RAT activity, complicating detection efforts. For instance, over 80% of malware in 2023 was delivered via encrypted channels (Google Transparency Report, 2023). This dual-use of encryption creates a critical paradox in cybersecurity: it secures legitimate data while simultaneously shielding malicious behavior.

The risk is especially acute in sectors such as finance, healthcare, and education, where data breaches can have severe consequences. Across Africa, organizations such as X Company (a regional firm involved in this study) report significant challenges in monitoring encrypted traffic. Up to 75% of suspicious events bypass conventional intrusion detection systems (IDS) due to encryption, signaling an urgent need for new detection strategies.

To address these challenges, this research—conducted at the University of Lay Adventists of Kigali (UNILAK)—explores the application of machine learning techniques to detect encrypted RAT traffic. Using a combination of publicly available datasets from Kaggle and private traffic logs from a local African firm, the study develops and evaluates models capable of identifying encrypted malicious activity without compromising user privacy. This work aims to contribute to the development of scalable, privacy-preserving cybersecurity solutions that are both practical and effective.

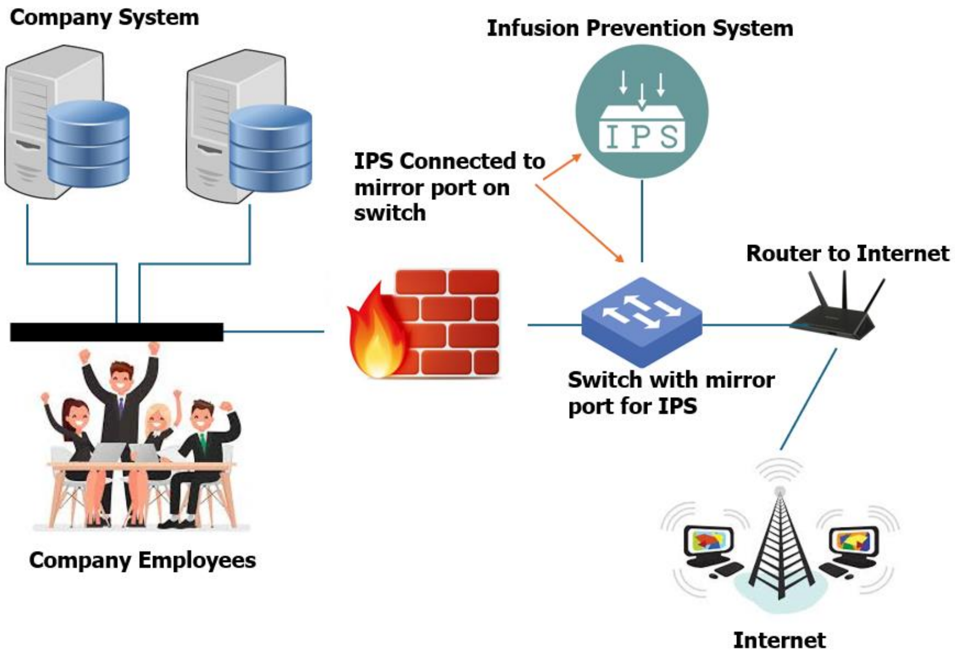
Company System
Infusion Prevention System
IPS
IPS Connected to mirror port on switch
Router to Internet
Switch with mirror port for IPS
Company Employees
Internet

Fig. 1: Diagram of a remote access trojan (RAT)

2. Literature Review

This review synthesizes empirical and theoretical insights into detecting encrypted Remote Access Trojans (RATs), drawing on peer-reviewed studies from Rwanda and globally.

A Trojan horse is a form of malware that masquerades as legitimate software to deliver malicious payloads, often establishing backdoors for unauthorized access (Aliyu et al., 2014; Wijayarathne, 2022). RATs have evolved from early tools like *Back Orifice* to sophisticated, often state-sponsored tools such as *Gh0st RAT* and *Xtreme RAT*, leveraging encrypted communication (e.g., HTTPS) to evade detection (Dimou et al., 2019).

Modern RATs exploit encryption (SSL/TLS) to obscure command-and-control (C2) traffic, making maliciously behavior indistinguishable from legitimate traffic (McDonald et al., 2022; Mokhtar et al., 2022). The proliferation of user-friendly malware creation tools and mobile platforms has further broadened the threat landscape (Eddy, 2014).

Trojan horses are classified into six categories: Remote Access, Data-Sending, Destructive, Proxy, FTP, and DoS Trojans—each exploiting different vectors and payloads (Spalka et al., 2002).

Detection of RATs primarily relies on signature-based and behavior-based methods. While signature-based detection is accurate for known threats, it falters against encryption and novel variants (Kwon et al., 2022). Behavioral and anomaly detection methods monitor deviations in network activity and system behavior, capable of flagging encrypted RATs but often suffer from false positives and high computational costs (Vasani et al., 2023).

Encryption, while essential for data confidentiality and integrity (Menders, 2019), paradoxically aids threat evasion, complicating detection efforts for cybersecurity systems (Johnson et al., 2016; Opderbeck, 2022). This necessitates techniques that analyze traffic behavior rather than content (Ozkan-Okay et al., 2023), with machine learning (ML) increasingly deployed to identify anomalies in encrypted streams (Mirza, 2024).

ML, a branch of AI, enables systems to learn from data and detect threats with minimal human input (Mosalam & Gao, 2024; Murphy, 2012). Models such as Random Forests, SVMs, Neural Networks, and KNN are widely applied in cybersecurity. Each offers strengths: Random Forests improve accuracy through ensemble learning, SVMs handle high-dimensional data, and Neural Networks capture complex patterns (Nigmatullin et al., 2020; Kwon et al., 2022; McDonald et al., 2022).

ML model development includes training, validation, and testing, and encompasses supervised, unsupervised, and reinforcement learning paradigms. Logistic regression is also used for binary classification (malicious vs. benign), with various algorithms offering trade-offs in interpretability and performance (Mirza, 2024).

Cybersecurity frameworks, grounded in socio-technical theories, advocate for integrated approaches that combine technical solutions with policy and human-centered strategies (Alshaikh et al., 2021; Tanwar, 2025; Johansen et al., 2022).

RATs often use evasion tactics like polymorphism, encrypted C2 channels, traffic mimicking (e.g., disguising as HTTPS or DNS), and rootkits, which complicate detection (Peter Szor, 2005; Gardiner et al., 2014; Zeltser, 2017; Cheruvu et al., 2019).

Recent studies propose innovative detection methods. Kwon et al. (2022) presented a hybrid statistical filter and autoencoder model that improved accuracy while reducing computational load. Vasani et al. (2023) compared anti-virus and adware detection systems, emphasizing the need for adaptive incident handling.

Despite advancements, research gaps persist—particularly in real-time detection of encrypted RATs and distinguishing maliciously encrypted metadata from benign activity. Existing models are often static or lack context-awareness (Kwon et al., 2022; Vasani et al., 2023).

This study addresses these gaps by proposing an adaptive detection framework combining behavioral heuristics and incremental learning. It incorporates features such as TLS handshake anomalies and packet burst patterns, enhanced by synthetic data for model robustness against zero-day attacks (Dimou et al., 2019). This approach distinguishes itself by unifying behavior-based insights and real-time ML adaptability, unlike prior works focused solely on anomaly or signature-based detection (McDonald et al., 2022; Kwon et al., 2022).

Guided by information security principles (Whitman & Mattord, 2011), the proposed framework includes data acquisition, preprocessing, feature engineering, and iterative model refinement to strengthen defense against sophisticated encrypted threats.

3. Methodology

3.1 Study Area and Dataset

This study focuses on detecting encrypted Remote Access Trojans (RATs) using machine learning techniques. Two datasets were utilized: a real-world network traffic dataset from an anonymized enterprise, referred to as "Company X," and a publicly available dataset from Kaggle. The combination of proprietary and open-source data ensured a wide range of traffic patterns and threat scenarios, improving the generalizability of the results.

3.2 Research Approach and Design

A quantitative, experimental research design was employed to evaluate the effectiveness of various machine learning models in detecting encrypted RATs. The process followed a structured sequence: data collection, preprocessing, model training, performance evaluation, and final benchmarking against established datasets.

3.3 Data Acquisition

Two datasets were used in this study: one from Kaggle (public) and the other from Company X (real-world). The Kaggle dataset contains 177,482 flow records, with a near-balanced distribution (51% Trojan, 49% benign). The Company X dataset contains 40,000 records with an imbalanced distribution (25.06% Trojan, 74.94% benign). No resampling or oversampling techniques were applied to maintain the inherent characteristics of real-world traffic.

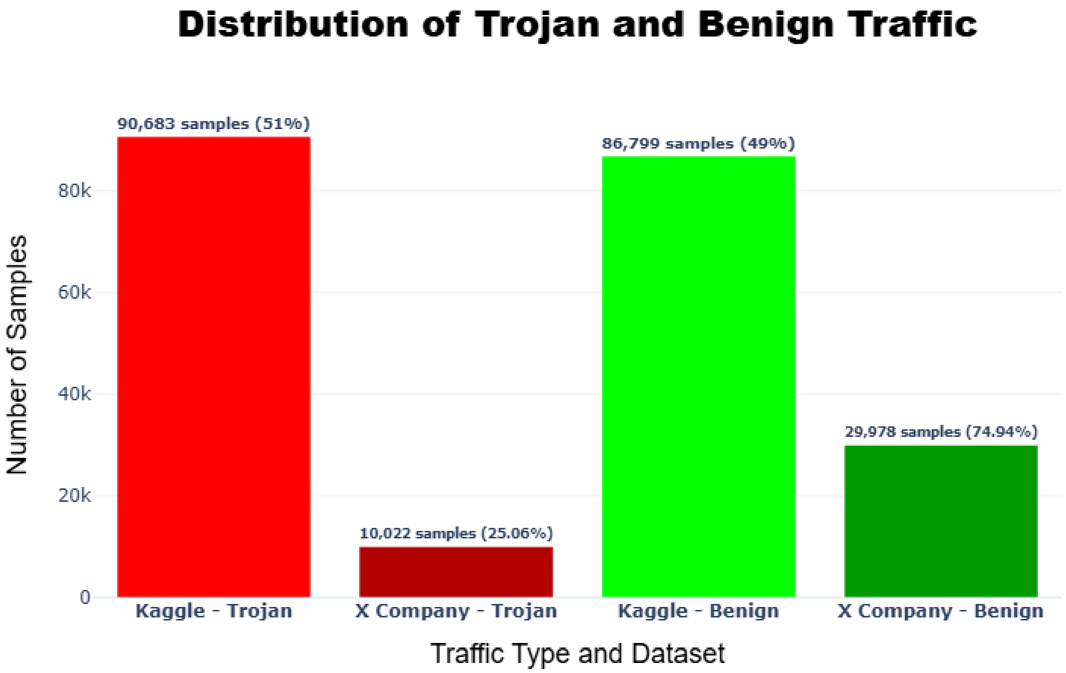

Fig. 2: Distribution of Trojan and Benign Traffic

3.4 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was performed to understand the structure and distribution of the data:

- **Descriptive Statistics**: Basic statistics (mean, median, standard deviation) were calculated for numerical features to understand data centrality and spread.
- **Class Distribution**: The distribution of Trojan vs. benign traffic was analyzed to identify any class imbalance that could impact model performance.

3.5 Data Preprocessing

Data preprocessing involves handling missing values, removing outliers, and encrypting sensitive data. For Company X, the columns Source IP (Srcip), Destination IP (Dstip), Session ID (Sessionid), Security Policy ID (Policyid), and Unique Policy ID (Poluuid) were encrypted using AES encryption. In the Kaggle dataset, similar encryption was applied to the Source IP, Destination IP, Source Port, Destination Port, Flow ID, and Timestamp. Feature selection was conducted using Lasso, null values were removed, and columns were renamed for clarity. The dataset was then split into training (80%) and testing (20%) subsets.

- **Missing Values**: We checked for and handled missing values as needed.
- **Data Type Consistency**: We validated that all features had the correct data type for consistency in further analysis.

3.6 Machine Learning Model Implementation

Four machine learning classification algorithms—Logistic Regression, Decision Tree, Random Forest, and XGBoost—were applied to both the Company X and Kaggle datasets to evaluate their effectiveness in detecting encrypted RATs.

3.7 Computational Resource Management

Experiments were conducted on the Kaggle platform using T4 GPU acceleration and high-memory settings, providing the necessary computational resources in a cost-efficient and scalable environment.

3.8 Model Evaluation

Model performance was assessed using classification metrics: accuracy (≥70%), precision, recall, F1-score, and Area Under the ROC Curve (AUC-ROC). Stratified k-fold cross-validation was employed to ensure robustness and mitigate variance. Error analysis was performed using confusion matrices, focusing on minimizing false positives and false negatives.

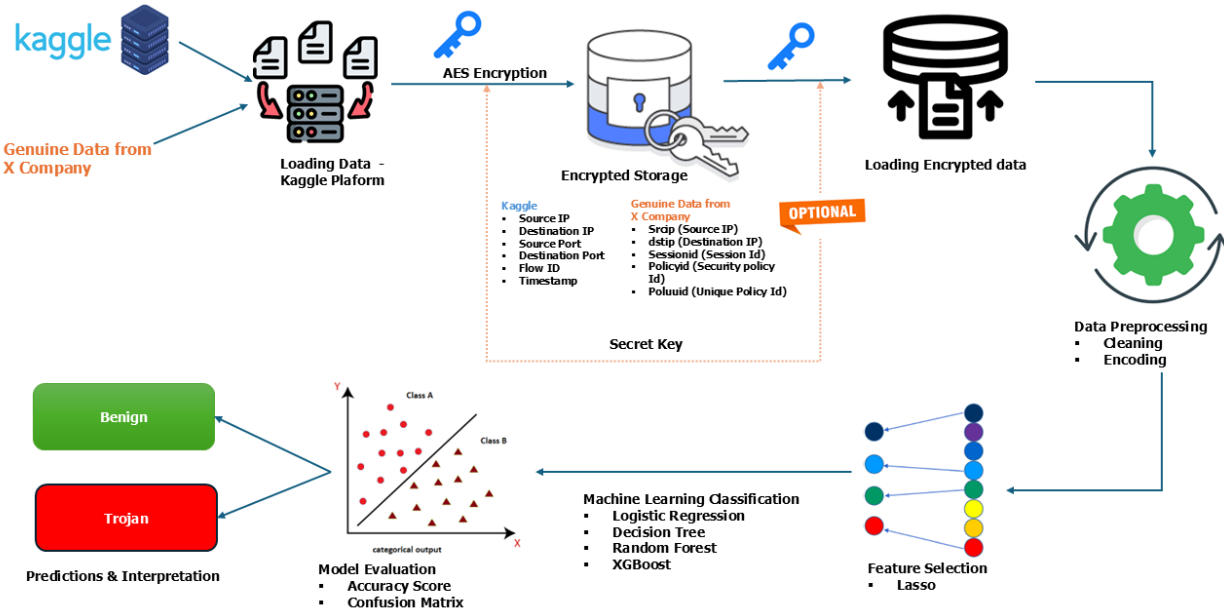

Fig. 3: System architecture and data processing workflow

4. Results and Discussion

4.1 Model Overview

The model evaluation process included several algorithms such as Logistic Regression, Decision Tree, Random Forest, and XGBoost, which were tested for classifying traffic as Benign or Trojan. Among these, the Random Forest model stood out as the top performer across both the Kaggle and X Company datasets. It excelled in various performance metrics, including accuracy, precision, recall, and F1-score, demonstrating its robustness and effectiveness in distinguishing between the two classes.

Table 1: Model performance comparison of Kaggle vs X Company

Metric	Kaggle	X Company	Interpretation
Accuracy	~0.72	~0.75	Slightly better accuracy on X Company data, indicating better overall predictions.
Precision	~0.71	0.88	Precision is much higher on X Company, meaning fewer false positives.
Recall	~0.73	0.76	Slight improvement in recall for X Company, meaning it catches more true positives.
F1-Score	~0.72	0.82	Stronger F1-Score for X Company, indicating a better balance between precision and recall.

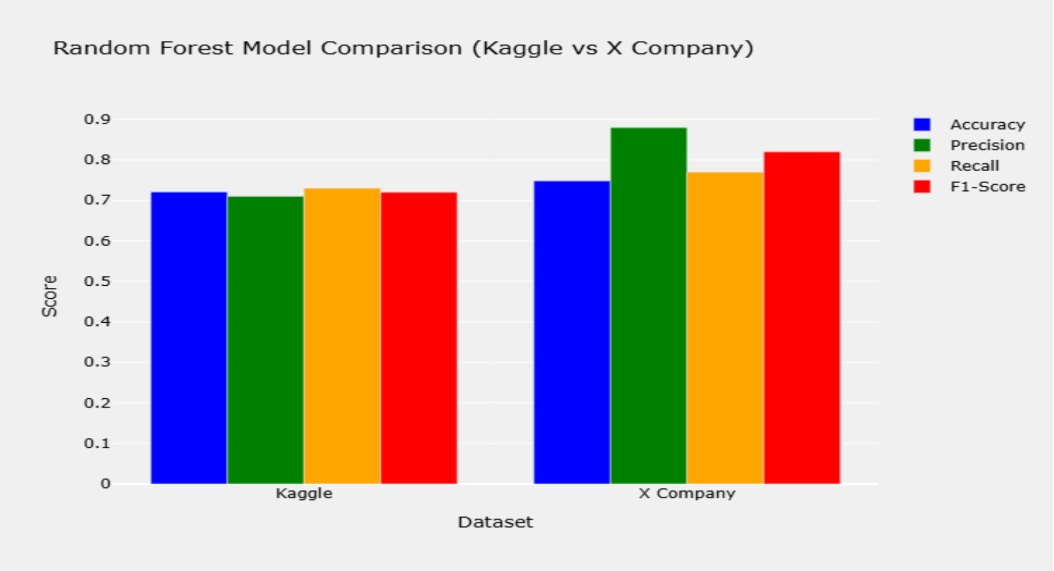

Fig. 4: Random Forest Model Comparison: Kaggle vs X Company

4.2 Confusion Matrix:

The confusion matrix was generated for the best model to provide a detailed understanding of its performance. It showed how well the model distinguished between benign and Trojan traffic.

Table 2: Confusion matrix analysis X company data confusion matrix

X company data confusion Matrix

Predicted	**Benign**	**Trojan**
Actual Benign	4603	1376
Actual Trojan	638	1383

Key Metrics:

- Accuracy: 74.83%
- Precision for Trojan: 50.12%
- Recall for Trojan: 68.44%

Kaggle data confusion matrix

Predicted	**Benign**	**Trojan**
Actual Benign	12791	4646
Actual Trojan	5251	12809

Key Metrics:

- Accuracy: 72.11%
- Precision for Trojan: 73.38%
- Recall for Trojan: 70.93%

4.3 ROC Curve and AUC:

Both datasets demonstrated strong performance, with AUC values indicating good discrimination between Benign and Trojan classes. The X company dataset achieved an AUC of 0.8326, while the Kaggle dataset had an AUC of 0.7996. This suggests that the Random Forest model generalizes well across both synthetic and real-world data, effectively distinguishing between the two classes.

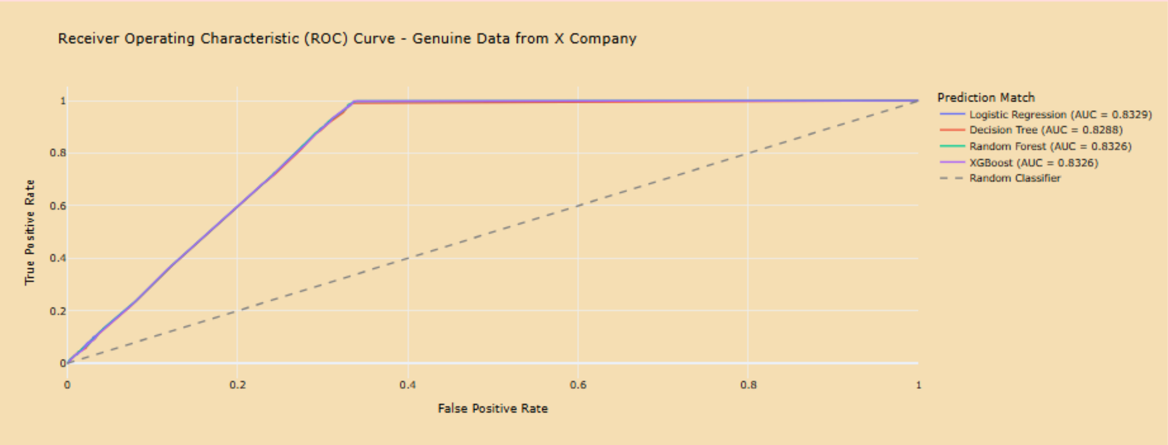

Fig. 5: Diagram visualization of all metrics of X Company data

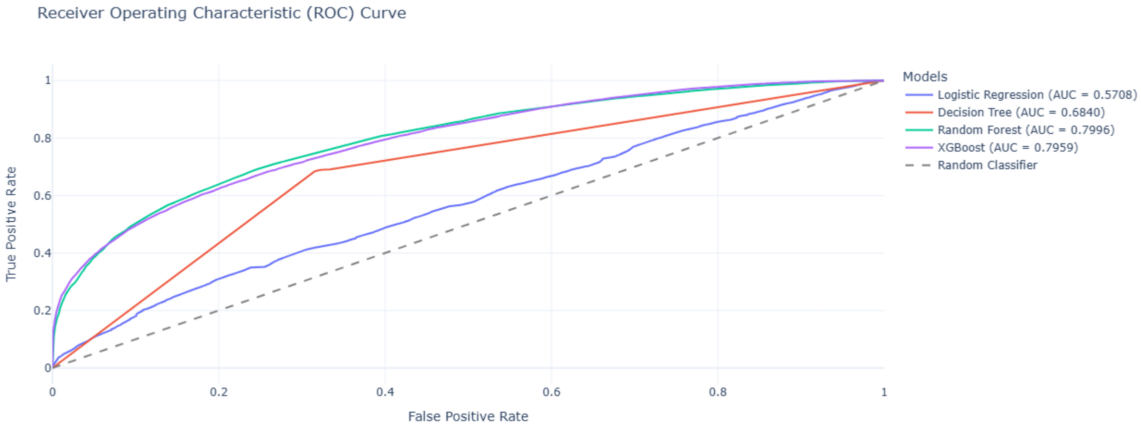

Fig. 6: Diagram visualization of all metrics of Kaggle data

4.4 Model Prediction Analysis:

In both the Kaggle and X company datasets, the models made 8 correct predictions, 1 false negative (Trojan → Benign), and 1 false positive (Benign → Trojan). Despite performing similarly in both datasets, the presence of false negatives, where a Trojan is missed, remains a major concern in cybersecurity, as it poses a significant risk by allowing malicious threats to go undetected.

Table 3: Sample Predictions

Genuine Data from X Company

Actual	Predicted	Interpretation
Trojan	Benign	False Negative
Benign	Benign	Correct
Benign	Benign	Correct
Trojan	Trojan	Correct
Benign	Benign	Correct
Benign	Benign	Correct
Benign	Trojan	False Positive
Benign	Benign	Correct
Benign	Benign	Correct
Benign	Benign	Correct

Sample Predictions – Kaggle Data

Actual	Predicted	Interpretation
Benign	Benign	Correct
Trojan	Trojan	Correct
Trojan	Trojan	Correct
Benign	Benign	Correct
Trojan	Benign	False Negative
Trojan	Trojan	Correct
Trojan	Trojan	Correct
Trojan	Trojan	Correct
Benign	Benign	Correct
Benign	Trojan	False Positive

5. Limitations

This study presents valuable insights, but several limitations need to be addressed. Model generalization is limited due to the lack of variability in the training data, and the absence of detailed vulnerability information restricts adaptability to real-world threats. Computational constraints prevented exploring more advanced architectures like Convolutional Neural Networks (CNNs) or transformer-based models. The study also lacks adversarial robustness, making it vulnerable to Trojan attacks, backdoor triggers, and data poisoning. Additionally, limited access to real-world datasets due to confidentiality and ethical concerns hindered comprehensive validation, affecting the study's transparency and applicability.

6. Conclusion

The Random Forest model has shown strong performance in detecting Trojan and Benign cases across multiple datasets. Its balance of precision and recall ensures minimal false positives while reliably identifying true threats. The model's robustness and resistance to overfitting make it well-suited for real-world deployment in cybersecurity applications. Given its accuracy, interpretability, and overall effectiveness, Random Forest is the most reliable model for this task and is highly recommended for broader implementation in the field.

8. Recommendations

We recommend adopting the Random Forest model for deployment due to its robust performance across both Kaggle and X Company datasets. Fine-tuning the model's threshold to reduce false negatives is essential, especially for early Trojan attack detection. Incorporating domain-specific features, such as network behavior, session length, and unusual port activity, will help strengthen the model's detection capabilities. Exploring ensemble strategies, such as combining Random Forest with XGBoost through a voting mechanism, will further enhance performance. Finally, real-time evaluation using confusion matrices and precision-recall curves will allow for ongoing adjustments to address any class imbalances.

Future research will focus on enhancing the model's capabilities by incorporating advanced deep learning architectures, such as CNNs, to improve detection accuracy. Leveraging Generative AI (LLMs) could help detect subtle Trojan behaviors and backdoor patterns. Real-world testing in collaboration with industry partners will refine the model while maintaining data confidentiality. Adversarial training techniques will be implemented to enhance model robustness against attacks like evasion or data poisoning. Additionally, Explainable AI (XAI) methods will increase transparency and trust in the model's decisions. Exploring cross-domain transfer learning will allow the model to be adapted for use in various sectors, such as healthcare, ensuring scalability and cost-efficiency. Ultimately, the goal is to transition from experimental models to real-time detection systems for critical applications like cybersecurity and fraud detection.

References

1. Aliyu, A., Danjuma, S., Dai, B., Waziri, U., & Ado, A. (2014). An Integrated Framework for Detecting and Prevention of Trojan Horse (BINGHE) in a Client-Server Network. ResearchGateate, 3(1), 8709–8716.
2. Wijayarathne, S. (2022). Trojan Horse Malware - Case Study. Sri Lanka Institute of Information Technology (SLIIT), Malabe, Sri Lanka, July.
3. Dimou, P., Fajfer, J., Müller, N., Papadogiannaki, E., & Střasák, F. (2019). Encrypted Traffic Analysis About Enisa. Enisa, November, 1–55.

4. McDonald, G., Papadopoulos, P., Pitropakis, N., Ahmad, J., & Buchanan, W. J. (2022). Ransomware: Analysing the Impact on Windows Active Directory Domain Services. Sensors, 22(3). https://doi.org/10.3390/s22030953
5. Mokhtar, B. I., Jurcut, A. D., ElSayed, M. S., & Azer, M. A. (2022). Active Directory Attacks—Steps, Types, and Signatures. Electronics (Switzerland), 11(16), 1–23. https://doi.org/10.3390/electronics11162629
6. Eddy, M. (2014). RATs Come to Android: It's Scary, But You're (Probably) Safe | PCMag.
7. Kwon, H. Y., Kim, T., & Lee, M. K. (2022). Advanced Intrusion Detection Combining Signature-Based and Behavior-Based Detection Methods. Electronics (Switzerland), 11(6), 1–19. https://doi.org/10.3390/electronics11060867
8. Vasani, V., Bairwa, A. K., Joshi, S., Pljonkin, A., Kaur, M., & Amoon, M. (2023). Comprehensive Analysis of Advanced Techniques and Vital Tools for Detecting Malware Intrusion. Electronics (Switzerland), 12(20), 1–30. https://doi.org/10.3390/electronics12204299
9. Johnson, C. S., Badger, M. L., Waltermire, D. A., Snyder, J., & Skorupka, C. (2016). Guide to Cyber Threat Information Sharing. https://doi.org/10.6028/NIST.SP.800-150
10. Opderbeck, D. W. (2022). Cybersecurity and Data Breach Harms: Theory and Reality. In SSRN Electronic Journal (Vol. 82, Issue 4). https://doi.org/10.2139/ssrn.4187263
11. Ozkan-okay, M., Yilmaz, A. A., & Akin, E. (2023). A Comprehensive Review of Cyber Security Vulnerabilities. MDPI.
12. Mirza, A. U. (2024). Exploring the Frontiers of Artificial Intelligence and Machine Learning Technologies. In Exploring the Frontiers of Artificial Intelligence and Machine Learning Technologies (Issue April). https://doi.org/10.59646/efaimlt/133
13. Mosalam, K. M., & Gao, Y. (2024). Basics of Machine Learning (pp. 31–56). Morgan & Claypool Publishers. https://doi.org/10.1007/978-3-031-52407-3_3
14. Murphy, K. P. (2012). Machine Learning A Probabilistic Perspective. In The MIT Press Cambridge, Massachusetts. The MIT Press Cambridge. https://doi.org/10.1007/978-94-011-3532-0_2
15. Nigmatullin, R., Ivchenko, A., & Dorokhin, S. (2020). Differentiation of sliding rescaled ranges: New approach to encrypted and VPN traffic detection. 2020 International Conference Engineering and Telecommunication, En and T 2020. https://doi.org/10.1109/EnT50437.2020.9431285
16. Altukruni, H., Maynard, S. B., Alshaikh, M., & Ahmad, A. (2021). Exploring knowledge leakage risk in knowledge-intensive organisations: behavioural aspects and key controls. arXiv preprint arXiv:2104.07140.
17. Tanwar, R. (2025). Cyber Security Challenges. International Journal For Science Technology And Engineering, 13(1), 564–566. https://doi.org/10.22214/ijraset.2025.66263
18. Johansen, M., Mass Soldal Lund, & Geir Olav Dyrkolbotn. (2022). Development of a customized remote access trojan (RAT) for educational purposes within the field of malware analysis. June, 1–64.
19. Peter Szor. (2005). COMPUTER VIRUS RESEARCH AND DEFENSE (K. Gettman (ed.)). Pearson Education, Inc.

20. Gardiner, J., Cova, M., & Nagaraja, S. (2014). Command & Control: Understanding, Denying and Detecting. ArXiv.Org, cs.CR(February), 1136.

21. Zeltser, L. (n.d.). When Bots Use Social Media for Command and Control.

22. Cheruvu, S., Smith, N., Kumar, A., & Wheeler, D. M. (2019). Demystifying Internet of Things Security: Successful IoT Device/Edge and Platform Security Deployment. Apress. https://doi.org/10.1007/978-1-4842-2896-8

23. Whitman, M. E., & Mattord, H. J. (2011). Principles of Information Security Fourth Edition. Learning, 269, 289.