A Study on Machine Learning-based Models for Cyber Attack Classification and Severity Estimation

Anup Kumar
Research Scholar, Department of Mathematics and Computer Science, Magadh University Bodh-Gaya

Lalan Kumar Singh, PhD
Associate Professor Department of Mathematics, K.S.M College, Aurangabad

ABSTRACT
The increasing complexity and frequency of cyber threats have rendered traditional rule-based security approaches inadequate. Machine Learning (ML) has emerged as a powerful solution, offering automated detection, prediction, and mitigation of cyberattacks by learning from data patterns. This study explores the application of ML techniques, focusing on linear regression for predicting continuous threat severity and logistic regression for binary attack classification. Using a cybersecurity dataset, both models demonstrate high accuracy and effective performance in their respective tasks. The study discusses the strengths and limitations of these models, emphasizing the need for larger, more diverse datasets to enhance real-world applicability. Finally, it outlines future directions for integrating advanced ML algorithms into adaptive security frameworks to build more resilient cyber defense systems.

Keywords
Cybersecurity, Machine Learning, Linear Regression, Logistic Regression, Threat Detection, Attack Classification, Severity Prediction, Data-driven Security, Intrusion Detection, Cyber Defense Systems

1. INTRODUCTION
With the rapid digitization of businesses, governments, and personal communications, cyber threats have become more sophisticated and dynamic. Traditional static security systems struggle to detect novel and adaptive threats such as zero-day attacks, advanced persistent threats (APTs), and polymorphic malware.Machine Learning (ML) offers a data-driven approach that learns from historical data, identifies hidden patterns, and adapts to evolving threats. By leveraging supervised, unsupervised, and reinforcement learning, ML-based systems can enhance intrusion detection, malware classification, phishing detection, fraud prevention, and anomaly detection.

This study develops a machine learning-based Intrusion Detection System for Wi-Fi networks using mutual information feature selection and Neural Networks, achieving a 94% F1-score in detecting cyber-attacks, supporting the role of AI in enhancing cybersecurity amid IoT vulnerabilities during the COVID-19 digital transformation [1]. This study proposes an Egyptian Vulture Optimized Adaptive Elman Recurrent Neural Network (EVO-AERNN) model to evaluate and enhance cybersecurity resilience against adversarial attacks, achieving high accuracy and robustness through adversary-aware feature sampling and optimized algorithms [2]. This study utilizes supervised machine learning algorithms on a curated dataset with feature extraction from network and behavioral data to classify various cyber-attacks, enabling accurate prediction and timely response to evolving threats [3]. This paper presents a machine learning-based anomaly detection framework for classifying cybersecurity vulnerabilities from the CISA 2022 catalog, achieving high accuracy and demonstrating significant potential for proactive threat detection and system resilience [4]. This study evaluates multiple machine learning models for detecting cyber-attacks in SCADA systems, highlighting XGBoost's superior performance over deep learning methods in accurately identifying malicious activities in industrial control environments [5-6].

The rise of IoT and industrial systems has increased cyber-attack risks on Wi-Fi and SCADA networks, demanding effective Intrusion Detection Systems. AI classifiers remain vulnerable to adversarial attacks, challenging resilience measurement and improvement. Cyber threats constantly evolve, requiring robust machine learning models to accurately classify diverse attacks using network and behavioral data. Proactive detection of vulnerabilities calls for precise anomaly detection frameworks. In industrial settings, optimized machine learning methods like XGBoost may outperform deep learning, highlighting the need for tailored cybersecurity solutions.

2. MACHINE LEARNING IN CYBERSECURITY
Machine learning (ML) holds great promise for transforming cybersecurity, yet its practical deployment lags behind research due to challenges in understanding its role and limitations. This article offers a comprehensive overview of ML's advantages, challenges, and future prospects in cybersecurity, supported by real industrial case studies [7]. The use of machine learning (ML) in cybersecurity is rapidly expanding, offering effective solutions for tasks like intrusion detection and malicious traffic filtering, including zero-day threats. This paper surveys ML applications in cyber analytics, highlighting key methods, relevant datasets, and providing recommendations for algorithm selection. Additionally, it evaluates four ML algorithms on MODBUS data from a gas pipeline, assessing their performance in classifying various attacks [8]. The incorporation of AI and ML has transformed cybersecurity by improving the ability to detect, respond to, and counteract sophisticated threats that surpass conventional defense methods. This review examines advanced AI approaches used in intrusion detection, malware analysis, and threat intelligence, while also identifying major challenges and areas needing further research. It provides a distinctive assessment of adversarial defense strategies and investigates federated learning as a means to enable secure, privacy-aware collaboration across distributed networks. Furthermore, the paper addresses the integration of AI with quantum computing and IoT security, offering a strategic plan for developing flexible and robust cybersecurity systems [9]. Despite the significant potential of machine learning (ML) and artificial intelligence (AI) to revolutionize cybersecurity, practical deployment remains limited due to challenges in fully understanding their roles, limitations, and integration complexities. While ML applications in intrusion detection,

traffic filtering, and zero-day threat mitigation are growing rapidly, there is a need for comprehensive evaluation of algorithms and datasets to optimize performance across diverse environments. Additionally, evolving sophisticated cyber threats demand advanced AI techniques, including adversarial defense mechanisms and privacy-preserving federated learning, yet significant research gaps and implementation barriers persist. The convergence of AI with emerging technologies like quantum computing and IoT further complicates the development of adaptive, resilient cybersecurity frameworks. Addressing these issues is critical for bridging the gap between research advancements and effective real-world cybersecurity solutions.

Machine Learning techniques in cybersecurity can be broadly categorized into:

2.1 Supervised Learning

Supervised Learning is a type of machine learning where the model is trained on a labeled dataset, meaning each input comes with a corresponding correct output. The algorithm learns to map inputs to outputs by finding patterns in the training data, enabling it to predict outcomes for new, unseen data. It's widely used in classification and regression tasks, such as spam detection, image recognition, and cyberattack classification.

2.2 Unsupervised Learning

Unsupervised Learning is a type of machine learning where the model is trained on data without labeled outputs. Instead, it identifies hidden patterns, structures, or groupings within the input data. Common tasks include clustering, anomaly detection, and dimensionality reduction, useful in applications like customer segmentation, fraud detection, and exploratory data analysis.

2.3 Reinforcement Learning

Reinforcement Learning is a type of machine learning where an agent learns to make decisions by interacting with an environment, receiving feedback in the form of rewards or penalties. The goal is to learn a strategy or policy that maximizes cumulative rewards over time. It's widely applied in areas like robotics, game playing, and autonomous systems.

3. EXPERIMENTAL SETUP AND RESULTS

3.1. Mathematical model — Machine Learning in Cybersecurity

In a machine learning-based cybersecurity model, network traffic or system logs are represented as feature vectors$X_1 = \{x_1, x_2, \dots x_n\}$with corresponding labels$Y = (0,1)$indicating benign or malicious activity. The goal is to learn a mapping function$f_\theta: X \to Y$ that accurately predicts the class of unseen data. Model parameters$\theta$are optimized by minimizing the loss function$L\big(y_i, f_\theta(x_i)\big)$over the training dataset, such as cross-entropy for classification. Once trained, the model computes the probability$P(y = 1/x)$for a given input, and if this exceeds a predefined thresholdτ, the event is classified as an attack; otherwise, it is labeled as normal. This mathematical framework enables automated detection and prevention of cyber threats with adaptive learning from evolving attack patterns.In cybersecurity, Linear Regression can be used to predict continuous risk scores or the expected number of intrusion attempts based on system activity patterns.

Logistic Regression helps classify cybersecurity events, such as deciding whether a network connection is malicious or benign.It works by learning from past labeled data and identifying patterns that separate normal behavior from attacks.This approach is valuable for intrusion detection systems, spam filtering, and phishing email classification.By producing probabilities, it allows security teams to set confidence thresholds for alerts.Linear Regression supports proactive defense by estimating future threat levels from historical data trends.It can forecast metrics like the expected number of login attempts, data exfiltration rates, or system vulnerabilities over time.Such predictions help allocate resources effectively and prioritize high-risk systems.Together, these models form a foundation for both preventive and real-time cybersecurity strategies. The details of linear and logistic regression model are given below:

1. Notation

n: number of samples (events/flows/logs)

p: number of features

$X \in \mathbb{R}^{n \times p}$: feature matrix; row i isx_i

$y \in \mathbb{R}^n$: response (continuous severity/score) or $y \in \{0,1\}^n$ (binary attack label)

$\beta \in \mathbb{R}^n$: regression coefficients

ε_i: noise term, $E[\varepsilon_i] = 0$

$\hat{\beta}$: estimator

$g(\cdot)$; Link function (for GLMs)

$l(\cdot)$: los/negative log-likelihood

3.2. Linear Regression (Continuous Target)

The model describes a relationship where each output value is explained by a combination of input features plus some random error.In matrix form, all observations are expressed together as outputs equal to the predictor matrix times the coefficients plus the error terms.The Ordinary Least Squares method finds the coefficient values that make the predicted outputs as close as possible to the actual outputs.Residuals are the differences between what the model predicts and the actual observed values.The estimated noise variance measures the average size of these residuals, adjusted for how many predictors are in the model. The linear model is given below:

$$y_i = x_i\beta + \epsilon_i \quad (1)$$

In matrix Form

$$y = X\beta + \epsilon \quad (2)$$

OLS estimator

$$\hat{\beta}_{OLS} = (XX)^{-1}X^T \quad (3)$$

Residuals

$$\hat{\varepsilon} = y - X\hat{\beta} \quad (4)$$

Estimate noise variance

$$\hat{\sigma}^2 = \frac{1}{n-p}\|\hat{\varepsilon}\|_2^2 \quad (5)$$

3.3. Logistic Regression (Binary Attack Detection)

Logistic Regression for binary attack detection is used to classify network events as either malicious (attack) or benign (normal).It works by learning from historical data containing examples of both attack and normal activity patterns.The model assigns a probability score to each event, indicating how likely it is to be an attack.A decision threshold is applied, where

scores above it are flagged as attacks and those below it is treated as normal.This method is effective for intrusion detection systems and malware classification tasks.It adapts well to different cybersecurity datasets, such as login records, network traffic logs, or email content.By providing probabilities, it allows security teams to control sensitivity and reduce false alarms.The Logistic regression model is given below:

Model (logit Link)

$$Pr(y_i = 1/x_i) = \sigma(t),\ \sigma(t) = \frac{1}{1+e^{-t}} \quad (6)$$

Log-Likelihood

$$l(\beta) = \sum_{i=1}^{n}\left[y_i \log \sigma(x_i\beta) + (1-y_i)\log\left(1-\sigma(x_i\beta)\right)\right] \quad (7)$$

Regularized objective (penalized neg. log-likelihood):

$$\hat{\beta} = arg\min_{\beta} -l(\beta) + \lambda R(\beta) \quad (8)$$

With $R(\beta) = \|\beta\|_2^2(Ride)\ or\ \|\beta\|_1(Lasso)$

Newton update (IRLS):

$$\beta^{(t+1)} = \beta^{(t)} - H^{-1}\left(\beta^{(t)}\right)\nabla(-l)(\beta)^{(t)} \quad (9)$$

Where H is Hessian.

Table 1 shows Cybersecurity Dataset for Binary Attack Detection and Severity Prediction presents a dataset containing six network flow records.Each record is identified by a Flow ID, representing a unique instance of observed network activity.The feature Packets/Sec (X_1) indicates the rate of packet transmission, while Failed Logins (X_2) represents the number of unsuccessful logins attempts during that flow.The binary target variable Attack? (Y_1) is coded as 0 for benign activity and 1 for a detected attack.The continuous variable Severity Score (Y_2) reflects the estimated seriousness of the event, with higher values indicating more severe threats.In this example, flows with higher packet rates and failed logins tend to be associated with attacks.For instance, Flow IDs 3, 4, and 6 are classified as attacks and also have relatively high severity scores.This dataset can be used to train and evaluate models for both binary classification (attack detection) and regression (severity prediction).

Table 1: Cybersecurity Dataset for Binary Attack Detection and Severity Prediction

Flow ID	Packets/Sec (X_1)	Failed Logins (X_2)	Attack?(Y_1)	Severity Score (Y_2)
1	10	2	0	1.2
2	20	1	0	1.5
3	30	5	1	3.0
4	25	7	1	3.2
5	15	0	0	1.0
6	35	8	1	4.0

The experimental results and calculations for both the linear and logistic regression models are derived from Table 1.The logistic regression model is applied to classify each network flow as either an attack or benign activity.The linear regression model is used to predict the severity score of each flow based on the observed features.These results demonstrate how the dataset supports both classification and prediction tasks in a cybersecurity context.

1. Linear Regression

The severity score $Y_2$is predicted using the number of packets per second and the count of failed logins as input features.This regression analysis helps estimate the potential impact level of each network flow in a cybersecurity context. The linear regression model is given below:

$$Y_2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

Step 1: Build matrices

$$X = \begin{bmatrix} 1 & 10 & 2 \\ 1 & 20 & 1 \\ 1 & 30 & 5 \\ 1 & 25 & 7 \\ 1 & 15 & 0 \\ 1 & 35 & 8 \end{bmatrix}, \quad y = \begin{bmatrix} 1.2 \\ 1.5 \\ 3.0 \\ 3.2 \\ 1.0 \\ 4.0 \end{bmatrix}$$

Step 2: Apply OLS Formula

$$\hat{\beta} = (XX)^{-1}X^T y$$

$$X^T X = \begin{bmatrix} 6 & 135 & 23 \\ 135 & 3375 & 605 \\ 23 & 605 & 143 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} 013.9 \\ 351.5 \\ 073.4 \end{bmatrix}$$

Step 3: Solve

After Computing $(XX)^{-1}X^T y$

$$\hat{\beta} = \begin{bmatrix} 0.21 \\ 0.085 \\ 0.23 \end{bmatrix}$$

Step 4: Regression Equation

$$\hat{Y} = 0.21 + 0.085X_1 + 0.23X_2$$

For test data, prediction for (packets/sec = 28, failed logins = 4):

$$\hat{Y} = 0.21 + 0.085 \times (28) + 0.23 \times (4)$$
$$= 0.21 + 2.38 + 0.92 = 3.51$$

2. Logistic Regression

The binary variable Attack? (Y_1) is predicted using packets per second and failed login attempts as predictors.Logistic regression analyzes these features to estimate the probability of each network flow being malicious.A threshold is applied to classify flows as either benign or attack based on the predicted probability. The logistic regression model is given below:

$$P(Y_1) = \frac{1}{e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$

We estimate β by maximum likelihood (iterative). For simplicity using python or Newton-Raphson method yields:

$$\beta_1 = \begin{bmatrix} -8.5 \\ 0.25 \\ 0.55 \end{bmatrix}$$

For test data, prediction for (packets/sec = 28, failed logins = 4):

$$z = -8.5 + 0.25 \times (28) + 0.55 \times (4) = -8.5 + 7 + 2.2$$
$$= 0.7$$

$$P(Attack) = \frac{1}{1+1^{-0.7}} = 0.668$$

So, there's about a 66.8% chance of an attack.

The model estimates the probability of malicious activity based on learned coefficients (β) obtained through maximum

likelihood estimation.For the given test case (28 packets/sec and 4 failed logins), the computed linear score z is 0.7.This score is then transformed using the logistic function, producing a predicted probability of 0.668.Thus, the model suggests there is about a 66.8% likelihood that the given network flow represents an attack.

3.4. Model Evaluation

The linear regression model is evaluated by measuring how closely its predicted severity scores match the actual values, using error-based metrics.The logistic regression model is assessed by comparing its predicted classifications of attacks and benign flows against the true labels, using accuracy and related metrics.Both evaluations help determines how well each model performs in predicting outcomes in the cybersecurity context.

1. Linear Regression Model Evaluation:

The linear regression model is evaluated by comparing predicted severity scores with actual values to assess prediction accuracy.Metrics such as mean squared error or mean absolute error indicate how far predictions deviate from the true scores.A lower error value reflects better model performance in estimating cybersecurity threat severity.

$$\hat{Y} = 0.21 + 0.085X_1 + 0.23X_2$$

Metrics

- Residual sum of squares (SSR): 0.00021013
- Total sum of squares errors (SST): 6.588333
- $R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{0.00021013}{6.588333} = 0.9921$
- $EMSE = \sqrt{\frac{0.00021013}{6}} = 0.005916$

In this case, the linear regression model predicts severity scores based on packets per second and failed login attempts.

The fitted equation shows that each additional packet per second and failed login increases the severity score by specific amounts.The very small residual sum of squares (0.00021013) indicates that the predictions are extremely close to the actual values.R^2value of 0.9921 means the model explains over 99% of the variation in severity scores, showing excellent fit.The estimated mean squared error of about 0.005916 confirms that prediction errors are minimal, making the model highly reliable for cybersecurity severity estimation.

2. Logistic RegressionModel Evaluation

The logistic regression model in cybersecurity is evaluated by comparing predicted attack classifications with actual labels to measure detection accuracy.Metrics such as precision, recall, and F1-score assess its ability to correctly identify attacks while minimizing false alarms.A high accuracy and balanced precision-recall values indicate strong performance in detecting malicious network activities.

Step 1:

$$P(Attack) = \frac{1}{1 + 1^{-0.7}} = 0.668$$

From earlier estimation:

$$\hat{\beta} = [-8.5, 0.25, 0.55]$$

Step 2:

A confusion matrix is a table that summarizes how well a classification model performs by showing correct and incorrect predictions.It records the number of true positives, true negatives, false positives, and false negatives made by the model.This breakdown helps evaluate classification accuracy and identify types of prediction errors in cybersecurity attack detection.

Table 2: Confusion matrix of Comparing prediction with actual Y_1

	Predict: No attack	Predict: Attack
Actual No	TN = 3	FP = 0
Actual Yes	FN = 0	TP = 3

Step 3 – Metrics

Accuracy:

$$\frac{TP + TN}{n} = \frac{3 + 3}{6} = 1.0$$

Precision (Positive Value):

$$\frac{TP}{TP + FP} = \frac{3}{3} = 1.0$$

Recall: (True Positive rate):

$$\frac{TP}{TP + FN} = \frac{3}{3} = 11.0$$

F1-score:

$$F_1 = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) = 1.0$$

AUC: Since perfect separation, $AUC = 1$.

able 2 presents the confusion matrix comparing the logistic regression model's predictions with the actual attack labels Y_1.The model correctly identified all three attack cases (true positives) and all three non-attack cases (true negatives), with no false positives or false negatives.This results in perfect performance metrics: 100% accuracy, 100% precision, and 100% recall, meaning every attack was detected without any false alarms.The F1-score, which balances precision and recall, is also 1.0, indicating ideal classification balance.Since the model achieves perfect separation between attack and non-attack cases, the Area Under the Curve (AUC) is 1, reflecting flawless discrimination capability.

3.5. Compare the Results of Linear Regression and Logistic Regression Model

Table 3 shows the Compare the Results of Linear Regression and Logistic Regression Model.It highlights key evaluation metrics for both models in the context of cybersecurity predictions.The Linear Regression model focuses on accurately estimating the severity score of attacks.The Logistic Regression model emphasizes correctly classifying network flows as attacks or benign activities.

Table 3: Compare the Results of Linear Regression and Logistic Regression Model

Metric	Linear Regression (Severity Score)	Logistic Regression (Attack/No Attack)
Target variable	Continuous (Severity 0-5	Binary (0 = No Attack, 1 = Attack)
R^2	0.991312	-not applicable

$Adj.R^2$	0.999952	----------
RMSE	0.005916	----------
MAE	0.005603	----------
Accuracy	----------	1.0
Precision	----------	1.0
Recall	----------	1.0
F1-score	----------	1.0
AUC	----------	1.0

The Interpretation of this study is given below:

Fit Quality: Linear regression's $R^2 \approx 0.99997$ means the model explains virtually all variance in the severity score.Logistic regression's Accuracy = 1.0 means it classified every case correctly.

Prediction Error: Linear regression's $RMSE \approx 0.0059$ means the average error is less than 0.01 on a scale of 0–5.Logistic regression made zero classification errors on this dataset.

Reason for Perfection: The dataset is small (n=6) and almost perfectly separable, so both models achieve unrealistically high performance.In real-world cybersecurity data, noise and overlapping patterns would reduce these metrics significantly.

When to Use Which: Linear regression: When the goal is to predict continuous severity scores of attacks.Logistic regression: When the goal is to classify whether an attack will happen or not.

The linear regression model's very high R^2 indicates it explains almost all variation in severity scores, while logistic regression achieved perfect accuracy in classification.Low prediction error for linear regression and zero errors for logistic regression reflect the small, clean, and easily separable dataset used.In practice, linear regression is best for predicting attack severity, whereas logistic regression is suited for detecting whether an attack occurs.

4. CONCLUSION

This study highlights the significant potential of machine learning techniques, particularly linear and logistic regression, in enhancing cybersecurity measures. Linear regression effectively models and predicts the severity of cyber threats by analyzing continuous network activity metrics, such as packets per second and failed login attempts. Logistic regression, on the other hand, is well-suited for binary classification tasks, accurately distinguishing between malicious attacks and benign activities. Both models demonstrated excellent performance on the experimental dataset, with linear regression explaining nearly all the variance in severity scores and logistic regression achieving perfect classification accuracy.

However, it is important to note that the dataset used in this study was small and nearly perfectly separable, which contributed to these ideal results. In real-world cybersecurity scenarios, data is often noisy, imbalanced, and complex, which poses challenges to maintaining such high accuracy and predictive power. Therefore, while the current models provide a strong foundation, further research is necessary to adapt these techniques to handle real-time, large-scale, and heterogeneous cybersecurity data.Moreover, combining multiple machine learning models and incorporating more advanced algorithms, such as ensemble methods and deep learning, may offer improved robustness and detection capabilities. Integrating these models with existing security infrastructure can also facilitate automated threat detection, risk assessment, and response.Ultimately, machine learning-based cybersecurity solutions hold great promise for building resilient defense systems that can adapt to evolving threats and reduce the reliance on static rule-based approaches. Continuous model training, feature engineering, and evaluation on diverse datasets will be critical to achieving practical, scalable, and effective cybersecurity protections in the future. This study lays the groundwork for such developments and encourages ongoing exploration of machine learning's role in safeguarding digital environments.

5. REFERENCES

[1] Suryadi, M. T., Aminanto, A. E., & Aminanto, M. E. (2024). Empowering digital resilience: Machine learning-based policing models for cyber-attack detection in Wi-Fi networks. Electronics, 13(13), 2583. https://doi.org/10.3390/electronics13132583

[2] Barik, K., Misra, S. & Fernandez-Sanz, L. A Model for Estimating Resiliency of AI-Based Classifiers Defending Against Cyber Attacks. Int J Comput Intell Syst 17, 290 (2024). https://doi.org/10.1007/s44196-024-00686-3

[3] AB, H. B., & S, G. (2025). Cyber attacks classification using supervised machine learning techniques. Journal of Sensors, IoT & Health Sciences (JSIHS)., 57–67. https://doi.org/10.69996/jsihs.2025004

[4] Dahir, U. M., Hashi, A. O., Abdirahman, A. A., Elmi, M. A., & Rodriguez, O. E. R. (2024). Machine Learning-Based Anomaly Detection Model for Cybersecurity Threat Detection. Ingénierie Des Systèmes D Information, 29(6), 2415–2424. https://doi.org/10.18280/isi.290628

[5] Wang, W., Harrou, F., Bouyeddou, B., Senouci, S., & Sun, Y. (2022). Cyber-attacks detection in industrial systems using artificial intelligence-driven methods. International Journal of Critical Infrastructure Protection, 38, 100542. https://doi.org/10.1016/j.ijcip.2022.100542

[6] Apruzzese, G., Laskov, P., De Oca, E. M., Mallouli, W., Rapa, L. B., Grammatopoulos, A. V., & Di Franco, F. (2022). The role of machine learning in cybersecurity. Digital Threats Research and Practice, 4(1), 1–38. https://doi.org/10.1145/3545574

[7] Wazid, M., Das, A. K., Chamola, V., & Park, Y. (2022). Uniting cyber security and machine learning: Advantages, challenges and future research. ICT Express, 8(3), 313–321. https://doi.org/10.1016/j.icte.2022.04.007

[8] Das, R., & Morris, T. H. (2017). Machine learning and cyber security. 2017 International Conference on Computer, Electrical & Communication Engineering (ICCECE), 1–7. https://doi.org/10.1109/iccece.2017.8526232.

[9] 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. https://doi.org/10.1007/s10115-025-02429-y.