Babylonian Journal of Machine Learning
Vol.2024, **pp**. 63–68
DOI: https://doi.org/10.58496/BJML/2024/006 ; ISSN: 3006–5429
https://mesopotamian.press/journals/index.php/BJML

Research Article

Artificial Intelligence Predictions in Cyber Security: Analysis and Early Detection of Cyber Attacks

Meaad Ali Khalaf [1,*], Amani Steiti [2]

[1] *Department of Computer Science, AUL University, Beirut, Lebanon*

[2] *Department of Computer Systems And Networks, Faculty of Information Engineering, University Tishreen, Latakia, Syria.*

ARTICLE INFO

Article History
Received 02 Mar 2024
Revised 01 Apr 2024
Accepted 20 Apr 2024
Published 09 May 2024

Keywords
Cyber security
Cyber Attacks
Machine Learning
Artificial Intelligence
Early Detection
Predictive Models

ABSTRACT

The landscape of cyber-attacks has changed due, to the upward push of digitalization and interconnected structures. This necessitates the need for revolutionary techniques to emerge as aware of and mitigate these threats at a degree. This studies delves into the correlation amongst cyber security and artificial intelligence (AI) with a focus on how AI can decorate detection of cyber-attacks via assessment, prediction and different strategies. By harnessing machine mastering, neural networks and records analytics predictive models driven with the useful resource of AI have emerged as an approach to deal with the ever evolving demanding situations posed through cyber threats. The number one goal of this observe is to look at the effectiveness of AI powered prediction fashions, in cyber security. It ambitions to evaluate how nicely those AI based systems carry out as compared to cyber security techniques emphasizing their capability to proactively locate and mitigate cyber threats as a way to minimize their effect. Additionally ability obstacles and ethical issues associated with AI based cyber security answers are also discussed. Also using AI algorithms to Analysis and Early Detection of Cyber Attacks using python programming language. The research's conclusions are extremely important for the field of cyber security since they provide information about how threat mitigation and incident response will develop in the future. This research helps to develop cutting-edge cyber security solutions by addressing the dynamic and constantly-evolving landscape of cyber threats.

1. INTRODUCTION

In the digital world the globally communicated earth offers never-before-seen chances because to the unrelenting advancement of technology but with increased connectivity also comes improved ability to a persistent and ever-evolving threat: cyber-attacks [1],[2]. These assaults have far-reaching effects that go well beyond the world of bits and bytes because they represent a genuine and growing risk to people, companies, and countries [3],[4]. Cybersecurity has become a vital line of defense in response to this dynamic digital battlefield, requiring constant innovation to protect our digital infrastructure [5][25].

The dynamic character of cyber risks demands that cybersecurity protocols via modified to counter new attack paths and highly skilled adversaries. The problem of always changing risks makes it difficult for conventional security systems to keep up with quick advancement [6][26].

A new era of cyber risks have produced near via the advent of the digital age. The growth of state-sponsored operations, the world scope of cyberwarfare and the level of style attacks are the hallmarks of this era [7]. This study aims to scaled light on the importance of early detection and defence methods for protecting cyber infrastructure as well as information integrity. Through an analysis of actual-life scenarios, this paper demonstrates the beneficial applications of AI in cybersecurity by providing particular instances when early detection and prediction models have been essential in averting serious security breaches[27].

The synthetic intelligence industry is leading the cybersecurity shift via offering scalable and flexible solutions due to its ability to self-analyze and adapt to new threats. This has led to the development of protection solutions in useful associations and based on AI[28]. This study aims to proposed combination of AI in cybersecurity, specific in the early detection of

*Corresponding author. Email: Mak105@live.aul.edu.lb

cyberattacks. AI shows predictive ability automates risk detection, and quickly varies to new attack vectors, transforming our understanding and reaction to cyber threats[29]. The object is to understand how AI-driven prediction models improve cybersecurity techniques, compare their effects to traditional strategies and predect any potential issues. This can improve cybersecurity strategies and enhance security.

2. RELATED WORK

This study [8] proposed a decentralized architecture for detecting and mitigating security attacks in an IoT environment utilizing a blockchain mechanism. The architecture is separated into three levels, with fog nodes detect attacks and communicating protocol changes to edge nodes. The sensory nodes monitor traffic, edge nodes manage it, and SDN controllers learn traffic patterns to detect malicious actions. The blockchain method aims to reduce edge node attacks, making the architecture efficient for future IoT platforms.

This study [9] proposed a HoneyNet approach to enhance AIoT security and flexibility via combining threat detection and situational awareness. The system utilizes Docker technology, images, and a deep learning model, and is validated utilizing the Site Where AIoT platform. It uses three honeypot types and four layer (Cloud, Fog, Edge, and Sensing).

This paper [10] presented a ML architecture for honeypot malware detection, utilizing SVM and decision algorithms. The architecture recognizes malware based on its actions and can educate itself based on detection outcomes. The honeypot system includes routers, data analysis, honeypot, and real system components, achieving high accuracy and efficiency through split testing and ten experiments.

This study [11] proposed a resource-optimized, fuzzy approach to detect and prevent spoofing attacks on low-interaction honeypots. Experimental data demonstrates that this method can recognize spoofing attempts. The study emphasizes the need for an intelligent system to optimize resources for spoofing detection. Fuzzy rules based on three fuzzy input variables demonstrate its potential.

3. RESEARCH METHODOLOGY

This section shows research methods utilized to applied ML to Predictions Cyber Security of Early Detection of Cyber Attacks." The aim of this paper is to apply ML models to predect cyberattacks. This section shows a summary of the research design, dataset utiliaed, experimental, data collection and preprocessing, and evaluation as shown in Figure 1.

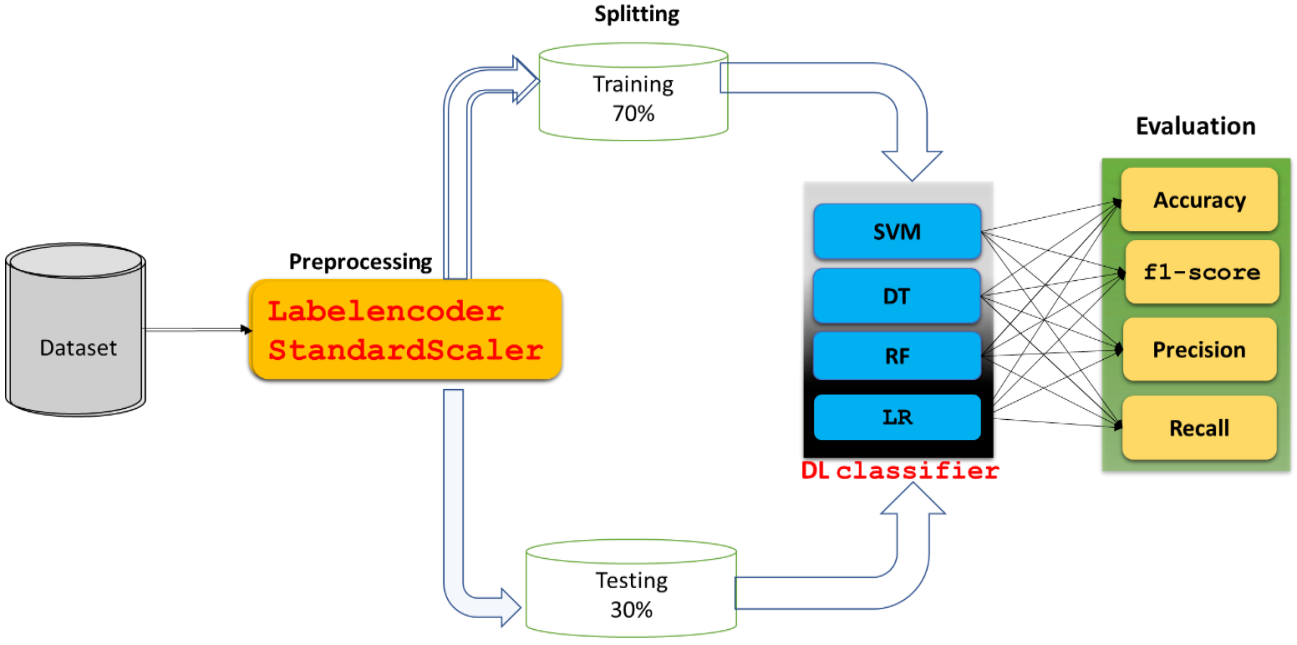

Fig. 1. Research Methodology

3.1 Data Collection

This study utilized UNSW-NB 15 dataset from Kaggle created using the IXIA PerfectStorm tool, combines real and synthetic attack behaviors. The dataset includes nine types of attacks, and uses tools like Argus and Bro-IDS to generate 49 features with class labels.

3.2 Labelencoder

LabelEncoder transforms categorical labels into numerical values, aiding machine learning models. It assigns a unique integer to each category, simplifying data handling. Caution is needed to avoid implying unintended ordinality [12].

3.3 StandardScaler

StandardScaler standardizes features by removing mean and scaling to unit variance, aiding algorithms sensitive to feature scales. It ensures consistency and robustness in machine learning models [13].

3.4 Splitting Dataset

Splitting a dataset into 70% training and 30% testing subsets aids in training and evaluating machine learning models. The training set is used for model training, while the testing set assesses its generalization to unseen data, ensuring reliable performance estimation.

3.5 Machine learning models

In this study, we used four machine learning models for the detection of cyber attacks such as SVM, DT, LR, and RF.

1. Decision Tree

A DT is a popular machine learning algorithm that builds a tree-like structure to make decisions based on input features [14]. It recursively splits the data into subsets based on the values of features, aiming to create homogeneous subsets with respect to the target variable [15]. At each split, the decision tree selects the feature that best separates the data according to certain criteria, such as Gini impurity or information gain. Decision trees are interpretable, versatile, and can handle both classification and regression tasks. However, they are prone to overfitting, especially with deep trees, which can be mitigated using techniques like pruning or ensemble methods [16].

2. Logistic Regression

LR is a statistical method used for binary classification tasks, where the outcome variable has two possible values. The LR algorithm is a classification method that models the outcome variable's probability based on predictor variables. It utilizes the logistic function to evaluate the coefficients of predictor variables, maximizing the likelihood of observed data. This method is large utilized due to its simplicity, interpretability, and efficiency, particularly when the relationship among predictors and outcomes is linear [17].

3. SVM

SVM is a good supervised learning algorithm for classification and regression tasks particularly effective in high-dimensional spaces and when features exceed samples [18]. It finds the optimal hyperplane to separate data points into classes while maximizing border. SVMs can handle linear and nonlinear tasks utilizing kernel functions like linear, polynomial, or RBF. They are strong against overfitting and large utilized in image, text, and bioinformatics fields.

4. Random Forest

RF is a useful ensemble learning algorithm that utilizes multiple DT to classify or predict data points [19]. It eases overfitting via averaging multiple trees, making it suitable for high-dimensional datasets. RF is simple, scalable, and effective across tasks like classification, regression, and outlier detection. It also shows evaluate of feature importance for feature selection and interpretation [20].

3.6 Evalusion

The evaluation of this study using the equations for accuracy, recall, F1 score, and precision.

Precision:- Positive Predictive Value [21]. As shown in equation 1.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (1)$$

Recall:- sensitivity or True Positive Rate [22]. As shown in equation 2.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Accuracy:- the percentage of accurate forecasts the model generates relative to all of the predictions it has made is known as its accuracy rate [23]. As shown in equation 3.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \quad (3)$$

F1 Score:- the harmonic mean of precision and recall [24]. As shown in equation 4.

$$\text{F} - \text{Score} = \frac{2 \times (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \quad (4)$$

Where:-
TN = correctly predicted negatives
TP = correctly predicted positives
FN = incorrectly predicted negatives
FP = incorrectly predicted positives

4. RESULTS AND DISCUSSION

This study proposed four ML algorithms SVM, LR, DT, and RF to evaluated detecting cyber attacks. The evaluation metrics included precision, recall, F1-score, and accuracy.

The results shows exceptional performance across all algorithms as shown in Table 1:

TABLE I. A COMPARATION RESULTS

Algorithms	precision	recall	F1-score	accuracy
SVM	98	98	98	98
Logistic Regression Classifier	98	98	98	98
Decision Tree	100	100	100	100
Random Forest Classifier	100	100	100	100

As shown in Figure 2 results of SVM, LR, DT, and RF and all achieved high precision, recall, F1-score, and accuracy of 98%,98%, 100%, 100% respectively. DT achieved perfect scores of 100% across all metrics, shows its ability to correctly identify cyber threats and minimize false positives and false negatives. The SVM and logistic regression classifiers also shows significant accuracy and consistency, demonstrating their superior performance in detecting cyber threats.

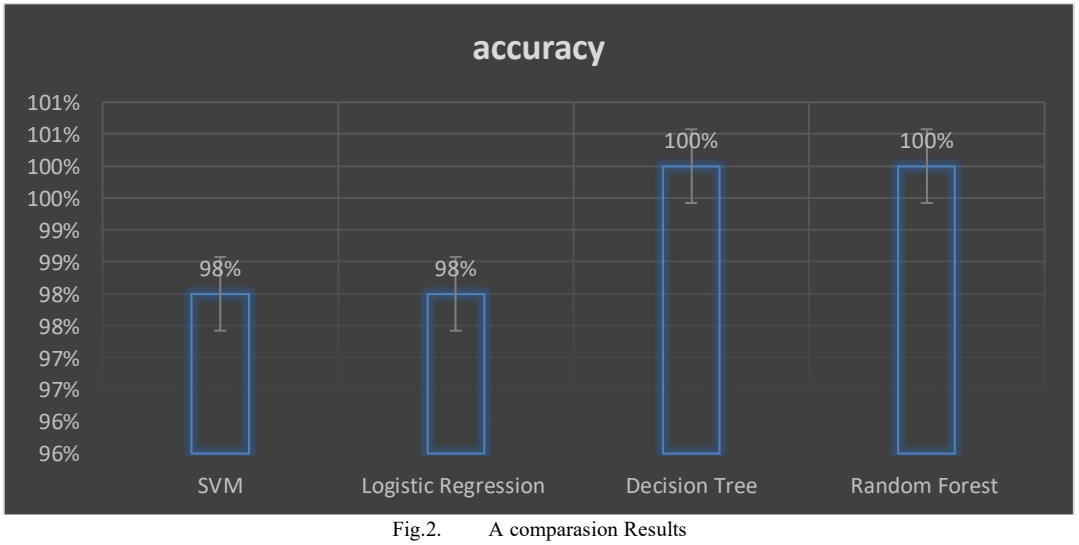

Fig.2. A comparasion Results

ML algorithms show ability in cybersecurity applications of detecting threats further validation and robustness testing are needed for reliability in real-world scenarios.

5. CONCLUSION

The study proposed four ML algorithms SVM, LR, DT and RF to detecting cyber attacks. The results shows a good performance across all algorithms, with the DT and RF classifiers shown perfect performance. The SVM and LR classifiers also shows good accuracy. The findings highlight the potential of ML algorithms in enhancing cybersecurity measures, accurately classifying cyber threats while minimizing false positives and false negatives. Future research should focus on diverse datasets, ensemble techniques and deep learning architectures for more accurate detection.

Conflicts of Interest

The author declares no conflict of interest in relation to the research presented in the paper.

Funding

None.

Acknowledgment

The author acknowledges the assistance and guidance received from the institution in various aspects of this study.

References

[1] Ö. Aslan, S. S. Aktuğ, M. Ozkan-Okay, A. A. Yilmaz, and E. Akin, "A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions," *Electronics*, vol. 12, no. 6, p. 1333, 2023.

[2] M. A. I. Mallick and R. Nath, "Navigating the Cyber security Landscape: A Comprehensive Review of Cyber-Attacks, Emerging Trends, and Recent Developments."

[3] N. S. Mohammed, O. A. Dawood, A. M. Sagheer, and A. A. Nafea, "Secure Smart Contract Based on Blockchain to Prevent the Non-Repudiation Phenomenon," *Baghdad Sci. J.*, vol. 21, no. 1, p. 234, 2024.

[4] Z. Balani and N. I. Mustafa, "Enhancing Cybersecurity Against Emerging Threats in the Future of Cyber Warfare," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 2s, pp. 204–209, 2024.

[5] T. Sobb, B. Turnbull, and N. Moustafa, "Supply chain 4.0: A survey of cyber security challenges, solutions and future directions," *Electronics*, vol. 9, no. 11, p. 1864, 2020.

[6] P. R. J. Trim and Y.-I. Lee, "The global cyber security model: counteracting cyber attacks through a resilient partnership arrangement," *Big Data Cogn. Comput.*, vol. 5, no. 3, p. 32, 2021.

[7] C. Whyte, "Cyber conflict or democracy 'hacked'? How cyber operations enhance information warfare," *J. Cybersecurity*, vol. 6, no. 1, p. tyaa013, 2020.

[8] D. Guha Roy and S. N. Srirama, "A blockchain-based cyber attack detection scheme for decentralized Internet of Things using software-defined network," *Softw. Pract. Exp.*, vol. 51, no. 7, pp. 1540–1556, 2021.

[9] L. Tan, K. Yu, F. Ming, X. Cheng, and G. Srivastava, "Secure and resilient artificial intelligence of things: a HoneyNet approach for threat detection and situational awareness," *IEEE Consum. Electron. Mag.*, vol. 11, no. 3, pp. 69–78, 2021.

[10] I. M. M. Matin and B. Rahardjo, "Malware detection using honeypot and machine learning," in *2019 7th international conference on cyber and IT service management (CITSM)*, 2019, vol. 7, pp. 1–4.

[11] N. Naik and P. Jenkins, "A fuzzy approach for detecting and defending against spoofing attacks on low interaction honeypots," in *2018 21st International Conference on Information Fusion (Fusion)*, 2018, pp. 904–910.

[12] Y. Ma, X. Zou, Q. Pan, M. Yan, and G. Li, "Target-Embedding Autoencoder With Knowledge Distillation for Multi-Label Classification," *IEEE Trans. Emerg. Top. Comput. Intell.*, 2024.

[13] F. Aldi, F. Hadi, N. A. Rahmi, and S. Defit, "Standardscaler's Potential in Enhancing Breast Cancer Accuracy Using Machine Learning," *J. Appl. Eng. Technol. Sci.*, vol. 5, no. 1, pp. 401–413, 2023.

[14] O. J. Kadhim, A. A. Nafea, S. A. S. Aliesawi, and M. M. Al-Ani, "Ensemble Model for Prostate Cancer Detection Using MRI Images," in *2023 16th International Conference on Developments in eSystems Engineering (DeSE)*, 2023, pp. 492–497.

[15] M. Wang *et al.*, "Wetland mapping in East Asia by two-stage object-based Random Forest and hierarchical decision tree algorithms on Sentinel-1/2 images," *Remote Sens. Environ.*, vol. 297, p. 113793, 2023.

[16] K. M. A. Alheeti, A. Alzahrani, M. Alamri, A. K. Kareem, and D. Al_Dosary, "A Comparative Study for SDN

Security Based on Machine Learning.," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 11, 2023.
[17] A. Das, "Logistic regression," in *Encyclopedia of Quality of Life and Well-Being Research*, Springer, 2024, pp. 3985–3986.
[18] N. N. Jamil and A. K. Kareem, "Comparative Analysis on Machine Learning and One-Dimensional Convolutional Neural Network to Predict Surface Enhanced Raman Spectroscopy," in *2023 3rd International Conference on Computing and Information Technology (ICCIT)*, 2023, pp. 216–221.
[19] H. Tao *et al.*, "Development of integrative data intelligence models for thermo-economic performances prediction of hybrid organic rankine plants," *Energy*, p. 130503, 2024.
[20] J. S. Rhodes, A. Cutler, and K. R. Moon, "Geometry-and accuracy-preserving random forest proximities," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023.
[21] A. A. Nafea, M. S. Ibrahim, M. M. Shwaysh, K. Abdul-Kadhim, H. R. Almamoori, and M. M. AL-Ani, "A Deep Learning Algorithm for Lung Cancer Detection Using EfficientNet-B3," *Wasit J. Comput. Math. Sci.*, vol. 2, no. 4, pp. 68–76, 2023.
[22] T. O. Omotehinwa, D. O. Oyewola, and E. G. Dada, "A light gradient-boosting machine algorithm with tree-structured parzen estimator for breast cancer diagnosis," *Healthc. Anal.*, vol. 4, p. 100218, 2023.
[23] A. K. Kareem and K. M. A. Alheeti, "Hybrid Approach for Fall Detection Based on Machine Learning," in *International Conference on New Trends in Information and Communications Technology Applications*, 2021, pp. 111–130.
[24] F. Farahnakian, J. Sheikh, F. Farahnakian, and J. Heikkonen, "A comparative study of state-of-the-art deep learning architectures for rice grain classification," *J. Agric. Food Res.*, vol. 15, p. 100890, 2024.
[25] Maryam Abdulsalam Ali and ALI ALQARAGHULI , Trans., "A Survey on the Significance of Artificial intelligence (AI) in Network cybersecurity", BJN, vol. 2023, pp. 21–29, Apr. 2023, doi: 10.58496/BJN/2023/004.
[26] L. Hussain, "Fortifying AI Against Cyber Threats Advancing Resilient Systems to Combat Adversarial Attacks", EDRAAK, vol. 2024, pp. 26–31, Mar. 2024, doi: 10.70470/EDRAAK/2024/004.
[27] A. S. . Bin Shibghatullah, "Mitigating Developed Persistent Threats (APTs) through Machine Learning-Based Intrusion Detection Systems: A Comprehensive Analysis", SHIFRA, vol. 2023, pp. 17–25, Mar. 2023, doi: 10.70470/SHIFRA/2023/003.
[28] D. Zaman and M. Mazinani, "Cybersecurity in Smart Grids: Protecting Critical Infrastructure from Cyber Attacks", SHIFRA, vol. 2023, pp. 86–94, Aug. 2023, doi: 10.70470/SHIFRA/2023/010.
[29] S. salman Qasim and S. M. NSAIF , Trans., "Advancements in Time Series-Based Detection Systems for Distributed Denial-of-Service (DDoS) Attacks: A Comprehensive Review", BJN, vol. 2024, pp. 9–17, Jan. 2024, doi: 10.58496/BJN/2024/002.