Applications of Artificial Intelligence for Cyber Security Group Project

Literature Review in Machine Learning Topics

Members:
Muhammad Nurfajri 45786305
Faisal Rahman 46261745
Sadia Nowshin 46295925

Table of Contents

1. Introduction

Machine learning topic is becoming popular in past years. The number of research papers based on this topic has increased significantly. The increase in the field happened due to the emergency of several new computing technologies and the abundance of available data [1]. Machine learning has developed numerous techniques and algorithms that can be applied and implemented in different information systems. For example, optimizing online search engines [2], filtering spam [3], creating personalized recommendations on a website [4], voice recognition [5], and detecting unusual patterns or anomalies for different types of systems [6].

Machine learning for detecting anomalies nowadays is becoming vital research for strengthening the security aspect of online systems [7]. Almost every developed online system in any big technology company already accommodates this potential advantage to build a more robust system that can be resilient to any cyber-attack before it occurs and generate destruction inside a system [8]. For example, machine learning can detect Malware, Probing or Port Scan, and Network Intrusion.

This research aims to provide a literature review on the implementation of machine learning for three different purposes: machine learning for detecting malware, machine learning for detecting Port Scan (Probing), and machine learning for detecting Network Intrusion.

1.1 Background

The significant increase of advanced technologies for numerous purposes and systems has prompted researchers to find effective and efficient methods or techniques to detect the anomaly inside those advanced systems. Reviewing literature from numerous sources related to anomaly detection in machine learning will generate helpful information about what methods or techniques can be implemented for specific systems or machines and eventually strengthen the security aspect of the system.

1.2 Context

The paper focuses on reviewing numerous research papers related to the specific anomaly detection of malware, probing or port scan, and network intrusion by extracting the method techniques that are being used and any novel proposed techniques, any limitations, and challenges regarding the techniques, and draw some conclusions.

1.3 Purposes

The main objective of the research paper is to review, discuss, evaluate, and conclude the algorithms or techniques that numerous researchers in this topic area have used. By understanding deeply, the implementation of the techniques for the specific tasks in machine learning, it will generate an essential piece of information that can be used as a consideration and option to solve any issue in real cases in the machine learning world. Moreover, limitations and conclusions are provided to obtain the extraction of the review.

1.4 Significant and Scope

This literature review paper aims to obtain a broader understanding of the implementation of machine learning for anomaly detection for specific purposes by reviewing an adequate number of research papers related to the topics, including challenges that researchers faced, and providing a brief evaluation of the used techniques. The scope of this paper only covers anomaly detection for specific purposes that have been mentioned.

1.5 Thesis Outline

This paper comprises sections as follows: Section 1 outlines the introduction of the research paper, Section 2 outlines the literature review from numerous research papers related to the topics, Section 3 outlines the limitations, Section 4 outlines the relationship between the reviewed papers and the analysis task, and Section 5 outlines the conclusion of the paper.

2. Literature Review: Machine Learning for Cyber Security Applications

This section will be used to elaborate extracted papers that have been reviewed related with the use of machine learning for cyber security applications.

2.1 Behavior-based Malware Detection Techniques

The researchers [9] analyzed the use of machine learning techniques to detect malware based on the behavior since manual analysis or inspection, or static malware analysis is no longer categorized as efficient and effective to deal with the high rate of spreading of modern malware nowadays. The literature aims to analyze which of the available techniques for malware detection is based on behavior that produces higher accuracy in detecting numerous kinds of malware.

The computer malware included in this analysis: Trojan Horses, Botnets, Backdoors, Worms, Viruses, and other malicious software. The classifier techniques used in the research paper are SVM or Support Vector Machine, K-Nearest Neighbors or KNN, J48 Decision Tree, Multilayer Perceptron Neural Network or MLP, and Naïve Bayes. This research paper argued that the best technique that produces the highest precision score is the J48 Decision Tree technique with 97.3% precision and roughly 2% rate of false-positive with an accuracy of 97%.

2.2 High Accuracy Machine Learning Framework for Malware Detection

A paper [10] conducted a research paper to propose a new framework for detecting malware using a machine learning approach. This research aims to build a framework that enables any machine learning algorithms to classify correctly between clean files and malware files with the highest accuracy possible. This approach uses three datasets: test dataset, training dataset, and scale-up dataset. The paper argued that the framework produced roughly 90% accuracy by using several different machine learning algorithms to detect malware.

The paper described that there is a limitation regarding the approach. Since the approach or framework used the non-stochastic version of the perceptron algorithm, handling a considerable dataset size will cause a problem. However, the researcher claimed that the solution for this limitation is to utilize distributed computing such as a grid system on a smart grid.

2.3 Machine Learning for Android Malware Detection

Another paper [11] conducted a research review based on machine learning to detect malware inside Android environments. This paper aims to help academics and researchers obtain a complete picture of malware detection on the Android mobile environment, which is currently the leading operating system for mobile.

The paper conducted a better research review on a wider range of aspects of this topic by analyzing and summarizing the status of research from vital perspectives such as the acquisition of samples, data preprocessing, feature selection, models on machine learning, algorithms used, and eventually evaluating the effectiveness of the malware detection. The researchers argued that the implementation of machine learning for detecting malware on the Android platform is dependent on the composition of the approach, such as the algorithms used and the training method. No single evaluation method can be used for all cases.

2.4 Machine Learning Intrusion Detection Anomaly-Based on Probing Attacks

The researchers [12] have conducted a machine learning research based for detecting anomaly on Probing Attack that was happened in an institutional network (case study). The paper explained the investigation on the network intrusion attempts with machine learning model based on anomaly to produce an improved protection than the conventional models. The paper used two type of machine learning models, convolutional neural network model and ensemble learning model. UNSW-NB15 and institutional datasets have been used for this research with final dataset included 100.000 flows per sessions and 38.804 unique hosts.

The researchers in this paper argued that the result of the research depicted that the convolutional neural network model (CNN) produced the highest scores compared to ensemble with different techniques such as SVM, KNN, Logistic Regression and Naïve Bayes. The CNN model overcomes other types of models by obtaining a 99% F1 score and ROC AUC Score.

2.5 Machine Learning for Sniffing Detection

Some researchers [13] have conducted a research paper that focused on sniffing detection based on network traffic and machine learning. This research aims to develop a novel detection method that detects whether any sniffing program is active on the specific (suspected) machine. The background research of this paper is to compete with the highly sophisticated methods executed by cybercriminals that can collect information on vulnerable devices (network probing).

The researchers mentioned several available methods that can be used to detect sniffing activities:

1. DNS-Based Approach
2. Forged Mac Addresses Approach
3. Combination of both

The research also compared the result generated by two different machine learning models XGBoost and LightGBM. The LightGBM was removed due to low performance during the tuning phase, while the XGBoost generated a high score for this scenario. XGBoost is a library for distributed gradient boosting that has been developed with high portability, efficiency, and flexibility [14].

2.6 High Accuracy Machine Learning Techniques for Probing Detection

The research paper [15] on probing detection focuses on how to detect probing attacks by utilizing several machine learning techniques such as Naïve Bayes, Decision Trees, Multilayer Perceptron, and Support Vector Machine. The research used KDDcup99 as a dataset. The main aim of this research is to obtain the result value from those machine learning techniques and compare them to conclude which techniques are considered as the best method for detecting a probing attack.

The KDDcup99 comprised several categories of attacks following the attack name:

1. Probes: nmap, ipsweep, Satan, portsweep
2. Denial of Service: teardrop, Back, Neptune, land, smurf, pod
3. Remote to Local: multihop, guess_passwd, spy, phf, ftp_write, warezclient, imap
4. User to Root: rootkit, Buffer_overflow, perl, loadmodule

The result of this research showed that the highest accuracy scores came from Decision Tree, Neural Network and Auto MLP techniques by generated more than 99% of accuracy.

2.7 Network Intrusion Detection Using Supervised Machine Learning

The researchers [16] conducted research on supervised machine learning by using feature selection to detect any intrusions inside a network or Network Intrusion Detection system. The aim of this research is to find the best model that produces the highest success rate possible by combining an algorithm of supervised learning with feature selection method. The paper was using NSL-KDD dataset and considering Support Vector Machine (SVM) and Artificial Neural Network (ANN) as the chosen supervised learning techniques that will be compared the result later on.

The researchers used 20% of the data set as training data comprised of 25.181 data instances with label. The approach that the researchers used is to make 2 models for each technique. For each 2 model, one was using 17 features, and another was using 35 features. The results of this research depicted that the ANN supervised learning algorithm outperform the SVM technique for classifying the traffic of the network with more than 94% detection accuracy, while SVM only produced roughly 82%.

2.8 Evaluation of Network Intrusion Detection Based on Machine Learning

Some researchers [17] conducted an evaluation of available Machine Learning techniques to be utilized as a detector for any intrusions inside a network based on its anomaly. The researchers argued that the old signature-based method is no longer effective as a new attack on it is already available and the lack of capability to detect the anomaly in rea-time. The paper used seven different machine learning methods and used the Kyoto 2006+ as the dataset. The aim of this paper is to evaluate the performance of those techniques.

The methods used for this research are Ensemble, Naïve Bayes, Fuzzy C-Means, KNN, K-Means, SVM, and Radial Basis Function. The researchers randomly picked and created two datasets separately from the Kyoto2006+ dataset. One set is used for testing and another set is for training. For clustering (unsupervised), test data set is utilized for validation. The result showed that most of the techniques generated higher than 90% for accuracy, recall, and precision. However, during the use of Receiver Operating Curve metric (ROC), the researchers found that the best algorithm for this scenario is Radial Basis Function (RBF) by producing 0.9741 ROC value.

3. Literature Review: Adversarial Machine Learning

This section will be used to elaborate extracted papers that have been reviewed related with adversarial machine learning topic

3.1 Machine Learning against Data Poisoning

Some researchers [18] conducted a research study on how machine learning can have a problem that dealt with data poisoning attack. The researcher argues that the trustworthiness of the models of machine learning can be compromised when the data is manipulated in order to mislead the process of learning by models. The aim of this research is to analyze how malicious manipulated dataset can reduce the model result, discussing how to mitigate this kind of attack on every step of model including before training, during, and after the training of the model.
The researchers divide the poison attack on machine learning into 3 different types:

1. Indiscriminate Poisoning Attacks
2. Backdoor Poisoning Attacks
3. Targeted Poisoning Attacks

The results for this research are covering the potential open challenges that may happen based on this attack, providing plausible method to defense the training models and producing techniques and benchmarks that can be used to assess and improve the trustworthiness of machine learning models related with data poisoning attacks.

3.2 Engineering Problems in Machine Learning Systems

Safety-critical systems that utilized deep learning and machine learning models such as autonomous vehicles need to be improved to avoid any fatal accidents in the real world. Thus, some researchers [19] have conducted research on how machine learning can cause engineering problems. The main objective objectives of this paper are to conduct identification, classification, and exploration to the open issue in safety-critical (engineer) system that utilized machine learning. This research covers the requirement, design and machine learning systems and models verification. The researcher also conducted a gap analysis which combined old conventional system and software quality standard SQuaRE to study the model's quality for system that utilized machine learning.

Furthermore, the paper also discussed about the direction of future quality models on machine learning models to overcome engineering problem. However, that still theoretical only and a subject for future research. The result of this paper showed that there is a lack of specification or requirements on machine learning model characteristics, lack of robustness of the model and last of interpretability. The researchers argue that the lack of specification of requirements and the lack of robustness on the model have a significant impact on conventional quality models.

3.3 The Risk of Machine Learning for Hardware Security

Some researchers [20] have conducted research to evaluate the opportunities and risks that might occur by implementing machine learning approach for hardware security purposes. The paper aims to explore deeply regarding attack and defense mechanism for hardware that are utilized machine learning. The researchers also cover an identification for finding the suitable algorithms of machine learning to be used for each category of hardware security problem.

The paper elaborates on several points related to the issues on hardware security that have been addressed by algorithms of machine learning as follow:

1. Reverse Engineering
2. IC Counterfeiting
3. Side Channel Analysis
4. Hardware Trojans
5. IC Overbuilding

Machine learning Models used for security purposes on hardware:

- Supervised learning
  1. Decision Tree (DT)
  2. Aritficial Neural Network (ANN)
  3. Convolution Neural Network (CNN)
  4. Random Forest (RF)
  5. Auto Encoders
- Unsupervised learning
  1. Partitioning around Medoids (PAM)
  2. K-means Clustering
  3. Ordering points to identify the clustering structure (OPTICS)

The paper's result showed the critical aspects regarding to the implementation of machine learning to the issues of hardware security and depicted how the practice of machine learning implementation has changes over a decade ago.

3.4 The Top 10 Risk of Machine Learning Security

A paper focuses on the evaluation and analysis on the risk that may or might occur in machine learning in terms of security perspective. The researchers argue that they have conducted deep analysis to identify architectural risk factors of machine learning system. The past researched shows that there are at least 78 specific risks related with nine components in the majority of machine learning systems. Thus, this paper aims to discuss and elaborate the top 10 most crucial risks of security. The following point are 5 of the most crucial risk on machine learning:

1. Adversarial examples: the idea of this attack is to deceive machine learning system by giving malicious input data.
2. Data Poisoning: manipulating the data that are being used by a machine learning system can compromise the entire system.
3. Online System Manipulation: a sophisticated attacker can encourage the “still-learning” machine learning system to walk on the wrong direction.

4. Transfer Learning Attack: the base system is already compromised by attacker, creating unanticipated results.
5. Data Confidentiality: attacks to extract confidential data or sensitive information from machine learning systems.

3.5 Security and Machine Learning in the Real World

In order to obtain promising defenses in man critical system in real world cases, a large number of academic research papers has urge to explore the causes of these gap or blind spots in the machine learning field. Although numerous researchers have developed been developed recently, there are still potential gap that need to be fixed. Thus, some researchers [21] conducted an evaluation on the security of a machine learning product on a vast scale. The authors also cover novel open challenges that may occur by implementing system security software in machine learning components.

The result of the paper showed that the authors successfully outline the direction for novel research into machine learning potential attacks and defenses mechanism that can be utilized to enhance the state of machine learning in the perspective of security. The authors also propose a list of the potential mitigation approaches that can be used for practitioners to secure their systems.

3.6 When Machine Learning Meets Security Issues

Due to the fact that machine learning is the one of the most utilized techniques in past decades which has been used in different fields, detect and defense any potential adversarial attacks using this approach become a promising solution in the cybersecurity world. Thus, some researchers [22] focus on reviewing numerous kinds of literature and surveys in two respects, the security and reliability of machine learning, and the implementation of machine learning for cybersecurity purposes. The authors categorize several categories of the security issue related to machine learning into 14 categories some of them also considered cybersecurity field concerns.

The authors provided several possible attacks and their possible remedies. For example, a Poisoning attack can be solved by using data sanitization, an Evasion attack and Impersonate attack can be avoided by using adversaries retraining, and an Inversion attack can be prevented by applying a Differential privacy approach. The authors argue that the limitation of this topic is dependent on the new emerging system that uses machine learning and give some conclusion that the topic deserves further discussion to fill the security gap in machine learning cyber security and propose the possible security issue on AI terminal.

3.7 Applications in Security and Evasions in Machine Learning

Machine learning has drawn the attention of both the public and academics due to its potential utilization to enhance any systems for any different kinds of fields. Machine learning is utilized to solve serious problems such as real-time detection, the assessment of data leakage vulnerability, and many more. Thus, some researchers [23] conducted research to focus on the comparison of different security applications' points of view where machine learning models become the essential part of the application.

The aim of this paper is to provide the results from the different machine learning application and assess the vulnerability in the model of machine learning to catchup with adversarial attacks when the application is being build. The authors also provided threat model mechanism and defenses strategies against different scenario of adversarial attacks. At the end of the paper, the authors concluded that the fast increase in security events in decision system that using machine learning approach, it always opens a potential new research area in the future.

3.8 Machine Learning for 5G Limitations

Driven by a huge number of mobile traffic nowadays, 5G is designed to be a main crucial enabler and a leading infrastructure on the communication and information technology industry by providing numerous services and various requirements. Machine learning is expected to be a vital role to help fulfill the 5G vision. Thus, the authors of the research [24] focus on the combination of 5G technology and Machine Learning technology to realizing a novel network called Beyond 5G (B5G). This paper also covers the limitations in terms of security in the implementation of machine learning for 5G environments. For example, using machine learning can create adversarial attacks due to the characteristic of the environment.

4. Challenges and Limitations

Talking about further expansion and the limitation about the metrics for Android application classifiers, such as robustness, confidence, and generalization capability. At present, most learning algorithms have a high computational cost, which limits their applicability in many practical scenarios. Therefore, the real-time performance of the classifier deserves further study. On the basis of the existing evaluation metrics, we can construct new evaluation metrics that are multi-level or multi-dimensional. For example, the work in shows that a two-tier evaluation.

combining the metrics of AUC and accuracy has a better performance than an evaluation of AUC or accuracy alone. In addition, the design of evaluation methods with lower computational cost and higher efficiency, systematic evaluation of the correlation between different evaluation metrics, and exploration of the relationship between evaluation methods and test adequacy are all areas that invite further study.

5. The Comparison the Reviewed Papers with Analysis Task

We have worked with 3 datasets, Probing, unsw-B15(IDS) and malware. For each of these datasets we have used 2 algorithms for anomaly detection. Before training we did some pre-processing to be able to achieve slightly better results. However, the accuracy of the papers we reviewed were considerably higher. Most of them are above 90%, which is considered standard for a model to be used in real scenarios.

On the other hand, the best result that we got was around 85%. While the rest ranged between 60-75%. This is in fact a very poorly trained model, our assumption is that we needed to do more pre-processing, or that our approach to processing the data was not accurate. One other point could be that our dataset was too small and not evenly distributed.

It should also be noted that, in the papers we researched, they had implemented CNN, which yields very good results for a small amount of data. Since we had no prior knowledge on this topic, we decided not to explore that area.

6. Conclusion

In conclusion, we learned about working on datasets from scratch. We also realized that the most crucial part of Artificial Intelligence (AI) is the pre-processing phase. If one can successfully pre-process data, then the results would undoubtedly be very high. We also learned that the slightest changes in the dataset can give a much better output, and it’s very difficult to get it right the first time. over trial and errors only can we find the correct method to process a dataset to yield satisfactory results.

Moreover, after reviewing numerous literatures related to machine learning application in different kind of system and fields, also the potential issue that might occur in the implementation of machine learning models, we obtained broader understanding on how the trends and developments of machine learning in within the past decades until today.

References

[1] K. Wagstaff, "Machine learning that matters," *arXiv preprint arXiv:1206.4656,* 2012.

[2] J. D. F. a. T. J. Boyan, "A machine learning architecture for optimizing web search engines," *AAAI Workshop on Internet Based Information Systems,* 1996.

[3] K. Tretyakov, "Machine learning techniques in spam filtering.," *Data Mining Problem-oriented Seminar, MTAT,* vol. 3, no. 177, pp. 60-79, 2004.

[4] L. M. M. M. G. G. G. K. D. a. D. E. Visuwasam, "Smart personalised recommendation system for wanderer using prediction analysis," *International Journal of Intelligence and Sustainable Computing 1,* vol. 1, no. 3, pp. 223-232, 2021.

[5] N. H. H. B. P. a. V. K. D. Tandel, "Voice recognition and voice comparison using machine learning techniques: A survey," *In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS),* pp. 459-465, 2020.

[6] D. K. a. J. K. K. Bhattacharyya, Network anomaly detection: A machine learning perspective, Crc Press, 2013.

[7] A. B. M. A. T. Q. N. a. F. M. D. Nassif, "Machine learning for anomaly detection: a systematic review," *IEEE Access,* 2021.

[8] M. M. F. B. K. A. T. R. B. a. A. M. Gander, "Anomaly detection in the cloud: Detecting security incidents via machine learning," *In International Workshop on Eternal Systems,* pp. 103-116, 2012.

[9] I. A. E. a. A. S. N. Firdausi, "Analysis of machine learning techniques used in behavior-based malware detection," *second international conference on advances in computing, control, and telecommunication technologies. IEEE,* 2010.

[10] D. M. C. D. A. a. L. C. Gavriluţ, "Malware detection using machine learning," *In 2009 International Multiconference on Computer Science and Information Technology,* pp. 735-741, 2009.

[11] K. S. X. G. X. M. Z. D. S. a. H. L. Liu, "A review of android malware detection approaches based on machine learning," *IEEE Access 8,* pp. 124579-124607, 2020.

[12] E. C. T. a. C. A. Tufan, "Anomaly-Based Intrusion Detection by Machine Learning: A Case Study on Probing Attacks to an Institutional Network," *IEEE Access 9,* pp. 50078-50092, 2021.

[13] M. P. Ż. P. N. K. C. a. W. M. Gregorczyk, "Sniffing detection based on network traffic probing and machine learning," *IEEE Access 8,* pp. 149255-149269, 2020.

[14] T. T. H. M. B. V. K. Y. T. H. C. a. K. C. Chen, "Xgboost: extreme gradient boosting," *R package version 0,* pp. 1-4, 2015.

[15] C. a. V. K. B. Ambedkar, "Detection of probe attacks using machine learning techniques," *International Journal of Research Studies in Computer Science and Engineering (IJRSCSE),* vol. 2, no. 3, pp. 25-29, 2015.

[16] K. A. B. M. Y. J. a. M. M. R. Taher, "Network intrusion detection using supervised machine learning technique with feature selection," *2019 International conference on robotics, electrical and signal processing techniques (ICREST),* pp. 643-646, 2019.

[17] M. a. C.-H. L. Zaman, "Evaluation of machine learning techniques for network intrusion detection," *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium,* pp. 1-5, 2018.

[18] A. E. K. G. A. D. B. B. F. R. a. M. P. Cinà, "Machine Learning Security against Data Poisoning: Are We There Yet?," *arXiv preprint arXiv:2204.05986,* 2022.

[19] H. H. Y. a. T. N. Kuwajima, "Engineering problems in machine learning systems," *Machine Learning,* vol. 109, no. 5, pp. 1103-1126, 2020.

[20] R. a. K. C. Elnaggar, "Machine learning for hardware security: Opportunities and risks," *Journal of Electronic Testing,* vol. 34, no. 2, pp. 183-201, 2018.

[21] I. W. C. E. K. E. K. T. K. a. J. L. Evtimov, "Security and machine learning in the real world," *arXiv preprint arXiv:2007.07205,* 2020.

[22] Z. L. B. T. S. a. J. L. Guan, "When machine learning meets security issues: A survey," *In 2018 IEEE international conference on intelligence and safety for robotics (ISR),* pp. 158-165, 2018.

[23] R. R. J. a. C. B. Sagar, "Applications in security and evasions in machine learning: a survey," *Electronics,* vol. 9, no. 1, p. 97, 2020.

[24] M. E. H. L. a. W. L. Morocho-Cayamcela, "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions," *IEEE Access 7,* pp. 137184-137206, 2019.