Assessing the Efficacy of Machine Learning and Deep Learning in the field of Cyber Security

Rajesh Eswarawaka
Department of Machine Learning and Computer Science
ACM Engineering College
Bangalore, Karnataka 560076, India.
rajesheminent@gmail.com

Mais Nijim
Department of Electrical Engineering and Computer Science
Texas A&M University-Kingsville
Kingsville, Texas 78363, U.S.A.
mais.nijim@tamuk.edu

Viswas Kanumuri
Department of Electrical Engineering and Computer Science
Texas A&M University-Kingsville
Kingsville, Texas 78363, U.S.A.
satya_viswas_varma.kanumuri@students.tamuk.edu

Hisham Albetaineh
Department of Electrical Engineering and Computer Science
Texas A&M University-Kingsville
Kingsville, Texas 78363, U.S.A.
hisham.albataineh@tamuk.edu

Abstract-- *The use of machine learning has become widespread across various fields because of its superior performance compared to conventional rule-based algorithms. As a result, these models have also been integrated into cyber security systems, Machine learning is being utilized to aid or possibly even supplant the role of human security analysts. However, it's important to evaluate the effectiveness of machine learning in cyber security with careful consideration, especially if complete automation of detection and analysis is being considered. This study provides an in-depth research focuses on machine learning techniques applied in intrusion, malware, and spam detection that are tailored towards security professionals. The primary objective of our study is to evaluate the degree of advancement or maturity of these techniques of ML-based cybersecurity solutions and to identify any limitations that could impede their effectiveness as detection mechanisms. To achieve this, we conducted a thorough literature review and performed experiments on enterprise systems and network traffic in real-world settings. Our goal is to gain understanding of the capabilities and limitations of ML solutions and provide actionable insights for their improvement.*

Keywords—*cyber security, machine learning and deep learning.*

I. Introduction

The usage and popularity of machine learning (ML) have been rapidly increasing due to its ability to solve real-world problems in various Machine learning (ML) has become increasingly popular and is now used in various Machine learning has also been implemented in a range of other fields, including computer vision, social media marketing[1] etc. Even human operators have been traditionally relied upon in various fields, ML algorithms are often more effective in certain scenarios[2]. This trend of adopting ML is also making an impact In the cybersecurity sector, some detection systems are being improved by incorporating machine learning components[3]. Although creating a completely automated cyber defense system remains a long-term goal, network and security operation center operators can take advantage from using ML-based detection and analysis tools[4].

In order to use the existing status of ML solutions in cybersecurity, we carried out a thorough review of existing literature and conducted real experiments using network traffic from real, large enterprises[5]. Our study is different from other papers that evaluate ML solutions for cybersecurity by concentrating on one specific application and targeting AI specialists rather than security operators. Instead, we have tailored our research to be of direct relevance to security operators. We excluded commercial products that utilize ML or the often overhyped term "AI" since vendors typically do not disclose their algorithms and may neglect to acknowledge limitations and problems[6].

Our research involves proposing an innovative way to classify the various ML methods that are utilized in cybersecurity. Our classification system is based on the three main areas where ML is commonly applied in this field: Our study specifically examines the use of machine learning in intrusion detection, malware analysis, and phishing detection. We assess the primary shortcomings of current machine learning techniques, including their advantages and disadvantages, particularly in regard to false positives and false negatives. Additionally, we underscore the challenges of managing machine learning architectures in the cybersecurity domain, as well as the necessity for ongoing fine-tuning, that is constantly evolving. Furthermore, we examine recent findings that demonstrate the efficacy of adversarial attacks in circumventing ML detection mechanisms.

Our study aims to identify areas for improvement in ML components used in cyber defense platforms. In conclusion, we present our findings in a structured manner, with Our paper is structured as follows: Section 2 introduces our novel classification system of machine learning algorithms used in the cybersecurity field, while Section 3 outlines the three categories of cybersecurity issues examined in this study. Section 4 provides a comparative analysis and assessment of various machine learning solutions for cybersecurity, including their strengths and limitations. Finally, Section 5 offers concluding remarks and reflections on the findings of our research.

II. Cyber Security: Application of ML algorithms

The field of ML is continuously evolving and encompasses various paradigms that have cross-relationships and weak boundaries. Due to this, there is no fully accepted taxonomy from literature, and It is important to note that different perspectives and use cases can lead to different categorizations

979-8-3503-2759-5/23/$31.00 ©2023 IEEE
DOI 10.1109/CSCE60160.2023.00388

of machine learning algorithms in the cybersecurity domain[41]. Therefore, we propose an original taxonomy in Figure 1 that can capture the differences between the many techniques used in cyber detection. Our proposed taxonomy is designed to be accessible to security operators and does not aim to provide a comprehensive classification system that would meet the needs of all AI experts and applications. Figure 1 introduces the primary differentiation between two types of ML algorithms: Shallow Learning (SL) and Deep Learning (DL)[42]. SL algorithms rely on domain expertise or feature engineering to identify matching data characteristics in prior running the algorithm. In contrast, DL algorithms use Representation learning is a newer development in the field of machine learning that involves a multi-layered representation of input data and the ability to autonomously select features[43].

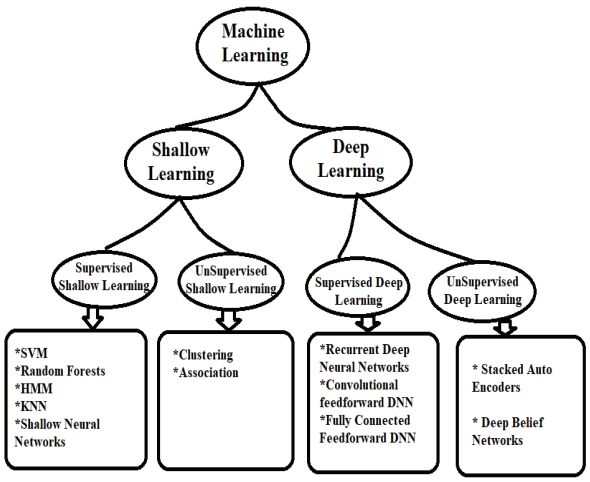

FIGURE 1. CLASSIFIED ML ALGORITHMS FOR CYBER SECURITY.

Machine learning techniques can be classified into SL and DL, with the latter being a more recent development. Shallow Learning involves a domain expert Shallow Learning algorithms require a domain expert or feature engineer to identify the Prior to executing an algorithm, it is necessary to identify relevant data characteristics when using traditional machine learning approaches. Deep learning techniques use a multi-layered approach to represent input data and can independently perform feature selection via representation learning.

In addition,Supervised and unsupervised algorithms can be further classified within both SL and DL approaches. Supervised techniques require a big and representative set of pre-labeled data for training, whereas unsupervised approaches do not require pre-labeled data.

In Figure 1, the leaves of the classification tree represent the most commonly used categories of machine learning algorithms.It's worth noting that each category can encompass multiple techniques. Our taxonomy is tailored to meet the needs of security operators and does not attempt to provide a definitive classification system that would satisfy all AI experts and applications.

A. Shallow Learning

1) Supervised SL algorithms

• Naïve Bayes is a probabilistic classification algorithm that assumes the input dataset features are independent of one another. This type of algorithm is scalable and can generate accurate results, even with small training datasets[19].

• Logistic Regression (LR) is a type of categorical classifier that utilizes a discriminative model.and make the same a-priori independence assumption as NB algorithms. The performance of LR methods depends more on the quantity of the training data.

• Support Vector Machines (SVM) are classifiers that map data samples into a feature space. to maximize the distance between each category of samples. SVMs perform well in binary classification tasks based on input features, but their performance is not as strong in multi-class classifications. and may have limited scalability.

• Random Forest (RF) is a machine learning algorithm that uses multiple decision trees to make predictions.. It's effective for handling large datasets and is particularly useful for multiclass problems. However, it's prone to overfitting with deep trees.

• Hidden Markov Models (HMM): Hidden Markov Models (HMM) are a type of probabilistic model that use a set of states, each of which has a probability distribution for producing outputs. They're useful for analyzing time-based data and predicting future events based on past observations. In cybersecurity, HMMs are commonly used to detect anomalies in network traffic and identify malicious behavior. HMMs can be trained on both labeled and unlabeled data, but their accuracy is improved when trained on labeled datasets.

• K-Nearest Neighbor (KNN): These classifiers are used for classification and can handle multi-class problems. However, they have computationally demanding training and test phases since each test sample is compared with all the training samples[8].

• Shallow Neural Network (SNN): These algorithms depends on neural networks. with a limited number of neurons and layers. SNNs are used for classification tasks in cybersecurity, despite the existence of unsupervised SNNs[9].

2. Unsupervised machine learning

(USL) algorithms are used to identify patterns and relationships in data without prior training or knowledge of the outcome. Two popular unsupervised SL algorithms are clustering and association.

Clustering is a technique that clusters similar data points together based on their characteristics. Common clustering methods K-means and hierarchical clustering are two widely used unsupervised machine learning algorithms. used for grouping data points based on their similarities. While

hierarchical clustering forms clusters in a hierarchical manner by iteratively merging smaller clusters into larger ones.

However, clustering methods have limitations in scalability and efficiency, and are not suitable for large datasets. Nonetheless, they are often used as a flexible solution in data analysis, k-means clustering can be used as a preliminary step before applying a supervised algorithm or as a standalone unsupervised learning technique.detecting anomalies. The choice of which algorithm to use depends on the specific data and analysis needs.

Association algorithms are a type of machine learning algorithm used for finding the unknown patterns within the data, which makes it useful for prediction. However, one downside of association algorithms is that they have a tendency to generate a big number of rules, which may not always be valid or meaningful. As a result, human expert inspection is often required to ensure the accuracy and relevance of the generated rules.

B. Deep Learning

Deep Learning (DL) algorithms depends on Deep Neural Networks (DNNs), which are highly complex networks with multiple layers that can autonomously learn data features. DL algorithms are typically categorized into two types, supervised and unsupervised[10].

1) Supervised Learning DL algorithms :Supervised Learning DL, There are various deep neural networks used in ML like (DNN, FNN, CNN), and Recurrent Deep Neural Networks (RNN)[11]. FNNs are neural networks in which each neuron will be connected to adjacent layers neuron[12]. CNNs, on the other hand, use convolutional layers to extract features from input data. RNNs are used for analyzing sequential data I.e given output relays on both, previous input, present input and outputs in the sequence[13]. These deep neural network algorithms have demonstrated high accuracy and efficiency in various applications[14]. FNNs are a flexible and general-purpose solution for classification problems, but are computationally expensive. CNNs are suitable for analyzing spatial data, but are less effective with non-spatial data. RNNs are more complex than FNNs, but excel at generating sequences[15].

2) Deep Learning Unsupervised algorithms: Deep Belief Networks (DBN) and Stacked Autoencoders (SAE). DBNs are made up of Restricted Boltzmann Machines (RBMs) and Deep Neural Networks (DNNs) can used for pre-training task, because of their strong feature extraction capabilities[16]. However, they need a training stage with unlabelled datasets, which can be time-consuming and resource-intensive. Which are composed of multiple Autoencoders, are also useful for pre-training and perform better with small datasets[17].

III. Machine Learning Algorithms Applications

In the realm of cyber security, machine learning (ML) algorithms are being applied to detect and prevent unauthorized access to computer systems and networks[19]. Malware analysis involves the detection, analysis, and classification of malicious software (malware) such as viruses, worms, and Trojans. Spam detection involves the identification and filtering of unwanted and unsolicited emails, often used for phishing attacks or spreading malware[20]. Machine learning algorithms are used in these areas to improve the accuracy and effectiveness of the detection and prevention processes[22].unauthorized activities in a computer or network. While traditional IDS relied on known attack patterns, modern IDS use anomaly detection, threat detection [7][35], and ML-based classification. Two specific challenges within intrusion detection are botnet detection and detection of Domain Generation Algorithms (DGA)[31]. Botnets are networks of infected machines used for nefarious activities, and DGA automatically generate domain names to communicate with external servers, thereby evading traditional defenses based on blacklists. ML techniques can be used to detect botnets and DGAs[36].

Malware analysis is a significant problem in cyber security, as modern malware can generate new variants that achine learning techniques could be employed to study various variants of malware and accurately attribute them to their respective malware families[16][23][32]. This approach is particularly effective in cases where traditional rule-based identification methods have failed.

Spam and phishing detection is essential for reducing unwanted emails that contain malware, which can provide a foothold for attackers[27]. Traditional filters are often evaded by advanced tacktics used by attackers. ML approaches can be used to improve the spam detection process[28].

	Intrusion Detection			Malware Analysis	Spam Detection
	Network	Botnet	DGA		
Supervised	RNN [8]	RNN [9]		FNN [10] CNN [11] RNN [12]	
Unsupervised	DBN [13] SAE [14]			DBN [15] SAE [16]	DBN [17] SAE [18]
Supervised	RF [3] NB [3] SVM [3] LR [3] HMM [3] KNN [3] SNN [3]	RF [19] NB [19] SVM [19] LR [20] KNN [21] SNN [22]	RF [23] HMM [23]	RF [24] NB [24] SVM [24] LR [24] HMM [25] KNN [24] SNN [26]	RF [27] NB [28] SVM [28] LR [27] KNN [27] SNN [27]
Unsupervised	Clustering [29] Association [30]	Clustering [5]	Clustering [31]	Clustering [24] Association [32]	Clustering [33] Association [34]

Figure 2: ML to Cyber Security Problems

Figure2 1 displays the main ML algorithms used for cyber security problems, categorized by the algorithm family and the type of issue. The rows indicate the fIn Section 2, a family of machine learning algorithms was discussed for various cybersecurity issues. A table was used to illustrate the algorithms used for each problem, with the rows representing the algorithms and the columns representing the cybersecurity issues. Each cell in the table indicates which ML algorithms were used for each problem. Any empty cells in the table indicate that no algorithms were used for that particular problem.suggest that no proposals have been made for that category of problems. The table indicates that SL algorithms are

used for all issues. Supervised DL algorithms are frequently applied In the field of cybersecurity, machine learning algorithms are commonly used for malware analysis and intrusion detection, but less frequently used for spam detection. In particular, unsupervised deep learning algorithms are often relied upon for spam detection. Despite the relevance of natural language processing in cybersecurity, it is not as widely used in practice as other machine learning techniques. to DGA detection, no DL algorithm has been applied to this problem. Overall, the number of DL algorithms used for cyber security problems is lower than that of SL algorithms, primarily due to the recent development of large neural networks for DL. Consequently, this creates several research opportunities to bridge this gap.

IV. EVALUTION

In this section, we highlight seven critical issues that organizations should consider before deploying machine learning (ML) algorithms in their Network Operations Center (NOC) or Security Operations Center (SOC). However, it's essential to note that, at present, no ML algorithm can function entirely autonomously without human supervision. To support our claims, we conducted experiments in the areas of DGA We utilizedThe machine learning algorithms Random Forest and Feedforward Fully Connected Deep Neural Network have been used for both DGA (Domain Generation Algorithm) detection and network intrusion detection. These algorithms have shown promising results in accurately detecting malicious activity and have the potential to improve the overall security of networks.
To develop our DGA detection model, we obtained two labeled datasets: one containing DGA generated using known techniques and the other with DGA created using newer approaches. We also randomly selected non-DGA domains from the top-1 million domains in Cisco Umbrella. Table 2 presents the key metrics of our training datasets for a study focused on developing a machine learning-based classifier to detect malicious domain names. Our testing dataset was comprised of 10,000 domains extracted from each of the training datasets, with an additional 20,000 domains used as an unlabeled dataset for further analysis.

DATA SET	DGA TECHNIQUE	NON-DGA COUNT	DGA COUNT
1	Known	20,486	21,178
2	Newer & Known	8017	37,563

TABLE 1. Training Datasets for DGA Detection Experiments

We assessed the effectiveness of two classifiers that we developed for network intrusion detection. One of the classifiers was based on shallow learning and used Random Forest algorithm, and the other on deep learning using a Feedforward Neural Network. To assess the classifiers' performance, we used common metrics such as Recall, F1-score and precision, as Accuracy is not reliable when legitimate events outnumber illegitimate events by several orders of magnitude in real organizations. To reduce bias, we computed each metric after 10-fold cross-validation.

Dataset	Malicious Flows	Bening Flows
1	1000	100000
2	2500	250000
3	5000	500000

TABLE 2: Trining Dataset to Network

Although deep learning has been successful in some fields, such as computer vision, it does not always outperform shallow learning algorithms in cybersecurity. In fact, well-tuned shallow learning algorithms can still perform better than deep learning methods. To assess the effectiveness of two classifiers, we utilized the third dataset from Table 3 and presented the outcome in Table 4. The F1-score obtained by the Random Forest classifier was approximately 0.8, indicating its high performance, while the FNN classifier only attained an F1-score of 0.6, even with the best topology of 1,024 neurons across 4 hidden layers. Hence, we recommend that security administrators not only consider the multi-layer neural approach provided by deep learning but carefully evaluate the performance of different classifiers to select the one that suits their organization's specific needs best.

A. Shallow vs Deep Learning

1) Tracking Deep Learning has been shown to be superior to Shallow Learning in certain applications, particularly in computer vision. However, in the field of cyber security, well-configured Shallow Learning algorithms may perform better than Deep Learning approaches. Although there are fewer proposals for Deep Learning techniques in this domain, they are not always the best option.

Classifier	F1-Score	Precision	Recall
Random Forest (SL)	0.7865	0.8652	0.7854
Fully Connected Feedforward DNN	0.6086	0.7508	0.5027

TABLE 3. Comparision Between DL and SL Classifiers

B. Specific detectors vs General

The effectiveness of machine learning (ML)-based security products is often exaggerated by vendors as a solution for a Unbiased experimental results suggest that machine learning (ML) algorithms can outperform traditional methods in detecting specific types of cyberattacks, rather than attempting to identify multiple threats at once. In response, researchers developed multiple intrusion detection systems that focused on specific attack types, using their own random forest classifiers. Each classifier was trained and tested using The performance of six attack-specific classifiers and one general-purpose classifier was evaluated using the third dataset from Table 4.

Classifier	Precision	Recall	F1-score
RandomForest (RF)	0.8652	0.7283	0.7935
Deep Neural nets	0.7723	0.5684	0.6072

TABLE4: Comparision between Supervised and Deep Learning

The results were compared to a baseline classifier, and Table 5 shows the Precision, Recall, and F1-scores of the classifiers. The five attack-specific classifiers achieved high F1-scores over 0.95, while the general-purpose classifier had poor performance. The study concluded that using a single ML detector for identifying malicious traffic is currently not feasible, but having multiple detectors that focus on particular attack types can improve detection capabilities significantly.

TABLE5: Results for Specific Classifiers and General Classifiers

Attack Name	Precision	Recall	F1-Score
General Approach	0.8739	0.7256	0.7895
Possible Malware Infection	0.9839	0.9357	0.9628
DOS Attempt	0.9883	0.9869	0.9864
Over Flow Attempt	0.9844	0.9857	0.9735
Cache Poisoning Attempt	0.9781	0.9417	0.9785

C. Vulnerability to adversarial attacks

Skilled attackers use various techniques to bypass machine learning-based detectors, also known as adversarial attacks [5]. These attacks can target the integrity, availability, or privacy of the targeted system [6]. Adversarial attacks aimed at integrity deceive classification or clustering algorithms by producing malicious activities labeled as benign. Availability attacks create a large number of normal events, which the detectors classify as attacks, resulting in many false positives. Privacy attacks allow attackers to extract information about Adversaries can attempt to bypass defensive machine learning algorithms by exploiting vulnerabilities in the target network. Additionally, the emergence of generative adversarial networks (GANs) in recent years has allowed for the automatic generation of adversarial samples against machine learning systems. GANs are a type of deep neural network that can be used to generate samples that can fool machine learning systems[29].

Adversarial learning is a technique that involves training machine learning models with adversarial samples to improve their robustness against adversarial attacks. Adversarial samples are intentionally crafted inputs designed to mislead the model and cause it to produce incorrect outputs. By including adversarial samples in the training dataset, the model can learn to better detect and handle such attacks.

D. *Selection of a machine learning algorithm*

This is an important point to consider when evaluating the effectiveness of machine learning algorithms in cybersecurity. The performance of an algorithm can be highly dependent on the specific dataset used for training and testing, as well as the features selected and any pre-processing steps taken. Therefore, it is essential to compare algorithms on the same datasets and under the same conditions to obtain meaningful and unbiased results. Security administrators should be cautious when interpreting results from different studies that do not follow this approach, as they may not be directly comparable or indicative of real-world performance.

Incorrect classification of cybersecurity threats can have serious consequences, thus minimizing false positives and false negatives is crucial. False positives in malware and intrusion detection can slow down remediation efforts, while false positives in phishing detection can prevent legitimate messages from reaching users. Failing to detect malware, network intrusions, or phishing emails can be disastrous for organizations[28]. To evaluate the effectiveness of machine learning (ML) techniques for malware analysis, we will use a method presented in [24], which compares various ML techniques on the same dataset. For phishing detection, we will refer to the results in [27], which compares different ML classifiers on a custom dataset of 3,000 phishing emails. While achieving high accuracy scores is important in intrusion detection, it may not be sufficient in modern solutions that generate many events. We will conduct an experiment using two DGA detectors based on Random Forest classifiers trained on datasets from Table 2 and validated on a real domain dataset, and the results will be summarized in Table 6.

TABLE6: DGA Detection Classifiers Performance on Real Datasets

Classifier	Traning Dataset	Domains Classified as DGA
1	Known Dataset	476 (2.19%)
2	Newer Dataset	396(1.98%)

V. CONCLUSION

The use of machine learning and deep learning techniques has gained significance in the domain of cyber security, especially for detecting and preventing intrusions, analyzing malware, and identifying spam. This paper has presented a taxonomy of the most commonly used ML algorithms and analyzed their application to these specific problems. However, there are challenges that need to be addressed for effective use of ML in cyber security. One challenge is that all ML techniques are susceptible to adversarial attacks, and it is crucial to continuously retrain the models and perform meticulous parameter tuning to ensure their robustness and effectiveness in

detecting and preventing malicious activities, are necessary to maintain effectiveness. Additionally, using the same classifier for different threats can result in inadequate detection performance. This will be resolved by taking multiple ML classifiers to find the specific threats. Although deep learning is in its early stages, significant advancements are expected, particularly in adversarial learning. While ML stratagies could support security operators and automate some tasks, it is important to recognize their limitations. It is critical not to overestimate The autonomous nature of ML algorithms can also pose a threat to organizations as attackers with expertise can exploit them to infiltrate, sabotage or steal data.

REFERENCES

[1] M.I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, 2015.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, 2015.

[3] A. Buczak and E.Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Communications Surveys & Tutorials, 2015.

[4] E. Blanzieri and A. Bryl, "A survey of learning-based techniques of email spam filtering," Artificial Intelligence Review, 2008.

[5] J. Gardiner and S. Nagaraja, "On the Security of Machine Learning in Malware C8C Detection," ACM Computing Surveys, 2016.

[6] L. Huang, A. D. Joseph, B. Nelson, B. I. P. Rubinstein, and J. D. Tygar, "Adversarial machine Learning," in ACM workshop on security and artificial intelligence, 2011.

[7] [7] F. Pierazzi, G. Apruzzese, M. Colajanni, A. Guido, and M. Marchetti, "Scalable-architecture for online prioritization of cyber threats," in International Conference on Cyber Conflict (CyCon), 2017.

[8] J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection," in IEEE International Conference on Platform Technology and Service (PlatCon), 2016.

[9] [9] P. Torres, C. Catania, S. Garcia, and C. G. Garino, "An analysis of Recurrent Neural Networks for Botnet detection behavior," in IEEE Biennial Congress of Argentina (ARGENCON), 2016.

[10] G. E. Dahl, J. W. Stokes, L. Deng, and D. Yu, "Large-scale malware classification using random projections and neural networks," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[11] G. D. Hill and X. J. Bellekens, "Deep Learning Based Cryptographic Primitive Classification," arXiv preprint, 2017.

[12] R. Pascanu, J. W. Stokes, H. Sanossian, M. Marinescu, and A. Thomas, "Malware classification with recurrent networks," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.

[13] M. Z. Alom, V. Bontupalli, and T. M. Taha, "Intrusion detection using deep belief networks," in IEEE National Aerospace and Electronics Conference (NAECON), 2015.

[14] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), 2016.

[15] Y. Li, R. Ma, and R. Jiao, "A hybrid malicious code detection method based on deep learning," International Journal of Security and Its Applications, 2015.

[16] W. Hardy, L. Chen, S. Hou, Y. Ye, and X. Li, "DL4MD: A Deep Learning Framework for Intelligent Malware Detection," in International Conference on Data Mining (DMIN), 2016.

[17] G. Tzortzis and A. Likas, "Deep belief networks for spam filtering," in IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2007.

[18] G. Mi, Y. Gao, and Y. Tan, "Apply stacked auto-encoder to spam detection," in International Conference in Swarm Intelligence, 2015.

[19] M. Stevanovic and J. M. Pedersen, "An efficient flow-based botnet detection using supervised machine learning," in IEEE International Conference on Computing, Networking and Communications (ICNC), 2014.

[20] S. Ranjan, Machine learning based botnet detection using real-time extracted traffic features, Google Patents, 2014.

[21] B. Rahbarinia, R. Perdisci, A. Lanzi, and K. Li, "Peerrush: mining for unwanted p2p traffic," in International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, 2013.

[22] A. Feizollah and e. al, "A study of machine learning classifiers for anomaly-based mobile botnet detection," in Malaysian Journal of Computer Science, 2013.

[23] M. Antonakakis, R. Perdisci, Y. Nadji, N. Vasiloglou, S. Abu-Nimeh, W. Lee, and D. Dagon, "From throw- away traffic to bots: detecting the rise of DGA-based malware," in USENIX Security Symposium, 2012.

[24] T. Chakraborty, F. Pierazzi, and V. Subrahmanian, "Ec2: Ensemble clustering and classification for predicting android malware families," IEEE Transactions on Dependable and Secure Computing, 2017.

[25] C. Annachhatre, T. H. Austin, and M. Stamp, "Hidden Markov models for malware classification," Journal of Computer Virology and Hacking Techniques, 2015.

[26] J. Demme, M. Maycock, J. Schmitz, A. Tang, A. Waksman, S. Sethumadhavan, and S. Stolfo, "On the feasibility of online malware detection with performance counters," in ACM SIGARCH Computer Architecture News, 2013.

[27] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "A comparison of machine learning techniques for phishing detection," in ACM Proceedings of the Anti-Phishing Working Groups, 2007.

[28] G. Xiang, J. Hong, C. P. Rose and, L. Cranor, "Cantina+: A feature-rich machine learning framework for detecting phishing web sites," ACM Transactions on Information and System Security (TISSEC), 2011.

[29] G. Apruzzese, M. Marchetti, M. Colajanni, G. Gambigliani Zoccoli, and A. Guido, "Identifying malicious hosts involved in periodic communications," in IEEE International Symposium on Network Computing and Applications (NCA), 2017.

[30] F. S. Tsai, "Network intrusion detection using association rules," International Journal of Recent Trends in Engineering, 2009.389

[31] F. Bisio, S. Saeli, L. Pierangelo, D. Bernardi, A. Perotti, and D. Massa, "Real-time behavioral DGA detection through machine learning," in IEEE International Carnahan Conference on Security Technology (ICCST), 2017.

[32] Y. Ye, D. Wang, T. Li, D. Ye, and Q. Jiang, "An intelligent PE-malware detection system based on association mining," Journal in computer virology, 2008.

[33] W.-F. Hsiao and T.-M. Chang, "An incremental cluster-based approach to spam filtering," Expert Systems with Applications, 2008.

[34] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based associative classification data mining," Expert Systems with Applications, 2014.

[35] K. Rieck, P. Trinius, C. Willems, and T. Holz, "Automatic Analysis of Malware Behavior Using Machine Learning," Journal of Computer Security, 2011.

[36] H. S. Anderson, J. Woodbridge, and B. Filar, "DeepDGA: Adversarially-Tuned Domain Generation and Detection," in ACM Workshop on Artificial Intelligence and Security, 2016.

[37] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in neural information processing systems, 2014.

[38] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: A review," Expert Systems with Applications, 2009.

[39] R. Zuech, T. M. Khoshgoftaar, and R. Wald, "Intrusion detection and big heterogeneous data: a survey," Journal of Big Data, 2015.

[40] A. Khan, B. Baharudin, L. H. Lee, and K. Khan, "A review of machine learning algorithms for text documents classification," Journal of advances in information technology, 2010.

[41] V. M. R. M, V. Khullar, A. A. Bhosle, M. D. Salunke, J. L. Bangare and A. Ingavale, "Streamed Incremental Learning for Cyber Attack Classification using Machine Learning," 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2022, pp. 1-5, doi: 10.1109/CISCT55310.2022.10046651

[42] [42]H. M. Farooq and N. M. Otaibi, "Optimal Machine Learning Algorithms for Cyber Threat Detection," 2018 UKSim-AMSS 20th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK, 2018, pp. 32-37, doi: 10.1109/UKSim.2018.00018.

[43] [43]R. -F. Hong, S. -C. Horng and S. -S. Lin, "Machine Learning in Cyber Security Analytics using NSL-KDD Dataset," 2021 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taichung, Taiwan, 2021, pp. 260-265, doi: 10.1109/TAAI54685.2021.00057.