Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 1 (2024)

Detecting Cyber Threats Utilizing Machine Learning Approaches: An Assessment of Performance Perspective

Bipin Kumar Singh[1], Manish Kumar[2], Tushar Rexwal[3], Dr. Anupriya Jain[4].

School of Computer Applications, Manav Rachna International Institute of Research and Studies, Faridabad, India.

Abstract:- In contemporary society, our extensive dependence on the internet for various facets of everyday life has led to a remarkable upswing in online activities. Nevertheless, this surge in internet usage has concurrently resulted in a heightened prevalence of cyber threats and cybercrimes. Cybercriminals persistently devise methods to evade security protocols, rendering conventional approaches insufficient for identifying attacks, particularly those exploiting undisclosed vulnerabilities. To confront this issue, a plethora of machine learning techniques has been devised to fortify cybersecurity and uncover instances of cybercrimes. This study specifically centers on the assessment of three widely adopted machine learning methodologies: Belief Networks, Decision Trees, and Support Vector Machines. Their efficacy in discerning spam messages, detecting intrusions into computer systems, and identifying malicious software is evaluated using established datasets commonly utilized for benchmarking purposes.

Keywords*: Cyber threats, Cybercrimes, Internet dependence, Machine learning techniques, Belief networks, Decision trees, Support vector machines, Spam detection, Intrusion detection, Malicious software identification.*

1. Introduction

Cyberspace, a worldwide platform that facilitates the global exchange of electronic resources, encompasses diverse data forms, including electronic documents, audiovisual content, and social media interactions. It integrates elements such as the Internet, proficient users, system resources, data, and newcomers. Over time, it has evolved into a pivotal channel for information exchange and resource accessibility, witnessing a substantial increase in internet usage, particularly in developed nations, where it has risen by 81% since 2017. Despite the significance of cyberspace, this heightened prominence has also given rise to increased risks associated with cybercrimes and threats.

In addressing the growing array of cyber threats, the field of cybersecurity has witnessed noteworthy progress to adeptly confront these challenges. It involves a range of technologies, expert knowledge, and processes designed to fortify cyberspace security against cybercriminal activities. The realm of cybersecurity typically categorizes into conventional and automated approaches. While conventional cybersecurity methods face limitations due to inexperienced users, system vulnerabilities, and constrained access to clean data, they unintentionally contribute to the reinforcement of cybercrimes. The trajectory of cybersecurity points towards automated techniques with adaptive learning capabilities, enabling the identification of novel, sophisticated cyberattacks and staying vigilant against evolving threats.

Cyber threats entail a broad spectrum of risks, encompassing endeavors to pilfer information, breach integrity rules, or disrupt computer devices or networks. These threats manifest diversely, ranging from phishing and malware to attacks on Internet of Things (IoT) devices, denial-of-service attacks, spam, intrusions into networks or mobile devices, financial fraud, and ransomware. This paper centrally concentrates on investigating malware detection, intrusion detection, and spam identification as pivotal elements within the realm of cybersecurity.

Unsolicited spam emails, often utilized for advertising or disseminating deceptive content, impose a substantial drain on network and computer resources, including bandwidth, memory, and processing time. In contrast,

malware, serving as a collective term for malicious software, refers to software installations on computer systems intended to disrupt operations and compromise electronic data. It encompasses diverse forms such as viruses, worms, ransomware, adware, spyware, malvertising, and Trojan horses. Intrusions into computer networks and devices, with the objective of identifying and exploiting vulnerabilities, pose a significant cyber threat. To counteract these threats, the role of Intrusion Detection Systems (IDS) becomes crucial, classified into signature/misuse-based, anomaly-based, and hybrid systems.

Machine Learning (ML) emerges as a foundational strategy to counter cyber threats and address the limitations inherent in traditional security measures. Situated within the domain of Artificial Intelligence (AI), ML techniques exhibit the notable capability to autonomously learn from experience without explicit programming, providing a versatile solution. The application of ML techniques is increasingly extending into various domains, encompassing cybersecurity, medical science, education, intrusion detection, spam detection, and malware detection. A plethora of ML techniques, such as decision trees, random forests, naive Bayes, support vector machines, K-nearest neighbors, deep belief networks, artificial neural networks, and K-means, have proven effective in detecting and categorizing diverse cyber threats. This article specifically concentrates on assessing the effectiveness of decision trees, deep belief networks, and support vector machines through a comparative analysis using well-established benchmark datasets.

2. Literature Review

In a research endeavour undertaken by scholars [26], an investigation was conducted into the application of widely adopted machine learning techniques to enhance cybersecurity and fortify defences against cybercriminal activities. The authors underscored several challenges encountered in the deployment of machine learning methods and concluded that, despite the myriad applications of these techniques in safeguarding cyberspace, significant strides are required to bolster classifiers against adversarial attacks. They emphasized the susceptibility of machine learning classifiers to cyber threats and adversarial attacks.

In an independent inquiry conducted by researchers [27], a brief overview was provided that encompassed various publications concentrating on the utilization of machine learning models to enhance cybersecurity. The study delved into prevalent challenges related to acquiring fitting datasets that exhibit optimal relevance to specific security issues.

In a study conducted by researchers [28], a succinct assessment was carried out to compare the performance of different machine learning techniques, specifically within the domain of anomaly detection. The authors examined the efficacy of feature selection in the context of Machine Learning for Intrusion Detection Systems (IDS). They emphasized that the convolutional neural network (CNN) classifier, if fully exploited, represents an underutilized resource with the capacity to notably propel advancements in cybersecurity.

In their research [29], the authors performed an examination of the functions fulfilled by different machine learning techniques in the realms of spam, malware, and intrusion detection. They highlighted that no machine learning technique is immune to cyberattacks, and all face challenges in keeping pace with the ever-changing landscape of cybercrimes.

Adopting an innovative strategy, scholars in [30] introduced a machine learning method for identifying spam in text messages, leveraging content-based features. Their research determined that the proposed averaged neural network, coupled with content-based feature selection, outperformed numerous contemporary machine learning techniques in terms of accuracy on the identical dataset. Conversely, in a separate investigation [31], the authors pinpointed constraints in signature-based classification methods for detecting mobile malware. They advocated for an image-based deep learning approach to differentiate between malicious and authentic attributes using grayscale images.

Furthermore, scholars in [32] introduced a statistical semi-supervised machine learning method for detecting intrusions in Android mobile devices. With the escalating volume of data traffic in the mobile domain, sophisticated cybercrimes have surfaced, prompting the need for the development of more advanced machine learning techniques to proficiently identify malicious activities.

In this paper, we conduct a thorough examination of commonly utilized machine learning techniques to assess their efficacy in detecting well-known cybercrimes. Our specific focus revolves around the analysis of three widely employed machine learning methodologies: decision trees, deep belief networks, and support vector machines. Diverging from the majority of review articles that concentrate on a singular threat, our study addresses three primary cyber threats: intrusion detection, spam detection, and malware detection. We present an in-depth comparison of each classifier's performance using frequently employed datasets and engage in discussions regarding their computational complexity.

The forthcoming segments of the paper will delve into the fundamentals of machine learning, providing an overview of the classifiers under consideration and delineating the criteria employed for evaluating classifier performance. The ensuing discussion section will scrutinize different cyber threats, presenting performance evaluations based on accuracy, recall, and precision. Ultimately, the concluding section will concisely summarize the principal findings of the study.

I. Fundamentals of Machine Learning:

Artificial Intelligence (AI) constitutes a field in computer science dedicated to replicating human brain functions through the utilization of artificial entities, automating a myriad of processes. Within the realm of AI, Machine Learning, a subset thereof, accomplishes specific objectives by learning from experience instead of relying on explicit programming, thus eliminating the necessity for pre-fed data [33]. Machine learning comprises three sub-branches: supervised learning, unsupervised learning, and semi-supervised learning. In supervised learning, the target class or label is known in advance, whereas unsupervised learning involves dealing with unknown classes and clustering data based on similarities among data objects. Semi-supervised learning amalgamates elements of both supervised and unsupervised learning.

Various machine learning techniques find extensive use in cyber threat detection, encompassing decision trees, random forests, naive Bayes, support vector machines, K-nearest neighbors, deep belief networks, artificial neural networks, and K-means [33]. However, our study zeroes in on three specific techniques: decision trees, deep belief networks, and support vector machines. Subsequently, we offer a succinct overview of each.

A deep belief network (DBN) comprises intricate layers derived from Restricted Boltzmann Machines (RBMs). Employing a greedy approach, DBNs facilitate communication between each layer and its adjacent layers, though nodes within a layer do not share information horizontally with other nodes. In a DBN, each layer undertakes both input and output tasks, excluding the initial and final layers, with the terminal layer serving as the classifier. The computational complexity of DBNs is denoted as O((n + N)k), where k signifies the number of iterations, n denotes the number of records, and N represents the count of parameters within the DBN [34].

Decision trees (DT) represent supervised machine learning techniques with nodes, paths, and leaf nodes as their primary constituents. Nodes can assume roles as either root or intermediate nodes, and decision trees adhere to the if-then rule to determine the most suitable root node at each level. The leaf node designates the decision class, serving as the conclusion. The time complexity associated with decision trees is expressed as O(mn^2), where n denotes the number of instances and m represents the number of attributes [36, 37].

II. Discussion and Performance Evaluation:

The digital landscape is rife with a broad spectrum of cybercrimes that persistently target user data privacy on computer networks and mobile devices. To thwart these continually evolving threats, a diverse array of machine learning techniques has been developed. However, despite these advancements, machine learning solutions still grapple with keeping pace in the dynamic realm of cybercrimes. In our thorough review, we specifically concentrate on detecting three critical cyber threats: Intrusion Detection Systems (IDS), malware detection, and spam detection. Our analysis revolves around three primary learning models: decision trees, support vector machines, and deep belief networks.

The selection of datasets is pivotal for conducting a robust evaluation of classifier performance, given that the quality and scale of the dataset significantly influence the results. Real-time and diverse datasets tend to produce more reliable outcomes than customized datasets. In our analysis, we specifically highlight frequently utilized and

benchmark datasets, which include KDD CUP 99 [41], Spambase [42], Twitter dataset [43], Enron [44], NSL-KDD [45], DARPA [46], and various malware datasets [47]. These datasets serve as the foundation for evaluating and contrasting the performance of machine learning models concerning the detection of these cyber threats.

TABLE I. Machine Learning Model Performance in Spam Detection:

Cyber Threat	Learning Model	Dataset	Reference	Published Year	Sub-Domain	Performance Results		
						Precision	Accuracy	Recall
Spam Detection	Support Vector Machine	Spambase	[48]	2011	Email Spam	93.12 %	96.90 %	95.00 %
			[49]	2015	Email Spam	79.02 %	79.50 %	68.67 %
		Twitter Dataset	[50]	2018	Spam Tweets	92.91 %	93.14 %	93.14 %
			[51]	2015	Spam Tweets		95.20 %	93.60 %
	Decision Tree	Enron	[52]	2016	Email Spam	98.00 %	96.00 %	94.00 %
			[52]	2016	Email Spam	98.00 %	96.00 %	94.00 %
		Spambase	[53]	2014	Email Spam	91.51 %	92.08 %	88.08 %
			[54]	2014	Email Spam	-	94.27 %	91.02 %
	DBN	Enron	[55]	2016	Email Spam	96.49 %	95.86 %	95.61 %
			[56]	2016	Email Spam	98.39 %	97.50 %	98.02 %
		Spambase	[57]	2007	Email Spam	94.94 %	97.43 %	96.47 %
			[58]	2018	Email Spam	96.00 %	89.20 %	-

We utilized accuracy, recall, and precision as the primary evaluation metrics to gauge the effectiveness of our classification models. Tables 2, 3, and 4 present a comprehensive overview of the performance of our three selected learning models in the domains of spam detection, malware detection, and intrusion detection.

The columns in these tables are largely self-explanatory. The "Cyber Threat" column specifies the particular type of threat under consideration, while the "Learning Model" column identifies the machine learning model used. The "Dataset" column indicates the dataset utilized, focusing on frequently used and benchmark datasets. The "Reference" column provides the citation of the respective paper containing the evaluation results.

The "Sub-domain" column displays variations for each cyber threat, emphasizing specific aspects of the threat addressed. The "Performance Results" column succinctly summarizes the performance results reported in the cited articles.

In the subsequent sub-sections, we will engage in a detailed discussion of each cyber threat, examining specific performance metrics, findings, and insights pertaining to spam detection, malware detection, and intrusion detection.

Spam Detection

Spam, posing a widespread threat to computer and network resources, appears in various forms, encompassing unwanted messages across diverse mediums like text messages, images, and videos on mobile devices [59]. In the realm of computing devices and networks, spam tweets and spam emails are the most frequently encountered manifestations. These unsolicited messages impose a toll on network resources, particularly bandwidth, and consume valuable time, especially when they manifest as unnecessary advertisements. Consequently, machine learning techniques have been extensively utilized to differentiate between legitimate emails and spam emails, as detailed in Table I.

Significantly, Support Vector Machines (SVM) and Decision Trees (DT) have exhibited commendable accuracy, achieving a rate of 96.90% [48]. However, Deep Belief Networks (DBN) have surpassed the competition, boasting a precision value of 98.39% when tested with the Enron dataset [56]. DBN's superior performance extends to recall and precision metrics compared to SVM and DT. Moreover, when using the Spambase dataset, SVM outperformed DT, securing an accuracy rate of 96.90% [48]. Conversely, when employing the Enron dataset, the decision tree exhibited superior precision relative to SVM and achieved precision levels similar to DBN [52]. Table 2 unequivocally illustrates DBN's excellence in handling these specific datasets. Based on these evaluation metrics, we recommend the utilization of DBN for spam detection.

Intrusion Detection

Intrusions, maliciously targeting computer networks and devices, present another formidable threat to cyberspace. These intrusions serve as a means to identify vulnerabilities within a network, pinpointing weaknesses in computer systems that can be exploited in subsequent attacks [60]. Intrusion detection systems (IDS) play a crucial role in safeguarding against these intrusions and can operate at either the network or host computer level. Intrusions typically fall into three classifications: signature/misuse-based, anomaly-based, and hybrid [61]. Traditional techniques often struggle to keep pace with the evolving landscape of intrusions.

Despite the valuable insights provided by commonly used datasets for intrusion detection, such as DARPA and KDD versions, these datasets are more than fifteen years old. Table II offers a comprehensive presentation of the evaluation results for intrusion detection, with DBN emerging as the top performer in terms of accuracy. DBN achieves remarkable accuracy results of 96.70% when tested with the NSL-KDD dataset [62]. The superior performance of DBN underscores its potential as a robust tool for intrusion detection.

TABLE II. Machine Learning Model Performance in Intrusion Detection System:

Cyber Threat	Learning Model	Dataset	Reference	Published Year	Sub-Domain	Performance Results		
						Precision	Accuracy	Recall
Intrusion Detection	Support Vector Machine	NSL-KDD	[63]	2019	Anomaly-Based	-	89.70 %	-
			[41]	2014	Hybrid-Based	74.00 %	82.37 %	82.00 %
		DARPA	[64]	2007	Hybrid-Based	-	69.80 %	-
			[65]	2014	Anomaly-Based	-	95.11 %	-
	Decision Tree	KDD	[66]	2018	Misuse-Based	-	99.96 %	-
			[67]	2017	Hybrid-Based	-	86.29 %	78.00 %
		NSL-KDD	[68]	2019	Hybrid-Based	-	93.40 %	-
			[69]	2017	Hybrid-Based	91.15 %	90.30 %	90.31 %
	DBN	KDD	[61]	2015	Anomaly-Based	-	97.50 %	-
		NSL-KDD	[62]	2015	Hybrid-Based	97.90 %	96.70 %	-
			[70]	2017	Anomaly-Based	88.60 %	90.40 %	95.30 %

TABLE III. Machine Learning Model Performance in Malware Detection:

Cyber Threat	Learning Model	Dataset	Reference	Published Year	Sub-Domain	Performance Results		
						Precision	Accuracy	Recall
Malware Detection	Support Vector Machine	Malware Dataset	[71]	2017	Static	-	94.37 %	-
			[72]	2013	Dynamic	-	95.00 %	-
		Enron	[73]	2015	Dynamic	-	97.10 %	-
			[52]	2016	Static	84.74 %	91.00 %	100 %
	Decision Tree	Custom	[74]	2016	Static	99.40 %	99.90 %	-
		Malware Dataset	[75]	2017	Static	-	84.70 %	-
			[76]	2014	Static	97.90 %	-	96.70 %
	DBN	Custom	[77]	2016	Dynamic	78.08 %	71.00 %	59.09 %
			[77]	2016	Static	83.00 %	89.03 %	98.18 %
		KDD CUP99	[77]	2016	Hybrid	95.77 %	96.76 %	97.84 %
			[78]	2015	Hybrid	-	91.40 %	95.34 %

Nevertheless, it is crucial to emphasize that the decision tree has showcased remarkable accuracy at 99.96%, outperforming both DBN and SVM when assessed with the KDD dataset [66]. Remarkably, the decision tree consistently surpasses other learning classifiers, demonstrating outstanding efficiency at 99.96% across diverse datasets [66]. In contrast, DBN has reported exceptional recall and precision values of 95.30% and 97.90%, respectively [62, 70].

Considering the evidence provided in Table 3 and the referenced articles, the decision tree emerges as the preferred learning classifier for intrusion detection.

Malware Detection

Malware, a fusion of malicious software, encompasses various software designed to disrupt computer operations and compromise electronic data. Notable types of malware include viruses, worms, ransomware, adware, spyware, malvertising, and Trojan horses [79]. Malware poses a significant threat, disrupting the normal operation of computers and mobile devices and compromising data integrity and the availability of computer and network resources. In response, machine learning techniques have been leveraged to detect and combat malware. Table 4 offers a comprehensive assessment of the performance of each learning classifier in this context.

Malware detection is further categorized into sub-domains: static detection, which examines applications for malware without executing them, and dynamic detection, which involves testing applications by executing them. Hybrid detection combines elements of both static and dynamic methods [80].

The decision tree demonstrates the highest overall accuracy of 99.90% when assessed on custom data collected by the authors [74]. However, on a malware dataset, SVM has surpassed the decision tree in terms of accuracy. SVM has also achieved the best recall value, reaching a perfect 100% [52]. Based on the collective findings presented in the referenced papers, SVM is recommended as the preferred learning classifier for detecting and categorizing applications vulnerable to malware.

3. Conclusion

The escalation of cyber threats presents a formidable challenge, rendering traditional security measures inadequate in tackling these ever-evolving risks. Machine learning techniques have emerged as crucial assets in enhancing cybersecurity defenses and, unfortunately, also in the arsenal of cyber attackers.

In this investigation, we conducted a comparative analysis of three learning models focused on detecting and categorizing intrusion, spam, and malware. Our assessment utilized widely recognized benchmark datasets, comparing the performance metrics of recall, precision, and accuracy.

It is crucial to underscore that recommending a single learning technique universally suitable for all cyber threat detection is not feasible. Each learning model is deployed for specific threat types, each carrying its unique strengths and advantages.

There is an urgent need to develop updated benchmark datasets that encompass the dynamic landscape of cyber threats. These datasets should cover diverse attack scenarios, addressing issues such as missing data values. Moreover, the creation of tailored learning models explicitly designed for security purposes is imperative to enhance detection capabilities.

Looking ahead, our future endeavours will delve into a broader spectrum of learning techniques dedicated to cyber threat detection. This contribution aims to fortify cybersecurity measures against the swift evolution of threats, contributing to the reinforcement of defenses for a safer digital landscape.

4. References

[1] "ICT Facts and Figures 2017." Telecommunication Development Bureau,International Telecommunication Union (ITU), Technical Report. https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx (accessed October 09, 2019).

[2] "What is Cyber-Security?" https://www.kaspersky.com.au/resourcecenter/definitions/what-is-cyber-security (accessed January 11, 2020).

[3] F. Farahmand, S. B. Navathe, P. H. Enslow, and G. P. Sharp, "Managing vulnerabilities of information systems to security incidents," in *Proceedings of the 5th international conference on Electronic commerce*, 2003: ACM, pp. 348354.

[4] P. Szor, *The Art of Computer Virus Research and Defense: ART COMP VIRUS RES DEFENSE _p1*. Pearson Education, 2005.

[5] M. Jump, "Fighting Cyberthreats with Technology Solutions," *Biomedical instrumentation & technology,* vol. 53, no. 1, pp. 38-43, 2019.

[6] N. Kostyuk and C. Wayne, "Communicating Cybersecurity: Citizen Risk Perception of Cyber Threats," 2019.

[7] A. K. Jain, D. Goel, S. Agarwal, Y. Singh, and G. Bajaj, "Predicting Spam Messages Using Back Propagation Neural Network," *Wireless Personal Communications,* vol. 110, no. 1, pp. 403-422, 2020.

[8] "Malware Types and Classifications." https://www.lastline.com/blog/malwaretypes-and-classifications/ (accessed April 18,2020).

[9] N. Sultana, N. Chilamkurti, W. Peng, and R. Alhadad, "Survey on SDN based network intrusion detection system using machine learning approaches," *Peer-toPeer Networking and Applications,* vol. 12, no. 2, pp. 493-501, 2019.

[10] M. Pradhan, C. K. Nayak, and S. K. Pradhan, "Intrusion Detection System (IDS) and Their Types," in *Securing the Internet of Things: Concepts, Methodologies, Tools, and Applications*: IGI Global, 2020, pp. 481-497.

[11] I. Firdausi, A. Erwin, and A. S. Nugroho, "Analysis of machine learning techniques used in behavior-based malware detection," in *2010 second international conference on advances in computing, control, and telecommunication technologies*, 2010: IEEE, pp. 201-203.

[12] A. V. Joshi, *Machine Learning and Artificial Intelligence*. Springer, 2020.

[13] D. Michie, D. J. Spiegelhalter, and C. Taylor, "Machine learning," *Neural and Statistical Classification,* vol. 13, 1994.

[14] K. Shaukat, A. Rubab, I. Shehzadi, and R. Iqbal, "A Socio-Technological analysis of Cyber Crime and Cyber Security in Pakistan," *Transylvanian Review,* vol. 1, no. 3, 2017.

[15] K. Shaukat, N. Masood, A. B. Shafaat, K. Jabbar, H. Shabbir, and S. Shabbir, "Dengue fever in perspective of clustering algorithms," *arXiv preprint arXiv:1511.07353,* 2015.

[16] K. Shaukat, N. Masood, S. Mehreen, and U. Azmeen, "Dengue fever prediction: A data mining problem," *Journal of Data Mining in Genomics & Proteomics,* vol. 2015, 2015.

[17] K. Shaukat, I. Nawaz, and S. Zaheer, *Students Performance: A Data Mining Perspective*. LAP Lambert Academic Publishing, 2017.

[18] K. Shaukat, I. Nawaz, S. Aslam, S. Zaheer, and U. Shaukat, "Student's performance in the context of data mining," in *2016 19th International MultiTopic Conference (INMIC)*, 2016: IEEE, pp. 1-8.

[19] S. Dey, Q. Ye, and S. Sampalli, "A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks," *Information Fusion,* vol. 49, pp. 205-215, 2019.

[20] B. Geluvaraj, P. Satwik, and T. A. Kumar, "The future of cybersecurity: Major role of artificial intelligence, machine learning, and deep learning in cyberspace," in *International Conference on Computer Networks and Communication Technologies*, 2019: Springer, pp. 739-747.

[21] A. A. Alurkar *et al.*, "A Comparative Analysis and Discussion of Email Spam Classification Methods Using Machine Learning Techniques," *Applied Machine Learning for Smart Data Analysis,* p. 185, 2019.

[22] E. G. Dada, J. S. Bassi, H. Chiroma, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems," *Heliyon,* vol. 5, no. 6, p. e01802, 2019.

[23] P. Jain, "Machine Learning versus Deep Learning for Malware Detection," 2019.

[24] P. Thiyagarajan, "A Review on Cyber Security Mechanisms Using Machine and Deep Learning Algorithms," in *Handbook of Research on Machine and Deep Learning Applications for Cyber Security*: IGI Global, 2020, pp. 23-41.

[25] S. S. Iyer and S. Rajagopal, "Applications of Machine Learning in Cyber Security Domain," in *Handbook of Research on Machine and Deep Learning Applications for Cyber Security*: IGI Global, 2020, pp. 64-82.

[26] V. Ford and A. Siraj, "Applications of machine learning in cyber security," in *Proceedings of the 27th International Conference on Computer Applications in Industry and Engineering*, 2014.

[27] H. Jiang, J. Nagra, and P. Ahammad, "Sok: Applying machine learning in security-a survey," *arXiv preprint arXiv:1611.03186,* 2016.

[28] E. Hodo, X. Bellekens, A. Hamilton, C. Tachtatzis, and R. Atkinson, "Shallow and deep networks intrusion detection system: A taxonomy and survey," *arXiv preprint arXiv:1701.02145,* 2017.

[29] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido, and M. Marchetti, "On the effectiveness of machine and deep learning for cyber security," in *2018 10th International Conference on Cyber Conflict (CyCon)*, 2018: IEEE, pp. 371-390.

[30] S. Sheikhi, M. Kheirabadi, and A. Bazzazi, "An Effective Model for SMS Spam Detection Using Content-based Features and Averaged Neural Network," *International Journal of Engineering,* vol. 33, no. 2, pp. 221-228, 2020.

[31] F. Mercaldo and A. Santone, "Deep learning for image-based mobile malware detection," *Journal of Computer Virology and Hacking Techniques,* pp. 1-15, 2020.

[32] J. Ribeiro, F. B. Saghezchi, G. Mantas, J. Rodriguez, S. J. Shepherd, and R. A. Abd-Alhameed, "An autonomous host-based intrusion detection system for android mobile devices," *Mobile Networks and Applications,* vol. 25, no. 1, pp. 164-172, 2020.

[33] C. Chen *et al.*, "A performance evaluation of machine learning-based streaming spam tweets detection," *IEEE Transactions on Computational social systems,* vol. 2, no. 3, pp. 65-76, 2015.

[34] Z. Chen, S. Liu, K. Jiang, H. Xu, and X. Cheng, "A data imputation method based on deep belief network," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2015: IEEE, pp. 1238-1243.

[35] D. M. Farid, N. Harbi, and M. Z. Rahman, "Combining naive bayes and decision tree for adaptive intrusion detection," *arXiv preprint arXiv:1005.4496,* 2010.

[36] Q. J. Ross, "C4. 5: programs for machine learning," *San Mateo, CA,* 1993.

[37] P. S. Oliveto, J. He, and X. Yao, "Time complexity of evolutionary algorithms for combinatorial optimization: A decade of results," *International Journal of Automation and Computing,* vol. 4, no. 3, pp. 281-293, 2007.

[38] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery,* vol. 2, no. 2, pp. 121-167, 1998.

[39] G. D. Forney, "The viterbi algorithm," *Proceedings of the IEEE,* vol. 61, no. 3, pp. 268-278, 1973.

[40] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Information Sciences,* vol. 340, pp. 250-261, 2016.

[41] M. S. Pervez and D. M. Farid, "Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs," in *The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)*, 2014: IEEE, pp. 1-6.

[42] "Spambase Dataset. Center for Machine Learning and Intelligent Systems at UC Irvine." https://archive.ics.uci.edu/ml/datasets/Spambase (accessed January 31, 2020).

[43] D. Gunawan, R. F. Rahmat, A. Putra, and M. F. Pasha, "Filtering Spam Text Messages by Using Twitter-LDA Algorithm," in *2018 IEEE International Conference on Communication, Networks and Satellite (Comnetsat)*, 2018: IEEE, pp. 1-6.

[44] B. Klimt and Y. Yang, "Introducing the Enron corpus," in *CEAS*, 2004.

[45] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in *2015 International Conference on Signal Processing and Communication Engineering Systems*, 2015: IEEE, pp. 92-96.

[46] A. Chahal and R. Nagpal, "Performance of Snort on Darpa Dataset and Diferent False Alert Reduction Techniques," in *3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS).*

[47] H. Kim, T. Cho, G.-J. Ahn, and J. H. Yi, "Risk assessment of mobile applications based on machine learned malware dataset," *Multimedia Tools and Applications,* vol. 77, no. 4, pp. 5027-5042, 2018.

[48] W. Awad, S. J. I. J. o. C. S. ELseuofi, and I. Technology, "Machine learning methods for spam e-mail classification," vol. 3, no. 1, pp. 173-184, 2011.

[49] R. Karthika and P. J. W. T. C. Visalakshi, "A hybrid ACO based feature selection method for email spam classification," vol. 14, pp. 171-177, 2015.

[50] G. Jain, M. Sharma, and B. J. I. J. o. K. D. i. B. Agarwal, "Spam detection on social media using semantic convolutional neural network," vol. 8, no. 1, pp. 1226, 2018.

[51] C. Chen *et al.*, "A performance evaluation of machine learning-based streaming spam tweets detection," vol. 2, no. 3, pp. 65-76, 2015.

[52] Z. Khan and U. Qamar, "Text Mining Approach to Detect Spam in Emails," in *The International Conference on Innovations in Intelligent Systems and Computing Technologies (ICIISCT2016)*, 2016, p. 45.

[53] S. A. Saab, N. Mitri, and M. Awad, "Ham or spam? A comparative study for some content-based classification algorithms for email filtering," in *MELECON 2014-2014 17th IEEE Mediterranean Electrotechnical Conference*, 2014: IEEE, pp. 339-343.

[54] Y. Zhang, S. Wang, P. Phillips, and G. J. K.-B. S. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," vol. 64, pp. 22-31, 2014.

[55] A. Tyagi, "Content Based Spam Classification-A Deep Learning Approach," University of Calgary, 2016.

[56] I. J. Alkaht and B. J. I. R. C. S. Al-Khatib, "Filtering SPAM Using Several Stages Neural Networks," vol. 11, p. 2, 2016.

[57] G. Tzortzis and A. Likas, "Deep belief networks for spam filtering," in *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, 2007, vol. 2: IEEE, pp. 306-309.

[58] Y. Rizk, N. Hajj, N. Mitri, M. J. A. C. Awad, and Informatics, "Deep belief networks and cortical algorithms: A comparative study for supervised classification," 2018.

[59] A. Sharaff, N. K. Nagwani, and A. Dhadse, "Comparative study of classification algorithms for spam email detection," in *Emerging research in computing, information, communication and applications*: Springer, 2016, pp. 237-244.

[60] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *Ieee Access,* vol. 5, pp. 21954-21961, 2017.

[61] M. Z. Alom, V. Bontupalli, and T. M. Taha, "Intrusion detection using deep belief networks," in *2015 National Aerospace and Electronics Conference (NAECON)*, 2015: IEEE, pp. 339-344.

[62] S. Jo, H. Sung, and B. Ahn, "A comparative study on the performance of intrusion detection using decision tree and artificial neural network models," *Journal of the Korea Society of Digital Industry and Information Management,* vol. 11, no. 4, pp. 33-45, 2015.

[63] J. Lee, J. Kim, I. Kim, and K. Han, "Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles," *IEEE Access,* vol. 7, pp. 165607165626, 2019.

[64] L. Khan, M. Awad, and B. Thuraisingham, "A new intrusion detection system using support vector machines and hierarchical clustering," *The VLDB journal,* vol. 16, no. 4, pp. 507-521, 2007.

[65] R. Kokila, S. T. Selvi, and K. Govindarajan, "DDoS detection and analysis in SDN-based environment using support vector machine classifier," in *2014 Sixth International Conference on Advanced Computing (ICoAC)*, 2014: IEEE, pp. 205-210.

[66] P. Mishra, V. Varadharajan, U. Tupakula, E. S. J. I. C. S. Pilli, and Tutorials, "A detailed investigation and analysis of using machine learning techniques for intrusion detection," vol. 21, no. 1, pp. 686-728, 2018.

[67] J. Kevric, S. Jukic, A. J. N. C. Subasi, and Applications, "An effective combining classifier approach using tree algorithms for network intrusion detection," vol. 28, no. 1, pp. 1051-1058, 2017.

[68] A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour, and H. Janicke, "A novel hierarchical intrusion detection system based on decision tree and rules-based models," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2019: IEEE, pp. 228-233.

[69] B. Ingre, A. Yadav, and A. K. Soni, "Decision tree based intrusion detection system for NSL-KDD dataset," in *International Conference on Information and Communication Technology for Intelligent Systems*, 2017: Springer, pp. 207-218.

[70] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of deep learning-based network anomaly detection," *Cluster Computing,* pp. 1-13, 2017.

[71] Y. Cheng, W. Fan, W. Huang, and J. An, "A Shellcode Detection Method Based on Full Native API Sequence and Support Vector Machine," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 242, no. 1: IOP Publishing, p. 012124.

[72] A. Mohaisen and O. Alrawi, "Unveiling zeus: automated classification of malware samples," in *Proceedings of the 22nd International Conference on World Wide Web*, 2013: ACM, pp. 829-832.

[73] P. Shijo and A. J. P. C. S. Salim, "Integrated static and dynamic analysis for malware detection," vol. 46, pp. 804-811, 2015.

[74] Q. Jamil and M. A. Shah, "Analysis of machine learning solutions to detect malware in android," in *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, 2016: IEEE, pp. 226-232.

[75] D. Moon, H. Im, I. Kim, and J. H. Park, "DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks," *The Journal of supercomputing,* vol. 73, no. 7, pp. 2881-2895, 2017.

[76] Z. Salehi, A. Sami, M. J. C. F. Ghiasi, and Security, "Using feature generation from API calls for malware detection," vol. 2014, no. 9, pp. 9-18, 2014.

[77] Z. Yuan, Y. Lu, Y. J. T. S. Xue, and Technology, "Droiddetector: android malware characterization and detection using deep learning," vol. 21, no. 1, pp. 114-123, 2016.

[78] Y. Li, R. Ma, R. J. I. J. o. S. Jiao, and I. Applications, "A hybrid malicious code detection method based on deep learning," vol. 9, no. 5, pp. 205-216, 2015.

[79] R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, and S. Venkatraman, "Robust intelligent malware detection using deep learning," *IEEE Access,* vol. 7, pp. 46717-46738, 2019.

[80] Damodaran, F. Di Troia, C. A. Visaggio, T. H. Austin, and M. Stamp, "A comparison of static, dynamic, and hybrid analysis for malware detection," Journal of Computer Virology and Hacking Techniques, vol. 13, no. 1, pp. 1-12, 2017.