Cybersecurity

Subjects: Computer Science, Information Systems

Contributor: Kamran Shaukat , Suhuai Luo , Vijay Varadharajan , Ibrahim A. Hameed , Shan Chen , Dongxi Liu , Jiaming Li

Cyberspace has become an indispensable factor for all areas of the modern world. The world is becoming more and more dependent on the internet for everyday living. The increasing dependency on the internet has also widened the risks of malicious threats. On account of growing cybersecurity risks, cybersecurity has become the most pivotal element in the cyber world to battle against all cyber threats, attacks, and frauds. The expanding cyberspace is highly exposed to the intensifying possibility of being attacked by interminable cyber threats. The objective of this survey is to bestow a brief review of different machine learning (ML) techniques to get to the bottom of all the developments made in detection methods for potential cybersecurity risks. These cybersecurity risk detection methods mainly comprise of fraud detection, intrusion detection, spam detection, and malware detection. In this review paper, we build upon the existing literature of applications of ML models in cybersecurity and provide a comprehensive review of ML techniques in cybersecurity. To the best of our knowledge, we have made the first attempt to give a comparison of the time complexity of commonly used ML models in cybersecurity. We have comprehensively compared each classifier's performance based on frequently used datasets and sub-domains of cyber threats. This work also provides a brief introduction of machine learning models besides commonly used security datasets. Despite having all the primary precedence, cybersecurity has its constraints compromises, and challenges. This work also expounds on the enormous current challenges and limitations faced during the application of machine learning techniques in cybersecurity.

cybersecurity machine learning malware detection intrusion detection system spam classification

1. Performance Comparison of Machine Learning Models Applied in Cybersecurity

Researchers are investigating machine learning techniques to detect different cybercrimes in cybersecurity. We have provided a detailed discussion of various cyber threats in Section 2. Furthermore, we have briefly presented an overview of frequently used security datasets in Section 2. This section provides a comprehensive survey of each ML model applied to deal with different cyber threats. Subsequent lines will explain the description of each column in Table 1, Table 2, Table 3, Table 4, Table 5 and Table 6. The ML technique columns describe the considered machine learning model. We have considered six ML models for this study: random forest, support vector machine, naïve Bayes, decision tree, artificial neural network, and deep belief network.

Table 1. Evaluation of SVM in Cybersecurity.

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
SVM	IDS	NSL-KDD	[1]	2019	Anomaly-Based	89.70%		
			[2]	2016	Anomaly-Based	98.89%	-	
			[3]	2014	Hybrid-Based	82.37%	74%	82%
		DARPA	[4]	2007	Hybrid-Based	69.80%		
			[5]	2014	Anomaly-Based	95.11%		-
		KDD CUP99	[6]	2011	Hybrid-Based	95.72%		
			[7]	2015	Hybrid-Based	96.08%	-	
			[8]	2014	Hybrid-Based	99.30%	-	
	Malware	Custom Dataset	[9]	2019	Static	95.17%	95.57%	95%
			[10]	2018	Static	89.91%	88.84%	
			[11]	2018	Dynamic	96.27%	96.16%	93.71%
		Malware Dataset	[12]	2017	Static	94.37%		
			[13]	2013	Dynamic	95%		
			[14]	2015	Dynamic	97.10%		
		Enron	[15]	2016	Static	91%	84.74%	100%
			[16]	2007	Static	96.92%	92.74%	97.27%
	Spam	SMS Collection	[17]	2014	SMS Spam	98.61%	98.60%	98.60%
		Spambase	[18]	2015	Email Spam	79.50%	79.02%	68.67%
			[19]	2011	Email Spam	96.90%	93.12%	95%
		Twitter Dataset	[20]	2018	Spam Tweets	93.14%	92.91%	93.14%
			[21]	2015	Spam Tweets	95.20%		93.60%
			[22]	2020	Spam Tweets	98.88%		94.47%

Table 2. Evaluation of Decision Tree in Cybersecurity.

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
Decision Tree	IDS	KDD	[23]	2018	Misuse-Based	99.96%		
			[24]	2005	Hybrid-Based	99.85%	99.70%	98.10%
			[25]	2017	Hybrid-Based	86.29%		78%
		NSL-KDD	[26]	2014	Anomaly-Based	99.64%		
			[27]	2017	Hybrid-Based	90.30%	91.15%	90.31%
			[28]	2019	Hybrid-Based	93.40%		
		KDD CUP99	[29]	2015	Misuse-Based	95.09%		
			[30]	2016	Hybrid-Based	99.62%		
			[31]	2018	Hybrid-Based	92.87%	99.90%	
	Malware	Custom	[32]	2016	Static	99.90%	99.40%	
			[33]	2017	Static	84.7%		
		Malware Dataset	[34]	2014	Static		97.90%	96.70%
			[35]	2013	Static	92.34%	-	93%
			[36]	2013	Dynamic	88.47%		
		SMOTE	[37]	2018	Dynamic	92.82%		
			[37]	2018	Dynamic	95.75%		
			[38]	2012	Static	96.62%		
	Spam	SMS Collection	[17]	2014	SMS Spam	96.60%	96.50%	96.60%
		Enron	[15]	2016	Email Spam	96%	98%	94%
			[15]	2016	Email Spam	96%	98%	94%
		Spambase	[39]	2014	Email Spam	92.08%	91.51%	88.08%
			[40]	2014	Email Spam	94.27%		91.02%
			[41]	2013	Email Spam	92.34%	93.90%	93.50%

Table 3. Evaluation of DBN in Cybersecurity.

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
DBN	IDS	KDD	[42]	2015	Anomaly-Based	97.50%		
			[43]	2015	Hybrid-Based	96.70%	97.90%	
		NSL-KDD	[44]	2017	Anomaly-Based	90.40%	88.60%	95.30%
			[45]	2019	Anomaly-Based	99.45%	99.20%	99.70%
		ISCX Dataset	[46]	2015	Misuse-Based	99.18%	-	-
	Malware	DLL	[47]	2008	Static	89.90%	87.40%	98.80%
		Custom	[48]	2016	Static	89.03%	83%	98.18%
			[48]	2016	Dynamic	71%	78.08%	59.09%
			[48]	2016	Hybrid	96.76%	95.77%	97.84%
		KDD CUP99	[49]	2015	Hybrid	91.40%	-	95.34%
	Spam	TARASSUL	[50]	2016	Email Spam	96.40%	95.31%	93.59%
			[50]	2016	Email Spam	97.50%	98.39%	98.02%
		Enron	[45]	2016	Email Spam	95.86%	96.49%	95.61%
			[16]	2007	Email Spam	97.43%	94.94%	96.47%
		Spambase	[51]	2018	Email Spam	89.20%	96%	
			[51]	2018	Email Spam	90.69%	97%	

Table 4. Evaluation of ANN in Cybersecurity.

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
ANN	IDS	NSL-KDD	[52]	2019	Anomaly-Based	94.50%	-	-
			[53]	2014	Anomaly-Based	97.53%	-	-
			[4]	2014	Hybrid-Based	97.06%	-	-
		DARPA	[54]	2015	Anomaly-Based	80%	-	80%
			[23]	2018	Misuse-Based	99.82%	-	-

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
	Malware	KDD CUP99	[55]	2009	Anomaly-Based	-	97.89%	98.94%
			[56]	2012	Anomaly-Based	62.90%	-	-
		VX Heavens	[57]	2012	Hybrid	88.89%	88.89%	-
			[58]	2012	Static	92.19%	-	-
			[59]	2013	Static	88.31%	-	-
		Enron	[60]	2018	Dynamic	82.79%	-	-
	Spam	Comodo	[61]	2016	Static	92.02%	-	-
		Spam-Archive	[62]	2011	Image Spam	93.70%	87%	94%
			[63]	2016	Email Spam	91%	-	-
		Spambase	[64]	2018	Email Spam	92.41%	92.40%	92.40%
			[65]	2013	Hybrid	93.71%	95%	-
		Twitter Dataset	[20]	2018	Spam Tweets	91.18%	91.80%	91.18%

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
Random Forest	IDS	KDD	[66]	2019	Anomaly-Based	99.95%		99.95%
			[67]	2016	Anomaly-Based	88.65%	-	94.62%
		NSL-KDD	[68]	2019	Anomaly-Based	95.10%	92.50%	
			[69]	2019	Hybrid-Based	75.30%	81.40%	75.30%
			[70]	2017	Hybrid-Based	97.10%		
		KDD CUP99	[68]	2019	Anomaly-Based	96.30%	99.80%	
			[71]	2016	Anomaly-Based	-	98.10%	98.10%
			[70]	2017	Hybrid-Based	98.10%	-	-
	Malware	Custom Dataset	[9]	2019	Static	98.63%	98.58%	98.69%
			[11]	2018	Dynamic	96.34%	96.59%	93.46%
		Malware Dataset	[72]	2016	Dynamic	96.14%		
			[34]	2014	Hybrid		96.50%	97.30%

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
			[73]	2017	Hybrid	91.40%	89.80%	91.10%
		VirusShare	[74]	2009	Static	95.60%	96%	
		SMS Collection	[17]	2014	SMS Spam	97.18%	97.30%	97.20%
			[75]	2013	Email Spam	99.54%		
		Spambase	[76]	2010	Email Spam	95.43%		
	Spam		[41]	2013	Email Spam	93.89%	95.87%	94.10%
			[77]	2011	Spam Tweets	95%	95.70%	95.70%
		Twitter Dataset	[78]	2016	Spam Tweets	96.20%	98.60%	75.50%
			[20]	2018	Spam Tweets	93.43%	93.25%	93.43%

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
Naïve Bayes		DARPA	[79]	2010	Anomaly-Based	91.60%		61.60%
			[80]	2007	Misuse-Based	99.90%	99.04%	99.50%
	IDS	NSL-KDD	[29]	2015	Misuse-Based	81.66%		
			[81]	2012	Anomaly-Based	36%	35%	80%
			[81]	2012	Anomaly-Based	99%	83%	78.90%
		KDD CUP99	[82]	2004	Anomaly-Based	99.27%		
			[80]	2007	Anomaly-Based		96%	99.80%
			[79]	2018	Signature-Based	99.72%		100%
	Malware	VX Heaven	[83]	2015	Static	88.80%		
		NSL-KDD	[84]	2013	Hybrid	99.50%		
			[85]	2007	Hybrid	99%		
		Malware Dataset	[35]	2013	Hybrid	89.81%	-	90%
			[86]	2015	Hybrid	95.90%	95.90%	95.90%
			[34]	2014	Hybrid		97.50%	67.40%

ML Technique	Domain	Dataset	Reference	Year	Approach/Domain	Results		
						Accuracy	Precision	Recall
	Spam	SMS Collection	[17]	2014	SMS Spam	97.52%	97.50%	97.50%
		Spambase	[19]	2011	Email Spam	99.46%	99.66%	98.46%
			[18]	2015	Email Spam	76.24%	70.59%	72.05%
			[87]	2015	Email Spam	84%	89%	78%
		Twitter Dataset	[41]	2013	Spam Tweets	92%	91.60%	91.4%
			[20]	2018	Spam Tweets	92.06%	91.69%	91.96%

domain column. The year columns depict the citation number of each article and published year, respectively. The values of approach or sub-domain columns are different for each cyber threat. IDS domain has three values that are anomaly-based, signature/misuse-based and hybrid-based. Malware has three further sub-classifications that are static, dynamic and hybrid. In the case of spam, sub-domains correspond to the medium in which the authors tried to identify the spam such as image, video, email, SMS and tweets. A description of each sub-domain/approach has been provided in Section 2. Finally, the result attribute presents the evaluation of each classifier applied in a particular sub-domain of cyber threat on a specific dataset and provided in the cited paper mentioned in the reference column.

2. Support Vector Machine

The principle superiority of support vector machine (SVM) is that it produces the most successful results for cybersecurity tasks. SVM distributes each data class on both sides of the hyperplane. SVM separates the classes based on the notation to the margin. Support vector points are those points that lie on the border of the hyperplane. The major drawback of the support vector machine is that it consumes an immense amount of space and time. SVM requires data trained on different time intervals to produce better results for a dynamic dataset [88].

SVM showed an accuracy of 99.30% with KDD Cup 99 dataset for IDS [8]. 96.92% is the best reported accuracy for malware detection using Enron dataset [16] and 96.90% with Spambase to classify spam emails [19]. The best reported recall for SVM to detect intrusion is 82% [3], malware is 100% [15], and spam is 98.60% [17]. SVM has obtained best precision while detecting the intrusion is 74% [24], malware is 96.16% [11], and spam is 98.60% [17]. A detailed performance comparison of SVM to various cyber threats on the frequently used dataset is presented in Table 1.

3. Decision Tree

Decision tree (DT) belongs to the category of supervised machine learning. DT consists of a path and two nodes: root/intermediate and leaf. Root or intermediate node presents an attribute that followed a path that corresponds to the possible value of an attribute. Leaf node represents the final decision/classification class. A decision tree is

used to find the best immediate node by following the if-then rule [89]. Further, 99.96% is the reported accuracy of DT while detecting the anomaly-based IDS with KDD dataset [23]. With standard SMOTE dataset, DT shows an outstanding accuracy of 96.62% for malware detection [38]. With the Enron dataset, DT correctly classified ham emails with an accuracy of 96% [15]. The best reported recall for DT to detect intrusion is 98.10% [24], malware is 96.70% [34], and spam is 96.60% [17]. DT has obtained best precision while detecting the intrusion is 99.70% [24], malware is 99.40% [32], and spam is 98% [15]. A detailed performance comparison of decision tree to various cyber threats on the frequently used dataset is presented in Table 2.

4. Deep Belief Network

A deep belief network (DBN) consists of various middle layers of restricted Boltzmann machine (RBM) organized greedily. Every layer communicates with the layers behind it and the layers ahead of it. There is no lateral communication between the nodes within a layer. Every layer serves as both an input layer and an output layer, except the first and the last layers. The last layer functions as a classifier. The primary purpose of a deep belief network is image clustering and image recognition. It deals with motion capture data. Deep belief network has shown the accuracy of 97.50% for IDS [42], 91.40% for malware detection [90] and 97.43% for spam detection [91] with KDD, KDD CUP99, and Spambase datasets, respectively. The best reported recall for DBN to detect intrusion is 99.70% [45], malware is 98.80% [47], and spam is 98.02% [50]. DBN obtained the best precision while detecting the intrusion is 99.20% [45], malware is 95.77% [48], and spam is 98.39% [50]. A detailed performance comparison of DBN to various cyber threats on the frequently used dataset is presented in Table 3.

5. Artificial Neural Network

An artificial neural network (ANN) classier consists of hidden neuron input and output layers and performs in two stages. The first stage is called feedforward. In this stage, each hidden layer receives some input nodes and based on the input layer and activation function, the error is calculated. In the second stage, namely feedback stage, the error is sent back to the input layer and process is continued in iterations until the correct result is gained [60]. The artificial neural network showed an accuracy of 97.53% for IDS [53], 92.19% for malware detection [58], and 92.41% for spam detection with NSL-KDD, VX Heavens, and Spambase datasets, respectively. The best reported recall for ANN to detect an intrusion is 98.94% [55], and spam is 94% [62]. ANN has obtained best precision while detecting the intrusion is 97.89% [55], malware is 88.89% [57], and spam is 95% [65]. A detailed performance comparison of ANN to various cyber threats on the frequently used dataset is presented in Table 7.

6. Random Forest

Random forest (RF) follows through the task by combing different predictions generated by joining different decision trees. RF raised a hypothesis to obtain a result [91]. RF falls under the category of ensemble learning. RF also termed as random decision forest. RF is considered as an improved version of CART that is a sub-type of a decision tree.

RF has shown an accuracy of 99.95% with IDS [66], 95.60% with malware detection [74] and 99.54% for spam detection [75] with KDD, VirusShare, and Spambase datasets, respectively. The best reported recall for RF to detect intrusion is 99.95% [66], malware is 97.30% [34], and spam is 97.20% [17]. RF obtained the best precision while detecting the intrusion is 99.80% [68], malware is 98.58% [9], and spam is 98.60% [78]. A detailed performance comparison of RF to various cyber threats on the frequently used dataset is presented in Table 5.

7. Naïve Bayes

The major limitation for Naïve Bayes (NB) classifier is that it assumes that every attribute is independent, and none of the attributes has a relationship with each other. This state of independence is technically impossible in cyberspace. Hidden NB is an advanced form of Naïve Bayes, and it gives 99.6% accuracy [92]. Naïve Bayes showed an accuracy of 99.90% with DARPA dataset for IDS [80]. 99.50% is the best reported accuracy for malware detection using NSL-KDD dataset [86]. With Spambase dataset, Naïve Bayes showed considerable accuracy of 96.46 % to classify spam or ham email [19]. The best reported recall for NB to detect intrusion is 100% [79], malware is 95.90% [86], and spam is 98.46% [19]. NB obtained the best precision while detecting the intrusion is 99.04% [80], malware is 97.50% [34], and spam is 99.66% [19]. A detailed performance comparison of NB to various cyber threats on the frequently used dataset is presented in Table 6.

References

1. Lee, J.; Kim, J.; Kim, I.; Han, K. Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles. IEEE Access 2019, 7, 165607–165626.
2. Sharma, R.K.; Kalita, H.K.; Borah, P. Analysis of machine learning techniques based intrusion detection systems. In Proceedings of the 3rd International Conference on Advanced Computing, Networking and Informatics; Springer: Berlin/Heidelberg, Germany, 2015; pp. 485–493.
3. Pervez, M.S.; Farid, D.M. Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs. In Proceedings of the 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014), Dhaka, Bangladesh, 18–20 December 2014; pp. 1–6.
4. Khan, L.; Awad, M.; Thuraisingham, B. A new intrusion detection system using support vector machines and hierarchical clustering. VLDB J. 2007, 16, 507–521.
5. Kokila, R.; Selvi, S.T.; Govindarajan, K. DDoS detection and analysis in SDN-based environment using support vector machine classifier. In Proceedings of the 2014 Sixth International Conference on Advanced Computing (ICoAC), Chennai, India, 17–19 Decmber 2014; pp. 205–210.

6. Horng, S.-J.; Su, M.-Y.; Chen, Y.-H.; Kao, T.-W.; Chen, R.-J.; Lai, J.-L.; Perkasa, C.D. A novel intrusion detection system based on hierarchical clustering and support vector machines. Expert Syst. Appl. 2011, 38, 306–313.
7. Masduki, B.W.; Ramli, K.; Saputra, F.A.; Sugiarto, D. Study on implementation of machine learning methods combination for improving attacks detection accuracy on Intrusion Detection System (IDS). In Proceedings of the 2015 International Conference on Quality in Research (QiR), Lombok, Indonesia, 10–13 August 2015; pp. 56–64.
8. Saxena, H.; Richariya, V. Intrusion detection in KDD99 dataset using SVM-PSO and feature reduction with information gain. Int. J. Comput. Appl. 2014, 98, 25–29.
9. Naz, S.; Singh, D.K. Review of Machine Learning Methods for Windows Malware Detection. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019; pp. 1–6.
10. Zhu, H.-J.; Jiang, T.-H.; Ma, B.; You, Z.-H.; Shi, W.-L.; Cheng, L.J.N.C. HEMD: A highly efficient random forest-based malware detection framework for Android. Neural Comput. Appl. 2018, 30, 3353–3361.
11. Feng, P.; Ma, J.; Sun, C.; Xu, X.; Ma, Y.J.I.A. A Novel Dynamic Android Malware Detection System With Ensemble Learning. IEEE Access 2018, 6, 30996–31011.
12. Cheng, Y.; Fan, W.; Huang, W.; An, J. A Shellcode Detection Method Based on Full Native API Sequence and Support Vector Machine. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Sanya, China, 12–15 November 2019; p. 012124.
13. Mohaisen, A.; Alrawi, O. Unveiling zeus: Automated classification of malware samples. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 829–832.
14. Shijo, P.; Salim, A.J.P.C.S. Integrated static and dynamic analysis for malware detection. Procedia Comput. Sci. 2015, 46, 804–811.
15. Khan, Z.; Qamar, U. Text Mining Approach to Detect Spam in Emails. In Proceedings of the International Conference on Innovations in Intelligent Systems and Computing Technologies (ICIISCT2016), Las Piñas, Philippines, 24–26 February 2016; p. 45.
16. Tzortzis, G.; Likas, A. Deep belief networks for spam filtering. In Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), Patras, Greece, 29–31 October 2007; pp. 306–309.
17. Najadat, H.; Abdulla, N.; Abooraig, R.; Nawasrah, S. Mobile sms spam filtering based on mixing classifiers. Int. J. Adv. Comput. Res. 2014, 1, 1–7.

18. Karthika, R.; Visalakshi, P.J.W.T.C. A hybrid ACO based feature selection method for email spam classification. WSEAS Trans. Comput. 2015, 14, 171–177.
19. Awad, W.; ELseuofi, S. Machine learning methods for spam e-mail classification. Int. J. Comput. Sci. Inf. Technol. 2011, 3, 173–184.
20. Jain, G.; Sharma, M.; Agarwal, B. Spam detection on social media using semantic convolutional neural network. Int. J. Knowl. Discov. Bioinform. 2018, 8, 12–26.
21. Chen, C.; Zhang, J.; Xie, Y.; Xiang, Y.; Zhou, W.; Hassan, M.M.; AlElaiwi, A.; Alrubaian, M. A performance evaluation of machine learning-based streaming spam tweets detection. IEEE Trans. Comput. Soc. Syst. 2015, 2, 65–76.
22. Sagar, R.; Jhaveri, R.; Borrego, C.J.E. Applications in Security and Evasions in Machine Learning: A Survey. Electronics 2020, 9, 97.
23. Mishra, P.; Varadharajan, V.; Tupakula, U.; Pilli, E.S. Tutorials. A detailed investigation and analysis of using machine learning techniques for intrusion detection. IEEE Commun. Surv. Tutor. 2018, 21, 686–728.
24. Stein, G.; Chen, B.; Wu, A.S.; Hua, K.A. Decision tree classifier for network intrusion detection with GA-based feature selection. In Proceedings of the 43rd Annual Southeast Regional Conference-Volume 2; ACM: New York, NY, USA, 2005; pp. 136–141.
25. Kevric, J.; Jukic, S.; Subasi, A.J.N.C. An effective combining classifier approach using tree algorithms for network intrusion detection. Applications 2017, 28, 1051–1058.
26. Gaikwad, D.; Thool, R.C. Intrusion detection system using ripple down rule learner and genetic algorithm. Int. J. Comput. Sci. Inf. Technol. 2014, 5, 6976–6980.
27. Ingre, B.; Yadav, A.; Soni, A.K. Decision tree based intrusion detection system for NSL-KDD dataset. In Proceedings of the International Conference on Information and Communication Technology for Intelligent Systems, Ahmedabad, India, 15–16 May 2020; pp. 207–218.
28. Ahmim, A.; Maglaras, L.; Ferrag, M.A.; Derdour, M.; Janicke, H. A novel hierarchical intrusion detection system based on decision tree and rules-based models. In Proceedings of the 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini Island, Greece, 29–31 May 2019; pp. 228–233.
29. Relan, N.G.; Patil, D.R. Implementation of network intrusion detection system using variant of decision tree algorithm. In Proceedings of the 2015 International Conference on Nascent Technologies in the Engineering Field (ICNTE), Navi Mumbai, India, 9–10 January 2015; pp. 1–5.
30. Goeschel, K. Reducing false positives in intrusion detection systems using data-mining techniques utilizing support vector machines, decision trees, and naive Bayes for off-line analysis. In Proceedings of the SoutheastCon 2016, Norfolk, VA, USA, 30 March–3 April 2016; pp. 1–6.

31. Malik, A.J.; Khan, F.A. A hybrid technique using binary particle swarm optimization and decision tree pruning for network intrusion detection. Clust. Comput. 2018, 21, 667–680.
32. Jamil, Q.; Shah, M.A. Analysis of machine learning solutions to detect malware in android. In Proceedings of the 2016 Sixth International Conference on Innovative Computing Technology (INTECH), Dublin, Ireland, 24–26 August 2016; pp. 226–232.
33. Moon, D.; Im, H.; Kim, I.; Park, J.H. DTB-IDS: An intrusion detection system based on decision tree using behavior analysis for preventing APT attacks. J. Supercomput. 2017, 73, 2881–2895.
34. Salehi, Z.; Sami, A.; Ghiasi, M.J.C.F. Using feature generation from API calls for malware detection. Security 2014, 2014, 9–18.
35. Santos, I.; Brezo, F.; Ugarte-Pedrero, X.; Bringas, P.G. Opcode sequences as representation of executables for data-mining-based unknown malware detection. Inf. Sci. 2013, 231, 64–82.
36. Islam, R.; Tian, R.; Batten, L.M.; Versteeg, S. Classification of malware based on integrated static and dynamic features. J. Netw. Comput. Appl. 2013, 36, 646–656.
37. Yan, P.; Yan, Z. A survey on dynamic mobile malware detection. Softw. Qual. J. 2018, 26, 891–919.
38. Kavzoglu, T.; Colkesen, I. The effects of training set size for performance of support vector machines and decision trees. In Proceedings of the 10th international symposium on spatial accuracy assessment in natural resources and environmental sciences, Florianópolis, Brazil, 10–13 July 2012; p. 1013.
39. Saab, S.A.; Mitri, N.; Awad, M. Ham or spam? A comparative study for some content-based classification algorithms for email filtering. In Proceedings of the MELECON 2014-2014 17th IEEE Mediterranean Electrotechnical Conference, Beirut, Lebanon, 13–16 April 2014; pp. 339–343.
40. Zhang, Y.; Wang, S.; Phillips, P.; Ji, G. Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. Knowl. -Based Syst. 2014, 64, 22–31.
41. Sharma, S.; Arora, A. Adaptive approach for spam detection. Int. J. Comput. Sci. Issues 2013, 10, 23.
42. Alom, M.Z.; Bontupalli, V.; Taha, T.M. Intrusion detection using deep belief networks. In Proceedings of the 2015 National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 June 2015; pp. 339–344.
43. Jo, S.; Sung, H.; Ahn, B. A comparative study on the performance of intrusion detection using decision tree and artificial neural network models. J. Korea Soc. Digit. Ind. Inf. Manag. 2015, 11, 33–45.
44. Kwon, D.; Kim, H.; Kim, J.; Suh, S.C.; Kim, I.; Kim, K.J. A survey of deep learning-based network anomaly detection. Clust. Comput. 2019, 22, 949–961.

45. Zhang, Y.; Li, P.; Wang, X.J.I.A. Intrusion detection for IoT based on improved genetic algorithm and deep belief network. IEEE Access 2019, 7, 31711–31722.
46. Ammar, A. A decision tree classifier for intrusion detection priority tagging. J. Comput. Commun. 2015, 3, 52.
47. Ye, Y.; Wang, D.; Li, T.; Ye, D.; Jiang, Q. An intelligent PE-malware detection system based on association mining. J. Comput. Virol. 2008, 4, 323–334.
48. Yuan, Z.; Lu, Y.; Xue, Y. Droiddetector: Android malware characterization and detection using deep learning. Tsinghua Sci. Technol. 2016, 21, 114–123.
49. Li, Y.; Ma, R.; Jiao, R. A hybrid malicious code detection method based on deep learning. J. Secur. Appl. 2015, 9, 205–216.
50. Alkaht, I.J.; Al-Khatib, B. Filtering SPAM Using Several Stages Neural Networks. Int. Rev. Comp. Softw. 2016, 11, 2.
51. Rizk, Y.; Hajj, N.; Mitri, N.; Awad, M. Deep belief networks and cortical algorithms: A comparative study for supervised classification. Appl. Comput. Inform. 2019, 15, 81–93.
52. Qureshi, A.-U.-H.; Larijani, H.; Mtetwa, N.; Javed, A.; Ahmad, J.J.C. RNN-ABC: A New Swarm Optimization Based Technique for Anomaly Detection. Computers 2019, 8, 59.
53. Shrivas, A.K.; Dewangan, A.K. An ensemble model for classification of attacks with feature selection based on KDD99 and NSL-KDD data set. Int. J. Comput. Appl. 2014, 99, 8–13.
54. Buczak, A.L.; Guven, E. A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun. Surv. Tutor. 2015, 18, 1153–1176.
55. Ahmad, I.; Abdullah, A.B.; Alghamdi, A.S. Artificial neural network approaches to intrusion detection: A review. In Proceedings of the 8th Wseas International Conference on Telecommunications and Informatics, Istanbul, Turkey, 30 May–1 June 2009.
56. Sheikhan, M.; Jadidi, Z.; Farrokhi, A. Intrusion detection using reduced-size RNN based on feature grouping. Neural Comput. Appl. 2012, 21, 1185–1190.
57. Chen, Y.; Narayanan, A.; Pang, S.; Tao, B. Multiple sequence alignment and artificial neural networks for malicious software detection. In Proceedings of the 2012 8th International Conference on Natural Computation, Chongqing, China, 29–31 May 2012; pp. 261–265.
58. Shabtai, A.; Moskovitch, R.; Feher, C.; Dolev, S.; Elovici, Y.J.S.I. Detecting unknown malicious code by applying classification techniques on opcode patterns. Secur. Inform. 2012, 1, 1.
59. Liangboonprakong, C.; Sornil, O. Classification of malware families based on n-grams sequential pattern features. In Proceedings of the 2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), Melbourne, Australia, 19–21 June 2013; pp. 777–782.

60. Phan, T.D.; Zincir-Heywood, N. User identification via neural network based language models. Int. J. Netw. Manag. 2019, 29, e2049.
61. Hardy, W.; Chen, L.; Hou, S.; Ye, Y.; Li, X. DL4MD: A deep learning framework for intelligent malware detection. In Proceedings of the International Conference on Data Mining (DMIN), Las Vegas, NV, USA, 12–15 July 2010; p. 61.
62. Soranamageswari, M.; Meena, C. A novel approach towards image spam classification. Int. J. Comput. Theory Eng. 2011, 3, 84.
63. Foqaha, M.A.M. Email spam classification using hybrid approach of RBF neural network and particle swarm optimization. Int. J. Netw. Secur. Appl. 2016, 8, 17–28.
64. Bassiouni, M.; Ali, M.; El-Dahshan, E.A. Ham and Spam E-Mails Classification Using Machine Learning Techniques. J. Appl. Secur. Res. 2018, 13, 315–331.
65. Arram, A.; Mousa, H.; Zainal, A. Spam detection using hybrid Artificial Neural Network and Genetic algorithm. In Proceedings of the 2013 13th International Conference on Intellient Systems Design and Applications, Salangor, Malaysia, 8–10 December 2013; pp. 336–340.
66. Gao, Y.; Wu, H.; Song, B.; Jin, Y.; Luo, X.; Zeng, X.J.I.A. A Distributed Network Intrusion Detection System for Distributed Denial of Service Attacks in Vehicular Ad Hoc Network. IEEE Access 2019, 7, 154560–154571.
67. Gupta, G.P.; Kulariya, M. A framework for fast and efficient cyber security network intrusion detection using apache spark. Procedia Comput. Sci. 2016, 93, 824–831.
68. Zhou, Y.-Y.; Cheng, G. An Efficient Network Intrusion Detection System Based on Feature Selection and Ensemble Classifier. arXiv 2019, arXiv:1904.01352.
69. Vinayakumar, R.; Alazab, M.; Soman, K.; Poornachandran, P.; Al-Nemrat, A.; Venkatraman, S.J.I.A. Deep Learning Approach for Intelligent Intrusion Detection System. IEEE Access 2019, 7, 41525–41550.
70. Prakash Chandra, P.U.K.; Lilhore, P.N.A. Network intrusion detection system based on modified Random forest classifiers for kdd cup-99 and nsl-kdd Dataset. Int. Res. J. Eng. Technol. 2017, 4, 786–791.
71. Vivek Nandan Tiwari, P.S.R. Enhanced Method for Intrusion Detection over KDD Cup 99 Dataset. Int. J. Curr. Trends Eng. Technol. 2016, 2, 218–224.
72. Galal, H.S.; Mahdy, Y.B.; Atiea, M.A. Behavior-based features model for malware detection. J. Comput. Virol. Hacking Tech. 2016, 12, 59–67.
73. Mosli, R.; Li, R.; Yuan, B.; Pan, Y. A behavior-based approach for malware detection. In Proceedings of the IFIP International Conference on Digital Forensics, Orlando, FL, USA, 30 January–1 February 2017.

74. Siddiqui, M.; Wang, M.C.; Lee, J. Detecting internet worms using data mining techniques. J. Syst. Cybern. Inform. 2009, 6, 48–53.
75. Rathi, M.; Pareek, V. Spam mail detection through data mining-A comparative performance analysis. Int. J. Mod. Educ. Comput. Sci. 2013, 5, 31.
76. Lee, S.M.; Kim, D.S.; Kim, J.H.; Park, J.S. Spam detection using feature selection and parameters optimization. In Proceedings of the 2010 International Conference on Complex, Intelligent and Software Intensive Systems, Krakow, Poland, 15–18 February 2010; pp. 883–888.
77. Mccord, M.; Chuah, M. Spam detection on twitter using traditional classifiers. In Proceedings of the International Conference on Autonomic and Trusted Computing, Banff, AB, Canada, 2–4 September 2011; pp. 175–186.
78. Xu, H.; Sun, W.; Javaid, A. Efficient spam detection across online social networks. In Proceedings of the 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, China, 12–14 March 2016; pp. 1–6.
79. Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine learning and deep learning methods for cybersecurity. IEEE Access 2018, 6, 35365–35381.
80. Panda, M.; Patra, M.R. Network intrusion detection using naive bayes. Int. J. Comput. Sci. Netw. Secur. 2007, 7, 258–263.
81. Sharma, S.K.; Pandey, P.; Tiwari, S.K.; Sisodia, M.S. An improved network intrusion detection technique based on k-means clustering via Naïve bayes classification. In Proceedings of the IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012), Nagapattinam, India, 30–31 March 2012; pp. 417–422.
82. Jackson, T.R.; Levine, J.G.; Grizzard, J.B.; Owen, H.L. An investigation of a compromised host on a honeynet being used to increase the security of a large enterprise network. In Proceedings of the Fifth Annual IEEE SMC Information Assurance Workshop, West Point, NY, USA, 10–11 June 2004; pp. 9–14.
83. Khammas, B.M.; Monemi, A.; Bassi, J.S.; Ismail, I.; Nor, S.M.; Marsono, M.N. Feature selection and machine learning classification for malware detection. J. Teknol. 2015, 77, 234–250.
84. Bhat, A.H.; Patra, S.; Jena, D. Machine learning approach for intrusion detection on cloud virtual machines. Int. J. Appl. Innov. Eng. Manag. 2013, 2, 56–66.
85. Gharibian, F.; Ghorbani, A.A. Comparative study of supervised machine learning techniques for intrusion detection. In Proceedings of the Fifth Annual Conference on Communication Networks and Services Research (CNSR'07), Frederlcton, NB, Canada, 14–17 May 2007; pp. 350–358.
86. Fan, C.-I.; Hsiao, H.-W.; Chou, C.-H.; Tseng, Y.-F. Malware detection systems based on API log data mining. In Proceedings of the 2015 IEEE 39th Annual Computer Software and Applications

Conference, Taichung, Taiwan, 1–5 July 2015; pp. 255–260.

87. Renuka, D.K.; Visalakshi, P.; Sankar, T.J.I.J.C.A. Improving E-mail spam classification using ant colony optimization algorithm. Int. J. Comput. Appl. 2015, 2, 22–26.
88. Iyer, S.S.; Rajagopal, S. Applications of Machine Learning in Cyber Security Domain. In Handbook of Research on Machine and Deep Learning Applications for Cyber Security; IGI Global: Hershey, PA, USA, 2020; pp. 64–82.
89. Quinlan, J.R. C4. 5: Programs for Machine Learning; Elsevier: Amsterdam, The Netherlands, 2014.
90. Tyagi, A. Content Based Spam Classification-A Deep Learning Approach; University of Calgary: Calgary, AB, Canada, 2016.
91. He, S.; Lee, G.M.; Han, S.; Whinston, A.B. How would information disclosure influence organizations' outbound spam volume? Evidence from a field experiment. J. Cybersecur. 2016, 2, 99–118.
92. Jiang, L.; Zhang, H.; Cai, Z. A novel Bayes model: Hidden naive Bayes. IEEE Trans. Knowl. Data Eng. 2008, 21, 1361–1371.