Cyber-Attack Detection in Cyber-Physical Systems Using Supervised Machine Learning

Prabhat Semwal and Akansha Handa

1 Introduction

The cyber-physical systems can be defined as the systems built by integrating sensors, computers, networks, communication, and other digital monitoring components into physicals infrastructure to control or monitor the infrastructure remotely and autonomously [1–3]. Some real-world examples of CPS include Smart grids, medical monitoring systems, robotics, autonomous vehicles, soil treatment plants, and water treatment plants [4–8]. The cyber-physical infrastructure operations include both cyber and physical aspects which make these systems vulnerable to both cyber and physical security threats. The attack on CPS can have a huge impact due to the diversity and scope of operations of these structures [6, 9–13]. Thus, the cyber aspect of such CPS has been studied in many pieces of research, which contributed their finds in detecting the cyber-attacks on CPS using machine learning [12, 14–17]. Advancement in Machine Learning and Deep Learning models has motivated the cybersecurity communities for leveraging these models so as to enhance the privacy and security of CPS [18–24]. During the past decade several models have been proposed for a diverse range of cybersecurity including malware detection [25–28], threat hunting [29–32] and privacy protection [33].

In this paper, we have used the SWat dataset which is the data collected from a Secure Water Treatment plant [34]. The data was collected for both normal operational days and few days with attacks on the water treatment. The dataset is processed and used to perform the cyber-attack detection on CPS systems using different supervised machine learning algorithms. We have performed the comparative analysis on the four models based on the major evaluation matrices:

P. Semwal (✉) · A. Handa
School of Computer Science, University of Guelph, Guelph, ON, Canada
e-mail: psemwal@uoguelph.ca; ahanda@uoguelph.ca

© Springer Nature Switzerland AG 2021

K.-K. R. Choo, A. Dehghantanha (eds.), *Handbook of Big Data Analytics and Forensics*, https://doi.org/10.1007/978-3-030-74753-4_9

Accuracy, True Positive Rate (TPR), False Positive Rate (FPR) and the Receiver Operating Characteristics (ROC) curve and Area Under ROC Curve (AUC).

2 Literature Review

In earlier studies, many computer scientists have proposed various approaches to resolve cyber threat hunting problems using different techniques of machine learning [35–39]. Cyber-attack detection is usually accomplished by grouping using power device data or measurements [40–44]. The involvement of risks or attacks is measured in various security and contact levels of the network. Cyber-attacks are observed by measurements by the improved state- estimation techniques using mode-based technique [45]. Numerous studies have presented network traffic-based intrusion detection Ghaeini et al. [46] employ this approach on the SWaT dataset used in our study. Similarly, [47] proposed an Enhanced SVM approach with combined features from two machine learning techniques demonstrated a low false-positive rate. Another paper [48] uses the Random forest Algorithm and achieves a significant accuracy of 94.0187% for cyber-attack detection. A behavior-based machine learning (ML) approach for the detection of any abnormal behavior or attack that may attempt to modify the behavior of the CPS [15]. This method not only recognizes the cyber-attack occurred on a layer of the physical process, but it also identifies the specific attack type. In This study [49] learns how to combine different machine learning methods with the IDS improving the accuracy of threat identification. A prototype IDS is expected in this study. This IDS prototype is equipped to improve accuracy in the identification of several attacks through a combination of machine learning methods. This method not only recognizes the cyber-attack occurred on a layer of the physical process, but it also identifies the specific attack type. In [50] the proposed cyberattack detection system has high detection accuracy and wide attack coverage in order to detect unrecognized attacks using network and host system information.

3 Methodology

This section will describe the process followed to build our supervised machine learning models which can detect the cyber-attack samples from the SWat dataset.

3.1 Dataset Processing

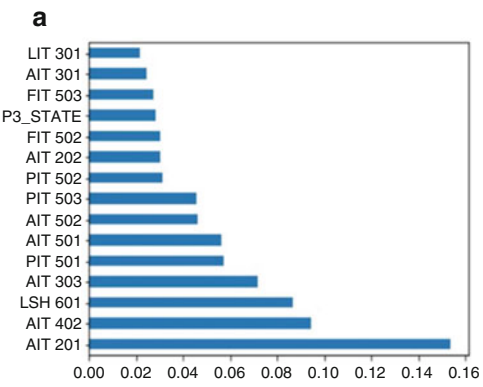
The Swat dataset consists of 77 features and a total of 14,995 data points, 9521 normal and 5474 attack data points. The few features like timestamp and other less

critical features were removed to process the dataset. The label feature (Target) was marked as 1 for attack and 0 for the normal activity data point.

3.1.1 Feature Selection

To reduce the overall dimensionality of the dataset, we performed the feature extraction process. The best feature that can contribute to the target variable was extracted by combining results of ExtraTreeClassifier and SelectKBest algorithms of Scikit-learn library and are shown in Fig. 1a, b respectively.

The most common and highest-ranked features were extracted and used for all the four-classification model. As shown in Table 1, the major operations of the water treatment plant was used as a major feature category set and the same category of features were used to identify the functionality of water plants at different levels process.



b

	Specs	Score
34	P3_STATE	62459.855584
38	AIT 402	53266.368895
64	PIT 502	16781.103371
6	AIT 201	4782.720040
29	LIT 301	3229.757381
40	LIT 401	3114.180748
49	AIT 502	2236.958556
27	DPIT 301	1548.854484
0	FIT 101	929.827097
26	AIT 303	431.844782
67	LSH 601	409.550393
28	FIT 301	318.947993
31	MV 302	225.104277
8	AIT 203	147.962507
35	P301 Status	58.015836

Fig. 1 (**a**) ExtraTreeClassifier (**b**) SelectKBest Result

Table 1 Feature category

Feature name	Description
FIT	Sensor: Inflow into the water tank
LIT	Sensor: Level transmitter
PIT	Sensor: Pressure meter
LSH	Alarm
AIT	Sensor

3.2 Machine Learning Classifiers

For the detection of cyber-attack samples, the KNN, SVM, Decision Tree, and Random Forest, classifiers were trained and tested on the transformed dataset, and results were recorded for comparative analysis.

3.2.1 KNN Model

In the KNN model, we used the processed dataset explained in Sect. 3.2.1. The KNN was implemented with the use of the Sckilearn library and in KNN we initialize the $K = 4$, but after trial and error K was finally set to 1 and the model was trained with $K = 1$ on the processed dataset.

3.2.2 SVM Model

The SVM model was trained and tested on the processed dataset. For SVM, kernel function was set to linear, and probability was set to True.

3.2.3 DT Model

Our DT Model was trained with the processed dataset. The DT model simply designs an inverted tree structure on the base of a trained dataset and then classify a sample by tracing the down designed tree.

3.2.4 RF Model

The RF model is like the DT model, but the RF model creates multiple decision trees instead of only one decision tree. In our RF model, the maximum depth was set to 2.

4 Results and Discussion

This section highlights the results achieved with different supervised machine learning techniques in detecting cyber-attack on a CPS system and will describe the comparative analysis results.

4.1 Evaluation Measures

To evaluate and compare the performance of the models, we have used the commonly used evaluation metrics. Table 2 contains a description of the used evaluation metrics for comparative analysis.

4.2 Experiment and Results

The processed dataset with the total samples of 14,994 and selected features was used to test all the models. All the models were trained on the processed dataset and the results observed on the basis of evaluation metrics (Table 2) are shown in Table 3.

4.3 Comparison of Models

In our experiment, the KNN model achieved an accuracy of 99%. The TPR received for KNN was 99.9% and the FPR was approximately 0%. With the SVM model, we received an accuracy of 98.7% and the average values of TPR and FPR were 99% and 0.01% respectively. Our DT model received 99% accuracy on the processed dataset and approximately 99% TPR and 0% FPR. Whereas, the RF model hit the accuracy of 96% with 98% APR and 0.01% FPR. According to the three-evaluation metrics values mentioned in Table 3, the DT model performed more effectively than other supervised machine learning models in classifying the cyber-attack samples in the Swat dataset.

Table 2 Description of evaluation metric used for comparative analysis

Evaluation metric	Description
TPR	$\frac{TP}{TP+FN}$
FPR	$\frac{FP}{FP+TN}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
ROC Curve	Formed by plotting TPR against FPR at various threshold settings
AUC	The area under the ROC curve

Table 3 Observed accuracy, TPR and FPR values

Model	Accuracy	TPR	FPR
KNN	99.65	99.9	0.008
SVM	98.70	99.03	0.018
DT	99.9	99.0	0.006
RF	96.3	98.6	0.013

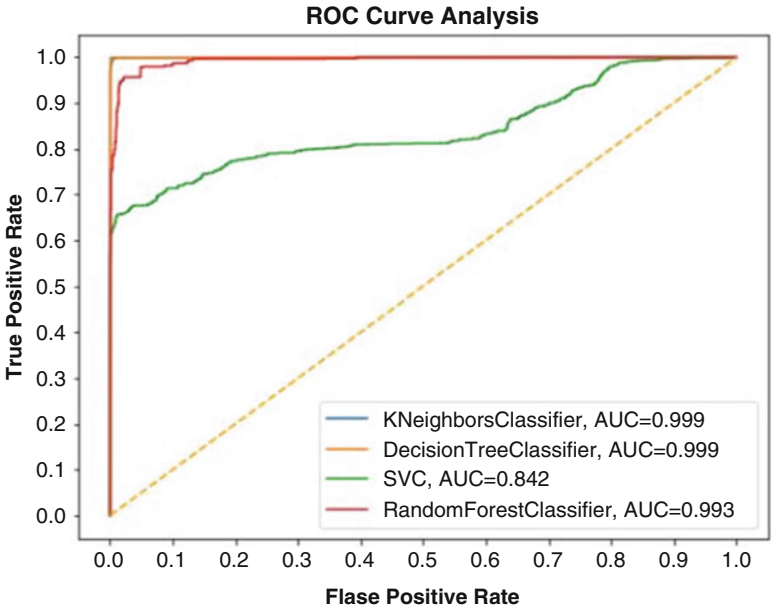


Fig. 2 ROC Curve for all four supervised machine learning models

4.4 ROC Curve

A ROC curve is a common graphical evaluation metric that is used for evaluating the performance of different machine learning classifiers. It allows us to analyze the binary classifier's capability of distinguishing between classes [51]. It is simply a plot of TPR and FPR at different threshold settings.

The ROC curve for classification of cyber-attack samples on a processed dataset, for all four supervised classification models, is shown in Fig. 2. As shown in the legend, the overall AUC value for KNN was 0.99, 0.99 for DT, 0.84 for SVM, and 0.99 for RF. The average AUC of all four models was extremely close to 1 which depicts that all four models perform better for binary classification of cyber-attack in a processed dataset. Although, the AUC value observed for all four models was close to 1. However, the AUC value for both KNN and DT is almost equal to 1 with AUC equal to 0.999 for both KNN and DT.

5 Conclusion

We were able to successfully design the four different machine learning models to classify the cyber-attack samples accurately from the Swat dataset. The results achieved using the critical evaluation metrics allowed us to perform effective

comparative analysis and propose the most suitable algorithm. Using these four supervised machine learning algorithms, we achieved an overall accuracy of 99% with KNN, 98% with SVM, 99% with DT, and 96% with RF. On the base of all the evaluation metrics, the DT outperforms the other classifier models with a reasonable high accuracy of 99.9% and other almost ideal evaluation metrics.

The future work will be to evaluate the other supervised machine learning algorithms and to experiment with the different cyber-physical system datasets.

References

1. H. Karimipour, H. Leung, Relaxation-based anomaly detection in cyber-physical systems using ensemble Kalman filter. IET Cyber-Phys. Syst. Theory Appl. **5**(1), 49–58 (2020)
2. A. Yazdinejad, A. Bohlooli, K. Jamshidi, Efficient design and hardware implementation of the OpenFlow v1.3 Switch on the Virtex-6 FPGA ML605. J. Supercomput. **74**(3), 1299–1320 (2018). https://doi.org/10.1007/s11227-017-2175-7
3. A. Yazdinejad, R.M. Parizi, A. Dehghantanha, K.-K.R. Choo, P4-to-blockchain: A secure blockchain-enabled packet parser for software defined networking. Comput. Secur. **88**, 101629 (2020). https://doi.org/10.1016/j.cose.2019.101629
4. H. Karimipour, A. Dehghantanha, R.M. Parizi, K.-K.R. Choo, H. Leung, A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. IEEE Access **7**, 80778–80788 (2019)
5. H. Karimipour, V. Dinavahi, Extended Kalman filter-based parallel dynamic state estimation. IEEE Trans. Smart Grid **6**(3), 1539–1549 (2015)
6. F. Daryabar, A. Dehghantanha, N.I. Udzir, S.B. Shamsuddin, Towards secure model for SCADA systems, in *Proceedings Title: 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec)*, (2012), pp. 60–64
7. S. Walker-Roberts, M. Hammoudeh, A. Dehghantanha, A systematic review of the availability and efficacy of countermeasures to internal threats in healthcare critical infrastructure. IEEE Access **6**, 25167–25177 (2018, March). https://doi.org/10.1109/ACCESS.2018.2817560
8. S. Nakhodchi, A. Dehghantanha, H. Karimipour, Privacy and security in smart and precision farming: A bibliometric analysis, in *Handbook of Big Data Privacy*, (Springer, Cham, 2020), pp. 305–318
9. S. Watson, A. Dehghantanha, Digital forensics: The missing piece of the internet of things promise. Comput. Fraud Secur. **2016**(6), 5–8 (2016). https://doi.org/10.1016/s1361-3723(15)30045-2
10. A. Al-Abassi, H. Karimipour, H.H. Pajouh, A. Dehghantanha, R.M. Parizi, Industrial big data analytics: Challenges and opportunities, in *Handbook of Big Data Privacy*, ed. by K.-K. R. Choo, A. Dehghantanha, (Springer, Cham, 2020), pp. 37–61
11. M. Alaeiyan, A. Dehghantanha, T. Dargahi, M. Conti, S. Parsa, A multilabel fuzzy relevance clustering system for malware attack attribution in the edge layer of cyber-physical networks. ACM Trans. Cyber-Phys. Syst. **4**(3), 1–22 (2020)
12. M. Saharkhizan, A. Azmoodeh, A. Dehghantanha, K.-K.R. Choo, R.M. Parizi, An ensemble of deep recurrent neural networks for detecting IoT cyber attacks using network traffic. IEEE Internet Things J. **7**(9), 8852–8859 (2020). https://doi.org/10.1109/jiot.2020.2996425
13. E.M. Dovom, A. Azmoodeh, A. Dehghantanha, D.E. Newton, R.M. Parizi, H. Karimipour, Fuzzy pattern tree for edge malware detection and categorization in IoT. J. Syst. Archit. **97**, 1–7 (2019)
14. M. Kravchik, A. Shabtai, Detecting cyber attacks in industrial control systems using convolutional neural networks, in *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and Privacy*, (2018), pp. 72–83

15. K.N. Junejo, J. Goh, Behaviour-based attack detection and classification in cyber physical systems using machine learning, in *Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security*, (2016), pp. 34–43
16. J. Inoue, Y. Yamagata, Y. Chen, C.M. Poskitt, J. Sun, Anomaly detection for a water treatment system using unsupervised machine learning, in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, (2017), pp. 1058–1065
17. M. Saharkhizan, A. Azmoodeh, H. HaddadPajouh, A. Dehghantanha, R.M. Parizi, G. Srivastava, A hybrid deep generative local metric learning method for intrusion detection, in *Handbook of Big Data Privacy*, (Springer, Cham, 2020), pp. 343–357. https://doi.org/10.1007/978-3-030-38557-6_16
18. A. Azmoodeh, A. Dehghantanha, Big data and privacy: Challenges and opportunities, in *Handbook of Big Data Privacy*, (Springer, Cham, 2020), pp. 1–5. https://doi.org/10.1007/978-3-030-38557-6_1
19. D.R. McKinnel, T. Dargahi, A. Dehghantanha, K.-K.R. Choo, A systematic literature review and meta-analysis on artificial intelligence in penetration testing and vulnerability assessment. Comput. Electr. Eng. **75**, 175–188 (2019)
20. J.C. Cabello, H. Karimipour, A.N. Jahromi, A. Dehghantanha, R.M. Parizi, Big-data and cyber- physical systems in healthcare: Challenges and opportunities, in *Handbook of Big Data Privacy*, ed. by K.-K. R. Choo, A. Dehghantanha, (Springer, Cham, 2020)
21. M. Conti, T. Dargahi, A. Dehghantanha, Cyber threat intelligence: Challenges and opportunities, in *Advances in Information Security*, (Springer, 2018), pp. 1–6. https://doi.org/10.1007/978-3-319-73951-9_1
22. StratoEnergetics, Slaughterbots (2017)
23. A. Yazdinejad, R.M. Parizi, A. Dehghantanha, G. Srivastava, S. Mohan, A.M. Rababah, Cost optimization of secure routing with untrusted devices in software defined networking. J. Parallel Distrib. Comput. **143**, 36–46 (2020)
24. A. Yazdinejad, R.M. Parizi, A. Dehghantanha, K.-K.R. Choo, Blockchain-enabled authentication handover with efficient privacy protection in SDN-based 5G networks. IEEE Trans. Netw. Sci. Eng., 1–1 (2019). https://doi.org/10.1109/tnse.2019.2937481
25. A. Azmoodeh, A. Dehghantanha, M. Conti, K.-K.R. Choo, Detecting crypto-ransomware in IoT networks based on energy consumption footprint. J. Ambient. Intell. Humaniz. Comput. **9**(4), 1141–1152 (2018)
26. A. Azmoodeh, A. Dehghantanha, K.-K.R. Choo, Robust malware detection for internet of (battlefield) things devices using deep eigenspace learning. IEEE Trans. Sustain. Comput. **4**(1), 88–95 (2018)
27. H. Haddadpajouh, A. Azmoodeh, A. Dehghantanha, R.M. Parizi, MVFCC: A multi-view fuzzy consensus clustering model for malware threat attribution. IEEE Access **8**, 139188–139198 (2020)
28. H. HaddadPajouh, A. Dehghantanha, R. Khayami, K.-K.R. Choo, A deep recurrent neural network based approach for internet of things malware threat hunting. Futur. Gener. Comput. Syst. **85**, 88–96 (2018). https://doi.org/10.1016/j.future.2018.03.007
29. A.N. Jahromi, S. Hashemi, A. Dehghantanha, R.M. Parizi, K.-K.R. Choo, An enhanced stacked LSTM method with no random initialization for malware threat hunting in safety and time-critical systems. IEEE Trans. Emerg. Top. Comput. Intell. **4**(5), 630–640 (2020). https://doi.org/10.1109/tetci.2019.2910243
30. S. Homayoun et al., Deep dive into ransomware threat hunting and intelligence at fog layer. Futur. Gener. Comput. Syst. **90**(Jan 19), 94–104 (2018)
31. S. Homayoun, A. Dehghantanha, M. Ahmadzadeh, S. Hashemi, R. Khayami, Know abnormal, find evil: Frequent pattern mining for ransomware threat hunting and intelligence. IEEE Trans. Emerg. Top. Comput. (2017). https://doi.org/10.1109/TETC.2017.2756908
32. S. Homayoun et al., DRTHIS: Deep ransomware threat hunting and intelligence system at the fog layer. Futur. Gener. Comput. Syst. **90**, 94–104 (2019). https://doi.org/10.1016/j.future.2018.07.045

33. A. Aminnezhad, A. Dehghantanha, M.T. Abdullah, A survey on privacy issues in digital forensics. Int. J. Cyber-Secur. Digit. Forensics **1**(4), 311–324 (2012)
34. J. Goh, S. Adepu, K.N. Junejo, A. Mathur, A dataset to support research in the design of secure water treatment systems, in *International Conference on Critical Information Infrastructures Security*, (2016), pp. 88–99
35. A. Yazdinejad, A. Bohlooli, K. Jamshidi, Performance improvement and hardware implementation of open flow switch using FPGA, in *IEEE 5th Conference on Knowledge Based Engineering and Innovation, KBEI 2019*, (2019), pp. 515–520
36. A. Yazdinejad, R.M. Parizi, A. Dehghantanha, H. Karimipour, G. Srivastava, M. Aledhari, Enabling drones in the internet of things with decentralized blockchain-based security. IEEE Internet Things J., 1 (2020). https://doi.org/10.1109/jiot.2020.3015382
37. A. Singh, K. Click, R.M. Parizi, Q. Zhang, A. Dehghantanha, K.-K.R. Choo, Sidechain technologies in blockchain networks: An examination and state-of-the-art review. J. Netw. Comput. Appl. **149**, 102471 (2020). https://doi.org/10.1016/j.jnca.2019.102471
38. A. Yazdinejad, R.M. Parizi, A. Dehghantanha, Q. Zhang, K.-K.R. Choo, An energy-efficient SDN controller architecture for IoT networks with blockchain-based security. IEEE Trans. Serv. Comput. **13**(4), 625–638 (2020)
39. D. Połap, G. Srivastava, A. Jolfaei, R.M. Parizi, Blockchain technology and neural networks for the internet of medical things, in *IEEE INFOCOM 2020 – IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, (2020), pp. 508–513. https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9162735
40. A. Yazdinejad, G. Srivastava, R.M. Parizi, A. Dehghantanha, K.-K.R. Choo, M. Aledhari, Decentralized authentication of distributed patients in hospital networks using blockchain. IEEE J. Biomed. Health Inform. **24**(8), 2146–2156 (2020)
41. Q. Chen, G. Srivastava, R.M. Parizi, M. Aloqaily, I. Al Ridhawi, An incentive-aware blockchain-based solution for internet of fake media things. Inf. Process. Manag., 102370 (2020). https://doi.org/10.1016/j.ipm.2020.102370
42. A. Yazdinejad, R.M. Parizi, A. Bohlooli, A. Dehghantanha, K.-K.R. Choo, A high-performance framework for a network programmable packet processor using P4 and FPGA. J. Netw. Comput. Appl. **156**, 102564 (2020)
43. R.M. Parizi, S. Homayoun, A. Yazdinejad, A. Dehghantanha, K.-K.R. Choo, Integrating privacy enhancing techniques into blockchains using sidechains, in *Proceedings of the 32nd IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2019)*, (2019). https://doi.org/10.1109/CCECE.2019.8861821
44. A. Yazdinejad, R.M. Parizi, G. Srivastava, A. Dehghantanha, K.-K.R. Choo, Energy efficient decentralized authentication in internet of underwater things using blockchain, in *2019 IEEE Globecom Workshops (GC Wkshps)*, (2019), pp. 1–6
45. J. Sakhnini, *Security of Smart Cyber-Physical Grids: A Deep Learning Approach* (2020), p. 83
46. World Health Organization et al., in *HAMIDS | Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy,* http://10.0.4.121/2994487.2994492?casa_token=fzc-QNOcjJkAAAAA:iKofJD9cHqHxMQjOxse0v8N4Au0fAwilQzYXDm0MO4aXMQHng4p3NHbqHNFgnwN8AIQNI6T2K5G (acc *Osteoarthr. Cartil*)
47. S. Singh, S. Silakari, An ensemble approach for cyber attack detection system: A generic framework, in *2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, (2013), pp. 79–84
48. M.T. Khorshed, N.A. Sharma, A.V. Dutt, A.B.M.S. Ali, Y. Xiang, Real time cyber attack analysis on Hadoop ecosystem using machine learning algorithms, in *2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, (2015), pp. 1–7
49. B.W. Masduki, K. Ramli, F.A. Saputra, D. Sugiarto, Study on implementation of machine learning methods combination for improving attacks detection accuracy on Intrusion Detection System (IDS), in *2015 International Conference on Quality in Research (QiR)*, (2015), pp. 56–64

50. F. Zhang, H.A.D.E. Kodituwakku, J.W. Hines, J. Coble, Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data. IEEE Trans. Ind. Inform. **15**(7), 4362–4369 (2019)
51. A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. **30**(7), 1145–1159 (1997)