Chapter 10
Anomaly Detection in Cyber-Physical Systems Using Machine Learning

Hossein Mohammadi Rouzbahani, Hadis Karimipour, Abolfazl Rahimnejad, Ali Dehghantanha, and Gautam Srivastava

1 Introduction

Cyber-Physical Systems are a result of an efficient combination of cyber systems and the physical world into an integrated structure for vital tasks which originated from advancements in digital electronics [1]. In these systems, physical components and computational resources are integrated through communication links for remote monitoring and control [2, 3].

The smart grid, as a cyber-physical system, emerged from the restructuring of traditional power networks [4]. These systems require smart tools not only for electrical flow, but also for better performance that has led to self-healing, adaptive protection, control, customer involvement, just to name a few [5–7].

Even though Cyber-Physical Systems develop system operator interaction with the consumer and other parties, many challenges have been created including security, reliability, stability, maintainability, safety, and predictability [8, 9]. Security is one of the most important challenges in cyber-physical systems due to the integration of many components which has made them vulnerable on both the physical and cyber sides. Malicious attacks have led to interrupt system operation or theft of arcane data which can be directed at the cyberinfrastructure or physical

H. Mohammadi Rouzbahani (✉) · H. Karimipour · A. Rahimnejad
School of Engineering, University of Guelph, Guelph, ON, Canada
e-mail: hmoham15@uoguelph.ca; hkarimi@uoguelph.ca; Canada-hkarimi@uoguelph.ca; arahimne@uoguelph.ca

A. Dehghantanha
Cyber Science Lab, School of Computer Science, University of Guelph, Guelph, ON, Canada
e-mail: ali@cybersciencelab.org

G. Srivastava
Department of Mathematics and Computer Science, Brandon University, Brandon, MB, Canada
e-mail: srivastavag@brandonu.ca

K.-K. R. Choo, A. Dehghantanha (eds.), *Handbook of Big Data Privacy*,
https://doi.org/10.1007/978-3-030-38557-6_10

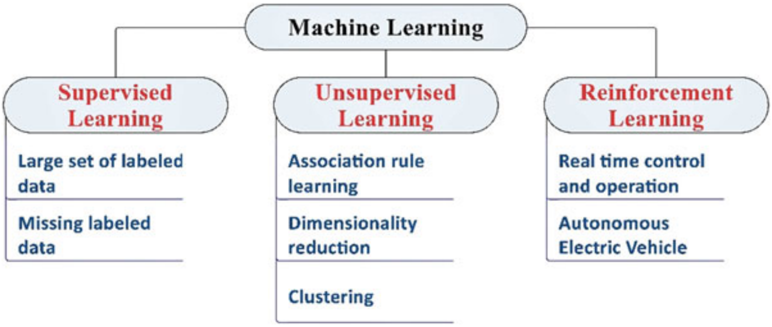


Fig. 10.1 Machine learning subsections in CPS security

components [8, 10–12]. Cyber-physical systems are facing a tsunami of generated data on different components which are too large and complex for real-time processing. Cloud computing techniques along with analytic methods such as machine learning (ML) can help generated information be secure whilst being processed, analyzed and stored [13, 14]. ML in the context of this chapter is referring to making predictions after learning from available data by a system. Figure 10.1 shows the application of ML in smart grid security.

There are many approaches such as ML to intrusion detection systems which are classified into supervised, unsupervised, and reinforcement learning and can build the requisite model based on training data [15]. In supervised ML, both normal and abnormal behaviors are provided to the model to learn trained labeled data. It is very difficult for attackers on cyber-physical systems to obtain labeled data [16, 17] while they do not need abnormal data in the training phase and it is a great advantage for unsupervised learning [2, 18, 19]. In reinforcement learning, there is no training data and as a result, the agent can learn from their own experience. In fact, it gathers the training examples by trial and error while it is attempting its tasks.

This chapter surveys ML methods for an anomaly attack detection framework for cyber-physical systems. Anomaly detection is defined as detecting patterns that do not fit into predictable behavior [20, 21]. Since the characteristics, structure, quantities, and patterns of research activities are understood by bibliometric analysis, the purpose of this chapter is to identify the state-of-the-art of anomaly attack detection in cyber-physical systems.

Web of Science is used as the search engine for this analysis. First, the related keywords are inputted for extracting publications. Then, we limit research time to the last 10 years. Finally, non-relevant and non-English publications were removed and the inquiry to collect the data for bibliometric analysis was as follows: (TS = ((anomaly detection OR outlier detection) AND (cyber-physical system OR cyber-physical system OR smart grid OR CPS cyber-physical systems))). As a result, in the primitive search, 389 publications were found which were reduced to 379 after the mentioned filters.

Results show that the greater number of the publications fall under computer science and engineering and most of them belong to the United States and China

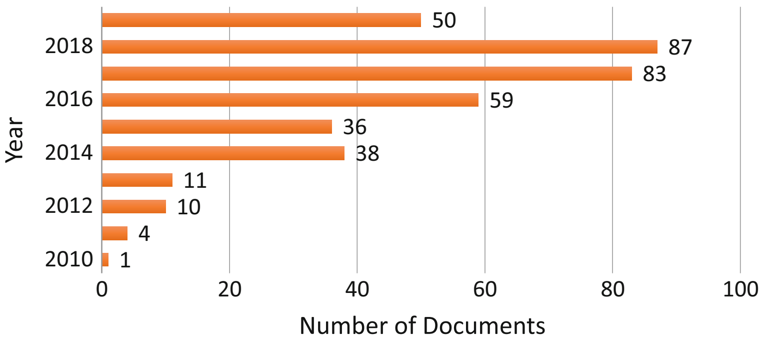


Fig. 10.2 The number of publications

(154 and 72 publications respectively). Iowa State University and the United States Department of Energy are the most productive institutions in this field of study, both being located in the United States.

Figure 10.2 shows 87 documents were published in 2018 while there was only one publication in 2010. Considering the fact that the study was conducted in August 2019, it is predictable to see the number of publications to be higher for 2019 compared to 2018.

The rest of this chapter is organized as follows. Section 2 presents an overall view of cyber-physical systems. Attack detection methods in CPS and anomaly detection are studied in Sects. 2 and 3 respectively. Section 4 provides a case study and Sect. 5 concludes the chapter.

2 Cyber-Physical Systems

According to the application of CPSs, these systems can be defined in different ways, such as deeply intertwined computation, communication, networking, advanced tools, and physical processes interacting with each other relying on IT systems, which are used to monitor and control the physical world [22, 23]. Figure 10.3 shows a holistic view of CPSs.

Different characterizations are presented for CPS which focus on different aspects of these systems including cyber capability, automation, dependability, networking, integration, complexity and reconfiguring [24] which we will briefly mention them [25].

Cyber-physical systems are the integrations of cyber capability and physical components which include distributed networks (i.e. Local Area Network, Bluetooth, Global System for Mobile Communications, etc.) and are severely limited by spatiality and real-time computation. Due to reliability and security necessities for CPS, there is a need to have adaptive capabilities with advanced feedback control technologies.

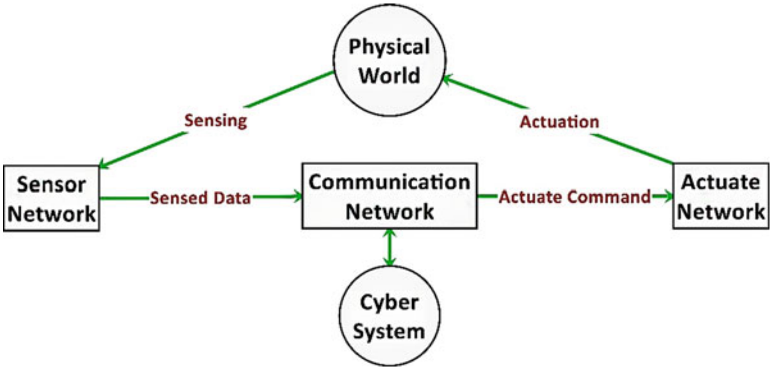
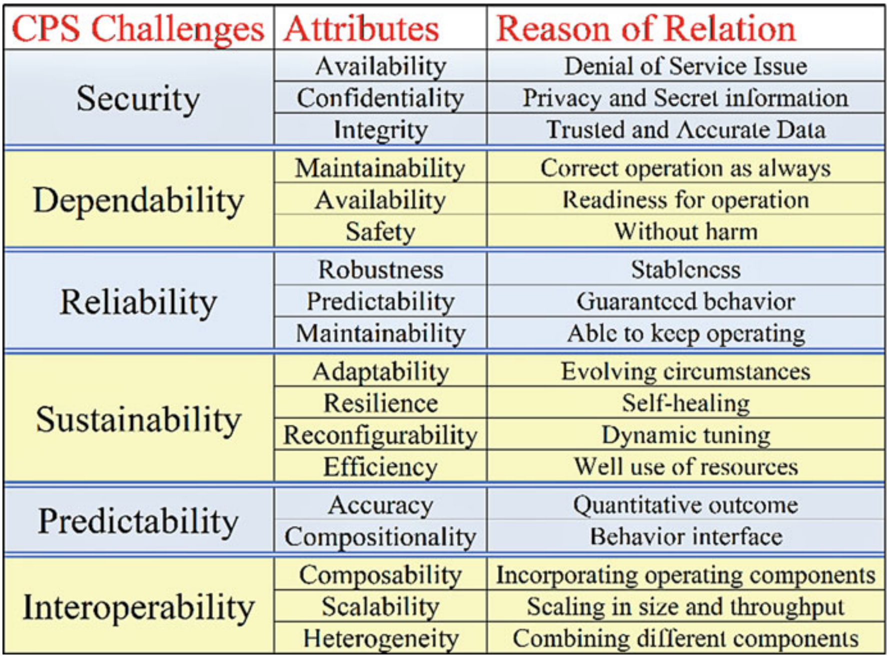


Fig. 10.3 Holistically view of CPSs



CPS Challenges	Attributes	Reason of Relation
Security	Availability	Denial of Service Issue
	Confidentiality	Privacy and Secret information
	Integrity	Trusted and Accurate Data
Dependability	Maintainability	Correct operation as always
	Availability	Readiness for operation
	Safety	Without harm
Reliability	Robustness	Stableness
	Predictability	Guaranteed behavior
	Maintainability	Able to keep operating
Sustainability	Adaptability	Evolving circumstances
	Resilience	Self-healing
	Reconfigurability	Dynamic tuning
	Efficiency	Well use of resources
Predictability	Accuracy	Quantitative outcome
	Compositionality	Behavior interface
Interoperability	Composability	Incorporating operating components
	Scalability	Scaling in size and throughput
	Heterogeneity	Combining different components

Fig. 10.4 CPS challenges

Since cyber-physical systems use distributed communication and smart tools and sensors, these systems are facing various challenges from different points of view which are presented in Fig. 10.4 [1]. However, in the rest of this part, we focus on security issues because CPSs are more vulnerable to cyber-physical malicious attacks [26–28].

Security solutions for cyber-physical systems are required and could be enhanced with Information Technology (IT) systems and techniques like cryptography, access

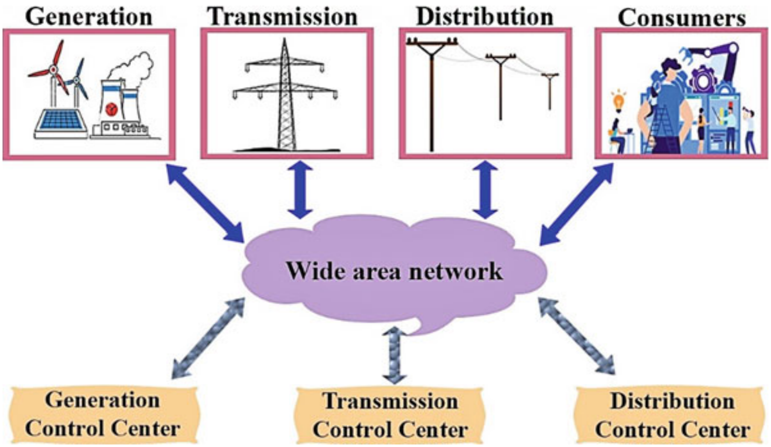

Generation
Transmission
Distribution
Consumers
Wide area network
Generation Control Center
Transmission Control Center
Distribution Control Center

Fig. 10.5 Smart grid as a CPS

control, attack detection, or others. Lack of security in CPSs (e.g. nuclear power station or medical devices) could cause a worldwide threat or disaster.

Security is also one of the most important challenges in the smart grid due to the high dependency of these systems on cyber information yielding new security vulnerabilities [12, 26, 27, 29]. These systems with extensive communication capabilities are good examples of CPSs, which provide the required infrastructure for handling new challenges. Rising electrical energy demand and several technological developments have motivated the advancement of the smart grid. From this, it can be seen that a comprehensive approach is needed for the realization of this issue to quantify attack impacts and assess the effectiveness of countermeasures. The smart grid view as a cyber-physical system is shown in Fig. 10.5.

3 Attack Detection in CPSs

There are three main security properties for a Cyber-Physical system including confidentiality, integrity, and availability [30]. So, attacks are classified considering the security properties as shown in Fig. 10.6.

The most efficient way of defending against network-based attacks is Network Intrusion Detection Systems (NIDS). NIDS are used in almost all Cyber-physical systems. Anomaly-based NIDS and signature-based NIDS are the two main kinds of these detection procedures [31]. Signature-based systems use pattern recognition methods while anomaly-based systems configure a statistical model defining the standard network traffic and flag any abnormal behavior that diverges from the model [32]. It should be noted here that the database of previous attack signatures are preserved and compared with analyzed information for signature-based systems

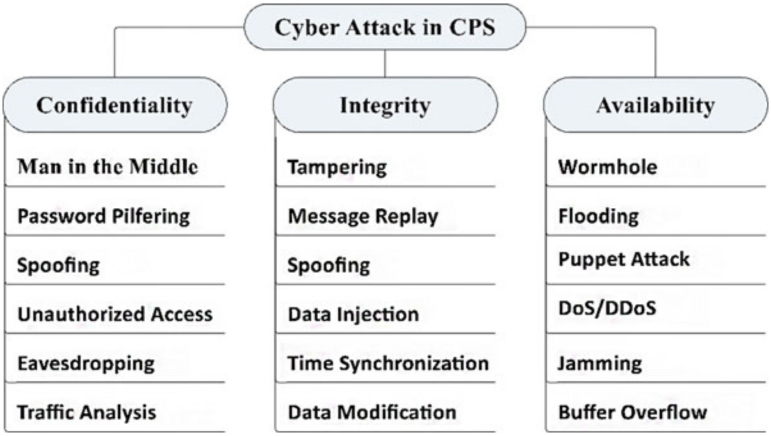


Fig. 10.6 Attack taxonomy for CPS

while in anomaly-based systems the database of general attacks a training phase is required, and it is a complex process due to the setting of a threshold level of detection. Since innovative attacks can be detected as soon as they take place, anomaly detection systems can detect zero-day attacks and it is a major advantage of this system in contrast to signature-based systems [33]. The rest of this section is focused on anomaly-based detection.

3.1 Anomaly Based Detection

The network's behavior is a very important factor and if it does not follow the predicted behavior which is learned by the specifications of the network managers, anomaly detection will commence. Given that various protocols are affected by the rule defining process, the ruleset can be recognized as the main drawback of anomaly detection. Rule definition becomes a difficult process when it is facing custom protocols. Network managers should be comprehensively familiar with the accepted network behavior because the malicious action goes unnoticed if it falls under the accepted behavior, while by defining the rules anomaly detection systems work properly [34]. Finally, anomaly detection is related to novel attacks without a signature which can be detected by anomaly-based method if it falls out of the usual traffic patterns [10]. This is a very big difference between anomaly and signature-based detection methods.

Anomaly detection could be matured upon a variety of general methods borrowed from various scientific fields including ML, statistics, artificial intelligence, clustering, pattern recognition, classification, system theory, signal processing, etc. Figure 10.7 shows a taxonomy for anomaly detection.

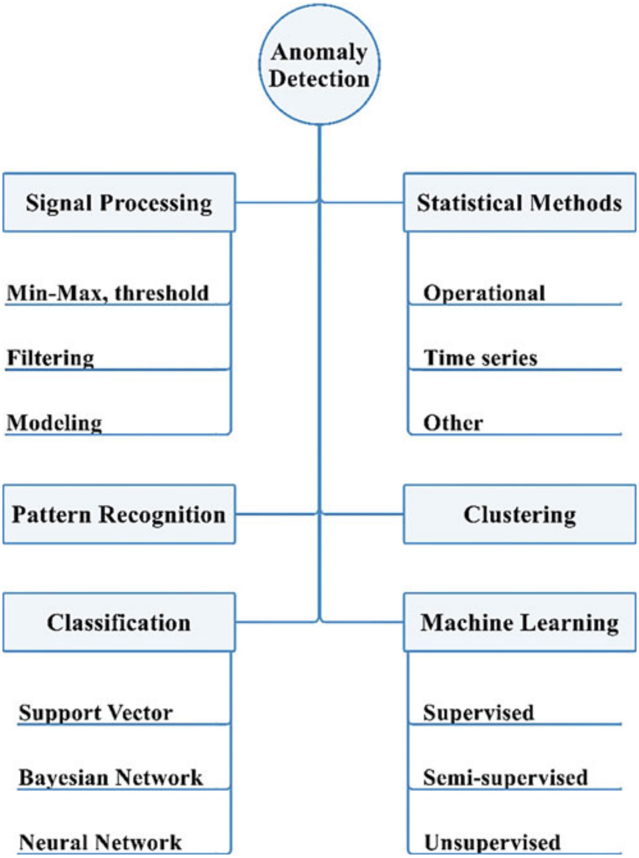


Fig. 10.7 Anomaly detection taxonomy

3.1.1 Statistical Methods

Anomaly detection methods have been advanced using statistical theories which are characterized and qualify the behavior of every component of the system. In these methods, the collected data should be given a probability distribution. The difference between current behavior and normal behavior is detected by using statistical properties such as mean, variance, etc. [35]. Corresponding to the currently observed and the previously trained profile are two different datasets during the anomaly detection which are used by statistical methods. There are many advantages for this method but the most important one is related to decreasing false detection rate because they can provide more accurate detection of malicious actions over a long duration. Given that the ability to learn from observation in statistical methods, detailed awareness about the standard activity of the system is not necessary [36, 37]. It should be noted here that these methods have some drawbacks. For example, the system can be attacked again by generating network traffic in such a way that looks similar to normal behavior. Another disadvantage is that if the system can be modeled in such a way that statistical methods cannot be used, it leaves the detection methods in a useless state [38].

3.1.2 Classification Based Methods

Each attack with a recognized outline and plan can be detected right away if it is dropped while the network administrator prepared details of the features to the detection system. That is why classification methods depend on administrators' substantial knowledge of the specifications of attacks [39]. If an attack signature has been provided previously by a network manager, the system is capable of detecting that because it can detect only what it knows is vulnerable to another new attack. Even if a new signature of attack is created and put into the system, the inflicted damages are not changeful and there are many losses likewise, the repair process is very expensive [40]. Finally, these methods are dependent on a standard traffic outline that makes the cognizance base and consider activities that stray from baseline outline as anomalous [41, 42].

3.1.3 Clustering Based Methods

One of the main subclasses of unsupervised ML is called classification. In this method, rules are found for grouping similar data examples without the need to labeled data [43]. There are many types of clustering methods but the two most important and functional ones are regular clustering and co-clustering [44, 45]. The difference between these methods is related to the method of clustering. In regular clustering, the rows of the data set are considered. In co-clustering the clusters are based on both rows and columns of the dataset simultaneously [46]. K-means is an example of regular clustering.

3.1.4 Signal Processing Approaches

Signal processing methods rely on time-series and spatial-temporal data [47, 48] which includes three sub-methods: Min-Max-Threshold, Filtering, and Modeling. The simplest form of anomaly detection is Min-Max-Threshold, where minimal, maximal or threshold values are defined from a series considered normal [49]. Filtering method compares a signal with a low-pass filtered version which gives an indication of an outlier value. Finally, the modeling method generates a model based on system identification techniques which are used to predict the next values.

3.1.5 Pattern Recognition

In this method, the difference between a normal and an abnormal state is made by the sequence of samples as the shape of the signal whereas the individual data alone is not important. Support vector machine, Neural networks, and Markov chains are trained in order to detect a difference between normal and abnormal shapes [50, 51].

3.1.6 Machine Learning

Machine learning aims to find patterns, make predictions, and make decisions based on historical information to perform a task [12]. Supervised, Unsupervised, Reinforcement and Semi-supervised learning are four types of ML. In supervised techniques, the rules are learned from different examples which are positive or negative and labeled data are used to find a model that explains the dataset. In unsupervised learning, a procedure cannot consider specified anomalies and the main objective is to find a pattern for unlabeled data. Finally, in semi-supervised learning, just the normal performance can be learned from positive examples so only a portion of data is labeled [16].

Machine learning approaches usually separate data into different categories: training and testing. Training data, which commonly is larger in size, is used for learning and providing a model for the system. Testing data, which is completely independent of the training, is used to assess the efficiency of the algorithm. In anomaly-based detection, the normal behavioral pattern is described and modeled by using a training set. Then, the model is applied to testing dataset in order to classify it as either normal or anomalous. In addition, some ML methods separate datasets into three categories instead of two, adding a validation dataset. The validation dataset is used to validate the testing dataset's accuracy when used as input to the given ML method. For illustration, the number of layers and nodes in Artificial Neural Network (ANN) can be varied and the best parameters are chosen that have less estimation of error and more efficient to be built depending on the performance on the validation dataset [12].

One important part of any anomaly detection method is evaluating the performance of ML algorithms. Classification accuracy is the most intuitive method in this evaluation, which measures the performance of the model by computing the ratio number of accurate predictions to the whole number of observations. The main drawback in this metric is that it works properly only when the dataset has equal values for false positives and false negatives [16, 18, 19].

F1 score is another metric in measuring the accuracy in uneven class distribution, which computes the balance between Precision and Recall. Precision is the ratio of correctly predicted positive observation compared to total positive observation, while Recall is the ratio of correct positive prediction to the total number of predictions in the same class (true positives and false negatives of the same class). As a result, F1-score can compute the performance by taking both false positives and false negatives into account. In multi-label ML algorithms, F1-score is usually used to evaluate the classification performance. Therefore, by maximizing the F1-score in multi-label classification, the performance of the algorithm can be considerably improved. Finally, ML is used in a wide range of cyber-physical systems due to the prediction and detection are the two most vital factors for these system operations. Anomaly detectors can be built based on ML algorithms, which could lead to secured cyber-physical systems [18, 19, 52].

4 Case Study

The use of ML techniques for the detection of anomalies can be exhibited through the following case study. Heuristic optimization algorithms are proposed as feature selection techniques to reduce the training time of the algorithms. Since one of the main concerns of the use of ML is computational efficiency, this case study aims to implement automated methods to reduce the dimensions of the data prior to training. This reduces the training and operating time of the ML algorithms for increased computational efficiency.

In this case study, ML classifiers are used to categorize the smart grid measurements as normal or malicious. A Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Naïve Bayesian (NB) classifier are implemented and compared in terms of classification accuracy. Each of the three classifiers is tested with three heuristic feature selection techniques, which are: Binary Cuckoo Search (BCS), Binary Particle Swarm Optimization (BPSO), and Genetic Algorithm (GA). These feature selection methods are optimization algorithms that find the ideal subset of features that produces the best accuracy. The classifiers are tested with each of the resultant subsets of features and evaluated based on its accuracy and F1 score.

Three different IEEE standard power systems are used in this experiment: The IEEE 14-bus system and the IEEE 118-bus system. The measurement data consists of power flow of branches and buses. For each power system, three sets of data were generated; a set of 1000 samples used for feature selection, a set of 40,000 samples used for training of the classification algorithms and a set of 10,000 samples used for testing and evaluation. Each set of data is divided in half into good and malicious data. The malicious data consists of measurements infected with a false data injection (FDI) attack.

Each of the classifiers, as well as the feature selection algorithms, consists of modifiable parameters that can affect the solution. As such, appropriate parameters must be chosen to ensure optimal solutions. For each of the classifiers, the parameters were chosen based on an accuracy test in which accuracy of the classifier was evaluated at varying parameters. Figure 10.8 shows the accuracy of the SVM with varying kernel coefficient (γ) and penalty parameter (C). Similarly, Fig. 10.9 shows the accuracy of the KNN with varying number of neighbors. These tests are performed on the smallest system, IEEE 14-bus system, due to their time-consuming nature. Based on these results, the parameters of the classifiers are chosen. The

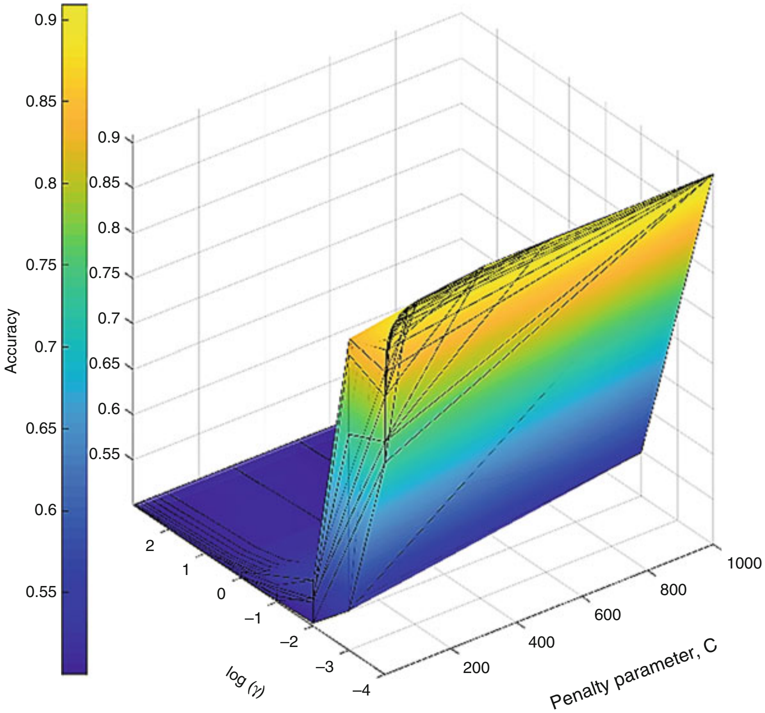


Fig. 10.8 Accuracy of SVM

penalty parameter and kernel coefficient of the SVM is chosen as 1000 and 0.0001 respectively, and the number of neighbors for the KNN algorithm is chosen to be 12. The Naïve Bayesian Classifier, however, was trained with the default smoothing rate of 1 × 10^{-9}.

Each machine learning classifier is tested with the subset of features produced by each of the feature selection algorithms. For each pair of classifier and feature selection algorithms, the classifier is used as the fitness of the solution for each of the heuristic feature selection techniques. The accuracy and F1-score of the classifiers are recorded for each of the resultant feature sets as well as without any feature selection. Furthermore, the runtime for each of the algorithms is recorded for analysis regarding computational efficiency.

The classification accuracy, F1-score, training time, and feature selection time are recorded for each combination of algorithms in Tables 10.1 and 10.2 for the IEEE 14-bus and IEEE 118-bus respectively. The results clearly demonstrate the trade-off between classification accuracy and runtime. The more simplistic classification algorithms like KNN and NB resulted in a much lower runtime; the associated feature selection time and training time is significantly lower than that of the SVM. The complex nature of the SVM algorithm results in a significantly longer feature selection time as well as training time. However, the resultant accuracy and

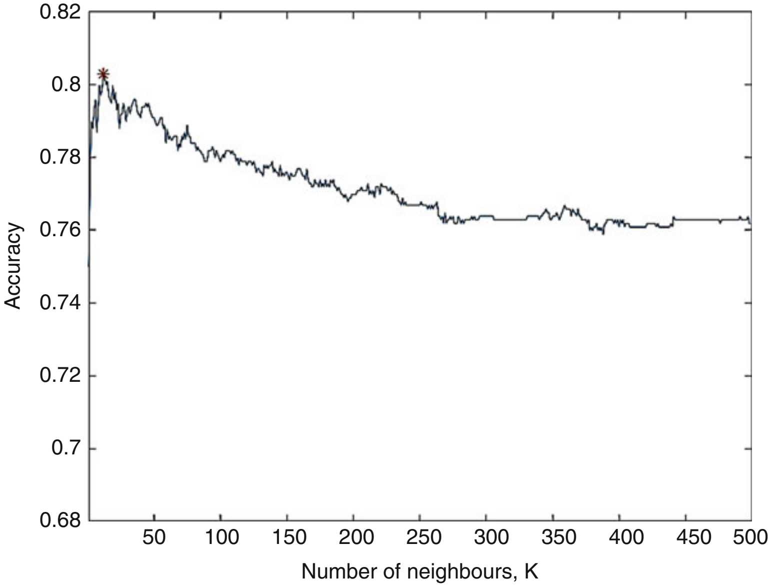


Fig. 10.9 Accuracy of KNN

Table 10.1 Results for the IEEE 14-bus system

Classifier	FSL	NF	CA (%)	F1-score (%)	FST	TT	TRT
SVM	None	34	88.89	88.85	0	1715.245	1715.245
SVM	BCS	10	90.16	90.07	109.855	522.311	632.17
SVM	BPSO	14	90.58	90.50	82.96	873.63	956.59
SVM	GA	8	89.64	89.53	1753.38	155.21	1908.59
KNN	None	34	73.33	73.10	0	0.0937	0.0937
KNN	BCS	21	74.23	74.03	19.59	0.0469	19.63
KNN	BPSO	5	75.62	75.53	15.05	0.0312	15.082
KNN	GA	11	75.06	74.97	265.78	0.0312	265.808
NB	None	34	77.73	77.20	0	0.0469	0.0469
NB	BCS	5	80.25	79.77	1.984	0.0156	2.000
NB	BPSO	5	81.00	80.56	1.509	0.0156	1.524
NB	GA	14	78.95	78.45	50.897	0.0156	50.91

FSL feature selection algorithm, *NF* number of features, *CA* classification accuracy, *FST* feature selection time, *TT* training time, *TRT* total runtime time

F1- score of the SVM algorithm is significantly higher. Furthermore, appropriate feature selection can significantly lower the overall runtime of the SVM, as can be seen from comparing SVM with no feature selection to that with BCS or BPSO for both power systems.

This case study demonstrates the effectiveness of ML techniques at classifying FDI attacks, which typically bypass the standard bad data detection systems.

Table 10.2 Results for the IEEE 118-bus system

Classifier	FSL	NF	CA (%)	F1-Score (%)	FST	TT	TRT
SVM	None	304	89.23	89.18	0	1584.34	1584.34
SVM	BCS	195	88.52	88.49	163.73	1172.16	1335.89
SVM	BPSO	185	87.50	87.47	154.04	1240.50	1394.55
SVM	GA	116	94.33	94.33	1816.94	1025.07	2842.01
KNN	None	304	76.07	74.95	0	0.7498	0.7498
KNN	BCS	189	75.90	74.83	68.308	0.4609	0.4609
KNN	BPSO	219	76.04	74.94	77.592	0.5166	0.5166
KNN	GA	125	78.43	77.67	776.94	0.2968	0.2968
NB	None	304	76.64	76.45	0	0.2656	0.2656
NB	BCS	91	79.17	78.97	6.341	0.0937	6.435
NB	BPSO	180	80.69	80.51	5.615	0.1718	5.786
NB	GA	146	81.11	80.93	106.53	0.1406	106.67

FSL feature selection algorithm, *NF* number of features, *CA* classification accuracy, *FST* feature selection time, *TT* training time, *TRT* total runtime time

Additionally, this study reveals the trade-off between computational time and performance. Furthermore, it was proven that heuristic feature selection can be successful at reducing the number of features and, as a result, reduce the training time of the classification algorithms. When combined with a computationally expensive classifier, heuristic feature selection can significantly reduce the overall runtime thus improving the computational efficiency of certain classifiers. This, however, was not exhibited in the more simplistic classifiers due to their much faster training time, which is reduced by less than the runtime of the feature selection algorithms. In realistic applications, with larger systems and larger data, the training time is expected to be significantly larger. As such, the reduction in runtime is expected to be much larger.

5 Conclusion

The main idea of cyber-physical systems is designing an integrated system instead of separate systems on cyber and physical systems. These systems could be a propitious paradigm for current and future engineered systems which are able to make an impressive impact on our interactions with physical components.

Security is one of the most important factors in CPSs because of the frequency of reported cyber-attacks. Although many detection methods have been proposed, new solutions are still expected against new threats and vulnerabilities. Many approaches are presented in this chapter for attack detection in CPSs such as anomaly detection by using ML including supervised, unsupervised, reinforcement, and semi-supervised methods. We also briefly introduce cyber-physical systems and security concerns about them. Then, detection methods were presented. Finally, a case study showing the effectiveness of different ML algorithms in classifying cyber-physical systems attack was given. Our results demonstrated that reducing the number of features can reduce the overall runtime of the program.

References

1. V. Gunes, S. Peter, T. Givargis, et al., A Survey on Concepts, Applications, and Challenges in Cyber-Physical Systems. Citeseer (2014). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.717.3807&rep=rep1&type=pdf
2. J. Goh, S. Adepu, M. Tan, et al., *Anomaly Detection in Cyber Physical Systems Using Recurrent Neural Networks* (2017). Ieeexplore.Ieee.Org. https://ieeexplore.ieee.org/abstract/document/7911887/
3. A. Jones, Z. Kong, C. Belta, Anomaly detection in cyber-physical systems: a formal methods approach, in *53rd IEEE Conference on Decision and Control* (2014). Ieeexplore.Ieee.Org. https://ieeexplore.ieee.org/abstract/document/7039487/
4. M. Cintuglu, O. Mohammed, K. Akkaya, A.S. Uluagac, *A Survey on Smart Grid Cyber-Physical System Testbeds* (2016). Ieeexplore.Ieee.Org. https://ieeexplore.ieee.org/abstract/document/7740849/
5. T. Agarwal, P. Niknejad, A. Rahimnejad, M.R. Barzegaran, L. Vanfretti, Cyber–physical microgrid components fault prognosis using electromagnetic sensors. IET Cyber-Phys Syst Theory Appl **4**(2), 173–178 (2019). https://doi.org/10.1049/iet-cps.2018.5043
6. H.M. Ruzbahani, H. Karimipour, Optimal incentive-based demand response management of smart households, in *2018 IEEE/IAS 54th Industrial and Commercial Power Systems Technical Conference* (*I&CPS*) (2018), pp. 1–7. https://doi.org/10.1109/ICPS.2018.8369971
7. H.M. Ruzbahani, A. Rahimnejad, H. Karimipour, Smart households demand response management with micro grid, in *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)* (2019), pp. 1–5. https://doi.org/10.1109/ISGT.2019.8791595
8. C.K. Keerthi, M.A. Jabbar, B. Seetharamulu, Cyber Physical Systems (CPS): security issues, challenges and solutions, in *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (2017), pp. 1–4. https://doi.org/10.1109/ICCIC.2017.8524312
9. A. Rahimneiad, I. Al-Omari, R. Barzegaran, H. Karimipour, Hybrid harmonic estimation based on least square method and bacterial foraging optimization, in *2018 IEEE Electrical Power and Energy Conference* (*EPEC*) (2018), pp. 1–6. https://doi.org/10.1109/EPEC.2018.8598450
10. 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 (2019). https://doi.org/10.1109/TSUSC.2018.2809665
11. A. Azmoodeh, A. Dehghantanha, R.M. Parizi, H. Karimipour, E. Modiri, D.E. Newton, Fuzzy pattern tree for edge malware detection and categorization in IoT zero trust distributed computing view project naive-Bayesian-based model for interoperability among heterogeneous Systems in Intelligent Buildings View project fuzzy pattern tree for edge malware detection and categorization in IoT. J. Syst. Archit. **97**, 1–7 (2019). https://doi.org/10.1016/j.sysarc.2019.01.017
12. 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 (2019). https://doi.org/10.1109/ACCESS.2019.2920326
13. R. Altawy, A.M. Youssef, Security tradeoffs in cyber physical systems: a case study survey on implantable medical devices. IEEE Access **4**, 959–979 (2016). https://doi.org/10.1109/ACCESS.2016.2521727

14. C.-W. Tsai, C.-F. Lai, M.-C. Chiang, L.T. Yang, Data mining for internet of things: a survey. IEEE Commun. Surv. Tutorials **16**(1), 77–97 (2014). https://doi.org/10.1109/SURV.2013.103013.00206
15. J. Sakhnini, H. Karimipour, A. Dehghantanha, *Smart Grid Cyber Attacks Detection Using Supervised Learning and Heuristic Feature Selection* (2019). http://arxiv.org/abs/1907.03313
16. O.M.K. Alhawi, J. Baldwin, A. Dehghantanha, Leveraging machine learning techniques for windows ransomware network traffic detection, in *Cyber Threat Intelligence*, (Springer, Cham, 2018), p. 70. https://doi.org/10.1007/978-3-319-73951-9_5
17. N. Milosevic, A. Dehghantanha, K.-K.R. Choo, Machine learning aided android malware classification. Comput. Elect. Eng. **61**, 266–274 (2017). https://doi.org/10.1016/J.COMPELECENG.2017.02.013
18. A. Shalaginov, S. Banin, et al., *Machine Learning Aided Static Malware Analysis: A Survey and Tutorial* (Springer, Berlin, 2018). https://link.springer.com/chapter/10.1007/978-3-319-73951-9_2
19. A. Shalaginov, S. Banin, A. Dehghantanha, K. Franke, *Machine Learning Aided Static Malware Analysis: A Survey and Tutorial* (2018). https://doi.org/10.1007/978-3-319-73951-9_2
20. V. Chandola, A. Banerjee, V. Kumar, Anomaly detection. ACM Comput. Surv. **41**(3), 1–58 (2009). https://doi.org/10.1145/1541880.1541882
21. S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, H. Karimipour, Cyber intrusion detection by combined feature selection algorithm. J. Inform. Secur. Appl. **44**, 80–88 (2019). https://doi.org/10.1016/J.JISA.2018.11.007
22. M. Conti, S. Das, C. Bisdikian, M. Kumar, et al., Looking ahead in pervasive computing: challenges and opportunities in the era of cyber–physical convergence. Pervasive Mob. Comput. **8**, 2–21 (2012). https://www.sciencedirect.com/science/article/pii/S1574119211001271
23. I. Horvath, B.H. Gerritsen, *Cyber-Physical Systems: Concepts, Technologies and Implementation Principles* (2012). Researchgate.Net. https://www.researchgate.net/profile/Imre_Horvath/publication/229441298_CYBER-PHYSICAL_SYSTEMS_CONCEPTS_TECHNOLOGIES_AND_IMPLEMENTATION_PRINCIPLES/links/0912f500e60008cd01000000.pdf
24. L. Miclea, et al., *About Dependability in Cyber-Physical Systems* (2011). *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/6116428/
25. J. Shi, J. Wan, H. Yan, H. Suo, A survey of cyber-physical systems, in *2011 International Conference on Wireless Communications and Signal Processing (WCSP)* (2011), pp. 1–6. https://doi.org/10.1109/WCSP.2011.6096958
26. F. Ghalavand, B. Alizade, H. Gaber, H. Karimipour, Microgrid islanding detection based on mathematical morphology. Energies **11**(10), 2696 (2018). https://doi.org/10.3390/en11102696
27. F. Ghalavand, B. Alizade, H. Gaber, H. Karimipour, F. Ghalavand, B.A.M. Alizade, et al., Microgrid islanding detection based on mathematical morphology. Energies **11**(10), 2696 (2018). https://doi.org/10.3390/en11102696
28. H. Karimipour, V. Dinavahi, On false data injection attack against dynamic state estimation on smart power grids, in *2017 IEEE International Conference on Smart Energy Grid Engineering (SEGE)* (2017), pp. 388–393. https://doi.org/10.1109/SEGE.2017.8052831
29. H. Karimipour, V. Dinavahi, Robust massively parallel dynamic state estimation of power systems against cyber-attack. IEEE Access **6**, 2984–2995 (2018). https://doi.org/10.1109/ACCESS.2017.2786584
30. S. Geris, H. Karimipour, A feature selection-based approach for joint cyber-attack detection and state estimation, in *IEEE International Conference on Smart Energy Grid Engineering (SEGE)* (2019), pp. 1–5. https://www.scpslab.org/publications.html
31. S. Mohammadi, V. Desai, H. Karimipour, Multivariate mutual information feature selection for intrusion detection, in *IEEE Canada Electrical Power and Energy Conference* (*EPEC*) (2018), pp. 1–6. https://www.scpslab.org/publications.html

32. H. Karimipour, S. Geris, A. Dehghantanha, Anomaly detection for large-scale smart grids (2019), pp. 1–4. https://www.scpslab.org/publications.html
33. M.R. Begli, F. Derakhshan, H. Karimipour, A layered intrusion detection system for critical infrastructure using machine learning, in *A Layered Intrusion Detection System for Critical Infrastructure Using Machine Learning* (2019), pp. 1–5. https://www.scpslab.org/publications.html
34. H. Pajouh, R. Javidan, et al., *A Two-Layer Dimension Reduction and Two-Tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks* (2016). *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/7762123/
35. G. Sebestyen, A. Hangan, et al., *A Taxonomy and Platform for Anomaly Detection* (2018). *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/8402710/
36. A. Patcha, J.-M. Park, An overview of anomaly detection techniques: existing solutions and latest technological trends. Comput. Netw. **51**(12), 3448–3470 (2007). https://doi.org/10.1016/J.COMNET.2007.02.001
37. N. Ye, Q. Chen, An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. Qual. Reliab. Eng. Int. **17**(2), 105–112 (2001). https://doi.org/10.1002/qre.392
38. P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, E. Vázquez, Anomaly-based network intrusion detection: techniques, systems and challenges. Comput. Secur. **28**(1–2), 18–28 (2009). https://doi.org/10.1016/J.COSE.2008.08.003
39. C.-I. Chang, S.-S. Chiang, Anomaly detection and classification for hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. **40**(6), 1314–1325 (2002). https://doi.org/10.1109/TGRS.2002.800280
40. M. Ahmed, A. Mahmood, J. Hu, A survey of network anomaly detection techniques. J. Network Comput. Appl. **60**, 19–31 (2016). https://www.sciencedirect.com/science/article/pii/S1084804515002891
41. W. Lee, X. Dong, Information-theoretic measures for anomaly detection, in *Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001* (2000), pp. 130–143. https://doi.org/10.1109/SECPRI.2001.924294
42. I. Steinwart, D. Hush, C. Scovel, A classification framework for anomaly detection. J. Mach. Learn. Res. **6**(Feb), 211–232 (2005). http://www.jmlr.org/papers/v6/steinwart05a.html
43. V. Estivil-Castro, ACM Digital Library, Proceedings of the twenty-eighth australasian conference on computer science, Newcastle, Australia, in *Proceedings of the Twenty-eighth Australasian Conference on Computer Science*, vol 38 (2005). https://dl.acm.org/citation.cfm?id=1082198
44. L. Portnoy, *Intrusion Detection with Unlabeled Data Using Clustering* (2000). https://doi.org/10.7916/D8MP5904
45. F. Zhouyu, W. Hu, T. Tan, Similarity based vehicle trajectory clustering and anomaly detection, in *IEEE International Conference on Image Processing 2005* (2005), pp. II–602. https://doi.org/10.1109/ICIP.2005.1530127
46. M. Ahmed, A. N. Mahmood, & M. J. Maher (2015). *Heart Disease Diagnosis Using Co-clustering*. https://doi.org/10.1007/978-3-319-16868-5_6
47. S. Agrawal, J. Agrawal, Survey on anomaly detection using data mining techniques. Proc. Comput. Sci. **60**, 708–713 (2015). https://www.sciencedirect.com/science/article/pii/S1877050915023479
48. M. Gupta, J. Gao, et al., *Outlier Detection for Temporal Data: A Survey* (2013). Ieeexplore.Ieee.Org. https://ieeexplore.ieee.org/abstract/document/6684530/
49. N. Laptev, S. Amizadeh, et al., *Generic and Scalable Framework for Automated Time-Series Anomaly Detection* (2015). *Dl.Acm.Org*. https://dl.acm.org/citation.cfm?id=2788611
50. S.-W. Joo, R. Chellappa, Attribute grammar-based event recognition and anomaly detection, in *2006 Conference on Computer Vision and Pattern Recognition Workshop* (*CVPRW'06*) (2016), p. 107. https://doi.org/10.1109/CVPRW.2006.32

51. L. Lankewicz, M. Benard, Real-time anomaly detection using a nonparametric pattern recognition approach, in *Proceedings Seventh Annual Computer Security Applications Conference* (n.d.), pp. 80–89. https://doi.org/10.1109/CSAC.1991.213016
52. M. Kakavand, M. Dabbagh, et al., *Application of Machine Learning Algorithms for Android Malware Detection* (2018). Researchgate.Net. https://www.researchgate.net/profile/Mohammad_Dabbagh3/publication/331216763_Application_of_Machine_Learning_Algorithms_for_Android_Malware_Detection/links/5c74adcb92851c69504146a9/Application-of-Machine-Learning-Algorithms-for-Android-Malware-Detection.pdf