Chapter 9
How Machine Learning Can Support Cyberattack Detection in Smart Grids

Bruno Bogaz Zarpelão, Sylvio Barbon Jr., Dilara Acarali, and Muttukrishnan Rajarajan

9.1 Introduction

The world's demand for electricity has been steadily growing due to several aspects of modern life, causing a push in industrial production and giving rise to new electricity-dependent technologies. At the same time, society has refused the idea of increasing the use of fossil fuels as power sources, given that they are responsible for several environmental problems we have faced. To cope with these challenges, power grids have been reshaped to become more resilient, reliable, and efficient. Renewable and alternative power sources have been increasingly adopted to reduce greenhouse gas emissions. The new power grids emerging from this modernisation process are named smart grids [1–3].

To reach their goals, smart grids rely on advanced control and communication technologies. Although these technologies have been used to make power grids more reliable, they are also responsible for introducing new vulnerabilities. Smart grids are complex and large-scale systems, composed of multiple domains involving customers, utilities, operators, and service providers. Attackers can target any part of these systems, from smart meters at customer premises to core devices at transmission networks or power plants. As smart grids are highly interconnected, an attack on a particular point, which at first sight does not seem to be significant, can escalate quickly to a massive disruption of the whole system [3–5].

B. B. Zarpelão (✉) · S. Barbon Jr.
Computer Science Department, State University of Londrina, Londrina, Paraná, Brazil
e-mail: brunozarpelao@uel.br; barbon@uel.br

D. Acarali · M. Rajarajan
School of Mathematics, Computer Science and Engineering, City, University of London, London, UK
e-mail: dilara.acarali@city.ac.uk; r.muttukrishnan@city.ac.uk

Cybersecurity measures must be set over the entire smart grid to ensure its reliability. Among all the available security solutions, attack detection systems are particularly important. As smart grids are complex and large, it is impossible to make sure that there are no security breaches in any part of the system. Researchers discover new vulnerabilities on a daily basis even in long-used devices, and well-known vulnerabilities may remain unpatched due to the lack of sufficient resources to cover such a huge attack surface. Therefore, attack detection systems are necessary to monitor the whole system continuously and alert the administrators when needed.

Attackers have evolved along with the defence tools. They are usually able to bypass or evade existing attack detection systems, and capable of developing smarter attacks that can adapt to new security measures. Machine learning is a promising solution for creating attack detection systems capable of dealing with these advanced adversaries in smart grids, having been successfully applied to detect attacks in other domains.

In this chapter, we present a survey about the application of machine learning for attack detection in smart grids. Our goal is to enable a better understanding of the attack types that affect smart grids, the aspects that drive detection systems development (detection methods, data collection, and system distribution), and how machine learning algorithms are employed in this context. Finally, we discuss open issues related to the current usage of machine learning-based detection in smart grids and point out some paths to address them.

The rest of the chapter is organised as follows. Section 9.2 presents an overview of smart grids to build a foundation for the rest of the study. Section 9.3 discusses the types of attacks that affect smart grids, while Sect. 9.4 shows the main aspects of detection systems. Section 9.5 details the foundations of the machine learning algorithms used for attack detection in smart grids. Section 9.6 presents the surveyed solutions, and Sect. 9.7 discusses their open issues and possible improvements. Finally, Sect. 9.8 presents the concluding remarks.

9.2 Smart Grids Overview

Smart grids are the convergence of power grids and Information and Communication Technology (ICT). They have been developed as a response to the growing demand for electrical power and the rise of renewable energy sources. In this context, ICT tools are used to improve the management and control of the whole cycle of power generation, transmission, and distribution, making sure that multiple power sources are explored and faults and outages are significantly reduced even with the system under constant pressure [1–3].

Power grid operation is divided into generation, transmission, and distribution. Energy is generated in power plants of different kinds (e.g. nuclear, thermal, wind, hydroelectric, or solar) and transmitted over long distances through high-voltage transmission lines to electrical substations. From electrical substations, energy is

distributed to end customers, according to their demand. As these systems spread over wide geographical areas, they are structured in a hierarchical fashion. A control centre monitors the power grid activity to ensure that multiple parameters like voltage, frequency, and current are within the expected range. Situational awareness is a key term to define the central control mission. Additionally, the power grid has some protection mechanisms, like breakers and relays, that take action when a fault occurs to keep the system up and avoid significant damage. Protection mechanisms can operate automatically or under the central control command. Summing up, power grids are huge and complex systems that operate under strict requirements and are monitored continuously to prevent outages that might have serious consequences [6–9].

In smart grids, ICT is used to enable two-way communication between the control centre and different parts of the power grid. Data about the power grid state are collected in real-time from all over the system, providing controllers with updated information that can be used to respond to unexpected behaviour, make demand predictions, and coordinate multiple power sources, among other tasks related to management and control. Most of these needs always existed in power grids. However, the reality has changed in recent years, making more sophisticated ICT solutions necessary to cope with rising challenges. For instance, renewable energy sources like wind power or solar power may generate energy intermittently, as less wind, cloudy weather or some other natural and unavoidable condition may affect their generation potential. In this sense, ICT solutions can help to forecast these occurrences and coordinate the use of these sources accordingly [6–9].

According to a conceptual model proposed by NIST (National Institute of Standards and Technology) [1], smart grids are organised into seven domains: customer, markets, service provider, operations, generation, transmission, and distribution. The customer domain encompasses electricity end-users. Smart grids include some differentials, such as dynamic pricing and generation of electricity by end-users, which add more complexity to the customer's role. For this reason, customers need a two-way communication interface with the grid, named ESI (Energy Services Interface). This interface sets the boundary between the customer and the utility and is usually deployed at the meter or local management system. Customers can have smart devices, which interact with the smart grid to provide details of their consumption and other energy parameters, while receiving commands from service providers that deliver management services.

The market domain consists of the operators responsible for commercialising grid assets, from bulk electricity suppliers to retailers that supply electricity to end-users. Organisations that deliver services such as billing, account management, and maintenance and installation to customers and utilities make up the service provider domain. The operations domain encompasses those who are responsible for ensuring that the smart grid's operations run smoothly. Their activities include grid monitoring and control, fault management, grid estimates calculation, analytics, planning, and maintenance. All of these tasks are performed from a control centre, which hosts some management systems, such as the EMS (Energy Management

System), dedicated to generation and transmission processes, and the DMS (Distribution Management System), responsible for distribution processes.

Electricity generation is the key process of the generation domain. Several energy sources such as nuclear fission, flowing water, wind, and solar radiation can be used to create electricity. The generation domain has a plant control system, which is used to monitor and control the power generation. It must report performance measures continuously, so the operators can predict possible issues and mitigate their effects. The transmission domain encompasses all the actors and functions needed to transmit the electrical power produced in the generation domain to the distribution domain. The transmission domain has the essential responsibility of balancing electricity generation and load. Any disturbance in this delicate balance can affect the grid frequency, leading to power outages or other kinds of damages to the system. The distribution domain is responsible for interconnecting the transmission domain and the customer domain. It also informs the operations domain about the power flow situation.

All of these domains have to interact and cooperate to reach their goals, and several technologies are available to support this need. Smart meters are deployed at the customers' side to measure their energy consumption and gather other management information in real-time (typically every 30 min) to report to other domains. These meters are part of a communication infrastructure referred to as AMI (Advanced Metering Infrastructure). In addition to smart meters, an AMI includes data concentrators for aggregating data collected from smart meters, and head-end systems, which are responsible for connecting smart meters and data concentrators to management information systems. Together, smart meters and AMIs behave as typical IoT (Internet of Things) systems, adding the many particularities of this paradigm [10].

SCADA (Supervisory Control and Data Acquisition) systems are also used to support data exchange among these domains. These systems are made up of three main components, RTUs (Remote Terminal Unit), MTUs (Master Terminal Unit), and HMI (Human Machine Interface). RTUs are deployed close to or at devices that are remotely controlled. MTUs (Master Terminal Units) are responsible for sending requests periodically to RTUs, asking for data about the monitored device, in a process referred to as polling. The polling frequency can range from multiple requests per second to one request every few minutes, depending on the importance of the monitored device. MTUs can also send commands to RTUs asking them to act over the controlled system. Human operators interact with these components through HMIs. In smart grids, RTUs can be deployed in the generation, transmission, and distribution domains. The control centre's management systems provide human operators with HMIs to monitor and control these RTUs.

Another solution to collect measurements from the transmission domain is the PMU (Phasor Measurement Unit). In a smart grid, PMUs are deployed at transmission substations to collect current and voltage phasor information. They operate at very high sampling rates, and are, therefore, able to collect many more measurements per second than a common RTU. All PMU measurements are timestamped, and GPS (Global Positioning System) devices are needed to

synchronise measurements from PMUs at different locations. PDCs (Phasor Data Concentrators) aggregate data from PMUs, perform quality checks, and then forward these measurements to EMSs, where the collected data is analysed for state estimation, monitoring, control, and protection.

Although smart meters, RTUs, and PMUs have some particularities that make them unlike one another, all of these devices share at least a common characteristic: they generate continuous data streams. Therefore, management and control systems for smart grids, which consume these data, have to be designed to handle continuous streams. This means they have to be able to learn incrementally, to manage the constant inflow of huge amounts of data, and to cope with real-time changes in the statistical distributions underlying the collected data.

Communication networks underpin all of these domains. As smart grid networks have to connect a great number of endpoints over wide geographical areas, they are organised hierarchically. At the customer end, there are HANs (Home Area Network), which connect smart devices within the customer's premises to the smart grid structure, enabling the energy usage management at the customer level. HANs are connected to the distribution system's networks, referred to as FAN (Field Area Network). These networks connect components such as RTUs in the distribution domain and smart meters to control centres. Networks in the distribution domain are also named NAN (Neighbour Area Network). Finally, WANs (Wide Area Network) connect distant sites, making up a backbone for the integration of the networks that compose smart grids. They are responsible for connecting the transmission and generation domains to the control centre and for transmitting information like PMU measurements and RTU readings. Also, they establish a communication path between FANs and control centres, which are usually separated by long distances.

Figure 9.1 presents an overview of the seven domains along with devices and communication networks that compose a smart grid.

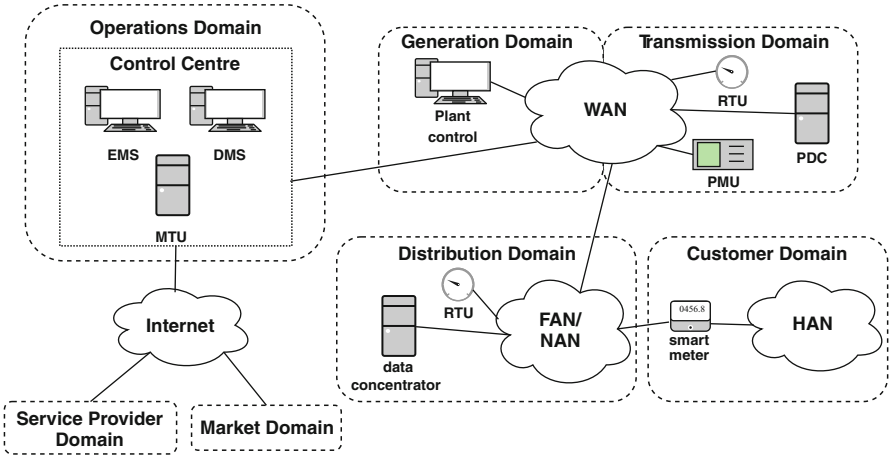

Fig. 9.1 Overview of main smart grid components and their relationships

9.3 Smart Grid Attacks

Smart grids are complex systems consisting of various specialised components working collaboratively to exchange sensitive data, process inputs, and make decisions, all in real-time. This combined complexity and sensitivity produces many vulnerabilities that can be exploited by malicious individuals. Furthermore, the accurate and sustained functionality of power infrastructures are non-negotiable; power is in constant demand. These issues are further exasperated by the vulnerabilities of wireless network technologies, and the presence of many potential access points (i.e. smart meters) [8]. All of this makes the smart grid a highly attractive target for those looking to cause large-scale disruption. A successful attack on the grid hinders everything in the affected region, as experienced in Ivano-Frankivsk in Ukraine in 2016, where thousands of people were left without electricity [11]. Despite the level of disruption caused in that incident, attacks on smart grids theoretically and feasibly have much a larger damage potential.

This section provides a taxonomy of smart grid attacks, along with detailed explanations of each category. In keeping with the basic principles of cybersecurity, the CIA triad is used to divide attacks into three main categories based on what they threaten: confidentiality, integrity, and availability. Each one is then further divided to distinguish between attack aims and methods. It should be noted that some categories inevitably have overlaps as attacks often interleave in complex campaigns. The outline given here aims to highlight individual malicious actions taken against the smart grid. In comparison to the taxonomy presented in [8], we consider "data attacks" and "device attacks" to fall under the integrity category, as the aim of both is to compromise the integrity of the grid network. Meanwhile, privacy attacks are directly analogous to attacks on confidentiality, and network availability attacks are captured in the same way.

Another important consideration is that attacks on smart grids can be considered over two planes: the cyber and the physical [8, 12–14]. This is because the grid is a digitised system that regulates and manages a physical utility. Hence, Cyber-Physical Threats (CPTs) are defined as attacks where a malicious action in the cyber plane has repercussions in the physical (and vice versa) [12, 13]. Examples of this include acts of remote sabotage (like the Stuxnet incident [13]), manipulation of the grid topology, and damage to hardware [13]. In [12], this idea is combined with big data concepts to categorise attacks by (a) data on the cyber plane, (b) data on the physical plane, and (c) metadata combining the cyber and physical planes. While Wu et al. [13] focus on manufacturing systems and Wang et al. [12] only consider false data injections (discussed in detailed in Sect. 9.3.2), the principles of CPTs can be applied across the attack spectrum. Therefore, the cross-plane nature of smart grid threats should be noted for the rest of the attacks discussed in this section.

Figure 9.2 presents the categories and attacks that are discussed in the rest of this section.

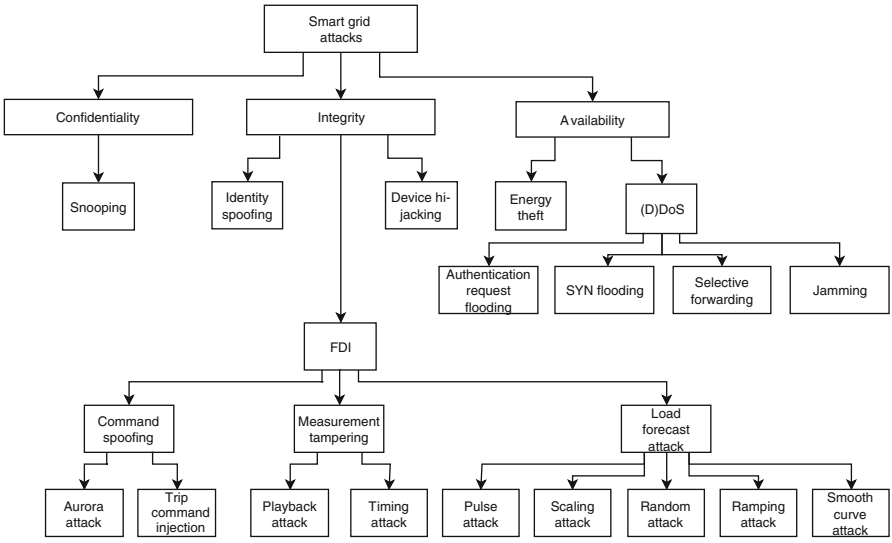

Fig. 9.2 Overview of the discussed attacks

9.3.1 Attacks on Confidentiality

Confidentiality is the quality of maintaining the privacy of data. By their nature, smart grids collect vast quantities of data that must be transmitted and processed in a secure manner. Recent regulations such as GDPR (General Data Protection Regulation) enshrine the privacy rights of users in law. Furthermore, grid devices generate very rich data, including user profiles, energy measurements, service specifications, telemetry details, and hardware information. Hence, the threat surface against confidentiality covers the whole of the smart grid infrastructure.

An example of a privacy-targeting attack is snooping. This is where malicious individuals aim to gain access to or visibility of data belonging to others. In the smart grid context, the communications between appliances, smart meters, and controllers are vulnerable to this. The power usage profiles of appliances are captured in readings and measurements made by smart meters; collectively these readings form a usage profile for the customers themselves [8]. An adversary may wish to capture this information to infer the activities and behaviours of users [8], which they may then use to plan intrusions or physical attacks on households/premises that appear to be unoccupied [8]. Similarly, such user profiling may form the basis of energy theft attacks (discussed in Sect. 9.3.3) to determine which accounts to steal electricity from with the least risk of detection [15].

Meanwhile, similar methods may also be used to infer the current topology of the smart grid. In their study of false data injections, Huang et al. [16] found that intelligence regarding the grid's structure could be mined from measurement data. To achieve this, the adversary requires some understanding of the grid's stochastic

behaviour [4], and a length of time over which to observe readings—a single set of measurements taken at one point in time is not sufficiently revealing [16]. However, using linear Independent Component Analysis (ICA) techniques, Huang et al. [16] were able to demonstrate that measurements collected over time can be used to build a model of the grid. Their approach was based on the principle that the physical topology and the load change independently. In other words, the variation in load (which changes more frequently) can be analysed given the relative stability of the topology (which changes less frequently) [4].

9.3.2 Attacks on Integrity

Integrity is the quality of maintaining the intended states of systems (and/or the data within them) so that those systems can continue to serve their intended purposes. In smart grids, the preservation of integrity ensures the timely and accurate exchange of the data signals used to make decisions about delivery and distribution. It also ensures that all grid components are truthful about their identities. This is crucial to the correctness of measurements, given that meters and sensors are numerous and distributed widely over geographic regions.

One way to damage integrity is to spoof the identities of grid components. This is where someone other than the legitimate device fraudulently claims to be that component, thus allowing an adversary to interact with the system under false pretences. For example, smart meters may be spoofed to send fake data to controllers [12, 15]. Similarly, spoofed devices can send incorrect timestamps to PMUs, disrupting grid synchronisation [17]. Another method is device hi-jacking. This is slightly different to spoofing because while the compromised device can be wielded by a potential attacker, its identity is still intact. The primary version of this attack is the recruitment of grid devices into a botnet. A bot binary is injected into devices via a virus or a worm [18]. This binary then automatically executes and connects to a remote command and control (C&C) network from which it receives attack instructions. Adversaries may also harvest data from devices via the same C&C network. Botnets, which provide foundations for other types of attack, are a known threat to WSNs (Wireless Sensor Network) and IoT networks [11]. A prominent example is the Mirai botnet, which hijacked IoT routers and cameras and was used to launch massive-scale DDoS attacks in multiple countries [11].

The biggest threat to smart grid integrity comes from False Data Injection (FDI) attacks. As the name suggests, this involves the introduction of maliciously crafted data into sensitive communication streams, with the aim of manipulating system outputs. Hence, FDI attacks are mainly targeted at data-reliant management processes [2, 4, 19, 20]. They may be considered analogous to man-in-the-middle attacks. Some possible scenarios explored in literature include attacks on state estimation systems [19], the EMS [4], AMIs [21], SCADA systems [22], local systems with clustered measurement hierarchies [23], and in wind farms [2]. Generally, attackers engaging in FDI will compromise a subset of grid components

but will not have visibility of the whole grid given its complexity and size [4]. However, due to the hierarchical structure, an FDI at one point in the network is still capable to causing widespread repercussions. Additionally, Anwar et al. [21] found that the impact of an FDI changes based on the characteristics of the targeted nodes, while Chen et al. [4] suggested that sophisticated script-based FDIs can learn the best injection approach through trial and error. This shows that even with incomplete information, this type of attack has great potential for damage and disruption.

FDI attacks are typically modelled using the formulation $z = Hx + a + n$ [2, 22, 24], where z is a set of measurements, x is the state vector (or bus voltage phase [24]), n is the measurement error or environmental noise [2, 24], and H is a Jacobian matrix of measurements that describe the current grid topology [2, 22]. Together these values determine the state of the grid. Then the attack vector a (describing the fake data and variables targeted) is added [2, 22, 24]. Attacks can be classified as normal or stealthy. For the latter, it is assumed that adversaries have some visibility of H, which allows them to set up their attacks intelligently so as to avoid threshold-based detection (i.e. residual test) mechanisms [2, 16, 19, 20, 22, 25]. Additionally, they may aim to manipulate the state variables corresponding to the targeted measurement variables to avoid noticeable anomalies [21]. An alternative approach is given by Chen et al. [4], who modelled FDIs as partially-observed Markov decision processes (POMDP), where the focus is on attackers (who have limited target visibility) aiming to learn the optimal setup [4].

The rest of the integrity-based attacks discussed in this section is specific sub-categories of FDI as identified in the smart grid literature.

Command spoofing is the intersection of identity spoofing and FDI attacks; fake data—styled as commands—is injected into the network, claiming to come from legitimate sources. Aurora is a type of command spoofing attack that targets the circuit breakers used to determine grid topology and the generators that they serve [26]. Specifically, fake control signals are sent to the breakers, instructing them constantly open and close at a high speed [5, 17]. Eventually, this causes the associated generators to desynchronise from the rest of the grid [17]. If a critical level is reached, this attack can cause physical damage to the generators [17], knocking them offline. Depending on the degree of physical damage and the positions of the affected breakers and generators, Aurora attacks can result in a significant drop in a smart grid's functional capacity and efficiency.

Another example of command spoofing is trip command injection attacks. These target protection relays (devices designed to respond to faults in power transmission lines) with fake relay trip commands, causing circuit breakers at the ends of transmission lines to open [5]. When this happens, additional strain is placed upon secondary transmission lines, as the system tries to meet demand [5]. Given the hierarchical nature of smart grid infrastructure, this can then result in cascading failures [5] and large-scale power outages. An alternative involves the disabling of relays so that faults do not trigger trip commands at all [5]. Meanwhile, fault replay attacks combine fake trip commands with fake transmission line faults [5]. To achieve this, measurements are altered to look like real-life faults either via

some hijacked devices or through data injections [5]. These false readings cause controllers to make incorrect management and distribution decisions.

Attacks on integrity may be directed at specific devices in the grid. An example is against PMUs, as explained by Wang et al. [12]. These devices are used to measure and synchronise phasor values collected from distributed sensors and meters; these measurements are then used to perform state estimations. FDIs can be applied directly to PMU data to manipulate the resulting state estimations [12]. Examples of how this may be achieved include playback attacks (where captured data is played in reverse order, giving incorrect readings) [12] and time attacks (where captured data is sampled at varying rates, distorting the true readings) [12].

Attacks may also target specific functionalities. For example, load forecast attacks hinder the grid's ability to determine where to distribute power and the correct load [27]. This is achieved using data injections designed specifically to distort these forecasts, applied continuously for the duration of the attack [27]. In implementation, there are several variations defined by Cui et al. [27]. Pulse attacks change the forecast values at regular intervals to be higher or lower than the true reading [27]. Scaling attacks tamper with values by multiplying them by a scaler [27]. Random attacks insert randomly-generated positive values [27]. Ramping attacks use a ramping function to either increase values over time ("up-ramping") or increase and decrease values repeatedly ("up and down-ramping") [27]. Finally, smooth curve attacks change forecasts' start and end points [27]. Given that each of these approaches causes a different impact on controller behaviours, adversaries are able to fine-tune attacks to suit their specific goals.

9.3.3 Attacks on Availability

Availability is the quality of maintaining the accessibility and functionality of a system to a satisfactory degree at all times. Power is a basic utility, and so power grids are fundamental parts of urban and rural infrastructure. Furthermore, smart grids require efficient feeds of real-time data and a high level of responsiveness from all components (e.g. controllers, synchronisers, smart meters, and sensors) for accurate decision making. In other words, components must have high availability for the grid's internal functionality.

The primary attack against availability is Denial-of-Service (DoS). This is where an attacker generates lots of traffic to overwhelm the capacity of target devices, causing them to crash and hence, rendering the services they provide unavailable. When this flood of traffic comes from multiple distributed sources, it is known as a distributed DoS (DDoS). Smart grids are highly susceptible to such attacks because they (a) house a large consumer device population [18], (b) consist of many low-power devices, and (c) have a hierarchical infrastructure [11]. This indicates a large potential attack surface of low-capacity devices, and many centralised control points to target. Devices may be compromised physically, have their identities spoofed, or engage in DDoS as part of a botnet [18]. As DoS attacks inject lots of malicious

data into the network, they may also be considered a form of FDI [28]. Typically, attacks take place between sensors and smart meters [18], or between smart meters and system controllers [3].

For example, sensors may be manipulated to send streams of malicious authentication requests to meters [18]. In the process of trying to verify request details, the capacity of the meter is exhausted, and so the authentication service is knocked offline [18]. A similar attack between meters and controllers would disrupt the collection of measurement data, causing controllers to make the wrong decisions about power management and distribution [3]. Smart grids are also vulnerable to typical application layer DDoS attacks like SYN flooding [29]. This is where 3-way TCP handshakes are intentionally left half-open (because the client never responds the server's SYN/ACK message), consuming server resources and causing their backlog queues to fill up so that new, legitimate requests are automatically dropped [29]. Choi et al. [29] demonstrated such an attack on the multicast-based communications of smart grid IEDs (Intelligent Electronic Device). Other documented examples of DDoS in the smart grid are those originating from buffer overflow attacks (where program code is tampered with) [29] and selective forwarding (where packets relating to a particular service are consistently dropped) [28].

Smart grids are based on wireless sensor technology, which uses broadcasting on open channels to enable the easy exchange of data between geographically distributed devices [30, 31]. This makes grids vulnerable to a special type of DDoS attack known as jamming, where attackers add random noise signals to wireless channels to corrupt the traffic exchanged between grid components [3, 31]. As with standard DDoS, this attack can disrupt traffic between appliances and meters or between meters and controllers [31]. In both cases, the accurate gathering of measurement data is denied, and where this causes controllers to make incorrect calculations about load, large-scale outages and mismanagement may result [31]. Additionally, jamming attacks may be easier to perform than conventional DDoS because they do not require a base of compromised devices to launch them [31].

To perform the attack, a jammer device or program selects a channel and then injects it with random noise [3, 30]. This is similar to the injection of fake requests into the network in DDoS. Generally, there are four jammer types identified in WSN literature, sorted into two categories. In the first category are "oblivious" jammers, i.e. those which operate only based on current channel availability [30]. These are static jammers (which always use the same channel) and random jammers (which switch channels randomly over time) [30, 31]. The second category consists of "intelligent" jammers, i.e. those that use historical data to make complex decisions [30]. These are myopic jammers (which learn users' channel usage patterns) and Multi-armed Bandits (MABs) (which use machine learning to predict user behaviours) [30]. In some cases, myopic and MAB jammers may be considered as one [31]. As suggested in [30], jamming attacks may be kept hidden by avoiding the use of licensed channels.

The attacks discussed so far cause disruptions in power distribution services by affecting particular functionalities provided or required by the grid. However,

the availability of power can also be attacked directly through resource exhaustion attacks and energy theft. In resource exhaustion, adversaries demand large quantities of electricity by sending many requests in quick succession [28]. This maximises the amount of power drawn from the grid and can eventually lead to the depletion of the available energy [28]. Such an attack can feasibly be launched at the appliance level by malicious "consumers" using energy-inefficient or high-consumption devices [28]. Meanwhile, energy theft involves the consumption of power without providing proper compensation for this service [15]. There are three types of energy theft: those launched by malicious consumers, those launched by industry insiders, and those conducted by organised criminals [15]. Malicious consumers may tamper with their appliances and meters to avoid making payments as due [15]. Industry insiders (i.e. utility company employees) may manipulate internal records and readings [15], either for their own benefit or as part of a larger campaign. Lastly, organised crime syndicates may use both of the previous methods to syphon energy from paying customers to sell illegally [15] or for further criminal activity.

Resource exhaustion and energy theft tend to occur alongside attacks on integrity (like FDI, tampering, and spoofing) and attacks on confidentiality (like user profiling). For example, the energy theft process requires a disruption of the communications to and from smart meters. This prevents the grid from learning consumers' energy usage levels. Then smart meters can be spoofed, and fake readings be sent to controllers [15]. To prevent tracing, existing audit logs and records may also be deleted from the meters [15]. Meanwhile, criminals targeting other consumers can use profiling techniques to infer their usage patterns from sensor data, allowing them to plan out their attacks.

9.4 Attack Detection

The area of attack detection has been driven by some core concepts that must be considered when this kind of solution is developed or deployed. The first one is the detection method, which is essential because it defines which situations the detection system sees as an attack. Alongside this, the types of data collected and the system distribution are also pillars of attack detection. They can impact various functions, including the system's processing performance and the attacks that can be detected. Figure 9.3 outlines these concepts and the related techniques, which are discussed in the context of smart grids in this section.

9.4.1 Detection Methods

Smart grid defence incorporates classical intrusion detection methods, broadly categorised as signature-based and anomaly-based. In the former, systems use templates derived from historic attack instances to recognise new instances of the

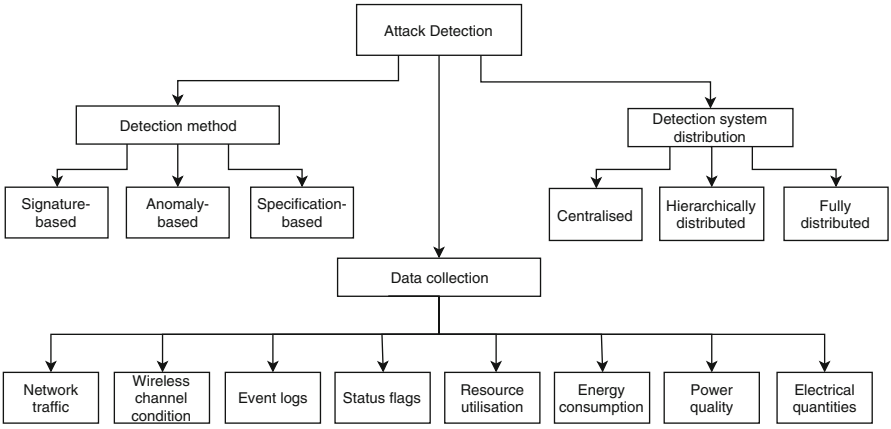

Fig. 9.3 Overview of attack detection core concepts and related techniques for smart grids

same or similar attacks. Typically, each attack on a system or network consists of a particular sequence of actions. These activities, observed in succession, can, therefore, be used to identify malicious activity. Signatures may be stored in a database and updated as new attacks are discovered. Popular attacks with generally well-understood methodologies (such as DDoS and FDI) may have strong patterns that make it easy and appropriate to model them [28]. Furthermore, one-to-one matching makes signatures very effective at detecting particular attacks. However, previously-unseen attacks will be missed, causing false negatives.

In contrast, anomaly-based detection observes the network or system for activity that deviates from a pre-defined norm. These norms may be derived from empirical baselines or heuristics. Given their malicious nature, many attacks fall outside the standard behavioural profile for systems or services, which results in unusual fluctuations in activity. Unlike the use of signatures, anomaly analysis is not limited by knowledge of historic incidents and is, therefore, capable of identifying day-zero attacks. Smart grids provide a wealth of data that can be used to generate complex and detailed activity profiles. However, not every anomaly corresponds to an attack attempt, and so a large number of false positives may be generated. This is especially true for smart grid sensors which are influenced by environmental factors. For example, high temperature readings may represent a sabotage attempt or may simply be caused by hot weather [32].

Novelty detection methods may be particularly useful for anomaly-based systems, once they can make the detection system reliable against previously-unseen attacks. The aim of these methods is to identify events that differ from the previously available data. New attack patterns, which were not present in the model induction, are placed out of the cluster of known patterns, and categorised as a novelty [33]. Novelty detection with multiple known classes is also widely applied in data stream classification, meaning that new classes may appear and previously known classes

may change [34]. The most traditional methods are based on clustering algorithms, such as k-means [35].

A more recent addition is specification-based detection. This is similar to anomaly-based approaches, but the baseline profile used is created to reflect the expected behaviours of a particular application or protocol [11, 26]. Each is given its own set of expected behaviours and the detection system flags up instances where related activity deviates from this set. This overcomes the limitation of signature-based detection because unknown attacks may be discovered. In addition to this, the granular definition of expected behaviour can improve upon the false positive rates of anomaly-based detection. This approach applies well to metering systems in smart grids where state monitoring is an integral process and there are known thresholds for safe operation [5, 11, 26]. Conversely, the need to generate multiple specifications may make specification-based detection unscalable in larger grid networks [5]. An alternative approach is to shift focus from expected behaviours to the characteristics of a medium to determine the best attack vector [36]. This can be difficult to achieve in complex cases but enables pre-emptive actions.

9.4.2 Data Collection

Attack detection is essentially a data-driven process. No matter which detection method is followed (anomaly, signature or specification-based), the attack detection system always gathers data from the protected system, analyses it, and determines whether there is an ongoing attack. Smart grids offer several data sources that attack detection systems can use.

Multiple features can be extracted from network traffic data [14, 29]. The contents of protocol headers and payloads, rate of packets of a particular type, number of malformed packets, time of packet round trips, average packet size, and average volume of bytes per second are all examples of information that can reveal some change as a consequence of an attack. Network packets can be gathered at different points in the smart grid's networks, but it is important to consider that this choice will define which types of attacks can be detected. For instance, if the attack detection system analyses the packets carrying measurements from smart meters to the control centre, it will be certainly able to detect attacks involving the smart meters or the AMI, like a DDoS. However, if only packet statistics are checked, but not the payload content, an FDI will be hardly detected. As many smart grid sites will be located in remote locations, wireless technologies are good candidates to connect these sites to the rest of the network. With this in view, measuring wireless channels' conditions may be useful to detect some attacks like jamming. A possible way of measuring a channel condition is to transmit control packets in selected channels and wait for ACK packets to analyse the channel performance [30]. Signal Strength Intensity (SSI) of smart meters and data concentrators can also be checked to detect jamming attacks [28]. If the sensed SSI is much higher

than an expected value, it can indicate a rogue or compromised device trying to jam legitimate transmissions.

Data about the status and events related to different devices in the smart grids can also be a good source of attack indicators. A power grid has several specialised devices (e.g. breakers) deployed across the system to control its operation and make sure that it is running correctly. Status flags collected periodically from these devices can reveal unexpected behaviours linked to command spoofing attacks, for example [26]. Several logs can also be processed to identify events that can help to uncover an attack episode [5]. Relay logs provide information about events involving breakers, while the control panel log can show whether there was scheduled maintenance for a particular grid component. Logs from Snort, a signature-based intrusion detection system for TCP/IP networks, can also be used to detect the presence of packets with some particular characteristics in the system. For example, it is possible to create a rule in Snort to trigger an alarm every time a packet carrying a command to change a breaker status is detected. Lastly, utilisation of hardware resources such as memory and CPU can also be monitored at the protected devices, as DoS/DDoS attacks may cause sudden changes in these measurements [29].

The data sources presented so far are particularly related to the operation of smart grids' infrastructure, such as breakers, relays, and networking devices. They encompass network packets, status flags, event logs, and resource utilisation data related to the daily routine of these devices. Smart grids also have another rich source of insights for attack detection: domain-specific data. Data about energy consumption and electrical quantities are already collected in real-time at multiple points of the grid for management and control purposes, and attacks (such as FDI and energy theft) can cause subtle but detectable changes in their behaviour.

FDI attacks typically affect the system state estimation. Therefore, measurements such as active and reactive power, current flow, voltage magnitude, and phase angles that feed the state estimation are used to detect these attacks [2–5, 12, 16, 19, 20, 22–26]. SCADA systems are usually employed to collect these measurements [4, 19, 22]. RTUs at different points in the power grid transmit these measurements every 2–5 s [19] to targets like the control centre. PMUs are also applied to collect this kind of data and send it to the control centre [5, 12, 23, 26]. They are much more precise than SCADA systems, reaching a sampling rate of 2880 samples per second [12]. However, this precision comes with a high computational cost, which poses a great challenge to management and control activities, including attack detection [5].

Alongside PMUs and SCADA systems, AMIs are also dedicated to collecting domain-specific data from smart grids for management and control. The main elements of AMIs are the smart meters, which are deployed at the customer end to send data related to energy consumption, power quality, and pricing to the utility provider [11]. Among them, energy consumption data has been used to detect energy theft [15], DoS [11], and FDI attacks [21]. The latter case is of particular interest here because it shows that AMI data improves state estimation, which usually takes energy consumption forecasts as input, instead of real consumption data. However, if the grid has an AMI, real-time energy consumption data is available and, consequently, consumption forecasts can be replaced by real data during state

estimation. It is worth noting that due to the huge number of customers and frequent collections, control centres face significant difficulties in storing and processing smart meter data [15].

There are yet other data sources that, despite being used infrequently, can also be helpful in attack detection. Some attacks can be directed to load forecasting data, which is important in enabling operators to foresee the system's conditions and get ready for upcoming events. Hence, data that is typically used by feed load forecasting processes, such as historical load data, weather data (e.g. temperature and humidity), and time data (e.g. time of the day and day of the week), becomes useful for attack detection [27].

9.4.3 Detection System Distribution

Attack detection is a process composed of data collection, system profiling, detection, and response. As smart grids are huge and hierarchically-arranged systems, data collection is usually distributed. In the case of detection based on network traffic data, packets must be sniffed at multiple points of the several networks that compose a smart grid communication infrastructure. Unlike traditional enterprise networks, where there is often a single point of connection to an external network (e.g. the Internet) which is monitored for attack detection, smart grid networks present a wider attack surface with several points to monitor. Likewise, when system logs are used as a data source, there are different critical systems to be monitored and, hence, multiple data collection points. Approaches based on SCADA systems or PMUs are naturally distributed in terms of data collection, as there are always several RTUs and PMUs deployed over the power grid to perform their primary function: controlling and monitoring the power grid.

The goal of system profiling changes according to the detection method. For anomaly and specification-based approaches, system profiling is responsible for building the notion of which activities are normal, while for signature-based ones it specifies what defines an abnormal activity. Then, during the detection task, data collected in real-time is analysed according to the knowledge built during the system profiling, and the response takes place when an attack is detected. The response can vary, from alerts that are presented in a management console to an action to mitigate the attack, such as blocking the attacker's access to the protected asset. For simplicity's sake, system profiling, detection, and response can be summarised with a single term: decision making.

In smart grids, decision making in attack detection can be centralised, partially (hierarchically) distributed, or fully distributed. For the rest of this chapter, when used to classify a detection approach architecture, the terms centralised, hierarchically distributed and fully distributed will refer to how decision making is performed.

In centralised architectures, all of the data collected is transmitted to the control centre, where decision making is performed. Attack detection approaches based

on data used for state estimation and event logs are usually centralised because this data is already transmitted to the central control as part of other control and management tasks [3, 12, 16, 19, 22, 26]. In these scenarios, data processing may rely on tools such as Hadoop, which runs in distributed computing environments [12]. Nevertheless, decision making is still centralised, meaning that there are no multiple instances concurrently determining whether an attack is occurring. Indeed, the main challenge for centralised architectures is to cope with the huge volume of data to be stored and processed, as smart grids have multiple data collection points operating at high sampling frequencies.

Hierarchically distributed architectures are directly linked to the hierarchical topology of smart grids. As mentioned in Sect. 9.2, smart grid networks are organised in three levels: HAN, FAN (or NAN), and WAN. Hierarchically distributed architectures seek to spread the decision-making process across network levels and, at the same time, keep some degree of central coordination. In other words, they transfer part of the burden of storing and analysing huge amounts of data from a central point to multiple points, while ensuring that a central element supervises attack detection. Hierarchically distributed solutions can rely on network traffic data [14], as well as on domain-specific data [11, 28]. Detection system agents are placed to monitor the communication traffic or the data collected from smart meters in HANs, FANs, and the WAN. Attacks detected in a particular network level can be checked in the next level up. For example, an attack detected in a smart meter after analysing data collected in a HAN can trigger an alert, that is sent to the detection system of the FAN where this HAN is connected. Then, that FAN's detection system analyses these alerts before confirming them [28]. Likewise, when a detection agent cannot decide if an attack is occurring based on the data it has, the decision can be passed up to the next level [14].

Alerts are not the only information that detection agents in lower levels send to their counterparts in upper levels. In some cases, the lower level detection agent can forward high-level statistics (e.g. a measure of anomaly evidence [11]) to the upper level agent. It is important to note that devices in lower network levels can face difficulties in hosting computationally costly processes because they are usually resource-constrained. A smart meter, for instance, may not be able to host a detection agent that runs a machine learning classifier. On the other hand, lower level monitoring can offer more detailed data for attack detection, particularly in situations where devices at the lower level are the targets. Therefore, those responsible for designing hierarchically distributed architectures have to consider this trade-off between computational capacity and data granularity.

Fully distributed architectures are not frequently proposed because the absence of any central control is seen as incompatible with the level of reliability demanded from smart grids. However, a fully distributed solution may be applicable in some specific situations. For example, to defend jamming attacks, distributed agents can be responsible for sensing communication channels and pointing out which ones are free from jamming attacks and, consequently, more suitable for transmission [30, 31].

9.5 Machine Learning

Machine Learning (ML) is a field that emerged from artificial intelligence and involves the use of algorithms belonging to different categories, such as supervised, semi-supervised, and unsupervised learning. In this chapter, we use the term ML to refer to algorithms where computers learn how to process their inputs, without this being explicitly implemented. In other words, with ML, computers are able to perform a task by making use of inference or based on observed patterns, and not by relying only on instructions that specify clearly how it should be done. Due to the vast number of ML algorithms, we focus on the widely used and most accurate ones in the smart grid field.

An ML model is a mathematical model that receives the description of a given problem as an input and delivers a generated solution as an output. This model is constantly updated by a data-driven induction towards making reliable predictions or decisions. Most ML models are induced using supervised algorithms, which demand a dependent variable. In classification problems, the dependent variable, commonly referred to as a label, is linked to the problem class of a given sample. For instance, to induce a supervised classification model to be embedded into a smart grid attack detection system, various examples of the attacks to be detected are needed, alongside instances of non-malicious behaviour. These examples are presented to the algorithm along with their respective labels, which inform the example's class. Examples and labels are then used to build a model capable of classifying new instances.

In an attack detection scenario, the classification problem is usually modelled as a binary classification task, as it supports two opposite classes: normal or anomalous behaviour. However, in some cases, there are more than two classes to be predicted, defining a multi-class problem. Multi-class detection systems are generally focused on identifying specific cyberattacks (energy theft, jamming, DoS, and FDI), supporting efficient countermeasures to minimise their damage and to combat the attack source. The most widely used algorithms for cyberattack detection in smart grids, for both binary and multi-class problems, are Support Vector Machine (SVM) [37], Artificial Neural Networks (ANN) [38], k Nearest Neighbours (kNN) [39], Naive Bayes (NB) [40], and Random Forest (RF) [41].

SVM is an algorithm developed to find a hyperplane in high-dimensional space from training samples, while attempting to maximise the minimum distance between that hyperplane and any training sample according to its class. The model (hyperplane) obtained by SVM is used for predicting new unlabelled samples. The default hyperparameters of SVM are the regularisation parameter (C) and the kernel. Some kernels, such as radial basis function (RBF) and polynomial, require additional hyperparameters.

The usage of ANN has been boosted by the advent of deep learning approaches. Its inducing architecture is based on connected artificial neurons used to simulate the learning process of a biological brain. From the shallow (Perceptron and Multilayer Perceptron) to the deep learning structure (Long Short-term Memory and Convolu-

tional Neural Networks), ANN has several architectures and hyperparameters to be defined.

Unlike SVM and ANN, kNN is a simple machine learning method, which predicts new samples based only on the distance between a given sample and the training pattern. Using the k hyperparameter as the number of relevant neighbours, kNN classifies a new sample based on its closest training examples in the feature space. Another simple ML algorithm is NB, grounded on the assumption of independence among features for modelling a classifier. Although the conditional independence premise is rarely true in most real-life applications, NB generates competitive models in practice. The reason for this is that an NB classifier will be successful as long as the actual and estimated distributions agree on the most probable class, regardless of feature independence.

RF is an algorithm based on classification trees. More precisely, it is an ensemble of decision trees created through bagging strategy, which combines multiple random predictors to generate its final result. RF presents some important advantages, such as the ranking of features and the reduced possibility of overfitting. Furthermore, as hyperparameters, it requires only the number of decision trees and the number of variables available for splitting at each tree node.

In addition to classification, another important application of supervised ML algorithms is grounded on regression problems. Instead of using categorical outputs (i.e. dependent variables), regression problems require the prediction of continuous values, e.g. power flow in smart grids. In this scenario, the attack detection relies on a threshold-based strategy and statistical control techniques, such as Cumulative Sum (CUSUM) [11, 16, 17]. CUSUM follows the premise that an attack modifies the typical behaviour of the evaluated stream. Detection based on CUSUM is usually performed by computing some stream signatures such as mean value, root mean square value, peak values, amplitude probability density function, rate of signal change variations, and zero crossings per unit time. When one or more signatures are changed, the cumulative sum is computed for detecting an increase in the mean value of a sequence of independent Gaussian random variables. If the CUSUM's value exceeds a threshold, an attack is characterised. The success of attack detection depends on proper hyperparameters setup, which are related to the tolerance interval, the probability of false alarms, and the detection delay from the observed stream.

In several cases, the dependent variable is not completely available due to the cost or complexity of its extraction. Attack detection in smart grids is an example, as it typically suffers from scarcity of labelled data. In smart grid scenarios, there are several security behaviours that should be simulated, studied, and stored to produce labelled data for supervised learning. Thus, when it is challenging to obtain labelled data, we can employ semi-supervised approaches [42].

Unsupervised clustering is a third category of machine learning approaches applied to smart grid attack detection. More precisely, it is tailored for scenarios with a total absence of labels [43]. The most used algorithm of this category is k-means, which, coupled with a heuristic algorithm (e.g. Particle Swarm Optimisation or PSO), is able to assume the likely number of clusters (k value) required to properly

distinguish attacks from normal situations. The k-means algorithm follows the premise that all evaluated instances can be associated with one of the k clusters. As the instances are grouped according to their behaviour, attack and normal instances will be clustered into different partitions. More recently, Bayesian clustering has also been used to address smart grid problems. For instance, Dasgupta et al. [44] made use of techniques from elastic shape analysis along with a Bayesian approach to cluster and evaluate electricity consumption curves according to their shapes.

Supervised, semi-supervised, and unsupervised categories cover most of the algorithms applied to cyberattack detection on smart grids. However, the current call for real-time (online) detection and high predictive performance paves the way for more recent paradigms of machine learning. On the one hand, dealing with real-time detection, we have the Hoeffding Tree (HT), an incremental decision tree for high-speed data streams classification [45]. On the other hand, focused on leveraging high predictive performance, we have Deep Learning (DL) methods [46].

HT is a supervised ML algorithm designed to induce models online in an incremental way (i.e. instance-by-instance) based on anytime learning as required for data stream processing. Therefore, the usage of HT for attack detection consists of the induction and classification of data flows from smart grids without *apriori* knowledge, i.e. there is no offline phase to train a model. Unlike offline learning, which assumes that all training data needed to create a model is already available, online learning assumes that new data can arrive at any time, which can make a model outdated [47]. Like in traditional ML processing, data stream mining can be performed by supervised and unsupervised algorithms. Considering high-speed stream scenarios, which may be the case for smart grids, unsupervised approaches have been reported as more feasible. In [48], an online unsupervised clustering algorithm was used for load profiling. The proposed approach takes advantage of the stream structure of the data, keeping the identified profiles updated in accordance with newly collected data. It is worth mentioning that the kernel of the proposed solution is based on the k-means algorithm.

In recent years, DL methods have drawn academic and industrial attention. These methods are grounded on discovering the intricate structure of inputs to learn representations of data with various levels of abstraction. Among all DL methods, some deep variants of multi-layer perceptron (MLP) were used to detect FDI. In [19], deep MLP was applied to identify attacks on smart grids using active power-flows, active power-injections, reactive power, and voltage measurements as features to induce the DL model.

All the algorithms, methods, and categories mentioned so far in this section are applied as unmixed or hybrid approaches when addressing attack detection in smart grids. Most of the works surveyed are designed as a pipeline composed of steps such as pre-processing, feature selection, and ML predictive algorithms. A hierarchical overview of ML algorithms and their combination is presented in Fig. 9.4.

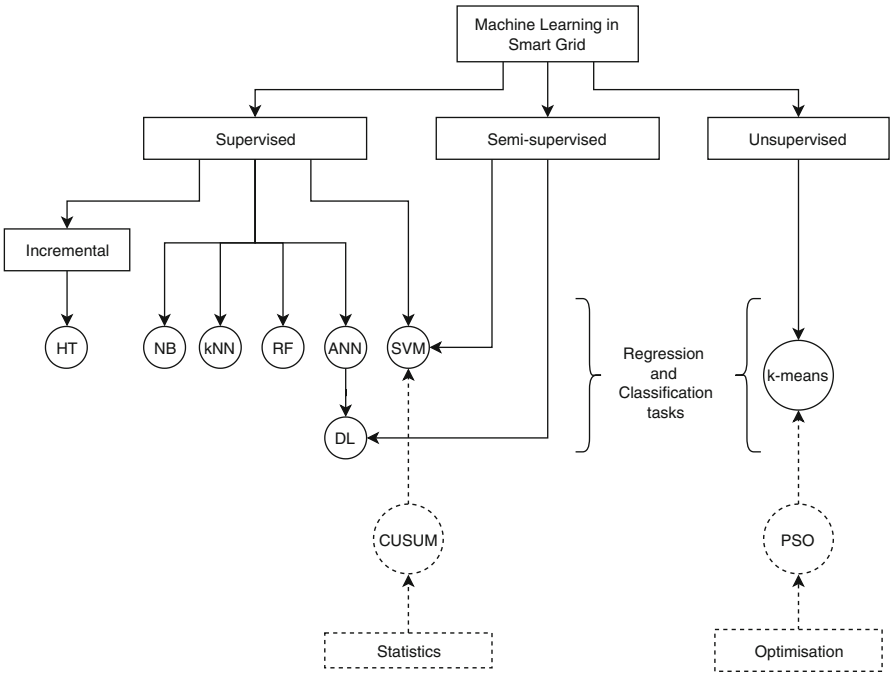

Fig. 9.4 Overview of machine learning algorithms (Hoeffding Tree, k Nearest Neighbours, Random Forest, Artificial Neural Networks, Deep Learning and Support Vector Machine) and their correlations with Statistics (Cumulative Sum) and Optimisation (Particle Swarm Optimisation) techniques

9.6 Existing Solutions

Solutions proposed in the literature for cyberattack detection in smart grids are diverse, which reveals the many decisions a researcher has to make when developing a new approach. Which attacks will be addressed, which data will be collected, how to distribute the system, and how to combine machine learning techniques are among the questions that must be answered. This section presents a literature review of proposals to tackle cyberattacks in smart grids using machine learning, providing a discussion of how the different authors addressed these issues.

FDI is the most addressed threat in works on attack detection in smart grids. Its high potential of disrupting smart grid operations is probably the leading cause for this concern. To identify both standard and stealthy FDIs, Huang et al. [16] contributed a centralised anomaly detection scheme applied to state estimation data. They define Gaussian-based vectors for observation and for an unknown data injection (commencing at a random time) with the aim of identifying the change in the observation vector's distribution from the idle state to the attack state. Based on the time of detection, the average run length (ARL) is calculated. Then, the detection

time and a threshold h are used in multi-thread processing (with a linear solver and Rao tests) to solve the problem recursively. The alarm is raised when a cumulative score reaches h. Based on tests on a 4-bus environment, the authors reported that the best detection is achieved at higher ARLs, and that the value of h influences the timeliness of detections.

Qiu et al. [49] addressed FDI attacks as part of their investigation into the application of cognitive radio networks for smart grids. They proposed the centralised use of Independent Component Analysis (ICA) to overcome FDI attacks, characterising them as instances of high interference. A data transmission matrix Z is defined, which contains a matrix X of source signals originating from smart meters. The aim is to fill out X with signal estimations. To do so, the attacker's signals must first be filtered out. This is achieved using the statistical properties of signals, with Principal Component Analysis (PCA) used to deal with differing power levels in the interference. Through simulations, the authors were able to demonstrate that ICA can effectively separate different signals.

Esmalifalak et al. [20] also used PCA to tackle FDI attacks but in a different way. The authors proposed two methods to detect stealth FDI attacks. The data used to detect the attacks consisted of measurements for state estimation, which is collected from multiple points in the power grid. These data can present some redundancies, and the number of dimensions in the detection problem is linked to the power system size. For instance, a 118 bus system used in this work for tests generated 304 dimensions. In the two methods, to avoid the curse of dimensionality, the authors employed PCA. Their underlying assumption is that normal data is generated according to physical laws, while tampered data is not, so these data should be separated in the projected space. The first detection method was based on a statistical anomaly detection technique, which made use of a threshold learnt from historical data. It is centralised and relies on data sent to the control centre. The second method was built upon distributed SVM, which, unlike the first method, requires labelled data from both classes (normal and attack) for training. To test their system, they used an IEEE 118-bus test system.

Ozay et al. made extensive use of machine learning classifiers to detect FDIs in two different works [23, 24]. In the first one [24], they proposed a centralised signature-based method for state estimation data. Supervised learning is used to classify samples as either "secure" or "attacked". Three machine learning algorithms were used: kNN, SVM, and sparse logic regression (SPR). kNN sorts feature vectors into neighbourhoods based on Euclidean distance. SVM identifies hyperplanes for the binary splitting of samples. SPR uses Alternating Direction Method of Multipliers (ADMM) with distributions for labels-to-samples matching. Testing on IEEE 9-, 30-, 57-, and 118-bus systems revealed that both kNN and SVM are negatively impacted by data sparsity, unlike SPR [24]. The authors suggested that kNN is suited for smaller systems, and SVM and sparse logistic regression for larger ones.

In [23], Ozai et al. presented a thorough study exploring multiple families of machine learning algorithms, including supervised, semi-supervised, and online machine learning. In their evaluations, the authors employed feature-level fusion

and ensemble methods. IEEE 9-, 57-, and 118-bus test systems were used again in the experiments. The authors pointed out that semi-supervised algorithms were more robust against sparse data than supervised ones. Also, feature-level fusion and ensemble methods were shown to be robust against changes in system size. Lastly, the performance of online classifiers was comparable to batch ones.

Yan et al. [25] also presented a comparative study exploring multiple machine learning classifiers. In their work, three supervised algorithms, namely SVM, kNN, and eNN, were used to detect FDI attacks on measurements for state estimation. The authors considered balanced and imbalanced cases, and analysed the impact of the magnitude of false data in the detection performance. Tests were based on the IEEE 30-bus test system, with the SVM classifier obtaining the best overall results.

In [12], the authors described the additional challenge of detecting FDIs in the vast amounts of data collected in smart grids. Based on experiments using a 6-bus power network in a wide area measurement system environment, these authors proposed a Margin Setting Algorithm (MSA). The proposed algorithm was compared to the SVM and ANN algorithms in a binary classification scenario for detecting playback and time attacks. Results demonstrated that the MSA achieved minimal errors and better accuracy than traditional machine learning algorithms with conventional hyperparameters.

Unlike the works presented so far, Hink et al. [26] approached FDI detection with three different classification schemes. They aimed to study the performance of multiple machine learning algorithms in distinguishing power system disturbances as malicious or natural. In the first classification scheme, each type of event was modelled as a class, meaning that it was a multi-class problem with 37 distinct classes. The second scheme took into account three classes: malicious event, non-malicious event, and no event; the latter corresponds to data related to normal operations. The last classification scheme had two classes: attack or normal. Seven machine learning algorithms were tested, namely OneR, NNge, RF, NB, SVM, JRipper, and AdaBoost. Although the results varied significantly for different classification schemes, the authors pointed out the combination of Adaboost, JRipper, and the 3-class model as the best solution among the studied ones.

Neural networks were also used as a feasible solution for FDI detection. Hamedani et al. [2] made use of Reservoir Computing (RC), an energy-efficient computing paradigm grounded on neural networks. The proposal was implemented by combining DFN (Delayed Feedback Network) and MLP to support spatio-temporal pattern recognition. Considering wind turbines as the major source of electrical power generation, collected measurements (i.e. temporal data) were encoded as feature vectors to be the input of the binary classification task, which distinguished instances between normal or under attack. Simulation results showed DFN + MLP could detect attacks under different conditions, such as different magnitudes and number of compromised meters, overcoming the performance of single MLP and SVM algorithms.

Another solution based on neural networks for detecting FDI was proposed by He et al. [22]. In this work, CDBN (Conditional Deep Belief Network) was explored to

extract high-dimensional temporal features for recognising the differences between the patterns in data compromised by FDI attacks and in normal data. The system architecture, composed of five hidden layers, obtained the best results when comparing three different numbers of layers. In comparison with SVM (Gaussian kernel) and ANN (MLP with a single hidden layer), CDBN achieved superior accuracy, followed by MLP and SVM. The authors claimed that they performed online attack detection, but it is important to mention that the training and updating were performed offline.

Following a different path, the proposal by Adhikari et al. [5] is based on pure online modelling. The authors proposed the use of the Hoeffding Tree coupled with a mechanism to handle concept drift when classifying binary and multi-class power system events and cyberattacks. A total of 45 classes of cyber-power issues were addressed using a combination of attributes from power and network transactions (such as voltage, current, and frequency), and logs from Snort. The authors put effort into tuning all the algorithms and deployed real-time analysis with a high level of accuracy. The main advantages of the proposed method are related to consuming less memory than traditional batch processing, as well as providing real-time analysis to classify a broad number of power system contingencies and cyberattacks. However, HT is a supervised machine learning algorithm, which depends on a labelling step. This becomes a pitfall for real-time applications.

Semi-supervised learning methods are an attractive alternative to ease the need for the labelling step. In [50], cyber-physical attacks on power systems were addressed using Reinforcement Learning (RL), more precisely a Q-learning semi-supervised algorithm. A contingency analysis system was proposed to handle sequential attacks in power transmission grids, such as blackout damage and hidden line failures. Based on simulated study cases with IEEE 5-bus, RTS-79, and 300-bus, it was possible to discover a more vulnerable target sequence in sequential attacks. Furthermore, when varying the blackout size and topology of attacks, the proposed solution was capable of reducing the number of successful attacks by excluding failed attack sequences.

Chen et al. [4] improved on the usage of the Q-learning algorithm in their proposal to enable the online learning of non-malicious and attack behaviour. Focused on detecting FDI attacks that affect the normal operation of a power system regulated by automatic voltage controls (AVCs), the authors proposed to model the attacks as a POMDP. An FDI mitigation method was developed, consisting of offline and online modules capable of detecting multiple attacks. The experiments performed on an IEEE 118-bus system assessed the scalability of the proposed solution and its ability to provide insights about attacks and their impact in the whole power system. The main contribution was the study of the RL usage, providing theoretical assumptions about scalability and feasibility of FDI detection, as well as further results from a mitigation system. However, the paper lacks real-life cases (very sparse attacks) and considers a naive virus spreading strategy.

Deep learning was the method used by Ashrafuzzaman et al. [19] to deal with FDI attacks. The experiment was carried out using a simulated IEEE 14-bus system. Four different architectures of deep learning models based on MLP were compared

to Gradient Boosting Machines (GBM), Generalised Linear Models (GLM), and Random Forest (RF). The authors explored 122 measurement features to find the 20 most important with RF importance. As an outcome, deep MLP structures obtained the best accuracy results. Also, the use of a smaller set of selected features resulted in training time speed-up. Lastly, the deep learning training cost was mentioned as a challenge that needs to be handled for speeding up the process. To obtain more confidence, the authors planned to use real-life datasets as future work.

Instead of detecting FDI attacks, Anwar et al. [21] just clustered AMI nodes according to their vulnerability to such attacks. Their idea is that some nodes, due to their inter-dependency to other ones, can cause more damage to the entire system when attacked. Therefore, these nodes should be identified to be better protected. To cluster the AMI nodes, the authors applied the k-means algorithm combined with CF-PSO over the nodes' voltage stability index. Three clusters were defined: one for the least vulnerable nodes, other for the nodes with moderate vulnerability, and the last one for the most vulnerable. Experiments were performed in a 33-bus and a 69-bus test systems.

Alongside FDI, DoS attacks are among the top concerns regarding smart grids cybersecurity. Fadlullah et al. [18] proposed a centralised Bayesian approach for early DoS detection. The DoS attack is modelled as an attacker with access to one or more smart meters (via a worm), which they use to generate many fake authentication requests to saturate the network and strain target devices. The system uses Gaussian process regression to create an attack forecast based on the current state of communications. A composite covariance function is used to analyse trends, and samples are taken to create a set of real observations. The method was tested in a simulated BAN (Building Area Network), with 50% of smart meters vulnerable to worm infection. The authors found the forecasting system to be effective with both long and short training times, noting that the BAN gateway can be impacted at different times depending on attack particularities.

Comparative studies on machine learning algorithms were also carried out for DoS detection. To achieve this, Choi et al. [29] simulated a SYN flood attack and a buffer overflow attack on the bay and the station levels of the grid. PCs were implemented to emulate IEDs, with the GOOSE protocol's publisher-to-subscriber multicast feature used to spread attack commands via a router. The data generated was then collected, and a set of traffic-based metric attributes extracted (consisting of both normal and attack state information). The authors used Weka's machine learning library to process the data using various algorithms including Bayes classifiers, neural networks, SVM, lazy classifiers, Voting Feature Intervals (VFI), rule-based classifiers, RF, and decisions trees. They reported that for both attack types, the use of key attributes improved detection ratings, and that decision trees produced the best results overall.

Yilmaz and Uludag [11] explored the online classification paradigm to develop the MIAMI-DIL (Minimally Invasive Attack Mitigation via Detection Isolation and Localization) approach. It focuses on detecting DoS attacks against nodes on the distribution and customer domains such as data concentrators, smart meters, and smart appliances. Their approach is based on an online and non-parametric

detector named ODIT, which combines features from GEM (Geometric Entropy Minimization) and CUSUM. ODIT is applied at different levels of the network, so an anomaly evidence score is computed for smart appliances (HAN level), smart meters, and data concentrators (FAN level). Anomaly evidence scores from different levels are gathered to decide whether any node in the system is under attack. If so, a mitigation procedure is carried out, which isolates the node involved in the detected attack.

As smart grid networks rely on wireless networks due to their capacity of covering long distances and reaching remote spots, jamming attacks are also a significant threat. Su et al. [31] and Niu et al. [30] proposed a distributed jamming-avoidance strategy where the efficient use of channels for secondary users (SU) is defined as a POMDP. Each SU uses the MAB algorithm to generate a set of possible strategies (i.e. channels it can sense), weighted by availability. It then selects a random strategy to try, calculating the distribution for the channel set. Estimated and actual success rates are then compared to update weightings. Simulations revealed that the more sophisticated the jammer, the more difficult the problem. However, the authors reported that over time, SUs could achieve a highly unified view of channel availability and were less likely to be affected by jammed channels. It is important to note that these works do not propose a jamming attack detection scheme, but a solution to avoid channels under this kind of attack.

As load forecasting is helpful to improve the smart grid operation and planning, attacks against this activity can lead operators to make wrong decisions. Cui et al. [27] employed some classical machine learning algorithms in a three-stage anomaly detection approach to address this issue. In the first stage, the data is reconstructed to deliver a suitable forecast based on feature selection. Afterwards, the attack template is detected via k-means clustering. Finally, in the third step, the identification of the occurrence of a cyberattack is performed using Naive Bayes algorithm and dynamic programming. Five different attack templates were studied: pulse attack, scaling attack, ramping attack, random attack, and smooth curve attack, with the latter ones being the more difficult to detect. The authors discussed the importance of feature selection in enhancing the accuracy of attack prediction. They also discussed the impact of adversaries in the anomaly detection model and detection performance, highlighting this topic as an important challenge for future works in cybersecurity.

Some works proposed schemes to address multiple types of attacks. Kurt et al. [3] explored RL and POMDP to detect FDI, DoS, and jamming attacks. They also claimed that the proposed solution would allow new unknown attack types to be detected. The authors implemented a framework to track slight deviations of measurements from normal system operation. To evaluate the results, the proposal was compared to a Euclidean detector and a Cosine-similarity detector. The best results were achieved by the RL proposed detector, followed by the Euclidean detector and the Cosine-similarity detector. The proposed solution achieved satisfactory results but, throughout the experiments, the authors had to handle several hyperparameters to tune the algorithms appropriately. This might present a challenge to this method's wide-scale adoption. Furthermore, when discussing the results, some concerns were

raised about the memory cost, and the possibility of improving performance with a DL algorithm was suggested.

Sedjelmaci and Senouci [28] also targeted multiple types of attack. They proposed a hierarchically distributed system to detect FDI, DoS, and energy theft by analysing data collected from smart meters. The system is composed of three agents, with one for each hierarchy level. The LLIDS (Low Level IDS) is deployed at smart meters, the MLIDS (Medium Level IDS) is embedded in data concentrators, and the control centre hosts the HLIDS (High-Level IDS). Firstly, a rule-based system analyses collected data. This system has a specific threshold for each kind of attack, FDI, DoS, or energy theft. When the rule-based system detects a threshold violation, it passes the analysis on to the IDS agent of the next upper level. Then, an SVM classifier is used to confirm whether the anomaly is an attack.

Works such as [32] and [14] do not specify the types of attacks they are aiming at. They build their approaches to detect anomalies or unusual behaviours that can signal an attack, but do not focus on any specific threat. Kher et al. [32] developed a hierarchical sensor model for anomaly detection using sensor data, covering both the lower node and the upper cluster head levels. The proposed protocol initially has all nodes in sleep mode (for synchronisation). Clusters are formed through the exchange of "Hello" messages (at the lower level), a cluster head is selected, and multiple cluster heads establish linear links with each other (at the upper level). Data from each cluster is integrated before being sent up the chain. Data received at towers is integrated again before being sent to the base station. This integrated data can then be analysed for anomalies. The authors used Weka-implemented supervised learning for this purpose, and reported that the decision tree classifier (J48) achieved the best performance compared to other algorithms like ZeroR, decision table, RF, and ADTree [32].

Zhang et al. [14] proposed a distributed IDS system that uses intelligent analysis modules (AMs) sitting across HAN, NAN, and WAN layers. AMs at each level work with other modules to form a self-contained IDS on that grid layer. At the NAN and WAN levels, the lower level IDS is used together with the local IDS, such that the overall system is formed hierarchically. For difficult decisions, data may be sent to higher layers for further analysis. Either unsupervised SVM (with a Gaussian radial basis function) or Artificial Immune System (AIS) algorithms (with a focus on clustering) are used for detection. Clonal selection algorithms CLONALG and AIRS2Parallel were tested, and the authors found that SVM had better overall performance, especially for remote-to-user (R2L) and user-to-root (U2R) attacks. They suggested that the detection accuracy of the AIS algorithms could be improved with a larger sample of attack data.

Table 9.1 presents a summary of all the reviewed works and their main characteristics in chronological order. In some cases, the reviewed work does not define the data source used or how the solution would be distributed. In these cases, the table presents "–" for the undefined feature.

Table 9.1 Surveyed solutions and their main features

Reference	Year	Data source	Distribution	Attack types	ML algorithms
[18]	2011	–	–	DoS	Other[a]
[49]	2011	–	–	FDI	Other[a]
[14]	2011	Network traffic	Hierarchically distributed	Anomalies	SVM
[32]	2012	–	–	Anomalies	RF
[29]	2012	Network traffic, CPU, and memory utilisation	–	DoS	SVM, ANN
[31]	2012	Wireless channel condition	Fully distributed	Jamming	Other[a]
[24]	2012	Measurements for state estimation	–	FDI	kNN, SVM
[16]	2013	Measurements for state estimation	Centralised	FDI	CUSUM
[20]	2014	Measurements for state estimation	Centralised and distributed	FDI	SVM
[26]	2014	PMU measurements and devices status flags	Centralised	FDI	RF, SVM
[30]	2015	Wireless channel condition	Fully distributed	Jamming	Other[a]
[21]	2015	Smart meter data (energy consumption)	–	FDI	PSO
[23]	2015	PMU measurements	–	FDI	ANN, SVM, kNN
[25]	2016	Measurements for state estimation	–	FDI	SVM, kNN
[11]	2016	Smart meter data (energy consumption)	Hierarchically distributed	DoS	CUSUM
[28]	2016	Smart meter data, amount of requested energy by customer, signal strength intensity	Hierarchically distributed	FDI, DoS, energy theft	SVM
[12]	2017	PMU measurements	Centralised	FDI	SVM, ANN
[5]	2017	PMU measurements, relay logs, central control logs, snort logs	Centralised	FDI	HT
[50]	2017	–	–	Sequential topology attack	RL
[22]	2017	Measurements for state estimation (SCADA)	Centralised	FDI	SVM, ANN
[19]	2018	Measurements for state estimation (SCADA)	Centralised	FDI	RF
[4]	2018	Measurements for state estimation (SCADA)	–	FDI	Other[a]
[3]	2018	Measurements for state estimation	Centralised	FDI, DoS, jamming	RL
[2]	2018	Measurements for state estimation	–	FDI	SVM, ANN
[27]	2019	Load data, weather info, temporal info	–	Load forecast attack	k-Means

[a]Refers to pure statistical approaches (e.g. maximum likelihood estimation) or probabilistic approaches (e.g. Markov models), outside of machine learning boundaries

9.7 Open Issues

Considerable progress has been made in smart grids scenarios when applying machine learning for cyberattack detection. However, several key issues need to be handled to allow the development of feasible solutions capable of achieving suitable performance in real-life scenarios.

The main challenges are related to the real-time nature of the problem, the need for labels in supervised learning, demand for more comprehensive and human-friendly models, and solution scalability. Additionally, some inherent challenges of machine learning (such as hyperparameter tuning and the capacity of algorithms for dealing with imbalanced data) open new paths for further research in applied smart grid security.

Real-time classification algorithms are often demanded in the literature. This type of algorithm could be implemented by an offline induction and online classification, as in most of the current proposals. However, these solutions require some additional effort to leverage reliable models since they become obsolete when the smart grid behaviour changes, culminating in the concept drift problem. Also, the cost of feature extraction needs to be suitable to support a real-time classification. The algorithms that meet the real-time classification requirements are grounded on stream mining. Stream mining algorithms are able to induce a model online, which eliminates the offline step and keeps the model updated. Some algorithms such as Very Fast Decision Tree [45] and Strict Very Fast Decision Tree [47] are important examples of stream algorithms.

Even though online classification algorithms pave the way for more useful solutions, their requirement for labelled instances is a hindrance. In other words, it is impossible to label each instance on the smart grid data flow. Thus, it is necessary to rely on semi-supervised approaches or unsupervised algorithms. The DenStream algorithm [51] is an unsupervised algorithm for stream clustering. Based on three types of clusters, DenStream can point out the core, potential, and outlier behaviours, giving insights into the smart grid's behaviour.

Changes are expected in smart grid behaviour during an attack, and consequently, the recognition of these pattern deviations allows a machine learning model to detect the attack. For this reason, the predictive performance of detection systems has been the main focus of current systems, avoiding false positives and improving the computational complexity of the designed solutions. However, some concerns on how the attacks happen, the importance of features used to describe the event, and a user-friendly model for supporting attack comprehension are also relevant demands from industry. In addition to being highly accurate, a machine learning model needs to produce meaningful results and help operators to make better decisions through the usage of more descriptive modelling.

Meaningful results from descriptive models support suitable incident comprehension and mitigation. Thus, choosing an algorithm that matches certain model legibility criteria is necessary. However, the amount of data collected from a smart grid environment poses an additional constraint: scalability. A highly accurate

algorithm that is able to output a user-friendly model considering the current smart grid scenario also needs to be scalable to handle huge amounts of data. Scalability is related to the parallelism inherent in an algorithm and can be measured according to its speed-up on a particular architecture [52]. Most of the current works are limited to experimental scenarios with controlled, finite, and synthetic data sets, which do not offer a close-to-reality challenge in terms of data volume.

Another problem related to synthetically-produced data is that they usually result in a balanced dataset, which promotes an unrealistically smooth model induction. Attack detection problems are usually imbalanced, since deviations caused by attacks are much less frequent than expected behaviour episodes. More precisely, the attack-related samples provided by a smart grid to induce a machine learning model are often much fewer than the non-malicious samples, making up an imbalanced dataset. Highly imbalanced problems generally present high non-uniform error, which compromises the overall performance when errors occur in the minority classes. There are several approaches to work around this issue. The most commonly used are based on undersampling the majority class or oversampling the minority one. Considering the possibility of losing important samples with the former approach, oversampling strategies, such as the Synthetic Minority Over-sampling Technique (SMOTE) [53], can be used to balance the original dataset and provide a reliable scenario for the machine learning algorithms.

However, if the synthetic data design is driven by simple constraints and deterministic behaviour, the performance achieved during the experiments can be biased by patterns that are easier to learn than they would be in real scenarios. Therefore, when applied to real-life scenarios, the solutions can demand a more complex pipeline or unfeasible modifications, which prevents their adoption. The same reasoning applies to the adversarial model design. In some works, researchers assume simple or very specific attack models, which can hinder the effective application of the proposed solutions in production environments.

The open issues mentioned so far are related to the application of machine learning to the smart grid domain. Nevertheless, the machine learning area has its own challenges, which are intrinsic to its algorithms and must be addressed regardless of the application domain. Hyperparameters tuning [54], temporal vulnerabilities [55], stream classification trade-offs [56], and more recent topics like adversarial machine learning attacks [57] are examples of these issues and pave the way for future studies.

Lastly, for some authors like [13, 17, 58], the use of multiple sources of data leads to improvements in detection performance. For example, the amalgamation of features extracted from computer network traffic and smart grid measurements can form a more robust feature vector, which covers a wider range of attacks. Working on the several possible combinations of smart grid data sources to assess how they enhance the range of detected attacks is another possible subject for future work.

9.8 Conclusion

As smart grids are critical infrastructures, cyberattacks against these systems have a high potential for causing large-scale disruption to electricity supplies. To assist in the fight against this threat, we have provided a study of how machine learning algorithms can be applied to detect attacks on smart grids. We outlined the possible attacks types, as well as the concepts that underpin the detection of such attacks. Then, we presented the machine learning algorithms that have been employed in proposed detection schemes. Following this discussion, a list of existing attack detection approaches based on machine learning was given, detailing how each one addressed the characteristics of this problem.

Some open issues were identified in the reviewed approaches, such as algorithms depending on a labelling process, approaches not prepared to deal with imbalanced datasets and real-time aspects of smart grids, algorithms producing poor descriptive models, experiments relying on poorly designed synthetic data, and testing with a limited range of attack behaviours. Among the recommendations for future work are suggestions to use stream mining algorithms and oversampling techniques, multiple data sources, and to invest more effort into producing more realistic data sets.

References

1. C. Greer, D.A. Wollman, D.E. Prochaska, P.A. Boynton, J.A. Mazer, C.T. Nguyen, G.J. FitzPatrick, T.L. Nelson, G.H. Koepke, A.R. Hefner Jr. et al., NIST framework and roadmap for smart grid interoperability standards, release 3.0. Technical Report (2014)
2. K. Hamedani, L. Liu, R. Atat, J. Wu, Y. Yi, Reservoir computing meets smart grids: attack detection using delayed feedback networks. IEEE Trans. Ind. Inf. **14**(2), 734–743 (2017)
3. M.N. Kurt, O. Ogundijo, C. Li, X. Wang, Online cyber-attack detection in smart grid: a reinforcement learning approach. IEEE Trans. Smart Grid **10**(5), 5174–5185 (2019)
4. Y. Chen, S. Huang, F. Liu, Z. Wang, X. Sun, Evaluation of reinforcement learning-based false data injection attack to automatic voltage control. IEEE Trans. Smart Grid **10**(2), 2158–2169 (2018)
5. U. Adhikari, T.H. Morris, S. Pan, Applying hoeffding adaptive trees for real-time cyber-power event and intrusion classification. IEEE Trans. Smart Grid **9**(5), 4049–4060 (2017)
6. W. Wang, Y. Xu, M. Khanna, A survey on the communication architectures in smart grid. Comput. Netw. **55**(15), 3604 – 3629 (2011) [Online]. Available: http://www.sciencedirect.com/science/article/pii/S138912861100260X
7. H. He, J. Yan, Cyber-physical attacks and defences in the smart grid: a survey. IET Cyber-Phys. Syst. Theory Appl. **1**(1), 13–27 (2016)
8. X. Li, X. Liang, R. Lu, X. Shen, X. Lin, H. Zhu, Securing smart grid: cyber attacks, countermeasures, and challenges. IEEE Commun. Mag. **50**(8), 38–45 (2012)
9. G. Loukas, *Cyber-Physical Attacks: A Growing Invisible Threat* (Butterworth-Heinemann, Oxford, 2015)
10. F. Samie, L. Bauer, J. Henkel, *Edge Computing for Smart Grid: An Overview on Architectures and Solutions* (Springer, Cham, 2019), pp. 21–42 [Online]. Available: https://doi.org/10.1007/978-3-030-03640-9_2

11. Y. Yilmaz, S. Uludag, Mitigating IoT-based cyberattacks on the smart grid, in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (IEEE, Piscataway, 2017), pp. 517–522
12. Y. Wang, M.M. Amin, J. Fu, H.B. Moussa, A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids. IEEE Access **5**, 26022–26033 (2017)
13. M. Wu, Z. Song, Y.B. Moon, Detecting cyber-physical attacks in cyber manufacturing systems with machine learning methods. J. Intell. Manuf. **30**(3), 1111–1123 (2019)
14. Y. Zhang, L. Wang, W. Sun, R.C. Green II, M. Alam, Distributed intrusion detection system in a multi-layer network architecture of smart grids. IEEE Trans. Smart Grid **2**(4), 796–808 (2011)
15. R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, X.S. Shen, Energy-theft detection issues for advanced metering infrastructure in smart grid. Tsinghua Sci. Technol. **19**(2), 105–120 (2014)
16. Y. Huang, M. Esmalifalak, H. Nguyen, R. Zheng, Z. Han, H. Li, L. Song, Bad data injection in smart grid: attack and defense mechanisms. IEEE Commun. Mag. **51**(1), 27–33 (2013)
17. H. He, J. Yan, Cyber-physical attacks and defences in the smart grid: a survey. IET Cyber-Phys. Syst. Theory Appl. **1**(1), 13–27 (2016)
18. Z.M. Fadlullah, M.M. Fouda, N. Kato, X. Shen, Y. Nozaki, An early warning system against malicious activities for smart grid communications. IEEE Netw **25**(5), 50–55 (2011)
19. M. Ashrafuzzaman, Y. Chakhchoukh, A.A. Jillepalli, P.T. Tosic, D.C. de Leon, F.T. Sheldon, B.K. Johnson, Detecting stealthy false data injection attacks in power grids using deep learning, in *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)* (IEEE, Piscataway, 2018), pp. 219–225
20. M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, Z. Han, Detecting stealthy false data injection using machine learning in smart grid. IEEE Syst. J. **11**(3), 1644–1652 (2014)
21. A. Anwar, A.N. Mahmood, Z. Tari, Identification of vulnerable node clusters against false data injection attack in an AMI based smart grid. Inf. Syst. **53**, 201–212 (2015)
22. Y. He, G.J. Mendis, J. Wei, Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism. IEEE Trans. Smart Grid **8**(5), 2505–2516 (2017)
23. M. Ozay, I. Esnaola, F.T.Y. Vural, S.R. Kulkarni, H.V. Poor, Machine learning methods for attack detection in the smart grid. IEEE Trans. Neural Netw. Learn. Syst. **27**(8), 1773–1786 (2015)
24. M. Ozay, I. Esnaola, F.T.Y. Vural, S.R. Kulkarni, P.H. Vincent, Smarter security in the smart grid, in *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)* (IEEE, Piscataway, 2012), pp. 312–317
25. J. Yan, B. Tang, H. He, Detection of false data attacks in smart grid with supervised learning, in *2016 International Joint Conference on Neural Networks (IJCNN)* (IEEE, Piscataway, 2016), pp. 1395–1402
26. R.C.B. Hink, J.M. Beaver, M.A. Buckner, T. Morris, U. Adhikari, S. Pan, Machine learning for power system disturbance and cyber-attack discrimination, in *2014 7th International Symposium on Resilient Control Systems (ISRCS)* (IEEE, Piscataway, 2014), pp. 1–8
27. M. Cui, J. Wang, M. Yue, Machine learning based anomaly detection for load forecasting under cyberattacks. IEEE Trans. Smart Grid **10**(5), 5724–5734 (2019)
28. H. Sedjelmaci, S.M. Senouci, Smart grid security: a new approach to detect intruders in a smart grid neighborhood area network, in *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)* (2016), pp. 6–11
29. K. Choi, X. Chen, S. Li, M. Kim, K. Chae, J. Na, Intrusion detection of NSM based DoS attacks using data mining in smart grid. Energies **5**(10), 4091–4109 (2012)
30. J. Niu, Z. Ming, M. Qiu, H. Su, Z. Gu, X. Qin, Defending jamming attack in wide-area monitoring system for smart grid. Telecommun. Syst. **60**(1), 159–167 (2015)
31. H. Su, M. Qiu, H. Wang, Secure wireless communication system for smart grid with rechargeable electric vehicles. IEEE Commun. Mag. **50**(8), 62–68 (2012)
32. S. Kher, V. Nutt, D. Dasgupta, H. Ali, P. Mixon, A detection model for anomalies in smart grid with sensor network, in *2012 Future of Instrumentation International Workshop (FIIW) Proceedings* (IEEE, Piscataway, 2012), pp. 1–4

33. J.L. Viegas, S.M. Vieira, Clustering-based novelty detection to uncover electricity theft, in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (IEEE, Piscataway, 2017), pp. 1–6
34. A.E. Lazzaretti, D.M.J. Tax, H.V. Neto, V.H. Ferreira, Novelty detection and multi-class classification in power distribution voltage waveforms. Expert Syst. Appl. **45** 322–330 (2016)
35. E.R. Faria, J. Gama, A.C. Carvalho, Novelty detection algorithm for data streams multi-class problems, in *Proceedings of the 28th Annual ACM Symposium on Applied Computing* (ACM, New York, 2013), pp. 795–800
36. R. Kaviani, K.W. Hedman, A detection mechanism against load-redistribution attacks in smart grids (2019). Preprint. arXiv:1907.13294
37. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
38. S. Haykin, *Neural Networks: A Comprehensive Foundation* (Prentice Hall PTR, Englewood Cliffs, 1994)
39. T.M. Cover, P. Hart et al., Nearest neighbor pattern classification. IEEE Trans. Inf. Theory **13**(1), 21–27 (1967)
40. I. Rish et al., An empirical study of the Naive Bayes classifier, in *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, vol. 3, no. 22 (2001), pp. 41–46
41. L. Breiman, Random forests. Mach. Learn. **45**(1), 5–32 (2001)
42. O. Chapelle, B. Scholkopf, A. Zien, Semi-supervised learning (Chapelle, O. et al., eds.; 2006) [Book reviews]. IEEE Trans. Neural Netw. **20**(3), 542–542 (2009)
43. S. Zanero, S.M. Savaresi, Unsupervised learning techniques for an intrusion detection system, in *Proceedings of the 2004 ACM Symposium on Applied Computing* (ACM, New York, 2004), pp. 412–419
44. S. Dasgupta, A. Srivastava, J. Cordova, R. Arghandeh, Clustering household electrical load profiles using elastic shape analysis, in *2019 IEEE Milan PowerTech* (IEEE, Piscataway, 2019), pp. 1–6
45. J. Gama, R. Fernandes, R. Rocha, Decision trees for mining data streams. Intell. Data Anal. **10**(1), 23–45 (2006)
46. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature **521**(7553), 436 (2015)
47. V.G.T. da Costa, A.C.P. de Leon Ferreira, S. Barbon Jr., et al., Strict very fast decision tree: a memory conservative algorithm for data stream mining. Pattern Recogn. Lett. **116**, 22–28 (2018)
48. G. Le Ray, P. Pinson, Online adaptive clustering algorithm for load profiling. Sustain. Energy Grids Netw. **17**, 100181 (2019)
49. R.C. Qiu, Z. Hu, Z. Chen, N. Guo, R. Ranganathan, S. Hou, G. Zheng, Cognitive Radio Network for the smart grid: experimental system architecture, control algorithms, security, and microgrid testbed. IEEE Trans. Smart Grid **2**(4), 724–740 (2011)
50. J. Yan, H. He, X. Zhong, Y. Tang, Q-Learning-based vulnerability analysis of smart grid against sequential topology attacks. IEEE Trans. Inf. Forensics Secur. **12**(1), 200–210 (2016)
51. F. Cao, M. Estert, W. Qian, A. Zhou, Density-based clustering over an evolving data stream with noise, in *Proceedings of the 2006 SIAM International Conference on Data Mining* (SIAM, Philadelphia, 2006), pp. 328–339
52. D. Nussbaum, A. Agarwal, Scalability of parallel machines. Commun. ACM **34**(3), 57–61 (1991)
53. N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
54. R.G. Mantovani, A.L. Rossi, J. Vanschoren, B. Bischl, A.C. Carvalho, To tune or not to tune: recommending when to adjust SVM hyper-parameters via meta-learning, in *2015 International Joint Conference on Neural Networks (IJCNN)* (IEEE, Piscataway, 2015), pp. 1–8
55. J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation. ACM Comput. Surv. **46**(4), 44 (2014)
56. V.G.T. da Costa, E.J. Santana, J.F. Lopes, S. Barbon, Evaluating the four-way performance trade-off for stream classification, in *International Conference on Green, Pervasive, and Cloud Computing* (Springer, Berlin, 2019), pp. 3–17

57. A. Kurakin, I. Goodfellow, S. Bengio, Adversarial machine learning at scale (2016). Preprint. arXiv:1611.01236
58. R.S. de Carvalho, S. Mohagheghi, Analyzing impact of communication network topologies on reconfiguration of networked microgrids, impact of communication system on smart grid reliability, security and operation, in *2016 North American Power Symposium (NAPS)* (IEEE, Piscataway, 2016), pp. 1–6