Journal of Intelligent Information Systems (2023) 60:377–405
https://doi.org/10.1007/s10844-022-00753-1

Machine learning in industrial control system (ICS) security: current landscape, opportunities and challenges

Abigail M. Y. Koay[1] · Ryan K. L Ko[1] · Hinne Hettema[1] · Kenneth Radke[1]

Received: 9 May 2022 / Revised: 16 September 2022 / Accepted: 27 September 2022 /
Published online: 12 October 2022

Abstract

The advent of Industry 4.0 has led to a rapid increase in cyber attacks on industrial systems and processes, particularly on Industrial Control Systems (ICS). These systems are increasingly becoming prime targets for cyber criminals and nation-states looking to extort large ransoms or cause disruptions due to their ability to cause devastating impact whenever they cease working or malfunction. Although myriads of cyber attack detection systems have been proposed and developed, these detection systems still face many challenges that are typically not found in traditional detection systems. Motivated by the need to better understand these challenges to improve current approaches, this paper aims to (1) understand the current vulnerability landscape in ICS, (2) survey current advancements of Machine Learning (ML) based methods with respect to the usage of ML base classifiers (3) provide insights to benefits and limitations of recent advancement with respect to two performance vectors; detection accuracy and attack variety. Based on our findings, we present key open challenges which will represent exciting research opportunities for the research community.

Keywords Critical infrastructure · Machine learning · Industrial control systems · Dataset · Cyber security · Operational technology

✉ Abigail M. Y. Koay
a.koay@uq.edu.au

Ryan K. L Ko
ryan.ko@uq.edu.au

Hinne Hettema
h.hettema@uq.edu.au

Kenneth Radke
k.radke@uq.edu.au

[1] School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane 4072, Queensland, Australia

1 Introduction

Industrial control systems (ICS) are commonly defined as a subset of operational technology (OT) that regulate critical industrial processes such as power generation, water utilities, oil & gas and transportation systems. Historically, ICS have not been designed with in-built security features (Hahn, 2016). However, with the rise in remote management capabilities facilitated by the increased use of the Internet over the last two decades, ICS began to connect to information technology (IT) systems. The so-called convergence of OT and IT in Industry 4.0 has brought vulnerabilities from IT systems (e.g insecure protocols, remote connections, etc) into OT systems. This has led to cyber attacks, previously only occurring in IT, now being seen in ICS. These cyber attacks have caused massive negative environmental, safety, societal, and physical impact to ICS asset owners, government and the general public.

One of the most notable early examples of ICS cyber attacks is the attack on the Iranian nuclear facility using the Stuxnet malware (Langner, 2011). The malware infected specific programmable logic controllers (PLCs), which led to the destruction of approximately 1,000 centrifuges used in Iran’s nuclear program. In 2015, the BlackEnergy3 malware compromised the Ukrainian power grid system, causing power outages in several areas, which affected nearly a quarter-million customers (Case, 2016). In May 2021, the Colonial pipelines, which supply 45% of the fuel consumption along the east coast of the United States of America, were hit by a cyber attack. As a result, the company was forced to shut down its operations for several days, which caused fuel shortages, price hikes and panic buying across the country (Sanger et al., 2021).

With the rising number of cyber attacks on ICS, the need for automated cybersecurity tools to augment current mostly-manual capabilities is becoming more apparent (Asghar et al., 2019; El Mrabet et al., 2018; Mehrfeld, 2020; Ko, 2020). Increasingly, researchers are achieving cybersecurity automation using machine learning (ML) techniques – learning patterns in ICS signals and data to make better security decisions (Beaver et al., 2013; Begli et al., 2019; Cui et al., 2020; Buczak & Guven, 2015; Torres et al., 2019; Apruzzese et al., 2018). In the literature, most of these ML techniques have shown promising results. However, adopting these approaches to the real environment still poses many challenges.

Motivated by the need to better understand these challenges to improve current approaches, this paper aims to investigate the issues affecting cybersecurity automation, particularly in cyber attack detection systems on ICS. Ultimately, this paper will answer the following research questions (RQ) with the associated key contributions (C) and expected takeaways (T):

RQ1: What are the vulnerabilities found in ICS? (Section 3)

- **C1**: We identify the well-known vulnerabilities in common ICS components and protocols that may still exist till this day in some critical infrastructure environments.
- **C2**: We also discuss the the common ICS cyber attacks that exploit these vulnerabilities and present three of the biggest security issues and challenges in ICS.
- **T1**: Understanding the well-known vulnerabilities commonly found in ICS components and the implications of these vulnerabilities in regards to cyber attacks can equip researchers with better knowledge of the ICS domain. This knowledge can be use to identify the potential root cause of cyber attacks and researchers can use the knowledge to developed effective detection/mitigation strategies.

RQ2: What are the current advancements in Machine learning for detecting cyber attacks in ICS with respect to detection accuracy, attack variety and type of learning classifiers? (Section 4)

- **C3**: We identify the contemporary machine learning algorithms which are commonly used as the basis of current state-of-the-art ML approaches and the type of cyber attacks they can detect.
- **C4**: We review and compare 30 recent ML approaches for ICS with respect to the type of base learning classifier used in the approach, types of cyber attacks and datasets the approaches have been evaluated upon, and performance metric used to evaluate the efficacy of the approach.
- **T2**: The key takeaway from C3 and C4 is the insight of ’which’ base ML algorithm is best suited for ’which’ cyber attack. This information may help researchers narrow down the best base ML classifier for developing new and better approaches. It also gives an overview of base ML algorithms that have or have not been used, alongside the type of datasets they have been evaluated upon.

RQ3: What are the current limitations and challenges that prevent real-world adoption of ML based approaches, and the research opportunities to address these limitations and challenges? (Sections 5 & 6)

- **C5**: We identify four critical challenges facing current ML research for detecting ICS cyber attacks based on the insights gained from C4.
- **C6**: We provide four recommendations for the research community to consider as the way forward and future research opportunities to overcome the challenges mentioned in C5.
- **T3**: The challenges mentioned in C4 highlight the important current research gaps in ICS security based on the findings of our literature reviews. Understanding these challenges could help generate new research directions and possibly draw focus to priority areas that would benefit critical infrastructures. In C5, we list several of the top priority areas and research directions that we believe should be the focal points in the development of ML in ICS security.

We begin by describing the vulnerabilities in ICS based on its components and protocols and present the cybersecurity challenges arising from these vulnerabilities. Subsequently, we reviewed current contemporary ML approaches performance focusing on the choice of base ML classifiers and cyber security datasets used within the development of the detection system. Based on the review, we summarise our observations and list key challenges ahead. Lastly, we highlight future research opportunities and recommendations to address the challenges in this increasingly important field. The contribution map is as shown in Fig. 1.

2 Related work

Buczak and Guven (2015) present a survey on data mining and ML methods for intrusion detection where most of their approaches are tested on the old KDDCup’99 dataset. Kwon et al. (2019) present a survey on deep learning methods which includes local experiments to show the effectiveness of deep learning methods in detecting network anomalies. Nguyen et al. (2018) surveyed deep learning techniques used to detect cyber attacks in mobile cloud computing. Unlike these surveys that focus only on network anomaly

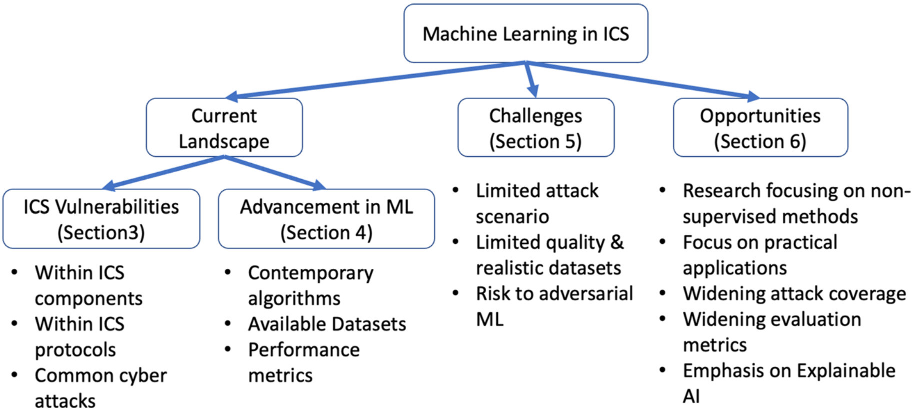
Machine Learning in ICS
Current Landscape
Challenges (Section 5)
Opportunities (Section 6)
ICS Vulnerabilities (Section3)
Advancement in ML (Section 4)
• Within ICS components
• Within ICS protocols
• Common cyber attacks
• Contemporary algorithms
• Available Datasets
• Performance metrics
• Limited attack scenario
• Limited quality & realistic datasets
• Risk to adversarial ML
• Research focusing on non-supervised methods
• Focus on practical applications
• Widening attack coverage
• Widening evaluation metrics
• Emphasis on Explainable AI

Fig. 1 Contribution map

detection or IT applications such as mobile cloud computing, our paper focuses on anomalies (or cyber attacks) found in the network and process layers in ICS.

Existing surveys usually focus on specific types of ICS cyber attacks. For example, Men et al. (2020) present a survey on ML methods focusing on addressing ICS protocol-based attacks. Cui et al. (2020) present a survey on detecting false data injection, replay and so-called ‘zero-dynamic’ attacks using ML techniques for smart grid. Zhang et al. (2021) review attack detection, estimation and control methods for two types of cyber attacks; denial-of-service and deception attacks where the latter mainly consist of spoofing and false data injection attacks. Tan et al. (2020) provide a brief survey on detection methods specific to false data injection attacks. Our work covers many types of ICS cyber attacks that existed in the real world and those that are used in cyber security research.

Some surveys look at ML approaches used for a wide range of purposes in industrial systems including anomaly detection and hardware/software fault detection. For example, Diez-Olivan et al. (2019) present a survey on the usage of ML for an industrial prognosis for a variety of critical infrastructure sectors. A recent survey (Umer et al., 2022) discusses a variety of ML methods that have been used in the ICS domain. Our survey only focuses on the advances of ML approaches by providing a comparison of the best ML methods for specific cyber security attacks and datasets.

Other surveys look at non-ML techniques for securing ICS. Hurst et al. (Hurst et al., 2014) present conventional methods for detecting and mitigating cyber threats on ICS in critical infrastructure. Maynard et al. (2020) study the various ICS cyber attacks and map them to the ‘so called’ ICS attack Kill Chain, which is popular in the industry (Assante and Lee, 2015). In our survey, we only include papers that use ML approaches to detect cyber attacks (or anomalies) and study their attack coverage, choice of classic ML algorithms as the foundation of their approach, quality of datasets and performance metrics used for cyber attack detection.

3 Vulnerabilities in ICS

Industrial control systems (ICS) are of programmable system which is used for monitoring, controlling and regulating industrial processes. Common types of ICS are supervisory control and data acquisition systems (SCADA) and distributed control systems (DCS). The

ICS environment contains a number of components specific to industrial control and management. Common types of ICS components are programmable logic controllers (PLCs), human machine interfaces (HMIs), sensors/actuators, safety instrumented systems (SIS), data historians, remote terminal units (RTUs) and engineering workstations.

The ICS environment is often (or 'may be') hierarchically ordered. The Purdue Enterprise Reference Architecture (PERA) (Williams, 1994) organises industrial control components into a six-tier architecture within specific network zones related to the broad ICS functionality required at each zone as shown in Fig. 2. ICS components are mainly found within Level 0 - 3.

We describe each of the ICS components in the OT network with their typically associated vulnerabilities in the next few paragraphs.

Programmable logic controllers (PLCs) are controller devices in Level 1 of PERA that are commonly programmed using ladder logic. Historically, PLC often has a customised operating system and a combination of function code and data blocks which may risk corruption,

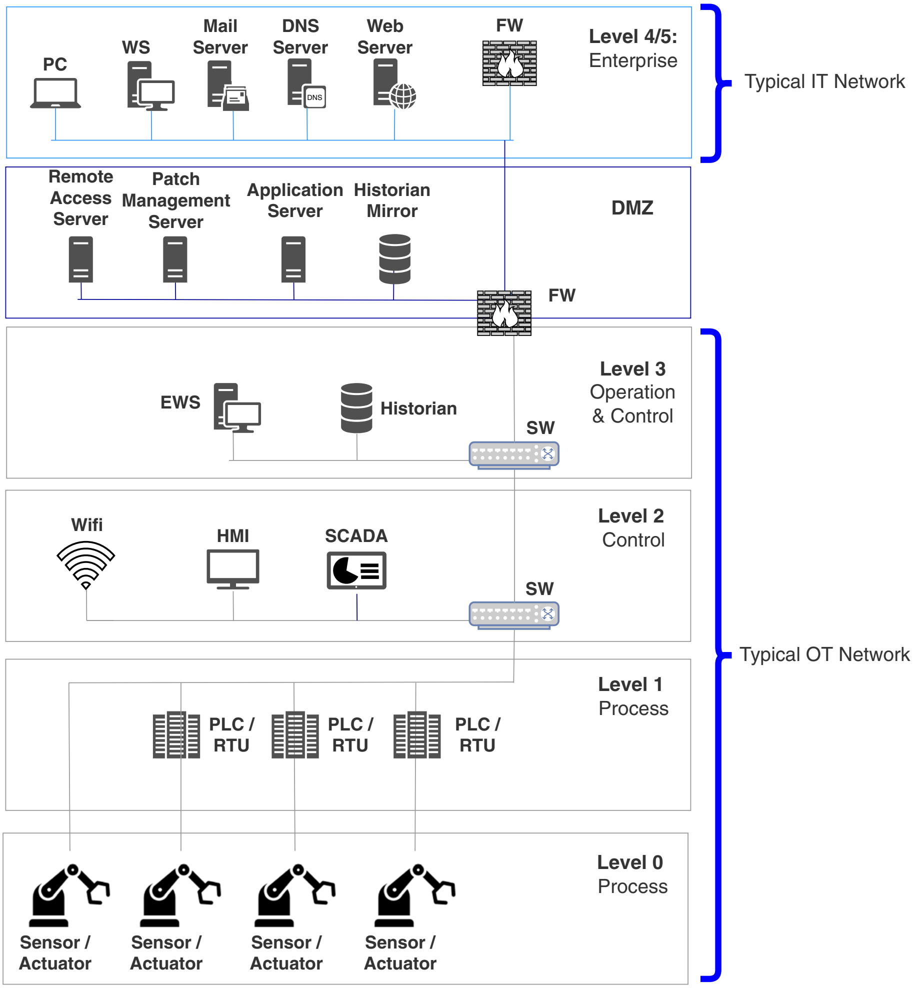

Fig. 2 ICS Architecture based on PERA model (Williams, 1994)

modification and configuration manipulation (Wu et al., 2019). An example an of attack that took advantage of PLC's known vulnerabilities is the Stuxnet attack (Langner, 2011).

Human Machine Interfaces (HMIs) are any device in a plant that requires human control in providing or in some cases displaying the state of a process or large piece of equipment. They are at level 2 of the Purdue architecture providing control panels to PLCs and often run commercially available lightweight operating systems such as Windows, but cannot usually be patched or secured (Chan et al., 2019). This makes them attractive targets for cyber attackers looking to gain operating systems access when onsite to install malicious software or gain control of other devices in the ICS environment.

Sensors/Actuators are at Level 0 of the Purdue architecture, and provide raw data feeds into PLCs (typically as data blocks). The key problem associated with sensors is that they are not capable of providing authentication or integrity guarantees to the data they provide, and PLCs in turn use sensor data to evaluate and execute control logic. The consequence is that control logic is based on unauthenticated inputs with no integrity controls such as 'unauthorised command' messages, thus compromising systems (Govil et al., 2017).

Safety Instrumented Systems (SIS) are designed and operated as independent systems that monitor the condition of the industrial process with the aim to shut it down should it enter a state in which the system itself may be damaged. The safety system itself is usually engineered to similar cybersecurity standards as the control system, with probably less monitoring on safety systems and can be compromised, as in the Trisis malware (Kanamaru, 2017).

Data Historians collect and maintain records of past events for analysis and display, usually in a database platform. It usually has the same vulnerabilities as common database platforms (Gonzalez et al., 2019).

Remote Terminal Units (RTUs) typically reside in remote locations to monitor field devices and transmit data back to a central monitoring station such as a Master Terminal Unit (MTU), a central PLC or an HMI. Like PLCs, RTUs suffer from poor security features and are vulnerable to attacks such as authentication bypass, data manipulation and malformed protocol messages (Graham et al., 2016). An example of known attacks on RTU is the Industroyer incident (Kshetri & Voas, 2017).

Engineering Workstations are placed at various locations in the plant to allow engineers to update components in the rest of the ICS systems. They are often poorly controlled from an IT security perspective, may run unsupported operating systems and run under generic administrator accounts, often allowing remote access. In addition, they are prone to software vulnerabilities, USB insertion of code or data on sites and may not run log monitoring and malware detection software (Antrobus et al., 2016). In the case of Stuxnet, the attacker utilises an engineering workstation as the initial access point.

To be able to visualise how these ICS components are connected and set up in the real-world, we present Fig. 3 which shows an example of an ICS environment in a water treatment plant. It contains the ICS components of a typical IT network, Demilitarized Zone (DMZ) and OT network based on the PERA model. From the figure, Zone A represents the SIS, Zone B represents the control systems where operators can control the sensors (A) and actuators (S) via engineering workstation and HMI, Zone C represents the Demilitarized Zone (DMZ) where external/untrusted devices (e.g remote operator console, smart

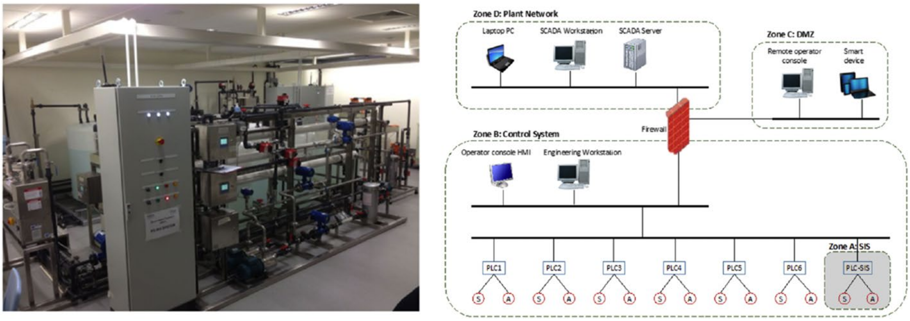

Fig. 3 Secure Water Treatment Tesbed (left) and its Architecture (right) for SWAT dataset from Goh et al. (2016)

devices) can connect to the plant and Zone D represents the plant network (e.g. Laptop PC, SCADA workstation. More details of the plant will be described in Section 4.2.

3.1 OT network protocols

In addition to vulnerabilities and weaknesses in ICS components, many ICS specific protocols are also vulnerable to cyber attacks. We list common protocols with their vulnerabilities:

Modbus is a de-facto communication protocol developed by Mobicon (now Schneider Electric) for PLCs and other ICS devices (Swales et al., 1999). It is insecure by design with known vulnerabilities that can lead to denial-of-service (Voyiatzis et al., 2015; Upadhyay & Sampalli, 2020).

PROFINET is an I/O protocol by PROFIBUS International (Feld, 2004). The protocol is based on ETHERNET standard and is vulnerable against attacks such as unauthorised access (Dias et al., 2018).

S7COMM is a proprietary protocol for Siemens PLCs (Beresford, 2011). The protocol lacks authentication and encryption which makes it vulnerable to spoofing and denial-of-service attacks (Alsmadi et al., 2021).

DNP3 is a reliable protocol that is used for communications between control system devices. In the default configuration it contains no authentication or encryption of the payload (East et al., 2009).

The main problem associated with ICS security protocols is many protocols currently in use do not implement message authentication and encryption, and have only weak or absent integrity protection. In consequence, adversaries have the ability to set up malicious control points and in some cases manipulate data in transit or through malicious drivers.

3.2 Common cyber attacks on ICS

cyber attacks on ICS may be *targeted* or *opportunistic*. Targeted attacks are defined as attacks that immediately target the *physical* infrastructure, whereas opportunistic attacks are classified as attacks that have an industrial attack as a byproduct rather than as the main objective.

The MITRE Corporation has recently released a MITRE ATT&CK for ICS framework to model the attack pathways to OT, in the form of tactics, techniques and procedures (Alexander et al., 2020). Some of the tactics in the ICS ATT&CK framework reappear in the general ATT&CK matrix, others are unique to industrial control systems. Among the unique tactics are inhibiting control functions, impairing process control and an impact category that lists the various forms of impacts that ICS cyber attacks may have. Attacks are usually not executed in a single step and with a single technique or procedure. Instead, they rely on a set of techniques executed in a sequence known as the kill chain. As an example, the ICS specific kill chain that underpins a lot of the impacts in the ATT&CK for ICS framework is developed in Assante and Lee (2015).

Existing ML approaches mainly have a more limited focus on specific techniques (in the technical sense related to the above). Our studies found that the most common cyber attacks can be categorised into four categories; denial-of-service (DoS) (Long et al., 2005), false data injection (FDI) (Mo & Sinopoli, 2010), reconnaissance (Rec) (Mazurczyk & Caviglione, 2021) and spoofing (Spo) (Hijazi & Obaidat, 2019).

3.3 ICS security issues and challenges

ICS environments are usually designed to rely on their environments for their security. Security weaknesses in industrial protocols, for instance, are usually addressed by running them on a separated network, with the presumption that access to such separated networks can be strictly controlled by the operator. In modern environments, such separation is less and less possible. Relative to the context in which ICS systems operate, three trends influence the degradation of cyber-security in ICS networks.

Convergence of IT and OT networks The advent of Industry 4.0 has led to a gradual convergence of IT and OT to allow process automation. ICS networks are a core part of OT networks. As a consequence, ICS networks are no longer isolated but are now exposed to automation components as well as increasingly the IT environment (and in some cases even the Internet), which increases their attack surface. For example, the Industrial Internet of Things often relies for its functionality on connections between critical infrastructure and a cloud platform that in turn is controlled via mobile phone apps. In many cases (like with intelligent lighting systems), these devices, apps and associated cloud infrastructure are deployed as an end-to-end third party solution over which the owner has little say, yet still ends up owning all of the risks.

Outdated Best Practices in the OT network It is considered best practice to maintain a separation between IT and OT networks as well as separate these networks from the Internet, as in the PERA model (Williams, 1994). Devices and applications in the OT environment are designed for long lifetimes and high availability, not for resistance to IT or Internet cyber threats. Notwithstanding recommended best practices, many OT environments have long had backdoors to enable remote support, often via insecure protocols such as FTP, TeamViewer, VNC and other remote access protocols. Such backdoors often existed without the knowledge of IT or cybersecurity departments and usually deployed consumer grade hardware and software standards.

Security is not a priority in the ICS infrastructure ICS infrastructure is usually not safe by design. There are many instances of processes running with elevated privileges in an 'always on' mode on devices that can be accessed by many users. A notable example is engineering workstations, where the software used to program the PLCs does not work well in a multi-user environment and needs to be available to contractors in case an update of the PLC programming is needed. Access to data blocks on PLCs is in turn required by industrial monitoring software and requires network access to the PLC over a programming port. In these situations, normal network controls, such as firewalls, are ineffective and to detect an intrusion, a protocol level understanding of the traffic is required.

3.4 ICS vulnerabilities in a nutshell

Overall, this section lists the well-known vulnerabilities found in ICS components and protocols, four common categories of cyber attacks, and the ICS security issues and challenges. In summary, the vulnerabilities mentioned are mostly around insecure authentication, risks to unauthorised modification to data/configuration, and outdated/unpatchable software. These are generally caused by poor design which did not consider the security aspect of the particular component or protocols. Therefore, vulnerabilities in ICS are not an easy fix because most of these components and protocols would require a design update or an extra layer of security added to them in order to make them secure. Alternatively, better detection strategies can be developed to detect cyber attacks that arise from these vulnerabilities.

4 Current advancement in machine learning

We review the performance of recent ML-based approaches for detecting ICS cyber attacks, particularly focusing on the last five years (2017-2022). We structured our review to provide insights on the following two key components (i.e. machine learning algorithms and datasets) used in the development of the ML-based detection systems:

1. ***machine learning algorithm*** - an algorithm that will 'learn' from input data and save the 'learned' information into a model. The model will then be used for classification, prediction or clustering tasks.
2. ***dataset*** - a set of data used for building and training the model. The data normally consists of both normal and attack samples. It will also be used to evaluate the machine learning model's performance.

We primarily queried from Google Scholar and Web of Science databases using a set of keywords from Table 1 and then manually filter the queried papers that are relevant to our survey and are highly cited or in the top ranked publications.

4.1 Contemporary machine learning algorithms

ML algorithms are used for learning the patterns from input data to build a model that can later be used to recognise the learned patterns from new data. This input data is also known as the training data. The ML model built can then be used on newer data for tasks such

Table 1 Search query keywords

Item	Keywords / Formula
Keyword 1	“Industrial Control System” OR “industrial control system” OR “ICS”
Keyword 2	“cybersecurity” OR “cyber” OR “security” OR “internet” OR “network”
Keyword 3	“detection” OR “detect” OR “detecting“
Keyword 4	“automating” OR “auto” OR “automatic” OR “automate”

as classification, prediction, clustering, dimensionality reduction and density estimation. In practice, ML model is periodically updated by adding newer data into the training data to ensure that the model can recognise newer patterns and maintain its accuracy.

In this section, we review contemporary ML algorithms used in ML-based detection approaches. These approaches typically use one of the contemporary ML algorithms with refined hyperparameters or input features. Some approaches combine two or more contemporary machine learning algorithms to improve performances. We divide these approaches into four main groups based on their learning characteristics, namely supervised learning, unsupervised learning, deep learning and ensemble learning.

Supervised learning use human intervention or ‘labels’ to learn the patterns. In attack detection tasks, binary-class labels (‘normal’, ‘attack’) are the common labels used to distinguish between benign data (“normal”) and malicious data (“attack”). Several types of supervised learning algorithms are *k*-Nearest Neighbour, Regressions (linear, logistic, Lasso, softmax), Bayes (Naive Bayes, Bayesian Network), Decision Trees (CART, J48, ID3, C4.5, REPTree), Artificial Neural Networks (NeuralNet, MLP, BPNN), Rule Induction (One-R, Zero-R, Ripper), Support Vector Machines (SVM), and Discriminant analysis (LDA, QDA).

Unsupervised learning requires no human intervention because it learns by grouping similar data together to form clusters or associations. This type of learning is desirable when labels are absent or insufficient from the training data. Common unsupervised learning algorithms found in the literature for ICS attack detection are Isolation Forest, One-Class Support Vector Machine (OCSVM) and Autoencoders such as Sparse Autoencoders (SpAE), Undercomplete Autoencoders (UAE), Variational Autoencoders (VAE) and Fair Clustering (FD). These algorithms only learn from normal data and any outliers or anomalies detected will be classified as ‘*attack*’.

Deep learning employs ‘multiple processing layers to learn representations of data with multiple levels of abstraction’(LeCun et al., 2015). Due to the *deep* learning from multiple representation levels, when trained properly, it can provide significantly better results than traditional ML algorithms. Deep learning algorithms can be a combination of both supervised and unsupervised learning techniques. Common deep learning algorithms are Deep Neural Networks (DNN), Convolutionary Neural Network (CNN), Deep Belief Network (DBN), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN, including Simple Recurrent Unit and Bi-directional Recurrent Unit), Stacked Autoencoder (StAE) and Gated Recurrent Units (GRU).

Ensemble Learning approaches learn from a single ML algorithm multiple times. At each time, a different parameter setting will be used. The results are then combined to form a single ML model. This approach is used to enhance existing models and provide better detection results. Commonly used ensemble learning algorithms for attack detection are Random Forest (RF), Bagging, Boosting (Adaptive Boosting, Gradient Boosting), ensemble deep learning, and ensemble neural network.

Table 2 presents a list of contemporary ML algorithms we reviewed under the four main categories mentioned above. We show the types of attacks each algorithm could detect; false data injection (FDI), denial of service (DoS), reconnaissance (Recon) and Spoofing (Spo) attacks. These algorithms can potentially detect other types of attacks, however, only the former four attacks have been evaluated in the papers we reviewed. This is partly due to the dataset limitations that will be explained in more detail in Section 4.2.

As shown in Table 2, existing approaches mostly focus on supervised learning methods such as Support Vector Machines and Decision Trees. However, supervised learning methods might not be suitable for real-world implementation in ICS due to the heavy reliance on labelled datasets for training. Generating labelled datasets is very expensive due to the need for human expertise to analyse and determine the label of each data.

Alternatively, unsupervised learning-based approaches seem to trump other approaches such as deep learning and supervised learning. Also, unsupervised learning-based approaches do not require labelled datasets for training and could identify cyber attacks by clustering data into groups. Data that does not belong to *normal* groups are considered *anomalies* or *attacks*. However, not many papers provide an evaluation of such approaches.

4.2 An overview of commonly-used ICS datasets with the best ML performance

Datasets are collections of past data that are used to train and build ML models. These datasets are normally collected from small-scale physical testbeds with processes mimicking the real-world environment. These datasets are also used to test and evaluate the performance of ML models. We provide a brief description of the commonly used, publicly available datasets. Table 3, presents a comparison of different ICS datasets used in evaluating ML-based approaches with their reported performance. From our review, we observed that most papers show the performance efficacy of their ML-based approaches through *accuracy* or *F1-score*. For simplicity, we only present the accuracy score of these approaches. However, not all of the surveyed papers have included accuracy in their evaluation results. In such cases, we reported their F1 score.

Secure Water Treatment (SWaT) (Goh et al., 2016) is a collection of data from a scaled-down real-world industrial treatment plant testbed implemented at the Singapore University of Technology and Design (SUTD). The dataset consists of 11-days of normal operational data from both physical properties and network traffic, and cyber and physical attack data recorded once every second. The normal operational data was collected in the first 7-days where the plant was running six stages of the filtration process normally without any deliberate interruption and attacks. In the last 4-days, 36 attacks, lasting between a few minutes to an hour were launched from multiple points in the plant.

Gas Pipeline (GP) (Turnipseed, 2015) is collected from the GP system provided by the Mississippi State University. It contains 274,627 instances of network communication

Table 2 Comparison of different ML approaches in ICS Attack detection

ML Algorithms	FDI	DoS	Rec	Spo	Reference
k-Nearest neighbour	✓	✓	✓	✓	(Süzen, 2021; Zhang et al., 2019; Khan et al., 2019; Robles-Durazno et al., 2018; Mokhtari et al., 2021)
Decision trees	✓	✓	✓	✓	(Anthi et al., 2021; Ling et al., 2021; Al-Abassi et al., 2020; Khan et al., 2019; Zhang et al., 2019; Sokolov et al., 2019; Chen et al., 2018; Yau & Chow, 2017; Mokhtari et al., 2021)
Support vector machines	✓	✓	✓	✓	(Anthi et al., 2021; Ling et al., 2021; Süzen, 2021; Xie et al., 2020; Wang et al., 2020; Anton et al., 2019; Sokolov et al., 2019; Robles-Durazno et al., 2018; Liu et al., 2018; Chen et al., 2018; Yau & Chow, 2017; Ribu Hassini et al., 2022)
Regressions	✓	✓	✓	✓	(Anthi et al., 2021; Sokolov et al., 2019; Ribu Hassini et al., 2022)
Bayes	✓	✓	✓	✓	(Anthi et al., 2021; Ling et al., 2021; Liu et al., 2018; Chen et al., 2018)
Discriminant analysis		✓		✓	(Khan et al., 2019)
Rule-based	✓	✓	✓		(Anthi et al., 2021; Chen et al., 2018)
Artificial neural networks	✓	✓	✓		(Anthi et al., 2021; Süzen, 2021; Ramotsoela et al., 2020; Xie et al., 2020; Sokolov et al., 2019; Khan et al., 2019)
Stochastic gradient descent	✓	✓	✓		(Süzen, 2021)
OCSVM	✓		✓	✓	(Inoue et al., 2017)
Isolation forest	✓	✓			(Elnour et al., 2020)
Fair clustering	✓				(Handa & Semwal, 2022)
Autoencoders	✓	✓	✓		(Kravchik & Shabtai, 2021; Wang et al., 2020; Yilmaz et al., 2019)
Deep neural networks	✓	✓	✓		(Al-Abassi et al., 2020; Xie et al., 2020; Wang et al., 2020; Yilmaz et al., 2019)
Convolutionary neural network	✓	✓		✓	(Kravchik & Shabtai, 2021; Ling et al., 2021; Krithivasan et al., 2020; Ramotsoela et al., 2020; Liu et al., 2018; Ribu Hassini et al., 2022)
Deep belief network	✓	✓	✓		(Süzen, 2021; Krithivasan et al., 2020)
Long short-term memory	✓	✓	✓	✓	(Ling et al., 2021; Xie et al., 2020; Ramotsoela et al., 2020; Chu et al., 2019; Sokolov et al., 2019; Ribu Hassini et al., 2022)
Recurrent neural network	✓	✓	✓		(Ling et al., 2021; Ramotsoela et al., 2020)
Gated recurrent units	✓	✓			(Ling et al., 2021; Xie et al., 2020; Ramotsoela et al., 2020; Sokolov et al., 2019)
Random forest	✓	✓	✓		(Anthi et al., 2021; Al-Abassi et al., 2020; Zhang et al., 2019; Anton et al., 2019; Chu et al., 2019; Sokolov et al., 2019; Khan et al., 2019; Robles-Durazno et al., 2018; Liu et al., 2018; Chen et al., 2018; Mokhtari et al., 2021; Ribu Hassini et al., 2022)

Table 2 (continued)

ML Algorithms	FDI	DoS	Rec	Spo	Reference
Bagging	✓	✓	✓		(Zhang et al., 2019; Chen et al., 2018)
Ensemble neural networks	✓	✓	✓		(Ramotsoela et al., 2020)
Boosting	✓	✓	✓		(Süzen, 2021; Khan et al., 2019; Sokolov et al., 2019; Chen et al., 2018; Ribu Hassini et al., 2022)
Ensemble deep learning	✓	✓	✓		(Al-Abassi et al., 2020; Zhang et al., 2019)
Majority voting	✓	✓	✓		(Chen et al., 2018)

FDI - False Data Injection, DoS - Denial of Service, Rec - Reconnaissance, Spo - Spoofing

Table 3 Comparison of different ICS Datasets used in Evaluating ML-based approaches

Name	No Classes	Type	cyber attacks	Best ML Algorithm	Accuracy
SWaT	5	Sensors/Actuators	FDI	EnsembleDL (DNN+DT)	0.9967 (Al-Abassi et al., 2020)
GP	8	Network	FDI, Recon, DoS	RF	0.9984 (Anton et al., 2019)
IUNO	3	sensors/actuators	FDI	RF	0.9998 (Anton et al., 2019)
BATADAL	15	sensors/actuators	FDI	AE	0.937* (Kravchik and Shabtai, 2021)
WST	8	Network	FDI, Recon, DoS	DBN	0.9972 (Süzen, 2021)
POWER	9	sensors/actuators	FDI	DNN	0.998 (Yilmaz et al., 2019)
WADI	17	sensors/actuators	FDI, Spo	AE	0.75*(Kravchik & Shabtai, 2021)
FESTO	2	sensors	FDI	RF	0.91*(Robles-Durazno et al., 2018)
HAI	2	sensors	FDI	RF	0.9976 (Mokhtari et al., 2021)
TEP	20	sensors	FDI	DNN	0.82 (Sokolov et al., 2019)
TLIGHT	2	logs/network	FDI	DT	0.9994 (Yau & Chow, 2017)

* F1 score is reported due to the absence of accuracy score

Some of these datasets such as IUNO, POWER, BATADAL, FESTO contain multiple sets of data. for simplicity, we reported the highest Accuracy / F1-score obtained among these sets of data

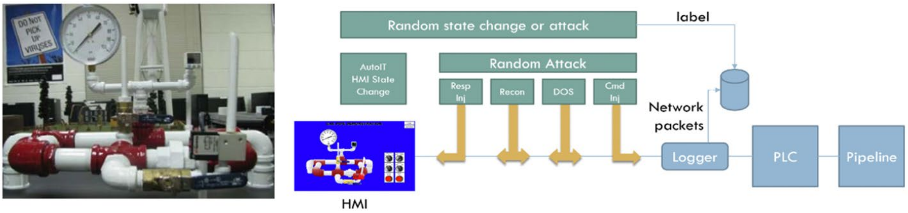

Fig. 4 Gas pipeline system and process framework for GP dataset from Turnipseed (2015)

between a Remote Terminal Unit (RTU) and a Master Control Unit (MTU) through the Modbus RTU protocol. Figure 4 shows the GP system and process framework. Based on the framework, network packet data is collected via a logger. The attacks are randomly executed from 35 cyber-attacks comprised of recon, FDI (respond injection(resp inj), command injection (cmd Inj)) and DoS attacks. These attacks constitute 21.9% of the total instances in the dataset.

IUNO datasets (Anton et al., 2019) are OPC UA based batch processing traffic. The data is generated with a Festo Didactic model representing a water pump environment, emptying and filling in the water tank. Three datasets were created where each dataset contains a specific approach of the false data injection attack.

BATtle of the Attack Detection Algorithms (BATADAL) datasets (Taormina et al., 2018) consist of three different simulated datasets based on a fictional C-Town water distribution system. These datasets were created for a cyber attack detection competition, where seven teams took part to develop solutions based on the simulated datasets. The datasets include two training datasets and a testing dataset. The first training dataset consists of 365 days of hourly normal data whereas the second training dataset consists of seven attacks spanned across 497 hourly records. The testing dataset consists of 407 hourly records with additional seven types of attacks. All of the 14 attacks were some form of False Data Injection attack such as replay, man in the middle and modification attacks.

Water Storage Tank and Gas Pipeline SCADA systems (WST) dataset (Morris & Gao, 2014) was collected from the laboratory-scale SCADA systems at Mississippi State University. Both datasets contain normal data and four types of attacks (two types of false data injection attacks, Denial of Service attack, and reconnaissance attack).

Power System Attack Datasets (Power) The Power System Attack datasets are three datasets made from one initial dataset created by Mississippi State University and Oak Ridge National Laboratory. The initial dataset was made from 15 sets of data containing 37 power systems scenarios that can be divided into three types of events: natural events, no events, and attack events. The attack events are false data injection attacks including remote command injection, and relay setting change attacks.

Water Distribution Testbed (WADI) WADI dataset (Ahmed et al., 2017) is collected from a scaled-down water distribution network in a city.

Festo MPA Process Control Rig (Festo) A clean water supply system was implemented using the Festo MPA process control rig at the Edinburgh Napier University (Robles-Durazno et al., 2018). It generated three datasets with false data injection attack to reduce the amount of water in the reservoir tank. The data is collected using the INA219 sensors via the I2C protocol.

Tennessee Eastman Process (TEP) simulation This is the oldest publicly available dataset in the ICS environment (Downs & Vogel, 1993). It involves simulating an actual industrial process plant in the chemical industry. Researchers have re-generated new datasets (Rieth et al., 2017) which include more examples for both training and testing data. The datasets contains 21 preprogrammed process faults that could simply be categorised as false data injection attacks.

Traffic Light Control System (TLIGHT) The dataset is based on an experimental setup using the Siemens S7 PLCs loaded with TLIGHT traffic light control program (Yau & Chow, 2017). Two set of datasets were created. The datasets contain seven types of normal operations that caused variation in the timers and output values of the traffic light control system. Attack data was created by altering some of the values using an open source program called Snap7.

HIL-based Augmented ICS Security (HAI) 1.0 This dataset (Shin et al., 2020) is collected from a simulated testbed which combines three physical systems; GE turbine, Emerson boiler and FESTO water treatment.

There are also recent datasets generated by the community that have not yet been widely used for evaluating ML-based approaches but are worth mentioning such as ELECTRA (Gómez et al., 2019).

Our review shows that existing datasets cover only a range of critical infrastructure sectors such as water and waste water, chemical, transportation system, and energy (gas pipelines), which allows approaches to be developed across different infrastructures. According to the Cybersecurity & Infrastructure Security Agency (CISA) of the United State government, there are 16 CI sectors. Unfortunately, most papers we reviewed were focused on evaluating their approaches in the water and waste water, and energy sectors and neglects 14 other equally important sectors. Moreover, these datasets contain only specific attacks which limit early detection capabilities. We also observed that the most common attack used for evaluation is the false data injection attack. However, this type of attack often caused an immediate impact on the critical infrastructure. For example, one of the attacks in SWaT dataset is ‘manually turning off the water tank’. The immediate result of this attack when any water tank (i.e. raw water tank, UF Feed tank) is switched off would naturally trigger an alarm and would require ML approaches for detection. Also, most datasets released publicly are using attack time as the indicator when labelling the data. Some of the normal data could be incorrectly labelled and impact the performance measure Fig. 5.

Based on our review, SWaT and GP are the most commonly used datasets. Due to the nature of the SWaT dataset, we can only evaluate ML approaches on a single type of attack (i.e., FDI) on this dataset. Conversely, on the GP dataset, we are able to perform the evaluation on several types of cyber attacks (i.e., DoS, FDI and Recon). However, these represent a small portion of the types cyber attacks found in the real-world environment.

As shown in Table 3, the best ML algorithms for most of the datasets are either Random Forest (RF), Decision Trees (DT) and Deep Neural Networks (DNN). These ML algorithms

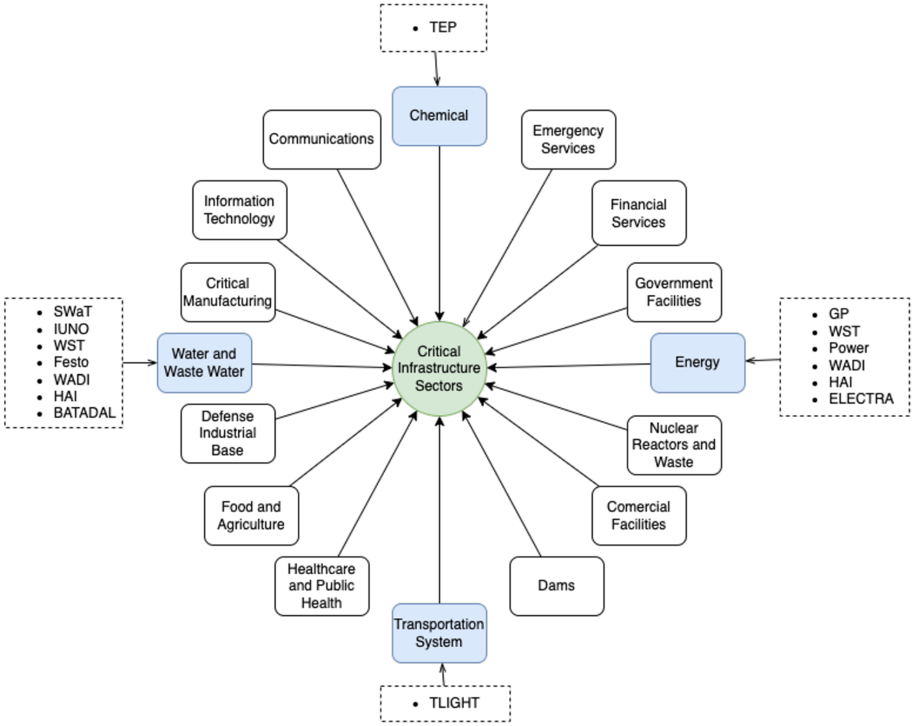

Fig. 5 16 Critical Infrastructure Sectors defined by CISA. We found that only four sectors (coloured in blue) have publicly available datasets that are used for developing ML-based approaches for ICS cyber attack detection from our review

are able to achieve near perfect detection accuracy — over 99% accuracy in most datasets. However, to our best knowledge, the ML algorithms applied to the datasets have not been adopted into real-world online scenarios, mostly due to the fact these datasets only covered a small subset of cyber attacks, particularly only FDI attacks. In addition, most datasets are built from either a scaled down version of a real-world environment or through simulation, which can be different from the actual real-world environment. Also, in some datasets such as WADI and TEP, the best ML algorithms have lower accuracy than the other datasets, which suggests that certain variations of FDI attacks might not be detected as successfully as other variations using the same base contemporary ML algorithms.

4.3 Performance evaluation metrics

Evaluation is an important component in determining the success of the ML-based approaches. In attack detection tasks, performance metrics are used to measure how well a particular approach can identify attacks correctly. The choice of metrics used is important because incorrect metrics can lead to biased evaluation and directly impact the reliability and generalisability of the approach (Juba & Le, 2019). We include the performance metrics that have been used to measure the performance of ML-based approaches in your reviewed papers as shown in Table 4.

Table 4 Comparison of performance metrics used in ML-based ICS attack detection research

Year-Publications	TP	TN	FP	FN	Acc	Pre	Rec	F_1	ROC	T_{tr}	T_{te}	MCC	Kappa
2022-Handa (Handa & Semwal, 2022)					✓	✓	✓	✓			✓		
2022-Hassini (Ribu Hassini et al., 2022)					✓	✓	✓	✓					
2021-Anthi (Anthi et al., 2021)						✓	✓	✓					
2021-Kravchik (Kravchik & Shabtai, 2021)						✓	✓	✓					
2021-Ling (Ling et al., 2021)	✓		✓	✓	✓							✓	
2021-Selim (Selim et al., 2021)					✓	✓	✓	✓					
2021-Süzen (Süzen, 2021)	✓		✓		✓			✓	✓				
2020-Al (Al-Abassi et al., 2020)					✓	✓	✓	✓					
2020-Elnour (Elnour et al., 2020)				✓	✓	✓	✓				✓		
2020-Krithivasan (Krithivasan et al., 2020)					✓	✓	✓	✓					
2020-Ramotsoela (Ramotsoela et al., 2020)	✓	✓				✓		✓			✓		
2020-Wang (Wang et al., 2020)					✓	✓	✓	✓					
2020-Xie (Xie et al., 2020)						✓	✓	✓			✓		
2019-Anton (Anton et al., 2019)				✓	✓	✓	✓				✓		
2019-Chu (Chu et al., 2019)			✓	✓	✓								
2019-Khan (Khan et al., 2019)					✓	✓	✓	✓			✓		✓
2019-Potluri (Potluri & Diedrich, 2019)					✓	✓	✓	✓					
2019-Sokolov (Sokolov et al., 2019)					✓	✓	✓						
2019-Yilmaz (Yilmaz et al., 2019)					✓	✓	✓	✓					
2019-Zhang (Zhang et al., 2019)	✓	✓	✓	✓									
2018-Chen (Chen et al., 2018)					✓	✓	✓	✓					
2018-Kravchik (Kravchik & Shabtai, 2018)								✓	✓	✓	✓		
2018-Liu (Liu et al., 2018)					✓	✓	✓	✓					
2018-Lin (Lin et al., 2018)	✓		✓							✓	✓		
2018-Robles (Robles-Durazno et al., 2018)					✓	✓	✓	✓					
2018-Wang (Wang et al., 2018)											✓		
2017-Terai (Terai et al., 2017)					✓*	✓	✓	✓					

Table 4 (continued)

Year-Publications	TP	TN	FP	FN	Acc	Pre	Rec	F_1	ROC	T_{tr}	T_{te}	MCC	Kappa
2017-Yau (Yau & Chow, 2017)					✓								
2017-Inoue (Inoue et al., 2017)						✓	✓	✓		✓	✓		
2017-Xiao (Xiao et al., 2017)			✓	✓	✓								

* Reported the complement of Accuracy (Error Rate)

Generally, the performance metrics are calculated using four base statistics. They are **True Positive** (*TP*) which represents the number of correctly detected attack instances, **True Negative** (*TN*) which represents the number of correctly detected normal instances, **False Positive** (*FP*) which defines the number of incorrectly detected attack instances, and **False Negative** (*FN*) which defines the number of incorrectly detected normal instances.

Accuracy (Acc) is the simplest and most widely used performance metric (Juba & Le, 2019). It measures the state of correctness where it measures the proportion of correctly detected attacks and normal instances. However, it does not take into account the proportion of incorrectly detected instances which renders it unsuitable for imbalanced datasets. For example, an accuracy of 90% can be achieved when all attack instances are incorrectly identified in a dataset with only 10% attack instances.

Precision (Pre) measures the proportion of attack (positive) instances identified was actually correct. A low value of precision indicates that the approach generates high amount of false positives. For example, a score of 30% means that only 30% of the attack instances identified are actual attack instances. The remaining 70% are actually normal instances.

Recall (Rec) measures the proportion of attack (positive) instances identified correctly. It is also known as the sensitivity metric or a measure of completeness. A low recall value indicates many missed identification (false negative). For example, a 20% recall value shows that only 20% of the attack instances in the dataset has been correctly identified.

F_1 −score (F_1) measures the accuracy of the approach by combining *precision* and *recall*. It is also known as the harmonic mean of *accuracy* which is highly recommended for measuring imbalanced datasets (Jeni et al., 2013).

Receiver Operating Curve (ROC) is a metric to examine the trade offs between TP and FP rates, where the X-axis represents true positive rate (TPR) and Y-axis represents false positive rate (FPR). A good ROC curve should be higher than the chance level ($>50\%$) .

Area under Curve (AUC) measures area under the ROC curve which represents the separability between attack and normal instances. The higher the AUC score, the better the approach's ability to distinguish between the two instances.

Testing Time is the time taken for an ML model or ML-based approach to perform detection. In some papers, it is also known as detection time.

Training Time is the time taken for an ML model or ML-based approach to learn the patterns and build the ML model for detection.

Matthew Correlation Coefficient (MCC) takes into account all four base statistics (TP, TN, FP, FN). A positive MCC value indicates good prediction, a zero value indicates random prediction and a negative value indicates adverse prediction.

Cohen's Kappa (Kappa) is used to measure the reliability of the detection results. It can also be seen as the measure of agreement with the positive (attack) label. For example, a Kappa score of 1 signifies complete agreement (highly reliable) whereas -1 signifies complete disagreement (highly unreliable). This measure is useful in determining the quality of the ML model especially when the data is highly imbalanced.

From Table 4, we observed that most papers show the efficacy of their ML-based approaches through *accuracy* accompanied by *precision*, *recall*, and *F1-score*. However, they are insufficient to determine the practicality of implementing their approaches to the real-world environment. To achieve effective research translation, time-based metrics need to be provided such as testing time and training time to show the time efficiency. It is unlikely for a ML-model that achieves a high F1 score (high TP, low FP) to be an effective detection tool if it has a high detection time.

4.4 Current advancement of ML in a nutshell

In this section, we have presented our findings based on our survey of the current advancement in ML and we found that state-of-the-art ML approaches for ICS still focus on particular cyber attacks, only have been evaluated on a small set of datasets, and their performance results can be biased due to the lack of coherent metrics used. We found that almost all of the ML algorithms listed have been evaluated on FDI attack, which is one of the most common cyber attacks in ICS. Among these ML algorithms, DT-based algorithms (including DT-based ensembles) followed by DNN-based algorithms provide the best results in detecting cyber attacks. However, these results are only based on a limited selection of available datasets and mostly measured according to the accuracy, precision, recall, and F_1 score of the algorithms. Time-based measurements (e.g. Training time) are hardly used as part of the evaluation, but these measurements are important to determine their suitability for real-time detection in critical infrastructure. Therefore, more comprehensive evaluation across different types of ICS environment and scenarios would be required to develop robust ML algorithms that can be put into production in the real-world environment.

5 Challenges in ML for ICS security

We have identified four critical challenges facing ML research for ICS security:

Limited attack scenarios for evaluation Despite cyber attacks on ICS in critical infrastructure being extremely damaging, highly targeted and specific attacks on them are not that common. The best-known attacks tend to be varieties of the Stuxnet, BlackEnergy, Trisis, Havex or Crashoverride malware families. These malware were highly targeted to specific environments, such as the Iranian uranium enrichment plant in the case of Stuxnet, or the Ukrainian power grid in the case of BlackEnergy. Besides these targeted attacks, some attacks can be classified as opportunistic such as ransomware. At the time of writing, most opportunistic ransomware has a variety of specific ‘kill lists’ added for ICS processes. While these attacks may be less targeted, they are nonetheless equally damaging to the critical infrastructure. This situation is in stark contrast to common IT infrastructures, where cyber attacks (e.g. malware) samples tend to be large and have a significant variety.

Limited good quality, realistic datasets Apart from having limited attack scenarios, available datasets used for training, testing and evaluations of ML-based approaches in ICS are outdated, unrealistic and may only reflect specific cyber attacks such as the KDDCup’99

(Hettich, 1999) and NSL-KDD (Tavallaee et al., 2009) datasets. Both datasets are still being used despite their weaknesses (Begli et al., 2019; Raman et al., 2019; Muna et al., 2018). For example, the KDD dataset has been criticised for having redundant records, missing values and outdated attacks (McHugh, 2000). Although the NSL-KDD removed the redundant records and missing values, it still contains the same outdated attacks as its predecessor. Newer datasets have been introduced for ICS research such as the Mississippi State University (MSU) Power, Gas, and Water datasets (Morris, 2018) and Singapore University of Technology and Design's Secure Water Treatment (SWaT) dataset (Goh et al., 2016). These datasets, however, capture data from specific components or protocols in their ICS environment which restricts the types of cyber attacks that are available for detection. Moreover, most of the cyber attacks in these datasets heavily rely on the assumption that attackers have gained access and control into the system or network which limits how early a cyber attack can be detected. The main issue for limited good quality datasets, especially real-world datasets is the risk of exposing sensitive information in the datasets even after the data is anonymised. Therefore, almost no one would share their dataset from real systems publicly.

Risk of adversarial attacks ML approaches rely heavily on the correctness and accuracy of training data and pre-trained models to be effective. However, a major weakness of such approach is that it provides opportunities for attackers to exploit these training data and pre-trained models to evade detection and reduces the effectiveness of the approaches. While adversarial attack in cybersecurity has been a well-known problem for over a decade (Biggio & Roli, 2018), it has only become more prominent in the recent years due to the rise of ML approaches for cybersecurity. Adversarial attacks are different from cyber attacks because they aim to confuse ML models into making incorrect classification rather than attacking cyber infrastructures (Kurakin et al., 2016). Several recent papers have presented or demonstrated new attack vectors and potential adversarial attacks on target ML models including the impact to ICS systems (Gómez et al., 2021; Umer et al., 2021; Zizzo et al., 2020; Erba et al., 2020). For example, Gómez et al. proposed a new method called Selective and Iterative Gradient Sign Method (SIGM) that selectively modify the data of certain features in ICS devices to fool the DNN model into miss-classification. At the same time, researchers have also came up with solutions and suggestions to addressing the issue, such as adversarial learning (Anthi et al., 2021), image transformation (Agarwal et al., 2020) and neural activation (Pawlicki et al., 2020). However, these methods are either specific to a particular attack (Anthi et al., 2021) or have not been specifically tested on ICS systems (Agarwal et al., 2020; Pawlicki et al., 2020). Hence, it is unknown if current ML approaches are resilient against adversarial attacks and are able to effectively detect all types of actual cyber attacks in ICS.

In summary, the combinations of these four challenges led to one of the biggest challenges in developing ML based approaches which is the evaluation of realistic attacks. The performance of these approaches could never truly be evaluated due to the limitation in realistic attacks and datasets. Moreover, there is not a standardised set of performance metrics to measure these approaches with. Because of this, it is hard for the industry to adopt these approaches to their systems especially in CI. Clearly, there is a strong need to address these challenges, not only to develop a more effective and scalable ML-based cyber attack detector, but to increase the trustworthiness of these new tools in the real-world.

6 Recommendations

To overcome the above challenges, we have the following recommendations:

More research focusing on unsupervised, deep, and ensemble learning methods A large volume of literature has evaluated supervised learning algorithms. However, these types of approaches relies heavily on labelled datasets. In the context of ICS security, more consideration is required to research other approaches, especially unsupervised learning or semi-supervised learning to reduce the reliance on labelled datasets (Wang et al., 2016).

More consideration towards practical application of the approach rather than focusing on accuracy alone There is currently a strong focus on building accurate ML models but a relative lack of consideration for the actual implementation of the approach itself. Researchers currently evaluate the approaches in an *offline* mode, using publicly available or private datasets where the data collection method varies from dataset to dataset. In some datasets, data from multiple sources are combined manually to become a single dataset. In real-world environments, attack detection needs to be *online* (real-time) to provide timely mitigation (Keliris et al., 2016) and better computational resource management (Li et al., 2019). Therefore, we should also consider where and how to implement such ML approaches so that data can be collected in an *online* mode to ensure similar performance can be achieved.

Attack coverage should be widened to include higher diversity of cyber attacks Our study has shown that the current attack coverage used for developing and evaluating detection tools is small, and might not reflect current real-world situations. While the MITRE ATT&CK framework (which describes recent attack types and related tactics, techniques and procedures) is becoming a standard in both industry and academia, there is room for more ML research applications into attacks described in the framework. However, the MITRE ATT&CK framework is still static in nature. Future research would require more types of attacks to be considered in both the development and evaluation of detection tools to enhance the attack coverage and detection performance in a dynamic environment.

Evaluation should include scalability, time and processing costs, and reliability, not just accuracy measures Researchers mostly evaluate their approaches using only accuracy based metrics such as accuracy, F1-score, precision, recall, etc, which may not be sufficient to determine the suitability in ICS. This is because these approaches might be slow or require high processing capabilities when processing the data. Including other factors such as scalability, time and processing costs will provide a better understanding of the overall performance of the approach.

ML-based approaches should include explainability to better understand classification results and make informed decisions Some existing ML algorithms, especially deep learning, behave like a *black-box* where the algorithm's internal workings are unknown. This hinders real-world applicability due to the 'unknown' nature of these approaches where there are possibility that the ML model could be bias and not work correctly during certain situation. Future development of ML-based approaches should consider including explainability via Explainable AI (XAI) methods (Holzinger et al., 2022).

In summary, the recommendations we have provided are based on the gaps we found in the literature and the findings we found in this paper. We believe that these recommendations could be a promising direction to address the challenges mentioned in Section 5. Also, these recommendations can help illuminate the priority areas and research directions for future research in ICS security, which may lead to the development of scalable and effective ML-based approaches in detecting ICS attacks.

7 Concluding remarks

This paper presented a review of the current advances in machine learning for detecting ICS cyber attacks. We described the current vulnerability landscape and the security issues and challenges faced in ICS. We also surveyed recent machine learning approaches and analysed their performance with respect to different datasets, base classifiers and attack variety to gather insights on the current advancement in the field. Our finding shows that only a handful of types of cyber attacks are included in the datasets, which only cover a small portion of the type of actual ICS cyber attacks. It was also shown that there was no clear one-size-fits-all type of machine learning algorithm which is suitable for all types. It is hence critical that ICS attack solutions adopt a cocktail of ML approaches for the variety of attacks faced by ICS. It is our hope that this paper will illuminate new research directions in ML approaches for scalable and effective ICS attack detection.

Acknowledgements Not applicable.

Author contributions Koay wrote the main manuscript text, and prepare figures and tables. Ko wrote some parts of Sections 1 and 6. Hettema wrote most part of Section 3. All authors reviewed the manuscripts.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions This work is funded by the University of Queensland Cyber Security Initiative Grant.

Data availability Not applicable.

Declarations

Human and animal Ethics Not applicable.

Ethics approval and consent to participate Not applicable.

Consent for publication Yes, we consent this paper to be published in the Journal of Intelligent Information System.

Competing interests No, I declare that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

References

Agarwal, A., Singh, R., Vatsa, M., & Ratha, N. (2020). Image transformation-based defense against adversarial perturbation on deep learning models. *IEEE Transactions on Dependable and Secure Computing*, *18* (5), 2106–2121.

Ahmed, C.M., Palleti, V.R., & Mathur, A.P. (2017). Wadi: a water distribution testbed for research in the design of secure cyber physical systems. In *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks* (pp. 25–28).

Al-Abassi, A., Karimipour, H., Dehghantanha, A., & Parizi, R.M. (2020). An ensemble deep learning-based cyber-attack detection in industrial control system. *IEEE Access*, *8*, 83965–83973.

Alexander, O., Belisle, M., & Steele, J. (2020). *Mitre att&ck®; for industrial control systems: Design and philosophy*. Bedford, MA, USA: The MITRE Corporation.

Alsmadi, I., Dwekat, Z., Cantu, R., & Al-Ahmad, B. (2021). Vulnerability assessment of industrial systems using shodan. Cluster Computing, 1–11.

Anthi, E., Williams, L., Burnap, P., & Jones, K. (2021). A three-tiered intrusion detection system for industrial control systems. *Journal of Cybersecurity*, *7*(1), 006.

Anthi, E., Williams, L., Rhode, M., Burnap, P., & Wedgbury, A. (2021). Adversarial attacks on machine learning cybersecurity defences in industrial control systems. *Journal of Information Security and Applications*, *58*, 102717.

Anton, S.D., Gundall, M., Fraunholz, D., & Schotten, H.D. (2019). Implementing scada scenarios and introducing attacks to obtain training data for intrusion detection methods. In *ICCWS 2019 14th International Conference on Cyber Warfare and Security: ICCWS 2019* (p. 56). Academic Conferences and publishing limited.

Anton, S.D.D., Sinha, S., & Schotten, H.D. (2019). Anomaly-based intrusion detection in industrial data with svm and random forests. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)* (pp. 1–6). IEEE.

Antrobus, R., Frey, S., Green, B., & Rashid, A. (2016). Simaticscan: Towards a specialised vulnerability scanner for industrial control systems. In *4Th international symposium for ICS & SCADA cyber security research 2016 4* (pp. 11–18).

Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. In *2018 10th International Conference on Cyber Conflict (CyCon)* (pp. 371–390). IEEE.

Asghar, M.R., Hu, Q., & Zeadally, S. (2019). Cybersecurity in industrial control systems: issues, technologies, and challenges. *Computer Networks*, *165*, 106946.

Assante, M.J., & Lee, R.M. (2015). The industrial control system cyber kill chain. SANS Institute InfoSec Reading Room 1.

Beaver, J.M., Borges-Hink, R.C., & Buckner, M.A. (2013). An evaluation of machine learning methods to detect malicious scada communications. In *2013 12th International Conference on Machine Learning and Applications*, (Vol. 2 pp. 54–59). IEEE.

Begli, M., Derakhshan, F., & Karimipour, H. (2019). A layered intrusion detection system for critical infrastructure using machine learning. In *2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE)* (pp. 120–124). IEEE.

Beresford, D. (2011). Exploiting siemens simatic s7 plcs. *Black Hat USA*, *16*(2), 723–733.

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, *84*, 317–331.

Buczak, A.L., & Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, *18*(2), 1153–1176.

Case, D.U. (2016). Analysis of the cyber attack on the ukrainian power grid. Electricity Information Sharing and Analysis Center (E-ISAC) 388.

Chan, R., Chow, K.-P., & Chan, C.-F. (2019). Defining attack patterns for industrial control systems. In *International Conference on Critical Infrastructure Protection* (pp. 289–309). Springer.

Chen, X., Zhang, L., Liu, Y., & Tang, C. (2018). Ensemble learning methods for power system cyber-attack detection. In *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 613–616). IEEE.

Chu, A., Lai, Y., & Liu, J. (2019). Industrial control intrusion detection approach based on multiclassification googlenet-lstm model. Security and Communication Networks 2019.

Cui, L., Qu, Y., Gao, L., Xie, G., & Yu, S. (2020). Detecting false data attacks using machine learning techniques in smart grid: A survey. Journal of Network and Computer Applications 102808.

Dias, A.L., Sestito, G.S., Turcato, A.C., & Brandão, D. (2018). Panorama, challenges and opportunities in profinet protocol research. In *2018 13th IEEE International Conference on Industry Applications (INDUSCON)* (pp. 186–193). IEEE.

Diez-Olivan, A., Del Ser, J., Galar, D., & Sierra, B. (2019). Data fusion and machine learning for industrial prognosis: Trends and perspectives towards industry 4.0. *Information Fusion*, *50*, 92–111.

Downs, J.J., & Vogel, E.F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering*, *17*(3), 245–255.

East, S., Butts, J., Papa, M., & Shenoi, S. (2009). A taxonomy of attacks on the dnp3 protocol. In *International Conference on Critical Infrastructure Protection* (pp. 67–81). Springer.

El Mrabet, Z., Kaabouch, N., El Ghazi, H., & El Ghazi, H. (2018). Cyber-security in smart grid: Survey and challenges. *Computers & Electrical Engineering*, *67*, 469–482.

Elnour, M., Meskin, N., Khan, K., & Jain, R. (2020). A dual-isolation-forests-based attack detection framework for industrial control systems. *IEEE Access*, *8*, 36639–36651.

Erba, A., Taormina, R., Galelli, S., Pogliani, M., Carminati, M., Zanero, S., & Tippenhauer, N.O. (2020). Constrained concealment attacks against reconstruction-based anomaly detectors in industrial control systems. In *Annual Computer Security Applications Conference* (pp. 480–495).

Feld, J. (2004). Profinet-scalable factory communication for all applications. In *IEEE International Workshop on Factory Communication Systems, 2004. Proceedings* (pp. 33–38). IEEE.

Goh, J., Adepu, S., Junejo, K.N., & Mathur, A. (2016). A dataset to support research in the design of secure water treatment systems. In *International Conference on Critical Information Infrastructures Security* (pp. 88–99). Springer.

Gómez, Á.L.P., Maimó, L.F., Celdrán, A.H., Clemente, F.J.G., & Cleary, F. (2021). Crafting adversarial samples for anomaly detectors in industrial control systems. *Procedia Computer Science*, *184*, 573–580.

Gómez, Á.L.P., Maimó, L.F., Celdran, A.H., Clemente, F.J.G., Sarmiento, C.C., Masa, C.J.D.C., & Nistal, R.M. (2019). On the generation of anomaly detection datasets in industrial control systems, (Vol. 7.

Gonzalez, D., Alhenaki, F., & Mirakhorli, M. (2019). Architectural security weaknesses in industrial control systems (ics) an empirical study based on disclosed software vulnerabilities. In *2019 IEEE International Conference on Software Architecture (ICSA)* (pp. 31–40). IEEE.

Govil, N., Agrawal, A., & Tippenhauer, N.O. (2017). On ladder logic bombs in industrial control systems, 110–126.

Graham, J., Hieb, J., & Naber, J. (2016). Improving cybersecurity for industrial control systems. In *2016 IEEE 25th International Symposium on Industrial Electronics (ISIE)* (pp. 618–623). IEEE.

Hahn, A. (2016). Operational technology and information technology in industrial control systems. In *Cyber-security of SCADA and other industrial control systems* (pp. 51–68). Cham: Springer.

Handa, A., & Semwal, P. (2022). Evaluating performance of scalable fair clustering machine learning techniques in detecting cyber attacks in industrial control systems. In *Handbook of Big Data Analytics and Forensics* (pp. 105–116). Cham: Springer.

Hettich, S. (1999). *Kdd cup 1999 data*. The UCI KDD Archive.

Hijazi, S., & Obaidat, M.S. (2019). Address resolution protocol spoofing attacks and security approaches: a survey. *Security and Privacy*, *2*(1), 49.

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., & Samek, W. (2022). Explainable ai methods-a brief overview. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers* (pp. 13–38). Springer.

Hurst, W., Merabti, M., & Fergus, P. (2014). A survey of critical infrastructure security. In *International Conference on Critical Infrastructure Protection* (pp. 127–138). Springer.

Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C.M., & Sun, J. (2017). Anomaly detection for a water treatment system using unsupervised machine learning. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 1058–1065). IEEE.

Jeni, L.A., Cohn, J.F., & De La Torre, F. (2013). Facing imbalanced data–recommendations for the use of performance metrics. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 245–251). IEEE.

Juba, B., & Le, H.S. (2019). Precision-recall versus accuracy and the role of large data sets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, (Vol. 33 pp. 4039–4048).

Kanamaru, H. (2017). Bridging functional safety and cyber security of sis/scs. In *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)* (pp. 279–284). IEEE.

Keliris, A., Salehghaffari, H., Cairl, B., Krishnamurthy, P., Maniatakos, M., & Khorrami, F. (2016). Machine learning-based defense against process-aware attacks on industrial control systems. In *2016 IEEE International Test Conference (ITC)* (pp. 1–10). IEEE.

Khan, I.A., Pi, D., Khan, Z.U., Hussain, Y., & Nawaz, A. (2019). Hml-ids: A hybrid-multilevel anomaly prediction approach for intrusion detection in scada systems. *IEEE Access*, *7*, 89507–89521.

Ko, R.K.L. (2020). Cyber autonomy: Automating the hacker-self-healing, self-adaptive, automatic cyber defense systems and their impact to the industry, society and national security, 173–191.

Kravchik, M., & Shabtai, A. (2018). Detecting cyber attacks in industrial control systems using convolutional neural networks. In *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy* (pp. 72–83).

Kravchik, M., & Shabtai, A. (2021). Efficient cyber attack detection in industrial control systems using lightweight neural networks and pca. IEEE Transactions on Dependable and Secure Computing.

Krithivasan, K., Pravinraj, S., VS, S.S., & et al. (2020). Detection of cyberattacks in industrial control systems using enhanced principal component analysis and hypergraph-based convolution neural network (epca-hg-cnn). *IEEE Transactions on Industry Applications*, *56*(4), 4394–4404.

Kshetri, N., & Voas, J. (2017). Hacking power grids: a current problem. *Computer*, *50*(12), 91–95.

Kurakin, A., Goodfellow, I., & Bengio, S. (2016). Adversarial machine learning at scale. https://doi.org/10.48550/arXiv.1611.01236

Kwon, D., Kim, H., Kim, J., Suh, S.C., Kim, I., & Kim, K.J. (2019). A survey of deep learning-based network anomaly detection. *Cluster Computing*, *22* (1), 949–961.

Langner, R. (2011). Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, *9*(3), 49–51.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Li, G., Shen, Y., Zhao, P., Lu, X., Liu, J., Liu, Y., & Hoi, S.C. (2019). Detecting cyberattacks in industrial control systems using online learning algorithms. *Neurocomputing*, *364*, 338–348.

Lin, Q., Adepu, S., Verwer, S., & Mathur, A. (2018). Tabor: a graphical model-based approach for anomaly detection in industrial control systems. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security* (pp. 525–536).

Ling, J., Zhu, Z., Luo, Y., & Wang, H. (2021). An intrusion detection method for industrial control systems based on bidirectional simple recurrent unit. *Computers & Electrical Engineering*, *91*, 107049.

Liu, J., Yin, L., Hu, Y., Lv, S., & Sun, L. (2018). A novel intrusion detection algorithm for industrial control systems based on cnn and process state transition. In *2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC)* (pp. 1–8). IEEE.

Long, M., Wu, C.-H., & Hung, J.Y. (2005). Denial of service attacks on network-based control systems: impact and mitigation. *IEEE Transactions on Industrial Informatics*, *1*(2), 85–96.

Maynard, P., McLaughlin, K., & Sezer, S. (2020). Decomposition and sequential-and analysis of known cyber-attacks on critical infrastructure control systems. *Journal of Cybersecurity*, *6*(1), 020.

Mazurczyk, W., & Caviglione, L. (2021). Cyber reconnaissance techniques. *Communications of the ACM*, *64*(3), 86–95.

McHugh, J. (2000). Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory. *ACM Transactions on Information and System Security (TISSEC)*, *3*(4), 262–294.

Mehrfeld, J. (2020). Cyber security threats and incidents in industrial control systems. In *International conference on human-computer interaction* (pp. 599–608). Cham: Springer.

Men, J., Lv, Z., Zhou, X., Han, Z., Xian, H., & Song, Y.-N. (2020). Machine learning methods for industrial protocol security analysis: Issues, taxonomy, and directions. *IEEE Access*, *8*, 83842–83857.

Mo, Y., & Sinopoli, B. (2010). False data injection attacks in control systems. In *Preprints of the 1st workshop on secure control systems* (pp. 1–6).

Mokhtari, S., Abbaspour, A., Yen, K.K., & Sargolzaei, A (2021). A machine learning approach for anomaly detection in industrial control systems based on measurement data. *Electronics*, *10*(4), 407.

Morris, T. (2018). Industrial control system (ICS) cyber attack datasets.

Morris, T., & Gao, W. (2014). Industrial control system traffic data sets for intrusion detection research. In *International Conference on Critical Infrastructure Protection* (pp. 65–78). Springer.

Muna, A. -H., Moustafa, N., & Sitnikova, E. (2018). Identification of malicious activities in industrial internet of things based on deep learning models. *Journal of information security and applications*, *41*, 1–11.

Nguyen, K.K., Hoang, D.T., Niyato, D., Wang, P., Nguyen, D., & Dutkiewicz, E. (2018). Cyberattack detection in mobile cloud computing: A deep learning approach. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1–6). IEEE.

Pawlicki, M., Choraś, M., & Kozik, R. (2020). Defending network intrusion detection systems against adversarial evasion attacks. *Future Generation Computer Systems*, *110*, 148–154.

Potluri, S., & Diedrich, C. (2019). Deep learning based efficient anomaly detection for securing process control systems against injection attacks. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)* (pp. 854–860). IEEE.

Raman, M.G., Somu, N., & Mathur, A.P. (2019). Anomaly detection in critical infrastructure using probabilistic neural network. In *International Conference on Applications and Techniques in Information Security* (pp. 129–141). Springer.

Ramotsoela, T.D., Hancke, G.P., & Abu-Mahfouz, A.M. (2020). Behavioural intrusion detection in water distribution systems using neural networks, (Vol. 8.

Ribu Hassini, S., Gireesh Kumar, T., & Kowshik Hurshan, S. (2022). A machine learning and deep neural network approach in industrial control systems. In *ICT Analysis and Applications* (pp. 525–536). Singapore: Springer.

Rieth, C., Amsel, B., Tran, R., & Cook, M. (2017). Additional tennessee eastman process simulation data for anomaly detection evaluation. Harvard Dataverse 1.

Robles-Durazno, A., Moradpoor, N., McWhinnie, J., & Russell, G. (2018). A supervised energy monitoring-based machine learning approach for anomaly detection in a clean water supply system. In *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)* (pp. 1–8). IEEE.

Sanger, D.E., Krauss, C., & Perlroth, N. (2021). Cyberattack forces a shutdown of a top us pipeline. The New York Times 8.

Selim, G.E.I., Hemdan, E.E.-D., Shehata, A.M., & El-Fishawy, N.A. (2021). Anomaly events classification and detection system in critical industrial internet of things infrastructure using machine learning algorithms. Multimedia Tools and Applications, 1–22.

Shin, H.-K., Lee, W., Yun, J.-H., & Kim, H. (2020). Hai 1.0: Hil-based augmented {ICS} security dataset. In *13Th USENIX workshop on cyber security experimentation and test (CSET 20).*

Sokolov, A.N., Pyatnitsky, I.A., & Alabugin, S.K. (2019). Applying methods of machine learning in the task of intrusion detection based on the analysis of industrial process state and ics networking. *FME Transactions*, *47*(4), 782–789.

Süzen, A.A. (2021). Developing a multi-level intrusion detection system using hybrid-dbn. *Journal of Ambient Intelligence and Humanized Computing*, *12*(2), 1913–1923.

Swales, A., et al. (1999). Open modbus/tcp specification. *Schneider Electric*, *29*, 3–19.

Tan, S., Guerrero, J.M., Xie, P., Han, R., & Vasquez, J.C. (2020). Brief survey on attack detection methods for cyber-physical systems. *IEEE Systems Journal*, *14*(4), 5329–5339.

Taormina, R., Galelli, S., Tippenhauer, N.O., Salomons, E., Ostfeld, A., Eliades, D.G., Aghashahi, M., Sundararajan, R., Pourahmadi, M., Banks, M.K., & et al. (2018). Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks. *Journal of Water Resources Planning and Management*, *144*(8), 04018048.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A.A. (2009). A detailed analysis of the kdd cup 99 data set. In *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications* (pp. 1–6). IEEE.

Terai, A., Abe, S., Kojima, S., Takano, Y., & Koshijima, I. (2017). Cyber-attack detection for industrial control system monitoring with support vector machine based on communication profile. In *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (pp. 132–138). IEEE.

Torres, J.M., Comesaña, C.I., & Garcia-Nieto, P.J. (2019). Machine learning techniques applied to cybersecurity. *International Journal of Machine Learning and Cybernetics*, *10*(10), 2823–2836.

Turnipseed, I.P. (2015). A new scada dataset for intrusion detection system research. PhD thesis, Mississippi State University.

Umer, M.A., Ahmed, C.M., Jilani, M.T., & Mathur, A.P. (2021). Attack rules: an adversarial approach to generate attacks for industrial control systems using machine learning. In *Proceedings of the 2th Workshop on CPS&IoT Security and Privacy* (pp. 35–40).

Umer, M.A., Junejo, K.N., Jilani, M.T., & Mathur, A.P. (2022). Machine learning for intrusion detection in industrial control systems: applications, challenges, and recommendations. International Journal of Critical Infrastructure Protection, 100516.

Upadhyay, D., & Sampalli, S. (2020). Scada (supervisory control and data acquisition) systems: Vulnerability assessment and security recommendations. *Computers & Security*, *89*, 101666.

Voyiatzis, A.G., Katsigiannis, K., & Koubias, S. (2015). A modbus/tcp fuzzer for testing internetworked industrial systems. In *2015 IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA)* (pp. 1–6). IEEE.

Wang, J., Miao, Y., Khamis, A., Karray, F., & Liang, J. (2016). Adaptation approaches in unsupervised learning: a survey of the state-of-the-art and future directions. In *International Conference on Image Analysis and Recognition*. Springer (pp. 3–11).

Wang, C., Wang, B., Liu, H., & Qu, H. (2020). Anomaly detection for industrial control system based on autoencoder neural network. Wireless Communications and Mobile Computing 2020.

Wang, W., Xie, Y., Ren, L., Zhu, X., Chang, R., & Yin, Q. (2018). Detection of data injection attack in industrial control system using long short term memory recurrent neural network. In *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 2710–2715). IEEE.

Williams, T.J. (1994). The purdue enterprise reference architecture. *Computers in Industry*, *24*(2-3), 141–158.

Wu, H., Geng, Y., Liu, K., & Liu, W. (2019). Research on programmable logic controller security. In *IOP Conference Series: Materials Science and Engineering*, (Vol. 569 p. 042031). IOP Publishing.

Xiao, Y.-j., Xu, W.-y., Jia, Z.-h., Ma, Z.-r., & Qi, D.-l. (2017). Nipad: a non-invasive power-based anomaly detection scheme for programmable logic controllers. *Frontiers of Information Technology & Electronic Engineering*, *18*(4), 519–534.

Xie, X., Wang, B., Wan, T., & Tang, W. (2020). Multivariate abnormal detection for industrial control systems using 1d cnn and gru. *IEEE Access*, *8*, 88348–88359.

Yau, K., & Chow, K.-P. (2017). Detecting anomalous programmable logic controller events using machine learning. In *IFIP International Conference on Digital Forensics* (pp. 81–94). Springer.

Yilmaz, M., Catak, F.O., & Gul, E. (2019). Sensor based cyber attack detections in critical infrastructures using deep learning algorithms. Computer Science 20.

Zhang, F., Kodituwakku, H.A.D.E., Hines, J.W., & Coble, J. (2019). Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data. *IEEE Transactions on Industrial Informatics*, *15*(7), 4362–4369.

Zhang, D., Wang, Q.-G., Feng, G., Shi, Y., & Vasilakos, A.V. (2021). A survey on attack detection, estimation and control of industrial cyber–physical systems. ISA transactions.

Zizzo, G., Hankin, C., Maffeis, S., & Jones, K. (2020). Adversarial attacks on time-series intrusion detection for industrial control systems. In *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (pp. 899–910). IEEE.