Anomalous behaviour detection for cyber defence in modern Industrial Control Systems

By

Gabriela Ahmadi-Assalemi

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for the degree of Doctor of Philosophy

October 2022

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or for any other purpose (unless otherwise indicated). Save for any express acknowledgments, references and/or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

Signature……Gabriela Ahmadi-Assalemi……

Date……11 October 2022……

Abstract

The fusion of pervasive internet connectivity and emerging technologies in smart cities creates fragile cyber-physical-natural ecosystems. Industrial Control Systems (ICS) are intrinsic parts of smart cities and critical to modern societies. Not designed for interconnectivity or security, disruptor technologies enable ubiquitous computing in modern ICS. Aided by artificial intelligence and the industrial internet of things they transform the ICS environment towards better automation, process control and monitoring. However, investigations reveal that leveraging disruptive technologies in ICS creates security challenges exposing critical infrastructure to sophisticated threat actors including increasingly hostile, well-organised cybercrimes and Advanced Persistent Threats. Besides external factors, the prevalence of insider threats includes malicious intent, accidental hazards and professional errors. The sensing capabilities create opportunities to capture various data types. Apart from operational use, this data combined with artificial intelligence can be innovatively utilised to model anomalous behaviour as part of defence-in-depth strategies. As such, this research aims to investigate and develop a security mechanism to improve cyber defence in ICS.

Firstly, this thesis contributes a Systematic Literature Review (SLR), which helps analyse frameworks and systems that address CPS' cyber resilience and digital forensic incident response in smart cities. The SLR uncovers emerging themes and concludes several key findings. For example, the chronological analysis reveals key influencing factors, whereas the data source analysis points to a lack of real CPS datasets with prevalent utilisation of software and infrastructure-based simulations.

Further in-depth analysis shows that cross-sector proposals or applications to improve digital forensics focusing on cyber resilience are addressed by a small number of research studies in some smart sectors.

Next, this research introduces a novel super learner ensemble anomaly detection and cyber risk quantification framework to profile anomalous behaviour in ICS and derive a cyber risk score. The proposed framework and associated learning models are experimentally validated. The produced results are promising and achieve an overall F1-score of 99.13%, and an anomalous recall score of 99% detecting anomalies lasting only 17 seconds ranging from 0.5% to 89% of the dataset.

Further, a one-class classification model is developed, leveraging stream rebalancing followed by adaptive machine learning algorithms and drift detection methods. The model is experimentally validated producing promising results including an overall Matthews Correlation Coefficient (MCC) score of 0.999 and the Cohen's Kappa (K) score of 0.9986 on limited variable single-type anomalous behaviour per data stream. Wide data streams achieve an MCC score of 0.981 and a K score of 0.9808 in the prevalence of multiple types of anomalous instances.

Additionally, the thesis scrutinises the applicability of the learning models to support digital forensic readiness. The research study presents the concept of digital witness and digital chain of custody in ICS. Following that, a use case integrating blockchain technologies into the design of ICS to support digital forensic readiness is discussed.

In conclusion, the contributions of this research thesis help towards developing the next generation of state-of-the-art methods for anomalous behaviour detection in ICS defence-in-depth.

Contents

List of Figures

List of Tables

Dedication

To my husband Masoud and my daughter Nazanin who have always supported me to continue my research and without their unconditional support, this journey would not have been possible.

Acknowledgements

I am deeply indebted to my supervisor and Director of Studies, Dr Haider M. al-Khateeb, for the opportunities, his advice, constructive criticism and continuous encouragement. I could progress this research due to his support and guidance.

I would like to express my sincere appreciation to Prof. Amar Aggoun for his support of my programme of study. Additionally, my gratitude and special thanks go to Dr Gregory Epiphaniou for his encouragement and advice during the early part of my research. Further, I am thankful to Prof. Prashant Pillai for his support. A big thank you to the viva voce examination Chair, Dr Martin Partridge, and the examiners Dr Liam Naughton and Dr Alexios Mylonas for their constructive feedback and insightful comments.

Finally, I am grateful to my friends and colleagues for their company and encouragement.

1. Chapter: Introduction

1.1 Introduction

Industry 4.0 has accelerated the systematic integration of the Internet of Things (IoT), Cyber-Physical Systems (CPS) and enabling technologies with aspects of smart cities, extending its initial vision beyond the production and manufacturing industries [1-5]. Smart cities have been enhanced by a fusion of ubiquitous internet connectivity with innovative applications of disruptive technologies creating highly fragile Cyber-Physical-Natural (CPN) ecosystems [6, 7]. CPS are a key component of smart cities that converge the physical and digital realms to achieve better value, innovation and sustainability [8-11]. Industrial Control Systems (ICS), a subset of CPS, are prevalent in critical infrastructures [12, 13]. The Industrial Internet of Things (IIoT) enable ICS to integrate disruptive technologies and innovative solutions to introduce process automation, monitoring and distributed control to support modern critical utility infrastructures in smart cities of the future [7, 14, 15]. Besides opportunities presented by enabling and emerging disruptive technologies such as 5G, edge computing, and Artificial Intelligence (AI), they increase the attack surface in ICS due to introducing new attack vectors [5, 15-17].

Against the backdrop of ICS environments transformed for better automation, process control, flexible and efficient administration, challenges such as complex interconnectivity and disparate priorities between Information Communication Technologies (ICT) and Operational Technologies (OT) make ICS vulnerable to cyber-attacks [18-20]. Throughout this thesis, the terms ICT and Information Technology (IT) are used interchangeably. Likewise, human behaviour is a fundamental part of ICS with a potentially profound impact. For example, an insider who has authorised access exposes a difficult-to-detect attack vector. Additionally,

the risks associated with insiders extend the threat landscape beyond the human factor to include Cyber-Physical Objects (CPO) that act as smart cyber insiders [5-7, 21, 22].

This research thesis recognises the major impact and profound consequences of disruption from cyber-attacks against ICS. The thesis researches and empirically evaluates the employment of Machine Learning (ML) techniques to develop and optimise the security mechanism to improve the proactive cyber defence in ICS. This research leverages data from physical plant sensors to identify anomalies and quantify cyber risk as part of a layered defence-in-depth approach in ICS. Furthermore, this research investigates support for reactive defence such as Digital Forensics and Incident Response (DFIR) as part of Digital Forensic (DF) readiness towards improving cyber resilience in ICS.

1.2 Context

CPS are vulnerable to many cybersecurity attacks due to the complexities introduced by growing networks of connected objects and human users. The integration of disruptive technologies creates fragile CPN ecosystems that expose the smart cities of the future to complex security challenges [5, 6, 23]. CPS should be secure-by-design to counter cybersecurity threats, resist cyber-attacks accurately, and function effectively under adverse conditions to minimise impact.

This is challenging to achieve in ICS due to the complexities posed by cyber components prevalent in OT such as Supervisory Control and Data Acquisition (SCADA), Human-Machine-Interface (HMI), Programmable Logic Controllers (PLC) and the quantum of sensors and actuators underpinned by communication networks

[24, 25]. The integration of insecure devices and disruptive technologies into ICS to achieve better automation, monitoring and distributed control increases the attack surface.

While traditionally ICS were isolated systems, modern ICS are complex, distributed and interdependent [19]. However, the cybersecurity measures that apply to Information Systems (IS) are not necessarily applicable to OT. Unlike ICT, operational infrastructure is typically highly automated with a focus on reliability, availability and safety. Whereas ICT is typically secure, centred around Confidentiality, Integrity and Availability (CIA) of the data [5].

1.3 Research Problem

ICS have a complex threat environment and are attractive targets vulnerable to sophisticated threat actors including organised cybercrime and Advanced Persistent Threats (APT) [5, 22, 26-31]. Besides external factors, ICS are vulnerable to insider threats including malicious intent, accidental hazards and professional errors. These threat actors have authorised access within the organisation and represent a difficult-to-detect cybersecurity challenge. According to CERT Insider Threat Centre, insider threats include unintentional or malicious actors, originating within the organisation where the insider has authorised access [32, 33]. In addition, the insider threat model extends the human element to include CPO such as robots and drones executing activities alongside human employees, which we consider smart cyber insiders. Besides targeted attacks, ICS are susceptible to challenges that result from the disparity of organisation-influenced priorities between ICT and OT further complicating the protection mechanisms in ICS [18, 19].

ICS were initially designed as isolated systems, with components not designed for security, combined with a decades-long lifecycle to support critical operations. With little security development, the continued reliance on security by obscurity is not a sustainable security defence mechanism [19, 34]. ICS are complex, interconnected and distributed networks that consist of segments including corporate networks, logical and physical control [24]. The complex interconnectivity and the prevalence of cyber components within these segments such as SCADA, HMI, PLC and the quantum of sensors underpinned by communication networks make ICS vulnerable to cyber-attacks [24, 25]. The increase in cybercrime in ICS is attributed to the fusion of devices, sensors and internet connectivity converging the physical and cyber domains [19, 30, 35-37].

ICS such as water treatment, water distribution plants, manufacturing, power grids, wind turbines, and transportation are critical to the functioning of industrial facilities and have emerged as integral parts of smart cities. ICS have a massive impact on the wider society which, if disrupted, could result in profoundly devastating consequences, real-world damage with a significant and hazardous impact on communities [6, 7, 25, 38, 39]. The impact could have a monetary impact on businesses, loss of Intellectual Property (IP), and threat to national security including socio-economic consequences on entire ecosystems [6, 23, 25, 39]. However, not every threat can be mitigated and not all cyber-attacks can be avoided, hence, DF readiness should be factored in as part of a defence-in-depth approach to ICS [5, 40-43]. Reactive mechanisms alone do not effectively mitigate these threats. Therefore, due to the complex security landscape in ICS, proactive and reactive cyber defence mechanisms are required to improve the cyber resilience in ICS as

part of a defence-in-depth capability, pointing to the significance of research to develop protective mechanisms.

ICS sensor-generated data is used for operational monitoring. Besides utilising ICS physical plant sensor data for operational monitoring, control and automation of processes, this data can be used innovatively to improve the defence-in-depth thus increasing resilience to cyber-attacks. There is little or no research contributing to other aspects enabled by anomaly detection. This research attempts to address the problem of a security process-driven protective mechanism to profile anomalous behaviour in ICS from sensors generated data and quantify cyber risk in the prevalence of anomalous behaviour. Anomalous behaviour detection from sensor data has key advantages. Anomalous behaviour detection is attainable from sensor data hence previously unknown attacks are detectable including external threat actors and smart-cyber insiders.

Furthermore, the fusion of ICT and OT environments, quantum and sources of data from sensors, actuators and controllers continually produce streams of data. Data streams present unique characteristics, they are continuous, mainly contain normal instances, evolve and the entire stream cannot be processed and analysed as a whole. Due to these characteristics, data streams are not suitable for traditional ML techniques, also known as batch learning. Batch learning is a group of techniques used for the detection, classification and prediction of anomalies **[**44**]**. ML techniques are broadly grouped into supervised, semi-supervised and unsupervised ML algorithms [45, 46]. Supervised learning algorithms are widely used in scientific research to implement security models and solutions **[**47**].** Summarily, supervised learning requires labelled data to train the model before making predictions on

unseen datasets. Semi-supervised learning algorithms utilise partially labelled data, where labels typically exist for data instances representing normal behaviour [45, 46, 48]. Whereas unsupervised learning algorithms leverage unlabelled data to train the model by understanding the underlying dataset's relationships and structures. ML techniques analyse static, finite datasets loaded into the computer's active memory to train the ML model offline, relying on historical data [49]. However, a new model is needed when data distribution changes. This approach is computationally expensive and has profound disadvantages. Thus, batch learning approach is unsuitable for mining continuous data streams in near-real-time to sustain effective cyber defence mechanisms. Specialist algorithms are required that are capable to process data dynamically and adapt to changing and scaled data [19, 50-52]. Hence, to address anomalous behaviour detection from ICS data streams in near-real-time, the research also leverages an adaptive form of supervised learning against the evolving threat landscape in ICS as part of a layered defence-in-depth approach.

1.4 Motivation

ICS are intrinsic elements of smart cities and critical to the operations of industrial facilities including water treatment systems, nuclear power plants, transport, electricity generation plants and gas pipelines. Due to the increase and sophistication of cybercrime, the countermeasures in ICS require a consistent and coordinated approach [28, 30, 40, 43]. Hence, modern ICS must adapt to emerging and evolving threats and develop defence measures with support for mission assurance.

A commitment to a more proactive approach to protecting operational infrastructure and the fact that ICS-related cyber defence is an active research area

is encouraging [5, 43, 53-55]. Besides automation, data from physical plant sensors coupled with ML techniques could help produce intelligence to develop countermeasures as part of a defence-in-depth approach to improve cyber resilience in ICS. Furthermore, the value of the data extracted from the sensors could contribute to post-incident investigations and reconstruction of events as part of DF readiness [6, 56].

This thesis is motivated by the opportunity offered by ML techniques to develop protective concepts to address the emerging and increasing threats prevalent in ICS and contribute to improving ICS cyber resilience by advancing knowledge in this field.

1.5 Research Aims and Objectives

1.5.1 Aims

Cyber defence can be broadly classified into two main groups: reactive defence and proactive defence. Hence, the thesis has two directions sharing the main goal to improve cyber resilience in ICS. The study aims to investigate and develop a security mechanism to improve the proactive cyber defence in ICS to support its mission objectives. The mechanism seeks to be security process-driven, integrate novel ML techniques and utilise data generated from physical plant sensors. The mechanism intends to be testable, trustworthy and repeatable. Secondly, the study aims to investigate how the security mechanism addresses the reactive defence as part of DF readiness in ICS.

1.5.2 Objectives

To advance the aims, this research sets out to achieve the following objectives in this thesis as outlined:

1) Conduct a Systematic Literature Review (SLR) of current cyber resilience and DFIR approaches in CPS in smart cities.
2) Investigate the current ML approaches to improve proactive cyber defence of ICS.
3) Develop an approach that leverages ML techniques to improve cyber resilience in ICS.
4) Develop a novel ML-based anomaly detection and cyber risk quantification mechanism, evaluating and analysing the efficacy.
5) Investigate the mechanism's support for post-incident investigations as part of DF readiness.

The Research Questions (RQs) outlined in Table 1 are mapped to the objectives and the relevant chapters.

Table 1 Mapping of Research Objectives chapters and research questions of related key publications

Research Questions (RQs)	Objectives				
	1	2	3	4	5
How do existing frameworks and systems that address CPSs in smart cities support cyber resilience and what empirical evidence has been reported?	✓	✓			
How do the identified frameworks and systems in smart cities address modern Digital Forensics and Incident Response (DFIR)?	✓				✓
What are the current cross-sector proposals or applications in smart cities that attempt to utilise interactions in CPS for the purpose of improving DFIR?	✓				✓
How can we form a framework which addresses anomaly detection in Cyber-Physical Systems (CPS) such that it is optimised and security-process driven?		✓	✓	✓	
How can this framework be utilised to quantify the cyber risk in CPS?			✓	✓	
And how can this framework support Digital Forensic and Incident Response (DFIR)?					✓
How can we form a model which addresses anomalous behaviour detection in ICS from data streams in near-real-time and maintains effective performance?		✓	✓	✓	
What is the effectiveness of an algorithm on anomalous behaviour detection in ICS utilising data streams?				✓	
And how does the performance of such a model compare with the performance of batch learning-based defence mechanisms?			✓	✓	
Chapters:	2	2,4,6	3,4,6	5	2, 5,6

1.6 Related Publications

The following publications are linked to the PhD thesis:

- Gabriela Ahmadi-Assalemi, Haider M. al-Khateeb, Meryem Ammi, “Adaptive Learning Anomaly Detection and Classification Model for Cyber and Physical Threats in Industrial Control Systems” submitted, under review.

- G. Ahmadi-Assalemi, and H. Al-Khateeb, “Blockchain technologies in the design of Industrial Control Systems for Smart Cities”, IEEE Blockchain Technical Briefs, vol. Q2 2022, 2022,[Online], Accessed: 04/09/2022, Available: https://blockchain.ieee.org/images/files/pdf/techbriefs-2022-q2/blockchain-technologies-in-the-design-of-industrial-control-systems-for-smart-cities.pdf

- G. Ahmadi-Assalemi, H. Al-Khateeb, G. Epiphaniou, and A. Aggoun, “Super Learner Ensemble for Anomaly Detection and Cyber risk Quantification in Industrial Control Systems”, IEEE Internet of Things Journal, pp. 1-1, 2022,[Online], https://doi.org/10.1109/JIOT.2022.3144127

- G. Ahmadi-Assalemi, H. Al-Khateeb, G. Epiphaniou, and C. Maple, “Cyber Resilience and Incident Response in Smart Cities: A Systematic Literature Review”, MDPI Smart Cities, vol. 3, no. 3, pp. 894-927, Aug 2020,[Online], https://doi.org/10.3390/smartcities3030046

The following publications are outputs from my broader PhD research and PhD-related MSc teaching activities delivered during my PhD programme:

- G. Ahmadi-Assalemi, H. Al-Khateeb, Amar Aggoun, “Privacy-enhancing technologies in the Design of Digital Twins for Smart Cities”, Elsevier Network Security, [Online], https://doi.org/10.12968/S1353-4858(22)70046-3

- R. Singh, H. Al-Khateeb, G. Ahmadi-Assalemi, and G. Epiphaniou, "Towards an IoT Community-Cluster Model for Burglar Intrusion Detection and Real-Time Reporting in Smart Homes", Challenges in the IoT and Smart Environments: A Practitioners' Guide to Security, Ethics and Criminal Threats, pp. 53-73, Cham: Springer International Publishing, 2021, https://doi.org/10.1007/978-3-030-87166-6_3

- G. Ahmadi-Assalemi, H. M. Al-Khateeb, C. Maple, G. Epiphaniou, Z. A. Alhaboby, S. Alkaabi, and D. Alhaboby, “Digital Twins for Precision Healthcare”, Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity, pp. 133,[Online], https://doi.org/10.1007/978-3-030-35746-7_8

- S. Alkaabi, S. Yussof, H. M. Al-Khateeb, G. Ahmadi-Assalemi, and G. Epiphaniou, “Deep Convolutional Neural Networks for Forensic Age Estimation: A Review”, Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity, pp. 375,[Online], https://doi.org/10.1007/978-3-030-35746-7_17

- G. Ahmadi-Assalemi, H. M. Al-Khateeb, G. Epiphaniou, J. Cosson, H. Jahankhani, and P. Pillai, "Federated Blockchain-Based Tracking and Liability Attribution Framework for Employees and Cyber-Physical Objects in a Smart

Workplace", in 2019 IEEE 12th ICGS3. London, UK, pp. 1-9, 16-18 Jan. 2019, https://doi.org/10.1109/ICGS3.2019.8688297

- T. Makonese, G. Ahmadi-Assalemi, H. M. Al-Khateeb, S Khan, M Patwary, "Advanced feature-driven anomalous behaviour classification from sensor networks for improved cyber defence in Industrial Control Systems", In preparation

The following publications contribute to my PhD research and are based on my MSc thesis:

- G. Ahmadi-Assalemi, H. M. al-Khateeb, C. Maple, G. Epiphaniou, M. Hammoudeh, H. Jahankhani, and P. Pillai, "Optimising driver profiling through behaviour modelling of in-car sensor and global positioning system data", Computers & Electrical Engineering, vol. 91, pp. 107047, 2021,[Online], https://doi.org/10.1016/j.compeleceng.2021.107047

- Project dataset online July 2018. DOI: https://doi.org/10.13140/RG.2.2.14505.49765

1.7 Scope, Limitations and Assumptions

1.7.1 Scope

The scope of this research covers ML model-driven security frameworks, systems, and data produced from CPS sensing technologies to address cyber resilience in smart cities, with a specific interest in ICS. Data generated from network traffic and data-driven models which require additional hardware to meet increased demand as a result of computational complexity are beyond the scope of the thesis.

Publicly available datasets generated from ICS testbeds will be used to evaluate the efficacy of the framework and associated ML models. In addition, a range of technologies and tooling will be utilised for developing and evaluating the models including Anaconda, R-Studio, Waikato Environment for Knowledge Analysis (WEKA), Massive Online Analysis (MOA), Python and Jupyter Notebook.

1.7.2 Limitations and Assumptions

The results produced throughout this thesis are based on laboratory studies generated from testbeds. While the dataset selection criteria reflect an environment of real-world characteristics, actual environment longitudinal field studies may challenge these results. Due to the time and cost constraints of the programme combined with access restrictions to datasets from ICS production environments, which often constitute critical infrastructure, it is not viable to carry out a longitudinal field-based case study. Hence, the applicability of the framework and models to the researched environment is evaluated in laboratory conditions using data from testbeds.

This thesis focuses on ML techniques in the generic IoT architecture [57] Sensing Layer, as shown in Figure 1, mapping to the Physical Control Layer in ICS as illustrated in Figure 2. This thesis does not focus on ML techniques applied at higher layers of the architecture. While attack vectors at the other layers are acknowledged, they are not addressed in detail by this thesis. In addition, the thesis develops mechanisms that are intended as improvements to existing models or presents new model concepts and therefore require further enhancements to be considered for production use.

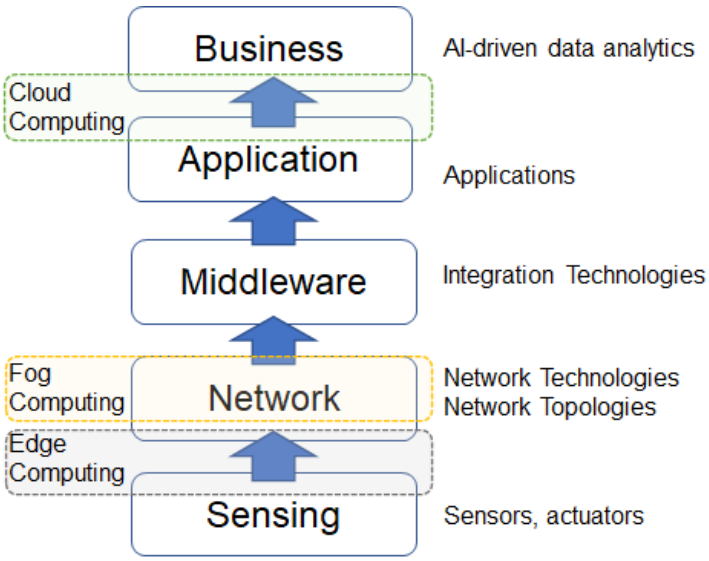


Figure 1 Smart City IoT Architecture Layers

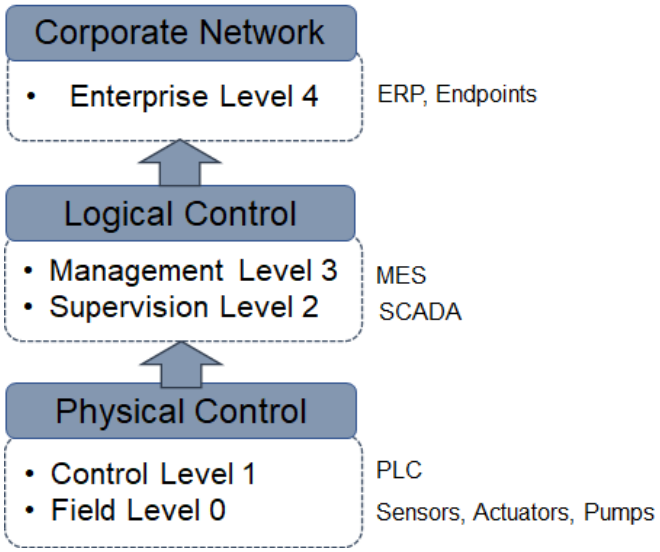


Figure 2 ICS Architecture Layers

1.8 Research Methodology

To achieve the aims and objectives of this research study, the following techniques are used systematically and iteratively:

- Systematic Literature Review (SLR)

An SLR approach is adopted that utilises the review protocol based on Kitchenham and Charters' guidelines for the computer engineering discipline coupled with the

PICOC (Population, Intervention, Comparison, Outcomes and Context) criteria [58]. The guidelines are based on 3 key phases, namely planning, conduction and reporting involving several discreet activities.

- Continuous research reviews and literature investigations

Besides SLR, continuous exhaustive literature investigation is employed to make use of pertinent research in the field and to factor in evolving trends. Existing research could help understand a specific challenge from an alternative perspective, and aid the discussion of the strengths and weaknesses of the researched approach based on the available evidence.

- Piloting

Piloting helps test and evaluate concepts, the appropriate instruments, ML techniques and develop the procedures.

- Case studies

Experimentation employed through the use of case studies helps evaluate the ML approaches and subsequently analyse their applicability given different circumstances. The case studies utilise publicly accessible datasets produced from testbeds. Key characteristics of a testbed environment include reproducibility, consistency and predictability which are critical to the evaluation of ML techniques.

1.9 Summary of Research Contributions

The research contributions to the field of cybersecurity, particularly the cyber defence of ICS in smart cities of the future include:

i. An SLR of primary studies that investigate empirical evidence reported for existing frameworks and systems that address the cyber resilience of CPS in smart cities (Chapter 2).

ii. An investigation of how current ML approaches applied to CPS in smart cities address modern DFIR (Chapter 2).

iii. An investigation of current cross-sector proposals or applications that leverage interactions in CPS in smart cities to improve DF (Chapter 2).

iv. Development of a novel super learner ensemble anomaly detection and cyber risk quantification (SPEAR) framework. The SPEAR framework provides ML-driven models for resilient profiling of anomalous behaviour in ICS from sensors-generated data and a cyber risk quantification model to produce a cyber risk value in the prevalence of anomalous behaviour (Chapter 4, Chapter 5).

v. A novel adaptive learning anomaly detection and classification (A-ADC) model for ICS physical plant sensor data streams (Chapter 6).

vi. Investigation of applicability of the SPEAR framework and the A-ADC model to support DFIR in the context of DF readiness to improve defence-in-depth in ICS (Chapter3, Chapter 4).

vii. Case studies demonstrating the applicability of the framework and the related models using ICS testbeds (Chapter 4, Chapter 6).

viii. A performance benchmark criteria framework is proposed to quantify the performance of classifiers across different levels of cyber-physical experimental environments (Chapter 6).

ix. Related published and cited research is outlined in detail in section 1.6.

1.10 Thesis Overview

The thesis is arranged into seven chapters. Beyond this chapter, the thesis is organised into six further chapters. Chapter 2 addresses Objective 1 and Objective 2. The chapter addresses Objective 1 by systematically identifying peer-reviewed

empirical primary studies providing evidence-based summary of key themes and possible future research directions and Objective 2 by investigating anomalies and ML approaches. Chapter 3 contains the overarching goal of the thesis covering cyber resilience in ICS and related threat modelling. Following that, Chapter 4 holds part of Objective 3 and presents the SPEAR novel security process-driven super learner ensemble for anomaly detection framework and related ML models. Chapter 5 is dedicated to the cyber risk value quantification model fulfilling Objective 4. Chapter 6 presents the A-ADC adaptive learning anomaly detection and classification model, the contribution required to achieve the second part of Objective 3. Objective 5 is embedded throughout and addressed in Chapters 2, 5 and 6. This research is concluded in Chapter 7 with an evaluation of the contributions of this research and finally addresses future works.

2. Chapter: Systematic Literature Review

The world is experiencing a rapid growth of smart cities accelerated by Industry 4.0, including the IoT, and enhanced by the application of emerging innovative technologies which in turn create highly fragile and complex CPN ecosystems. This chapter systematically identifies peer-reviewed literature. Then, it explicitly investigates empirical primary studies that address cyber resilience and DFIR aspects of CPS in smart cities. The findings show that CPS addressing cyber resilience and support for modern DFIR are a recent paradigm. Most of the primary studies are focused on a subset of the Incident Response (IR) process, the “detection and analysis” phase while attempts to address other parts of the DFIR process remain limited. Further analysis shows that research focused on smart healthcare and smart citizen were addressed only by a small number of primary studies. Additionally, the findings identify a lack of available real CPS-generated datasets limiting the experiments to mostly testbed-type environments or in some cases authors relied on simulation software. Therefore, contributing to this SLR, a search protocol is used to provide an evidence-based summary of the key themes and main focus domains investigating cyber resilience and DFIR addressed by CPS frameworks and systems. This SLR also provides scientific evidence of the gaps in the literature for possible future directions for research within the CPS’ cybersecurity realm. In total, 600 papers were surveyed from which 52 primary studies were included and analysed.

2.1 Introduction

Industry 4.0, synonymously known as Cyber-Physical-Production-Systems (CPPSs), is a concept formed in 2011 at the Hannover Fair to describe how CPS can be applied within production and manufacturing industries with enabled automation [1-4]. From the inception of the visionary notion specifically for factories and large-scale enterprises, CPS' reach has extended beyond production enterprises linking the Industry 4.0 concept with aspects of smart city initiatives [3, 4]. A key component of smart cities, CPS can be described as smart, embedded and networked systems within production systems [59]. CPS consist of a tangible element that is not completely controlled by an automated system and a cyber element that focuses on digital information. Together these elements form CPS entities capable of autonomous interaction regardless of human supervision [60]. Furthermore, these complex and growing networks of connected objects incorporate human users and form complex CPN ecosystems interrelating systems, software, people and services. As such, a problem within this complex cyberspace, including cybersecurity challenges, can have a cascading effect on the entirety of the ecosystem [6, 23].

The motivation and tactics of the cyber threats landscape shifted from individuals' hobby hacking to gaining kudos amongst their peers toward well-organised cybercrime [28-30]. The motivations are often intensified by the possibility to gain sensitive information, which can be used in subsequent attacks including cyber-attacks against ICS or Critical National Infrastructure (CNI). Verizon reported in their 2016 Data Breach Investigation Report the outcome of the investigation of 500 cybersecurity incidents in over 40 countries. In 89% of the cases, the key motives

reported were described as “financial” and “espionage” fixated on targets including manufacturing, healthcare, utilities and public services by organised crime and state-affiliated groups. Many of these attacks had a secondary motive to aid an intrusion of another target [61, 62]. This class of attacks known as APT characterise a well-resourced group of attackers that carry out multi-stage and often multi-year persistent targeted campaigns. Traditional IR methods fail in mitigating APT because they assume successful intrusion before IR takes place. A kill chain model enables one to map such campaigns, identify patterns linking individual intrusions and through iterative intelligence gathering enables the development of a resilient intelligence-driven mitigation approach [63]. In 2018, although the key motives remained largely unchanged, the most noteworthy attack vectors reported by the European Union Agency for Network and Information Security (ENISA) included malicious attachments, URLs in emails targeting the human element, web browser-based malicious scripts, malvertising, exploit kits and password reuse or weak service credentials in Internet exposed assets [28]. In 2019, law enforcement agencies responded to more attacks on CNI than ever before; this trend was highlighted as a key emerging threat by Europol [40]. CNI such as smart energy, water or transport are complex, large-scale interconnected CPS converging physical and cyber domains. They utilise geographically dispersed ubiquitous distribution networks, which extend beyond the boundary of a smart city, often across national borders and legal jurisdictions.

The rise of cybercrime has been greatly facilitated by the proliferation of modern advanced electronic communication technologies and the integration of IoT with physical systems [30, 35, 36]. High-profile cyber-attacks on ICS have been well

reported for some years, such as the Stuxnet malware targeting the Iranian nuclear plant [64], the attack on the Ukrainian power grid [65] or Norsk Hydro, a renewable energy supplier and the world's largest aluminium producer, which was compromised by the LockerGoga ransomware [40]. In case of a successful cyberattack, the disruption of power, water or fuel supplies to these facilities could have a potentially serious socio-economic impact including civilian unrest; however, consequences could be more profound. For example, in the widely reported Kemuri Water Company attack, the mixture of chemicals used to treat a water plant was altered. In this attack, the sensors responsible for monitoring the water treatment plants were compromised [61]. Due to the distributed nature and heterogeneity of CPS, human interactions and the omnipresence of the underpinning technologies create hugely diverse attack vectors which increase the threat of cyber-attacks on critical systems.

Due to the attacks becoming more sophisticated and targeted, the countermeasures also need consistency and coordination [28, 30, 40, 66]. Therefore, a new paradigm must address cyber threats and cybercrime. Formulating cyber resilience to counter cybersecurity threats is required to resist cyber-attacks and continue to function effectively under adverse conditions [67]. Accepting that not all cyber-attacks are avoidable and computer-related crime is on the increase, the IR becomes an important component of CPS' security management [40] including the need for Digital Evidence (DE). Forensic DE gathering must be carried out without compromising the integrity and authenticity of the DE to ensure admissibility in a court of law [56]. Therefore, the cybersecurity paradigm needs to shift to withstand cyber-attacks, function effectively under adverse conditions and support DF investigations by producing DE that is admissible in a court of law. Collaborative

practice and interdisciplinary approaches across smart sectors based on threat information sharing could increase situational awareness and help deal with potential threats or incidents more effectively.

Although CPS-related research is an active area, there seems to be substantially less empirical research available on frameworks and systems that address CPS in smart cities. Therefore, to make a meaningful contribution, this study uses a broad definition of frameworks as a common carefully designed organising structure of multiple approaches [68-70]. To help discover contributions in the literature of the specific research area this study includes systems to gain a deeper understanding of addressing support for cyber resilience across the physical, cyber and people dimensions in cross-sector applications within smart cities [71, 72]. Specifically, concerning frameworks and systems that address cyber resilience and modern DFIR, there appears to be a lack of available SLR based on recognised methodology, comprehensive protocols and quality assessment. For instance, to identify how CPS-related frameworks and systems support cyber resilience and to determine the support for modern DFIR in smart cities it is important to conclude what research has been published and systematically review relevant and available studies. Therefore, one of the key objectives of this chapter is to identify the current gaps in this research area. Overarchingly, the focus of this chapter is on reported empirical evidence in existing literature concerning cyber resilience and DFIR support in CPS across smart city sectors. Traditionally, “resilience” in a mechanical context was the materials’ resistance to shock, in the conventional networking context resilience focused on fault tolerance; however, the scope of this term extends to the cybersecurity discipline. This SLR considers cyber resilience as the ability of the frameworks

addressing smart cities to resist cyber-attacks across the physical and digital domains regardless of an external or insider attack [67, 73-75].

A small number of SLR studies in the realm of CPS have been published. These are outlined to examine the difference between the authors' focus on topics and this research. The author of [76] performed an SLR focusing on a smart grid and related cybersecurity. In this chapter, the presented results are aimed at addressing cybersecurity by identifying all standards which define cybersecurity requirements for smart grids and reviewing applicable standards and guidelines. In reference [77], the authors provide analysis to address cybersecurity issues in an Industry 4.0 context and focus on the physical Internet-connected systems. The authors concentrated on four areas, the definition of concepts relevant to Industry 4.0 and cybersecurity, the industrial focus, the characterization of cybersecurity and the management of cybersecurity issues. Authors in reference [78] presented their SLR findings concerning smart cities focusing on instrumented, interconnected and intelligent systems investigating four areas including security. One of the authors' conclusions was that little was mentioned in the newly emerging security and privacy challenges. Although the studies into this growing area of research provide valuable knowledge consolidation, they answer questions about the wider use of CPS and related cybersecurity. No other SLR on this research topic was found during the preparation of the thesis. The focus of this SLR remains specifically on CPS-related cyber resilience and modern DFIR informed by Cyber Threat Intelligence (CTI) to strengthen and accelerate cyber defence in smart cities.

Narrative reviews, such as [79-82], were found to focus on various IoT aspects and applications addressing challenges, threats and solutions. However, these studies address broader aspects related to the IoT but do not specifically investigate CPS with a focus on improving cyber resilience, or the value of CTI- or CPS-specific DFIR support in smart cities. The field of research related to CPS is still emerging, but the advancement is accelerating. Therefore, a comprehensive SLR is required focusing on ways that current CPS deal with cyber resilience and DFIR to guide future research.

This chapter critically examines existing research and uses the insights to conclude with suggestions for future research. The remainder of this chapter covers the Related Works in Section 2.2 follow by Methodology in Section 2.3. Section 2.4 contains the results, analysis and key findings from the included primary studies followed by a discussion in Section 2.5.

2.2 Related Work

2.2.1 Anomalous Behaviour Detection in ICS

Anomalous behaviour detection can be articulated as solving a problem of a group of data points that do not conform to the expected norm [48]. Several challenges impact anomaly detection including consideration of what is regarded as normal in a given domain, the evolving nature of normal behaviours and the prevalence of noise in datasets. This complexity is exasperated by threat actors who attempt to make anomalies appear as normal observations. Hence, anomaly detection is a complex problem and factoring in the different types of anomalies is important when applying

anomaly detection techniques. The following paragraphs identify and describe the types of anomalies in further detail.

- **Point-based anomalies**, as illustrated in Figure 3, contain anomalous instances that are few and different from the rest of the dataset [48]. The way ML techniques address point-based anomalies varies. Some approaches rely on defining the normal behaviour profile in the first instance by utilising clustering, statistical or classification techniques while other approaches use isolation methods. For example, as illustrated in Figure 3, a point-based anomaly is generated utilising Random Forest (RF) supervised learning ensemble in driver profiling where the driver performs an emergency brake as shown in Figure 3 a) [83] and a blocked sensor in ICS liquid distribution testbed as shown in Figure 3 b) [84].
- **Collective anomalies** are illustrated in Figure 4. The data point on its own is not considered anomalous, it is the collection of the data points together concerning the entire dataset that amounts to anomalous behaviour [48]. For example, Figure 4, illustrates a DoS attack in ICS liquid distribution testbed [84]. The highlighted area denotes the anomaly because the change in the data points values follows an abnormal pattern. However, the data point in itself is not an anomaly.
- **Conditional anomalies** also referred to as contextual anomalies, consider the data point within the context. The data points are anomalous within a given context, however, the data point itself is not anomalous, see highlighted area in Figure 5, [48]. Context requires contextual and behavioural attributes. For example, in Figure 5 the contextual attributes would be the longitude, latitude and speed limit in the given route section whereas the behavioural attribute would be the recorded speed. An identical data point in a different context may not be considered an anomaly.

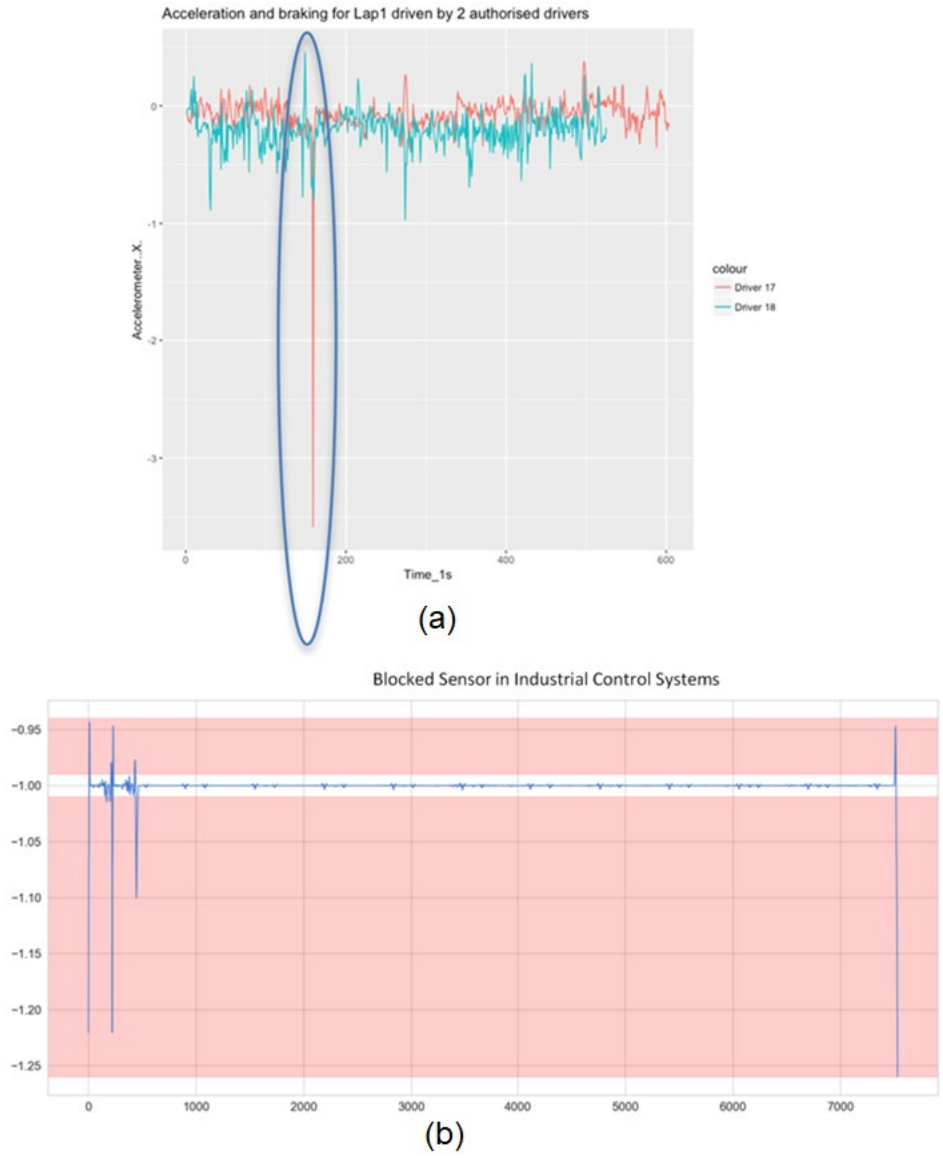


Figure 3 Point-based anomaly. Figure (a) shows an emergency break from a driver classification dataset [83]. Figure (b) shows an industrial control system blocked sensor from an ICS anomalous behaviour classification dataset [84].

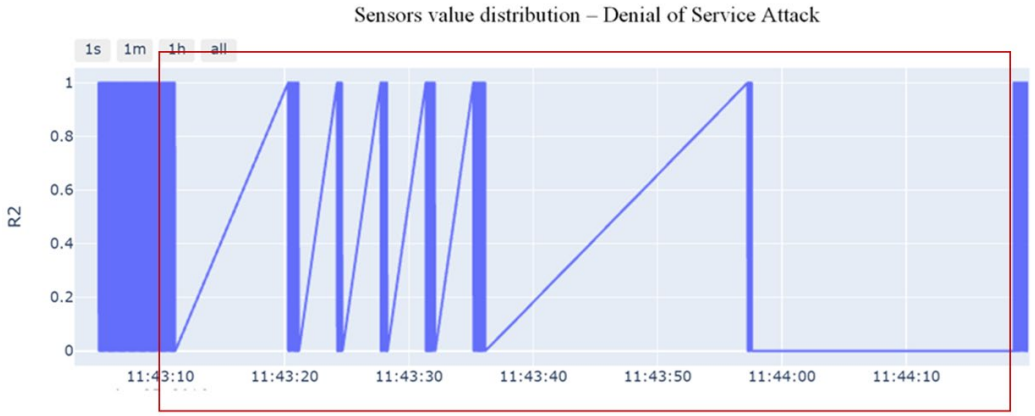


Figure 4 An example of a collective anomaly, a denial-of-service (DoS) attack from an ICS anomalous behaviour classification dataset [225].

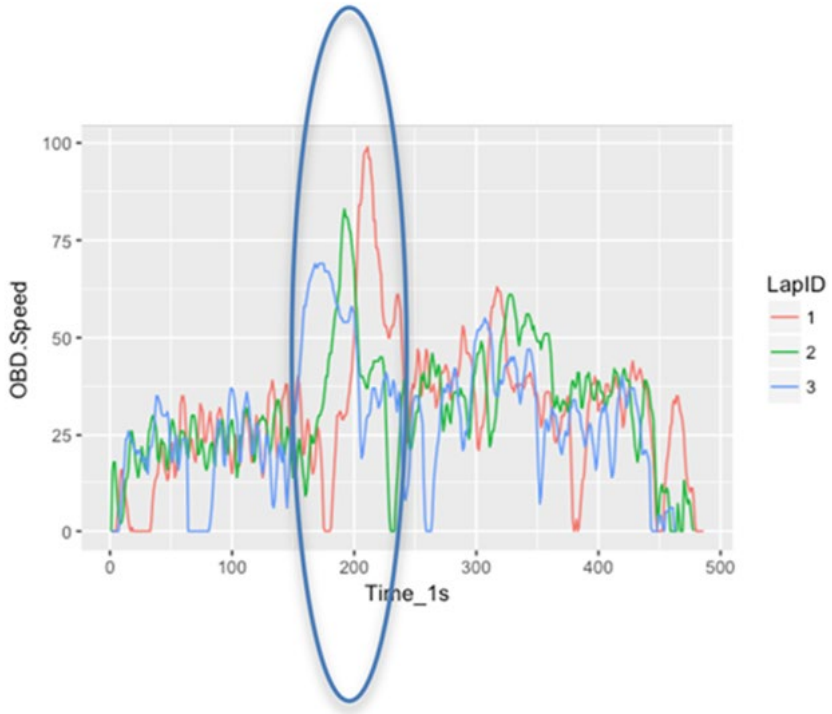


Figure 5 An example of a contextual anomaly, a speeding driver from a driver classification dataset [83].

2.2.2 *Application of Learning Techniques*

ML techniques utilised by domains including social media, medical analysis, computer vision and gaming are applied in cyber defence measures in smart city sectors such as transportation, healthcare, buildings and ICS [6, 85-89]. For example, ML techniques are utilised as cyber defence measures for anomalous behaviour detection. Figure 6 illustrates the data collection process from the recent research on the innovative use of in-car sensors data that aimed to profile drivers based on their behaviour using [83]. The gained insights can contribute to a range of applications such as authentication, authorisation and accounting models including the emergence of transferable solutions across smart environments [6, 90].

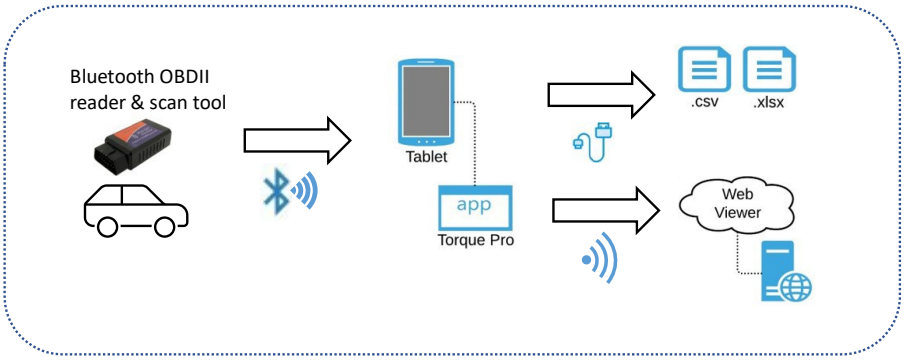


Figure 6 The driver profiling data collection process

Such an example from another recent research is illustrated in Figure 7. The figure shows the data collection process that innovatively leveraged sensors-based data in a smart workplace. The study proposed a tracking and liability attribution framework as part of a cyber defence mechanism [6]. The gained insights could contribute to the authorisation and detection of anomalies in near-real-time of employees and cyber-physical objects in a smart workplace.

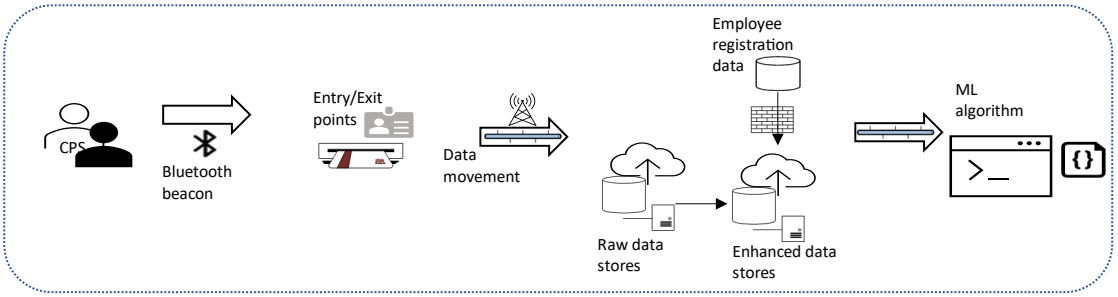


Figure 7 Tracking and liability attribution in a smart workplace.

Therefore, besides data produced at the network layer, from device and applications logs, ICS physical plant sensor data can be used to detect anomalous behaviour to increase resilience to cyber-attacks. One of the advantages of ML techniques over the signature, statistical or rule-based approaches is the detection of previously unseen attacks. According to [88], ML is frequently applied in intrusion detection, malware analysis, phishing and spam detection. The utilised ML approaches are

categorised into two main domains, shallow and deep learning. They both include supervised, unsupervised, and semi-supervised learning models, see Figure 8.

- Supervised Learning

In supervised learning, each instance has a pre-assigned class. The classifier is trained to apply the labelling of the target feature to new unseen data [45, 46].

- Semi-Supervised Learning

Semi-supervised learning algorithms utilise partially labelled data, where labels typically exist for data instances representing normal behaviour [45, 46, 48].

- Unsupervised Learning

In unsupervised learning, the classifier is looking for the presence of patterns [45, 46]. The algorithms leverage unlabelled data to train the model by understanding the underlying dataset's relationships and structures.

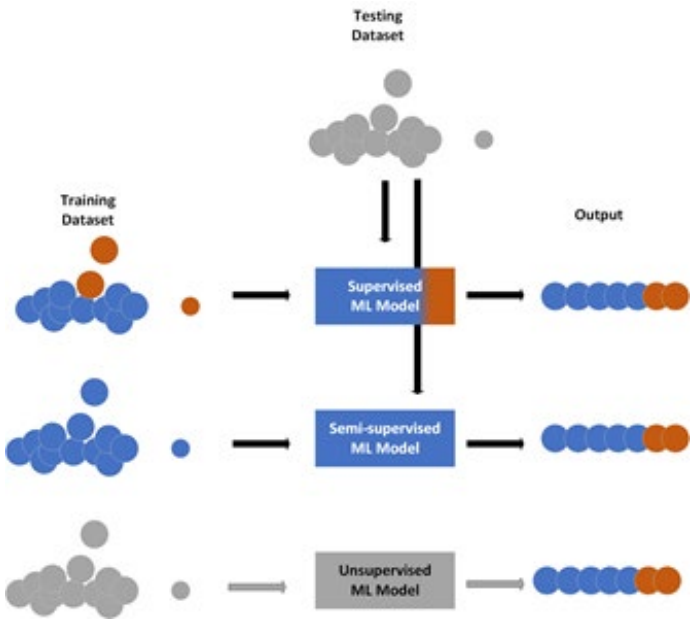


Figure 8 Machine Learning approaches based on dataset labels

This poses an important question; which one is the most suitable learning method? There is no ultimate de facto classifier, the choice depends on several factors not least the problem being solved. Other factors include distinct types of

classifiers that perform differently [45, 46], the types of datasets available, and organisational business and risk models. The following study [91] proposes a statistical testing procedure for algorithm comparison. Repeated training and testing are asserted in other scientific literature [46]. Another study [92] proposed an ensemble anomaly detection generic framework using Rolling Windows (RW) for energy consumption in buildings. To protect IoT network traffic, [93] utilised an ensemble learning method. The proposed method consisted of three ML techniques; Naïve Bayes (NB), Decision Tree (DT) and Artificial Neural Network (ANN) NB based on the AdaBoost classifier with majority voting. Furthermore, consideration should be given to the type of classifier for the scale and range of the investigated cyber-attacks, the classifier's performance in detecting the anomaly [88], and the algorithm's generalisation ability [45]. Different approaches were proposed for anomalous behaviour detection. Algorithms were utilised individually or as part of an ensemble such as Support Vector Machine (SVM) [92, 94-97], Principal Component Analysis (PCA) [92], Random Forest (RF) [92, 98-102], Autoencoder (AC) [92], ANN [93], DT [93, 103-105], NB [93], Isolation Forest (IF) [53, 106]. It is not the aim of this thesis to solve the classifier problem but to apply a robust model to the research problem outlined in this thesis and present a direction for future research.

Consideration is given to ML approaches leveraged for classification in SCADA systems including distance, density, isolation and ensemble approaches. These approaches are outlined in further detail in the following paragraphs.

- Distance-based algorithm: SVM

As proposed by Vapnik [107], the SVM algorithm is extensively applied in classification and regression problems [108]. Most SCADA system operations can be reduced and represented as a binary classification problem thus making the use of the SVM algorithm suitable for these tasks. Existing literature attests to the effective use of SVM for such problems [95-97].

While several hyperplanes can be used to separate the data points (vectors), the SVM algorithm operates by choosing the optimal hyperplane in an N-dimensional space with distinct data points, see Figure 9. The hyperplane is a decision boundary that is placed at the centre of the support vectors where the distance is the maximum from the closest points of either class giving the least test errors. Support vectors are the subset of data points closest to the hyperplane that impact its position and orientation. This shortest distance between the hyperplane and the support vector with weight *w* and bias *b* at either side is referred to as a margin. Hard margins see Figure 10 (a), typically leveraged by linearly separable data points can result in overfitting. Whereas soft margins as shown in Figure 10 (b), typical in non-linearly separable data points are robust to outliers and overfitting, allowing for misclassification, however, could result in underfitting.

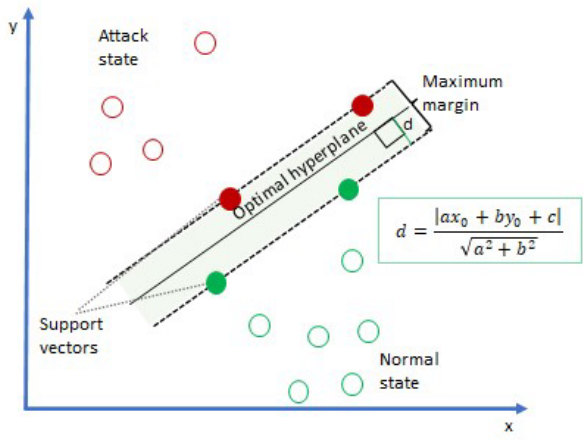


Figure 9 Simplified SVM structure

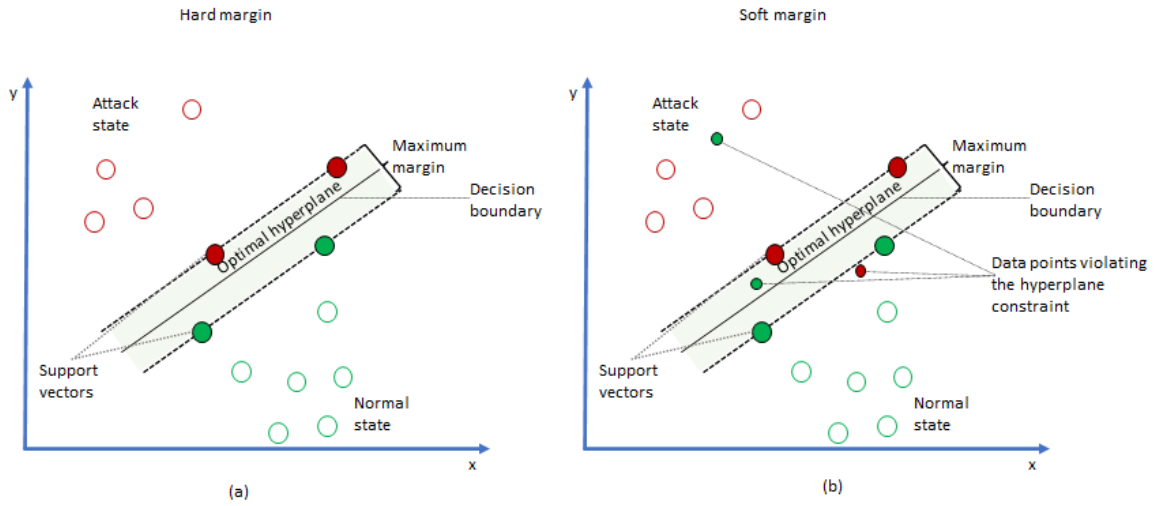


Figure 10 SVM algorithm Hyperplane hard-margin (a) and soft margin (b)

The equation of the optimal hyperplane equation is given by:

$$\mathrm{H}: \mathrm{w}^{T}(\mathrm{x}) + \mathrm{b} = 0 \tag{3.1}$$

where b is the intercept and bias term of the hyperplane equation in N-dimensional space. SVM segregates the data points to minimise the misclassification errors by computing the distance between data points and the hyperplane. The distance of a hyperplane equation $w^{T}\Phi(x) + b = 0$ from a given point vector $\Phi(x_0)$ is given as:

$$\mathrm{d}_{H}\left(\Phi(\mathrm{x}_0)\right) = \frac{\left|\mathrm{w}^{T}(\Phi(\mathrm{x}_0))+\mathrm{b}\right|}{\|\mathrm{w}\|_2} \tag{3.2}$$

where $\|w\|_2$ is the Euclidean norm for the length of w given by:

$$\|\mathrm{w}\|_2 =: \sqrt{w_1^2 + w_2^2 + \cdots . + w_n^2} \tag{3.3}$$

In linearly separable data points, the hyperplane is distinct, separating the classes. Solving this type of problem requires a linear ML classifier such as Logistic Regression (LR). Non-linear data points use Kernel functions such as Polynomial and Gaussian Radial Basis Function (RBF) functions in a higher-dimensional space to identify the optimal hyperplane. The polynomial function is given by

$$K\left(x_i, x_j\right) = \langle x_i, x_j \rangle^{p} \tag{3.4}$$

where x_i, x_j represent the classes of observations, p is the polynomial degree and K is the polynomial coefficient. The RBF function is given by

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \sigma > 0 \quad (3.5)$$

where σ is the Gaussian width parameter [45].

- Distance-based algorithm: k- Nearest Neighbour (kNN)

As illustrated in Figure 11, k-NN is a distance-based supervised ML algorithm, which is simple to design, easy to interpret and incurs low computation time [47]. K-NN classifies unknown data points based on their distance from the neighbouring, known data points [109]. Known as a lazy learning algorithm, it classifies data points and computes the class label based on the k nearest points. Instead of approximating the target function *f(x) = y* globally, during each prediction, the k-NN algorithm approximates the target function locally using the datasets closest in proximity as it is easier to learn to approximate a function locally than globally. To classify the data into either normal or attacked mode, the Euclidean Norm is used to calculate the distances between the datasets and the neighbours. The Euclidean Distance *d* is given by

$$\mathrm{d}(p, q) = \sqrt{\sum_{\mathrm{i}=1}^{\mathrm{n}} (q_i - p_i)^2} \quad (3.6)$$

where *p,q* are two points in Euclidean n-space *n* and q_i, p_i are the Euclidean vectors. This technique has been used for modelling intrusion detection on the SCADA dataset in **[**47, 110**]**.

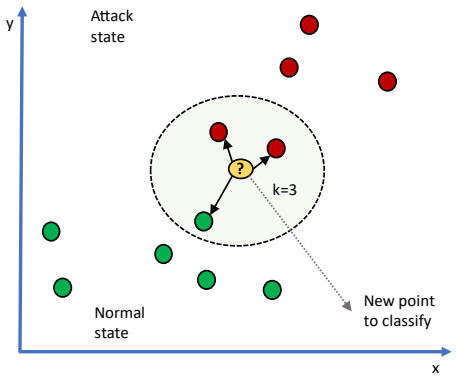


Figure 11 Simplified k-NN algorithm

- Decision Trees

DT, as illustrated in Figure 12, is a popular classification method in supervised learning [47], including SCADA studies [103-105]. The tree-node indicates a feature and the tree-branches a test outcome. Each non-leaf node is linked with a split, also called a feature test. The tree-leaves, also called leaf nodes signify a class label. A set of decision rules inferred from the data features are used to predict the target variable starting at the root node with the results produced at the leaf node [45, 111]. DT is a recursive process, where the dataset is split at each level and divided into data subsets that form the next split's input dataset. The split selection varies across DT algorithms, the Classification and Regression Trees (CART) algorithm uses the Gini index to select the split maximising the Gini [45, 46] given by:

$$G_{gini}\,(D; D_1..D_n) = I(D) - \sum_{i-1}^{n} \frac{|D_n|}{|D|} I(D_n) \tag{3.7}$$

where

$$I(D) = 1 - \sum_{y \in Y} P(y|D)^2 \tag{3.8}$$

Where *I(D)* is the Gini impurity, *D* is the training dataset, $D_1 \ldots D_n$ the training data subsets is the data.

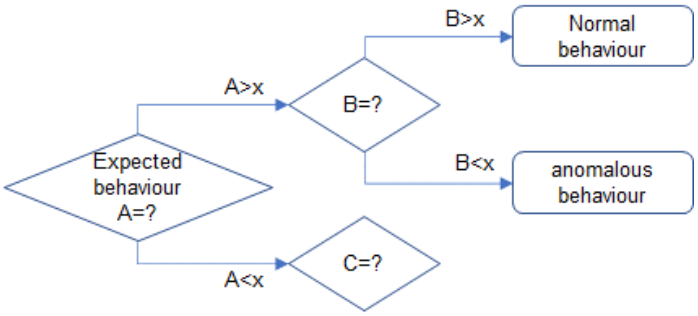


Figure 12 Simplified Decision Trees algorithm. A, B and C represent the features and 'x' is the threshold on which the tree splits into branches

DT is non-parametric, hence outliers do not impact the model and it is expected to perform well on linearly inseparable data [47, 112]. However, DT can struggle with high-dimensional data and can become computationally demanding while without adequate pruning it can overfit. Hence, ensembles of DT such as RF are employed [47].

- Ensembles

Ensembles are ML methods that consist of multiple learners trained to solve the same problem to obtain better predictive performance [32, 45, 46, 113]. An ensemble consists of several learners referred to as base learners with their individual decisions combined to classify new data points based on voting [45, 46, 113]. Ensembles have a comparable computational cost to constructing a single learner and often generalise better than an individual learner [45].

As illustrated in Figure 13, CART's extension, RF developed by Breiman [114] is a popular ensemble classifier including SCADA studies [98-102]. It is constructed from several DT base classifiers often linked with imbalanced learning, robust and scalable to large datasets [46, 47, 88, 89, 115].

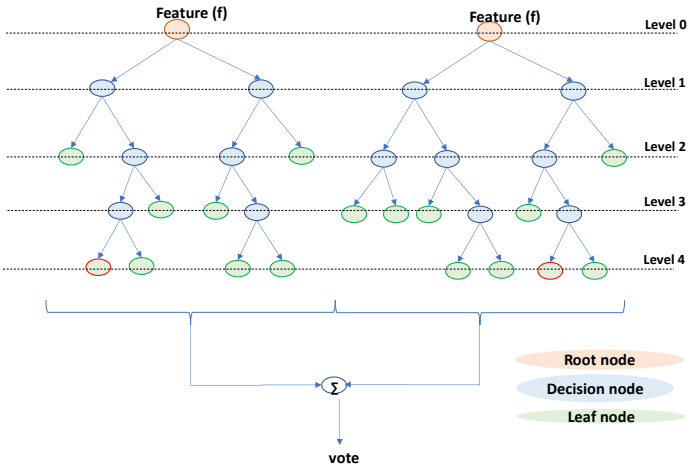


Figure 13 Simplified Random Forest Classification structure consisting of two trees.

Extra Tree Classifier (ETC), also called an extremely randomised tree classifier is similar to RF, however, in ETC the features and splits are selected randomly whereas RF uses a greedy algorithm [116].

AdaBoost is an iterative ensemble meta-algorithm combining weak learners [117]. It is considered adaptive as the succeeding learners focus on the previous learner's misclassified instances [46, 112]. While simple to implement and resistant to overfitting, AdaBoost is sensitive to noise.

Bagging is an ensemble meta-estimator fitting the base classifier on a random subset of the original dataset followed by aggregating the individual predictions to produce a final prediction [118]. While bagging can improve the model's accuracy without substantially compromising the variance, it requires more training data than some other techniques.

- Isolation-based classifier

Besides distance and density, isolation is another indicator of anomalies. The IF classifier detects anomalies by isolating instances [119]. Anomalies could have high density being clustered in smaller groups, as illustrated in Figure 14a, and low density at the border of normal instances, see Figure 14b.

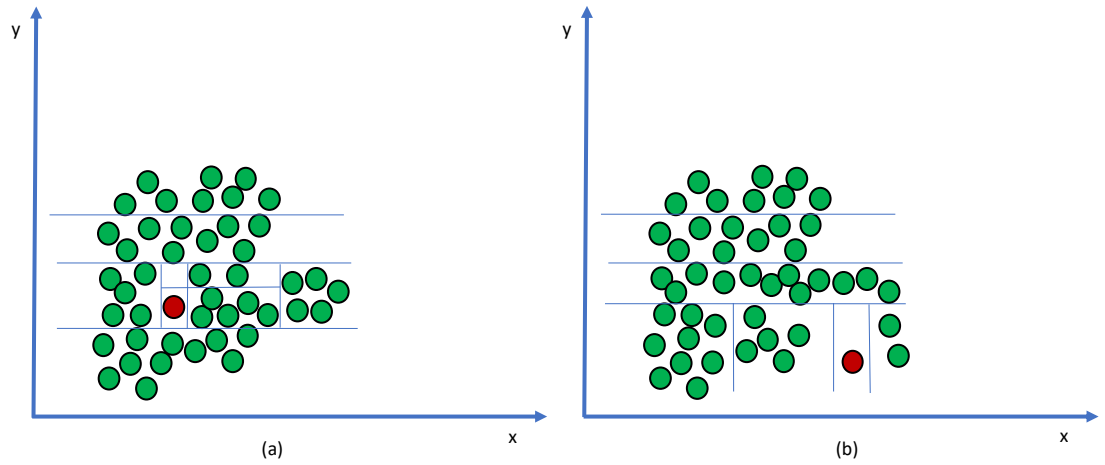


Figure 14 Simplified Isolation Forest algorithm

The IF algorithm is based on the characteristic that anomalies are few and different from normal observations within datasets, hence sensitive to isolation from normal instances [119]. According to [119], fewer partitions are required to isolate the anomalous point, hence a shorter path length from the root node to the leaf node, where the average path length of a tree with *n* nodes is given by

$$c(n) = 2H(n-1) - \left(2\,\frac{(n-1)}{n}\right) \tag{3.9}$$

Where *H(z)* is the harmonic number estimated as ln(z) + 0.5772156649 (Euler's constant), *c(n)* is the average of *h(x)* (the external node termination given *n*). Then, the anomaly score is given by

$$s(x,n) = 2 - \frac{E(h(x))}{c(n)} \tag{3.10}$$

Where *E(h(x))* is the average of *h(x)* from a set of iTrees. To scale, to large datasets, the random trees as created from smaller data subsets rather than the original dataset [119].

- Statistical Classifier

Logistic Regression (LR) is one of the simplest statistical ML algorithms for predicting binary classes, also known as Logit Regression and Maximum Entropy Classification. LR is suitable as a baseline method in binary classification problems, where the target variable is dichotomous. As illustrated in Figure 15, LR is a special case of linear regression using log-odds where *y* is the dependent variable and $X_1 \ldots X_n$ are the explanatory variables:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (3.11).$$

The probabilities of the outcome are modelled using a logistic function also called the sigmoid curve:

$$\text{Sigmoid function } f(x) = \frac{1}{1+e^{-x}} \quad (3.12)$$

where *e* is the base of natural logarithms. With the sigmoid equation applied to the linear regression is expressed as

$$\text{Sigmoid function } f(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}} \quad (3.13)$$

mapping to values between 0 and 1 with positive infinity being 1 and negative infinity becoming 0.

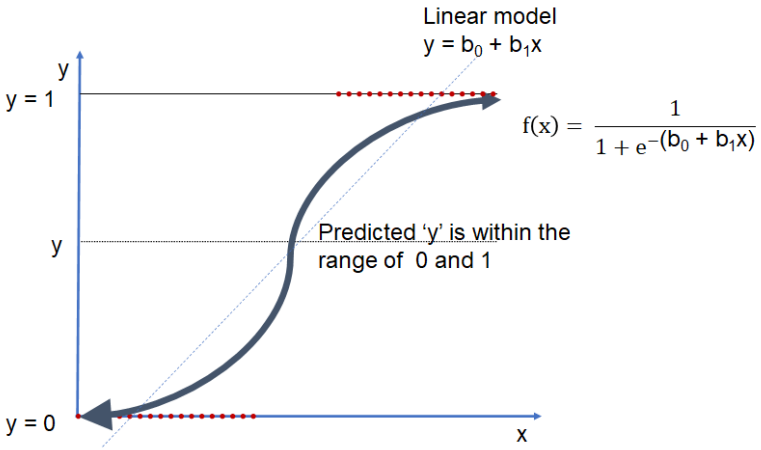


Figure 15 Simplified Logistic Regression algorithm

- Shallow Neural Network

The ANN algorithm imitates biological neurons by using interconnected basic processing units, also known as neurons, and nodes [45, 120, 121]. ANN is similar to the SVM in its ability to handle multi-dimensional data with sound generalisation. As illustrated in Figure 16, in its simplest form, ANN uses perceptrons for identifying linearly separable patterns [47]. The binary nature is the key feature of a perceptron. Classification models with proven effectiveness for SCADA systems have been developed and validated [122-124].

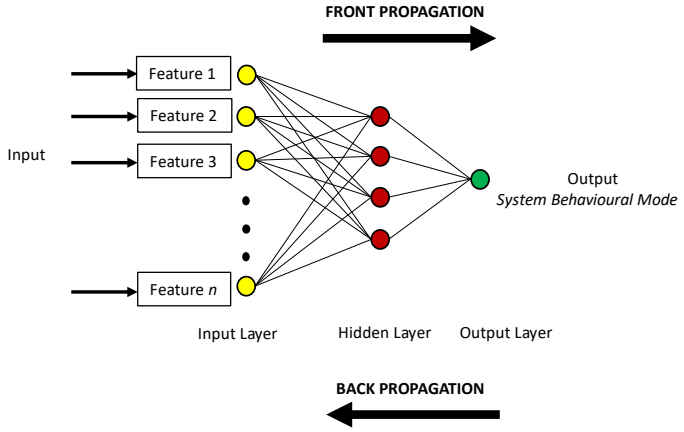


Figure 16 Simplified ANN algorithm

The ANN architecture consists of an input layer, a hidden layer and an output layer. ANN creates a fully connected network given that the neurons in each layer are interconnected with each other. The input layer typically corresponds to a feature, where the input activation function is expressed as

$$f(z) = z \tag{3.14}$$

In a linear operation, each neuron performs a weighted sum of the inputs adding a bias term. However, as real-life applications are non-linear, the weighted sum requires an activation function expressed as

$$z = f\left(\sum_{i=0}^{k} X_i W_i\right) \tag{3.15}$$

where *z*, is the output for a neuron, and *k* are the inputs. In the output layer, the neuron typically represents a label. The hidden and output layers are functional units that require an activation function such as

$$\text{Sigmoid function } f(z) = \frac{1}{1+e^{-z}} \tag{3.16}$$

$$\text{Tanh function f(z)} = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} \tag{3.17}$$

$$\text{Rectified Linear Unit (ReLu) function f}(z) = \max(0, y) \tag{3.18}$$

The choice of the activation function impacts the performance, convergence while training and computational cost of the ANN [45, 121]. The output of the sigmoid function, also called the threshold function, is 0 to 1 and is considered differentiable. Tanh is a hyperbolic tangent function with an output of -1 to 1. It centres the data with zero Means. The vanishing gradient is prevalent in both the sigmoid and tanh functions. The ReLu function solves the vanishing gradient problem present in the sigmoid and the tanh functions, the training is more efficient and simpler to compute.

- Density-based algorithms

An example of a density-based algorithm is the Local Outlier Factor (LOF), which is an unsupervised anomaly detection algorithm applied as an outlier or novelty detection [48, 119]. For outlier detection, LOF computes the local density deviation of a given data point from its neighbours where the outliers are considered to have a lower density than the neighbour [48, 125], see Figure 17.

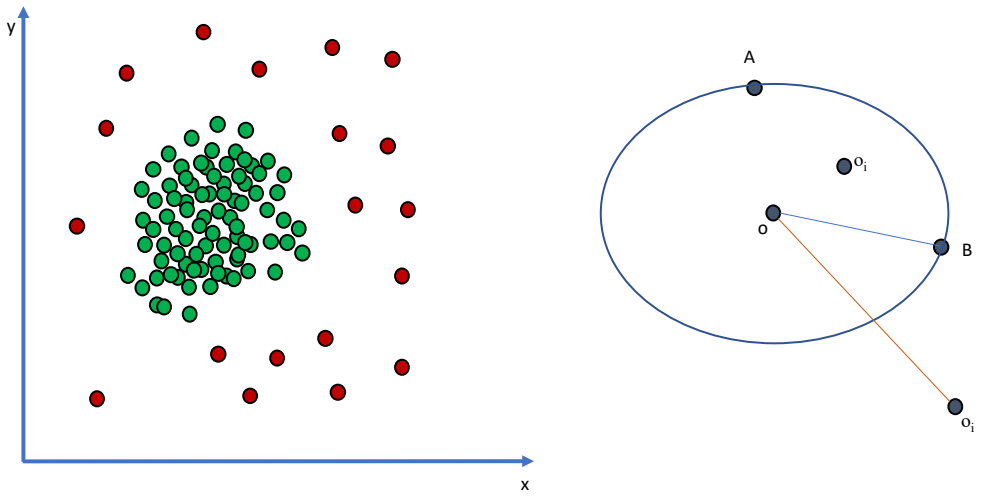


Figure 17 Simplified LOF algorithm and reachability distance where k=2

Firstly, the kNN is found for each data point, *x*, followed by the local density for a data point estimated by computing the Local Reachability Density (LRD) using the k-nearest neighbours, *N_k*, finally producing the LOF score by comparison of the LRD of the record and that of its k-neighbour [32, 125]. The LOF score is given by:

$$LOF\ (x) = \frac{\sum_{o \in N_k(x)} \frac{LRD_{k(o)}}{LRD_{k(x)}}}{|N_k(x)|} \tag{3.19}$$

2.2.3 *Data Stream Paradigm*

ML has evolved over the last two decades as a practical technology with commercial use. It is a relatively young scientific discipline that continues to expand being driven by new applications of ML techniques, data availability and decreasing computational cost [85]. As an active research area, ML methods outperform rule-based algorithms and human operators in several domains leading to the adoption of ML in fields including computer vision, transport, gaming, healthcare and cybersecurity [88].

Traditionally, ML algorithms subdivide large datasets into smaller subsets that are processed separately [115]. However, real-world applications such as ubiquitous sensor networks, traffic control and financial prediction do not deal with finite datasets and continuously generate large volumes of data sequentially. Continuously generated data has a substantial demand for storage and computational power for processing. Hence, it is not viable to store the data in the main computer memory all at once. Stored data may not reflect the current data distribution and can exceed the available computer memory [126, 127]. Furthermore, data streams can evolve and alter the normal distribution of attributes. Such changes can be sudden, referred to in the literature as concept shifts, while gradual or incremental changes are known as the concept of drift [128, 129]. Batch learning cannot learn from continuous data streams incrementally or deal with the concept of changes. The model's performance could be adversely impacted, and the model has to be updated, replaced or retrained. Therefore, dealing with ubiquitous data streams requires new or adapted ML methods [115, 129]. Adaptive learning is a technique consisting of algorithms that are capable of processing data streams, and require a limited amount of memory and time with an ability to adapt to changes in data distribution [130]. However, besides analysing big data for knowledge gain [128], according to Gartner 2021 technology trends, learning methods should include small and wide data to produce value and impact [131].

2.2.4 *Application of Online Learning Techniques*

Recent research has focused on the application of algorithms for streaming data including in ICS. Although the use of data streaming is an emerging area in the field of cybersecurity, several research studies apply online learning to data streams for

anomaly detection [37, 132-134] while other studies focused on ICS in particular [16, 19, 50, 52, 123, 126, 129, 135].

The problem of large volume and high-frequency data from heterogeneous sources streamed in real-time is investigated by [126] who acknowledges that not all observations can be loaded into active memory in real-time to process. The study standardises the kappa architecture and develops a multi-task learning model for real-time and large-scale data combined with tasks overlapping processing. They proposed an architecture that utilised k-NN classifier with Self Adjusting Memory (SAM) and sliding windows to produce data stream chunks to identify deviations in ICS. The choice of the k-NNSAM algorithm factored in the new data relevant to the current forecast and prior knowledge that is required for accurate classification. The concept of the approach is similar to the short-term-long-term memory model [126]. The aim of combining the two algorithms was to minimise the error and increase the classification precision. As illustrated in Figure 11, k-NN is a distance-based supervised ML algorithm that can be applied without parametrization to deal with heterogeneous concept drift of various types and rates [136].

While k-NN is conceptually simple, easy to interpret and incurs low computation time [47, 137, 138], scientific literature highlights this technique could suffer from increased CPU, time and memory consumption [130, 132]. K-NN computes an observation in the dataset by computing all other observations considering data with similar features to have the same classification [37]. K-NN has been extensively applied in batch learning to model the detection of anomalies for SCADA systems and APT detection [47, 112, 139]. K-NN is popular in data stream classification of

faults, detection in the underlying data distribution and intrusion detection [37, 130, 140].

Another study [129] combined the volume and velocity sensor data from power systems with the cyber segment of the power system for cyber-power events and intrusion detection. The study proposed an online learning method utilising HAT augmented with the Drift Detection Method (DDM) and Adaptive Windowing (ADWIN) mechanism. HAT was evaluated on synthetic datasets and extended to power systems [129, 141] to distinguish between faults with different characteristics, cyber-attacks and normal operations using data from heterogeneous sources. The HAT algorithm was proposed by [142], evolving from the Hoeffding Window Tree (HWT) to adaptively learn from data streams. Hoeffding Trees (HT) are incremental DT algorithms, see Figure 12. HAT is resilient to change with the ability to utilise change detectors including Linear Incremental Estimator (LIE), Exponential Weight Moving Average (EWMA), ADWIN and DDM [129, 141, 142].

Drift detection methods such as Cumulative Sum (CUSUM), Page-Hinkley Test (PHT), DDM [143], ADWIN [142], Early Drift Detection Method (EDDM) [144] and Exponentially weighted moving average Concept Drift Detector (ECDD) [140] are utilised to detect changes in data streams. The CUSUM and PHT are based on one-sided tests and only raise alarms when the mean increases. Whereas DDM relies on the data distribution being stationary where an increase in the prediction error is taken as evidence of change. ADWIN is a change detector and an estimation algorithm based on an exponential histogram. ADWIN manages the trade-off between the window length to produce robust results with low FP values and short

window length to detect the change efficiently by checking the change at many scales simultaneously. This approach makes it computationally costly compared to simpler methods such as EWMA or CUSUM. HDDM is an online method based on Hoeffding bounds coupled with two sided test that monitors error increments and decrements leveraging the average as the estimator [145]. HDDM receives data stream inputs monitoring its performance change estimating the stream value to be stable, warning or drift. The A-Test consists of the two-sample statistical test of moving averages for abrupt changes which monitors error increments and decrements.

The challenges of class imbalance are characteristic of real-world problems where the majority of the data belongs to the normal class and the minority class instances are rare [115]. This imbalance between classes can introduce a bias and skew the ML algorithms' performance. Sampling, ensembles and cost-sensitive methods address the problem of class imbalance [115, 146]. As illustrated in Figure 18, to balance datasets, sampling techniques are employed to change the balance of the data classes. To achieve sampling, the dataset needs to be cached in the memory, which does not satisfy the data stream paradigm [146] whereas cost-sensitive methods factor in the cost of misclassifying instances [45, 146, 147]. For example, [19] presents a method to solve the class-imbalance problem prevalent in ICS. The study focuses on an adaptive regularised cost-sensitive online learning method for multiclass classification. It asserts the technique demonstrating precise and efficient attack discrimination. Another study [115] introduces an approach to deal with imbalanced data streams utilising ARF with resampling to improve the performance of the minority class.

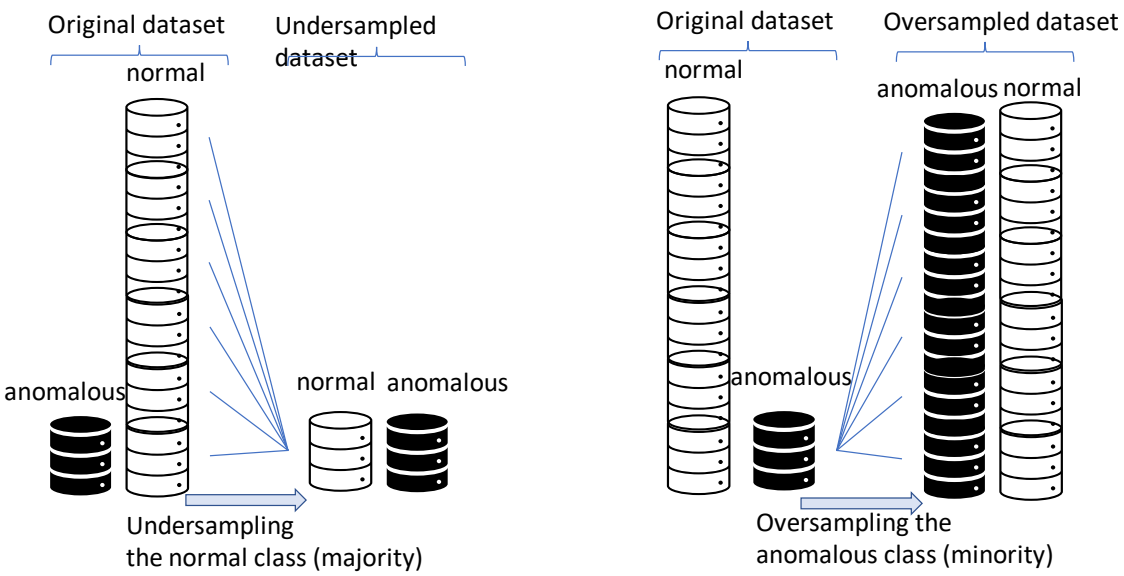


Figure 18 Simplified concept of dataset oversampling and undersampling

Recent research suggests that ensemble methods are effective in addressing the problem of concept drift. RF is a popular batch learning ensemble classifier constructed from several DT often linked with imbalanced learning [115], see Figure 13. RF has been adapted for stationary data streams [148] and dynamic streaming with entropy-based DDM [149]. Ensemble-based online learning is utilised in network intrusion detection by [37] who explored online homogeneous and heterogeneous ensembles. Their research concluded that an ensemble comprising an ARF of HT combined with HAT was best suited for the concept drift phenomena. The study achieved better run-time compared with other tested ensembles. The authors asserted their approach was more suited to the constraints of online training. ARF is not bound to a specific DDM, the following study [127] evaluated ARF with ADWIN and PHT while [115] presented ARF with resampling to factor in imbalanced class distribution.

Other studies include [135], which applied a combination of a k-NN and ARF combined with a Primal Estimated SubGradient Solver (PESGS) for SVM. Whereas

[52, 123] proposed a hybrid approach consisting of a spiking neural network and Restricted Boltzmann Machines (RBM). Another study proposed an attack detection framework utilising multimodal data and adaptive learning for critical water infrastructure [16]. The authors asserted that the model's performance was correlated with the data modalities combining red-green-blue fusion with thermal, Wi-Fi reflection data and ICS sensing information. The authors proposed a model that consisted of an adaptive Tapped Delay Line Convolutional Neural Network (TDL-CNN) combining deep learning with autoregressive and moving-average attributes. Whereas the following research study [50] focused on controlling network traffic utilising Online Sequential Extreme Learning Machines (OS-ELM) and RBM in a hybrid mode with a final aim of vulnerabilities detection for APT attacks.

2.3 Methodology

The aim of this chapter is achieved with an evidence-based SLR as the means to objectively address the RQs. The protocol is based on the SLR guidelines for the computer engineering discipline proposed by Kitchenham and Charters [58]. These guidelines, which aim to present a rigorous and credible methodology, are based on three key phases: planning, conducting and reporting, as demonstrated in Figure 19. The discreet activities in each phase are detailed in the Appendix in section 9.4 to allow replication of findings. Summarily, the core aspects of the systematic review protocol, the key contributions and the RQs are identified within the planning phase. The conducting phase consists of identifying the search strategy including the selection criteria for the primary studies, the selection procedure, the search strings and the quality assessment criteria. This phase involves the development of the data

extraction strategy, data synthesis and critical analysis. Finally, the information dissemination strategy is considered in the reporting phase. Each phase of the SLR is conducted iteratively to ensure a comprehensive evaluation. To maintain objectivity and mitigate bias, each phase was subject to a review and an approval process between the team before moving.

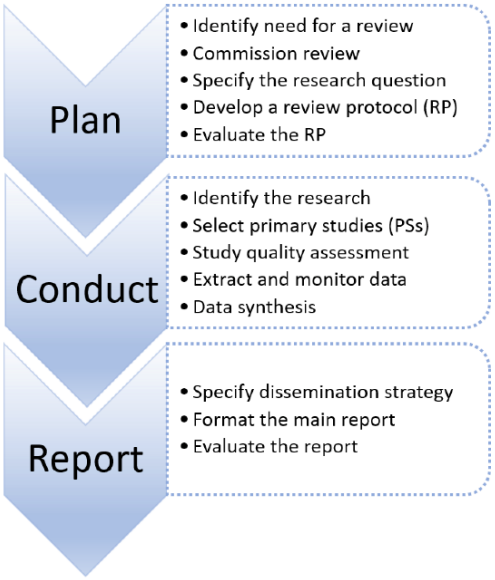


Figure 19 Phases conducted in this systematic literature review (SLR).

The main aim of the SLR in this chapter is to identify and present scientific evidence of gaps in current research and help inform the direction for further research. The aim can be achieved by answering the following three RQ:

RQ1: How do existing frameworks and systems that address CPS in smart cities support cyber resilience and what empirical evidence has been reported? Use cases and applications of CPS have diversified, and the complexities of these ecosystems have evolved. In addition to frameworks, the investigation focuses on how complex systems support cyber resilience by identifying commonalities. Within the many diverse definitions used in existing studies addressing smart cities [150-155] and the numerous terminologies used in literature to describe frameworks and systems [25,

72, 73, 156, 157], providing an answer to RQ1 helps us conclude a list of all existing and relevant frameworks and systems that address CPS in smart cities supporting cyber resilience as defined by the scope of this SLR.

RQ2: How do the identified frameworks and systems in smart cities address modern DFIR? The application of DFIR in the context of a smart city is a new field of study [41]. Whilst the research focuses on the applications of IoT-enabled CPS, smart cities are found to be vulnerable to cyber-attacks [158]. It is acknowledged that DFIR methodologies are lacking in smart city sectors [35, 42] and research suggests that DFIR faces more challenges in smart cities than other forms of digital breach investigations [159]. However, apart from the complexity of cyberspace, the IoT enabled CPS to create opportunities to facilitate modern DFIR [56]. RQ2 investigates how the components of the CPS frameworks help address modern DFIR.

RQ3: What are the current cross-sector proposals or applications in smart cities that attempt to utilise interactions in CPS for the purpose of improving DFIR? This RQ explores the transferable solutions and cross-sector interactions between smart buildings, smart homes, smart healthcare, smart energy and others as illustrated in Figure 20. Despite digitalisation in smart cities, information security strategies are limited to the sector boundary with little evidence of cross-sector information security practice sharing [30]. This study draws on the use of the term cross-sector partnerships in reference [160], as intensive and long-term interactions between organisations from at least two sectors such as business and healthcare. Throughout this chapter, cross-sector collaborations are used as interactions to adopt, share or coordinate cyber defence practices between at least two different smart city sectors. To address the existing and emerging cyber-attacks, transferable and innovative solutions should emerge from individual sectors within a smart environment to

support modern DF [30, 160]. RQ1, RQ2 and RQ3 help uncover key themes and gaps in current literature and suggestions for future research direction.

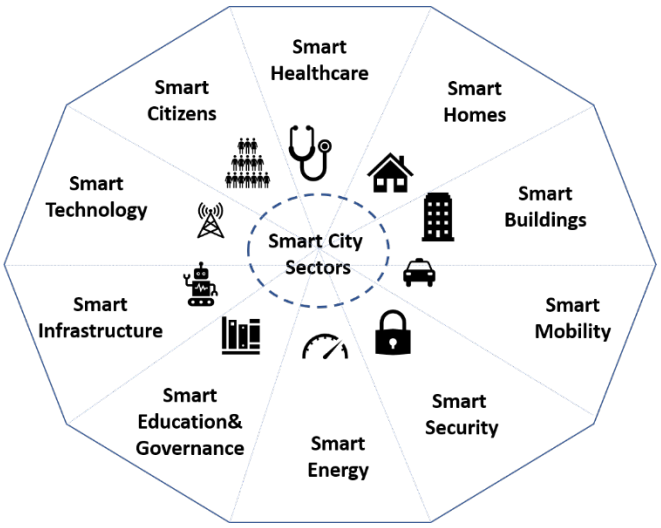


Figure 20 Core smart city sectors.

2.4 Results Analysis

2.4.1 Primary Studies

Applying the protocol revealed that no primary studies were published before 2011, suggesting that cyber resilience and DFIR addressed by CPS frameworks and systems in smart cities is a recent paradigm. Nevertheless, as Figure 21 shows, there is an upward trend in CPS-related research within smart cities addressing cyber resilience and modern DFIR, which indicates that this has emerged as an active research area. This trend will likely continue as the First Quarter (Q1) of 2019 is just over half of the studies published in 2018, as demonstrated in Table 2.

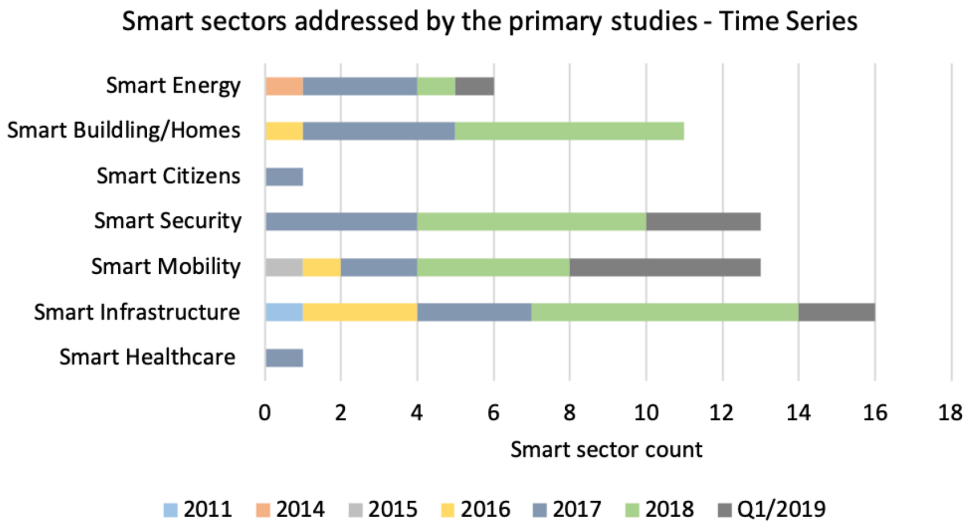


Figure 21 Smart sectors addressed by the primary studies time series.

Table 2 Primary studies' distribution by type. Journal—J or conference—C and publication year.

Year	11	14	15	16	17	18	Q1/19	Total Studies
%/year	2%	2%	2%	10%	23%	40%	21%	52
J			1			13	10	24
C	1	1		5	12	8	1	28

2.4.2 *Keyword Analysis*

To help establish common themes amongst the primary studies, a keyword analysis including all 52 primary studies was carried out. The frequency of specific keywords appearing in the primary studies is shown in Table 3. As the table captures, the second most frequently used keyword in the dataset is "System", closely followed by "Security", "Internet of Things" and "Cyber-Physical Systems". This shows an increasing research interest in the security of CPS in the context of the IoT. Furthermore, the keyword "framework" indicates that it is an active but still emerging area of research interest in the context of CPS cyber resilience and support for DFIR. The dataset also demonstrates that there is a significant disparity in the research interests in "detection" compared to other aspects of CPS security. The keywords used in established investigation models and frameworks to define these

investigation phases including “Response”, “Recovery” or “Prevention” rank lowest in the dataset. In addition, “Forensics” and “Cyber Resilience” rank also low in the dataset indicating potential areas for further research requirements.

Table 3 Primary studies’ keyword analysis.

Keyword	Occurrence	In Studies
Attacks	4165	50
System(s)	3650	52
Security	2272	51
Internet of Things/IoT	2024	36
Model(s)(ing)	2002	52
Cyber-Physical Systems	1857	52
Smart	1750	52
Device(s)	1610	50
Detection	1193	47
Approach(es)	589	50
Method(s)	579	49
Analysis	579	52
Framework(s)	491	44
Technique(s)	461	44
Cyber * resilience/resilience	251	26
Processing	242	38
Architecture	239	43
Forensic(s)	214	16
Cyber * security	179	37
Response	156	33
Incident(s)	41	15
Prevention	38	20
Recovery	32	10

[1] The asterisk (*) in this table is used to represent the variants considered during the keyword search: space, dash or continuous word without any space i.e., ‘cyber resilience’, ‘cyber-resilience’, ‘cyber resilience’ and ‘cyber security’, ‘cyber-security’, ‘cybersecurity’.

2.4.3 Key Themes

Our analysis of the primary studies shows several emerging themes and main focus domains, each of which is discussed within sections a - g.

a. Chronological Analysis of Key Events

The purpose of the chronological analysis is to examine the main determinants and the time correlation for the research distribution addressing CPS cyber resilience and modern DFIR in smart cities concerning the defined scope. To achieve this, the primary studies were organised in chronological order and classified depending on the year published and the type of publication, as shown in Table 2. The trend shows that the first empirical study concerning this topic is dated from 2011 from a conference proceeding. It is not until 2016 that there is an 8% increase in research for this subject area through conference proceedings as the main outlet for research publications. By 2017, the number of articles doubled and increased again in 2018. The differentiating factor was the high proportion of journal articles over publications from conference proceedings whilst by the first calendar quarter (Q1) of 2019 and the articles published in journals reached over 75% of studies published throughout the entire of 2018.

Further investigating the results from the chronological analysis, the following key years were highlighted as potential influencing factors concerning the investigated CPS-related research developments: 2011, 2016, and 2018.

2011. This year was defined by the Hannover Messe Fair, where the term “Industry 4.0” was born to describe the next industrial revolution, a vision of three German engineers. While the first industrial revolution dates back to the end of the 18th century introducing water and steam power, the second industrial revolution at the turn of the 20th century was centred around mass production using electricity and the third industrial revolution integrated IT and electronics into production systems, the 4th industrial revolution introduces digital processing and

implementation of the IoT into production. In this context, the concept and the vision have been established for CPS for production systems. Industrie 4.0, a German origin of the Industry 4.0 term, is used synonymously with cyber-physical production systems [1, 161]. In the post-recession output fall, the vision of Industry 4.0 elevated the German manufacturers and economy back into the spotlight [162, 163].

2016. The creation of the UK's National Cyber Security Centre (NCSC) as the technical cybersecurity lead was a feature of this year. Furthermore, the investment and economic infrastructure plans announced in the National Infrastructure Delivery Plan in the UK [164] and the announcement of the significant cybersecurity fund as part of the USA's Cybersecurity National Action Plan also took place in 2016 [165]. The World Economic Forum (WEF) was also held in Davos. The WEF used the motto: "Mastering the Fourth Industrial Revolution" [166]. The event was attended by 2500 participants and 40 heads of states from 140 different countries discussing ideas to tackle global challenges sustainably with the aid of technology and the economic impact of Industry 4.0.

2018. In the USA, there was the notable creation of the Cybersecurity and Infrastructure Security Agency (CISA) responsible for national critical infrastructure from physical and cyber threats. Australia released an update for its cybersecurity sector competitiveness plan outlining Australia's significant economic opportunities to become a "global cybersecurity powerhouse" [167]. Despite Industry 4.0 being a global phenomenon, the acceleration of efforts by countries in the race of Industry 4.0 is local to lead the change and be the face of the new digital transformation. This era is characterized by high-capacity and low-latency 5G networks that will catapult

digitalisation, which is predicted to create significant opportunities in many economic sectors. Furthermore, in terms of cybersecurity, the NCSC reported on the growing cybercrime threat, recording 34 significant cyber-attacks that typically required cross-government responses over two years [168]. The government has explicitly acknowledged the need to improve the resilience of the UK's CNI [169]. The consequence of the transformation not having peaked yet results in a continued increase in investment, grants and financial incentives; therefore, research efforts continue [170, 171].

Relating the primary studies' trends with the key events, this study identifies a link between the technological and economic landscape and cyber-resilience-centric research that addresses CPS in smart cities. From the primary studies, it emerges that the trend in the increase of papers has been influenced by a strategic focus on cybersecurity; improving the cybersecurity defence landscape, including the creation of NCSC and CISA; significant investment in improvements and strengthening of the national critical infrastructures. Coupled with efforts and initiatives exclusively focused on digital transformations to gain economic advantage could explain the surge in research studies published from 2016 onwards.

b. Cyber Resilience Analysis

To address the question of how existing frameworks and systems that address CPS in smart cities support cyber resilience, this study considers the scope of resilience within the cybersecurity discipline and the evidence reported in the primary studies. To achieve this, the primary studies were organised in order of the reported

evidence of how the cyber-attacks across the physical and digital domains were addressed and how the external or insider threats were approached.

Cyber resilience is widely acknowledged by governments including the UK's National Cyber Security Strategy 2016–2021. Although NCSC promotes cyberspace resilience by shaping technical standards that govern emerging technologies and promotes best practices [66], the Joint Committee on National Security and Strategy in their report acknowledged that the UK Government must do more to improve the cyber resilience of the CNI [169]. Cyber resilience has been acknowledged as a challenge in the IoT; President Obama issued Executive Order (EO) 13636 to strengthen critical infrastructure cybersecurity resilience. Likewise, improving cyber resilience is at the forefront of the Australian Government [167].

CPS resilience is accepted as an important aspect by the scientific community, governments and industry, it is a multi-dimensional and multi-disciplinary facet. However, despite many efforts to define the term "resilience", it has no clear and uniform definition or performance metrics [172, 173]. The term resilience is described by the National Institute of Standards and Technology (NIST) as "the ability to quickly adapt and recover from any known or unknown changes to the environment through a holistic implementation of risk management, contingency, and continuity planning" [174]. Furthermore, to evaluate CPS resilience, several areas of CPS resilience were studied including policy [175], the correlation of resilience on probability and impact of performance under adverse conditions [176] and risk and resilience correlation [173].

The nature of CPS is multi-dimensional, converging physical and cyber domains in a highly complex ecosystem integrating systems, software, people and services. In this study's approach to establishing how CPS in smart cities support cyber resilience, this research investigated the primary studies according to specific layers within the TCP/IP model—a standard model used in computer networks, based on modern DFIR general-purpose frameworks, adversary type and by the smart sector covered by each study.

Layers were identified concerning the TCP/IP (Transmission Control Protocol/Internet Protocol) model described in RFC 1122 [177]. The TCP/IP model consists of four layers, which, from the lowest to the highest, are the link layer, the internet layer (network), the transport layer, and the application layer. The primary studies can be categorised into three layers: physical, communication (aligns to the Internet and transport layers of the TCP/IP model) and application. A similar categorization approach was taken by authors [172] to define CPS resilience. For example, the physical layer includes physical faults, component failure and the delivery of the attacks through access within the security perimeter including attacks on CPS controllers, sensors and actuators. The communication category includes communication-environment-based disruptions and attacks such as Denial of Service (DoS), Man-in-the-Middle (MiM), the user to root type buffer overflow or remote to user ftp write. The application category included False Data Injection (FDI), malware and other services such as cloud storage and web application-based attacks. Some incidents can fit into more than one category [178].

DFIR Support was investigated concerning the phases that form the basic foundation of an IR plan from preparation to post-incident activities to identify how the primary studies address this process. The DFIR support was studied accordingly

to general-purpose DFIR frameworks and standards such as the Digital Forensic Research Workshop (DFRWS), Abstract Digital Forensic Model (ADFM), NIST 800-61 and International Organization for Standardization and International Electrotechnical Commission ISO/IEC 27050.

Adversary Type was identified within each layer, where the threat can be caused by external or internal factors. This study considers an internal threat to be a threat by an adversary initiated inside the security perimeter. In this thesis, the terms internal and insider threat are used interchangeably. Such an entity is authorised to access the systems or resources within the security perimeter but acts in a way that is not authorised. Examples include malicious or disgruntled employees or contractors who have direct access and sufficient knowledge of the system or the resource. In contrast, an external threat is initiated by an adversary from outside the security perimeter. Such an entity is not authorised to access or use the systems or resources and gains access through unauthorised or illegitimate attack vectors. This study investigates how the primary studies address this aspect; a similar emphasis on this approach was followed by reference [172].

Smart Sectors will leverage CPS performance and resilience differently. CPS operate across different smart sectors, therefore this SLR identifies the smart sectors as reported in the primary studies. Several studies specifically focus on the applicability of resilience in terms of the CPS's ability to withstand disruptions, recover from and adapt to known and unknown threats, as shown in Table 4. For example, in their approach, reference [158] argued that optimisation between smartness and cyber resilience in a CPS is required for a balance between functionality and cybersecurity without compromising the systems' resilience. In this study, the percolation theory was used as the basis for evaluating the stress caused

by disruptions. The authors in reference [179] argued that the absence of common security standards and flexible methods to assess IoT security requires dedicated testbeds to systematically evaluate the devices' resilience under various conditions. The study developed a security testbed framework for the IoT. The testbed consists of standard security testing predominantly based on well-established vulnerability scans and penetration testing methodologies including port scanning, process enumeration, fuzzing and fingerprinting. The advanced testing capabilities of the testbed are based on techniques and tools including ML, traffic-based IoT device type identification, automatic anomaly detection and environment simulations. The number of test scenarios demonstrated the effectiveness of the testbed in detecting the IoT devices' resilience against attacks including DoS. Another study [180] focused on CPS resilience mechanisms that can be applied during runtime to sustain resilience utilising self-healing structural adaptation. In the following study [181], the authors argued the importance of an interdisciplinary integrated approach between the cyber and physical layers. They asserted that cyber resilience-by-design must address two scopes to achieve overall resilience, the security controls, communication scope and the power engineers' scope to reinforce the weak points during the design. The study proposed an integrated cyber-physical sustainability metric framework to assess CPS cyber resilience.

Table 4 Primary studies focusing on aspects of cyber resilience(-by-design).

Year	Primary Study	Smart Sector
2011	[182]	Infrastructure
2012		
2013		
2014	[73]	Energy
2015	[183]	Mobility-Automotive
2016		
2017	[184]	Infrastructure
2018	[181, 185]	Energy, Mobility-Aviation
Q1 2019	[158, 179, 180]	Security, Mobility-Aviation

Further analysis investigating possible correlations with the emerging key themes discussed in this chapter shows no clear geographical correlation. The studies, categorised in Table 4, except for [181], acknowledged grant funding. Time correlation was observed with a continued trend in the increase of primary studies focusing on cyber resilience in 2018 and Q1 2019. This trend could indicate a response to the emergence of new and diverse types of security-related incidents that have the potential to be damaging and disruptive.

The author in reference [182] argued that the key difference between control and IT systems is the control systems' interaction with the physical world and concludes that to withstand cyber-attacks, systems should be resilient by design. The author asserts that the risk to control systems is higher due to the exposure and availability of vulnerabilities combined with the increasing motivations and capabilities of the attackers. The paper focuses on sensor attacks and addresses ways of prioritising sensors. Attack types were studied using the Tennessee-Eastman process control system (TE_PCS) model [186]. An automatic response mechanism was introduced based on various system states taking into consideration a false alarm response. The author's main conclusion was the strength of the TE-PCS's design resilience. Although the proposed principles and techniques could be applied to other physical processes and the false positive rate at 1000 simulation cycles was 0%, the automated response may not be appropriate for all control systems. The author cautions of a likely lack of resilience by design in large-scale control systems which could remain vulnerable to several attack vectors. Further, the author in reference [184] defined a trustworthy service as one which secures against cyber-attacks and operates normally despite faults or attacks. The authors proposed an IoT framework

to integrate Smart Water Systems (SWS) with the IoT using a multilayer architecture trustworthy service and proposed that security issues should be addressed systematically by developers during the design and development of each IoT layer. Anomaly Behaviour Analysis (ABA) Intrusion Detection System (IDS) methodology was applied to protect the secure gateway from attacks utilising the SWS Testbed. The secure gateway is part of the communication layer. The general detection rate of the ABA-IDS approach was over 90% for 600 packets/second intensity, with less than 3.5% recorded false alarm rate, with the fastest detection of 1 s and the slowest detection of a 4 s interval.

Other studies [73, 181] focused on CNI such as power grids whilst urban systems were investigated by reference [158]. In reference [73], the resilience of five classical routing protocols applied in distributed large-scale networks was studied through simulation. Resilient techniques using route diversification were introduced to enhance the protocols' resilience against cyber-attacks. The resilience was evaluated based on metrics consisting of five performance parameters which showed promising results. The communication layer was also the focus of the [181] study, which proposed a new metric system framework to assess the reliability of large-scale distributed power systems. The author asserts the importance of combining the communication layer's cyber vulnerabilities with the physical layers' resilience for a meaningful assessment of the system's sustainability. The following study [158] developed a network efficiency and resilience evaluation method for Intelligent Transportation Systems (ITS) in response to random and targeted attacks in urban areas. The author maintains that although the use of sensors is beneficial for automation, the infrastructure through their use becomes complex and liable to

unknown and little understood vulnerabilities. The article concludes that the system's relative resilience was not sensitive to the levels of disruption. Integrity attacks were investigated by reference [185] who proposed a global attack detection system for resilience against attacks on the railway traction systems. Resilience mechanisms that can be applied during runtime and are adaptable to the changing environment were studied by reference [179]. It is argued by reference [158] that the rate of integration of smartness in many systems proliferates at a greater rate than the ability to develop resilience. Whereas reference [180] identified resilience in the IoT as a significant challenge with research often focused only on one aspect or a single attribute of resilience. Our results, as shown in Table 5, support this notion, for example, 46% of the primary studies considered the communication layer, while only 5% considered all three layers. This SLR found that the communication layer had the most significant incremental trend in 2018, as presented in Figure 5, generally with an utmost focus across the smart industry and smart mobility sectors, Figure 6.

Table 5 Primary studies categorisation by the reference model layers.

Threat Layers	Primary Studies
Physical, Communication and Application	[68, 69, 185]
Physical and Communication	[2, 72, 125, 156, 157, 181, 183, 184, 187-194]
Physical	[41, 159, 195-197]
Communication	[36, 41, 73, 158, 179, 180, 196, 198-213]
Application	[70, 195, 214-217]

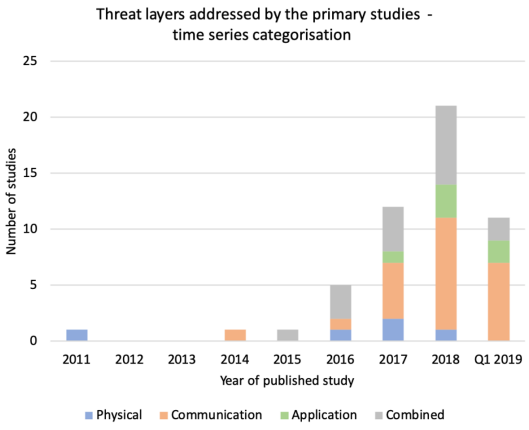


Figure 22 Time series categorization of the threat layers addressed by the primary studies.

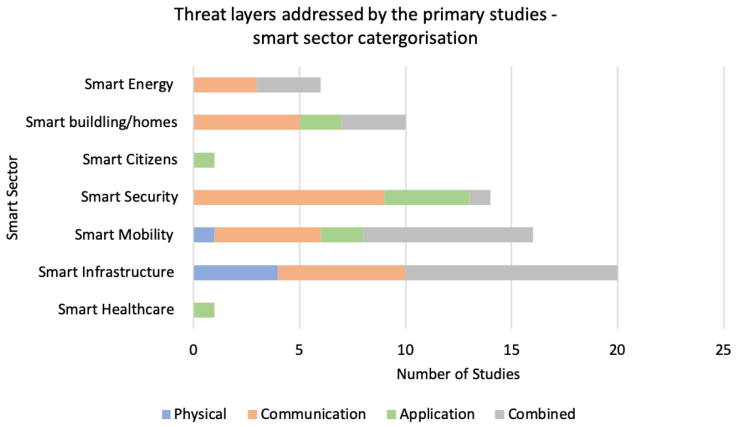


Figure 23 Reference model layers categorization of smart sectors addressed by the primary studies. In this graph, multiple sectors addressed in a single study are reported individually to preserve sector visibility.

When investigating the adversary type, the results show that 19% of the primary studies considered internal and external threats in their research, as presented in Table 6. In 45% of the studies, the threat type was not sufficiently clarified. However, this research observed a continued increase in studies focused on a combination of external and internal threats, as presented in Figure 24, generally with the greatest aggregation of studies in the smart infrastructure and smart mobility sectors, as illustrated in Figure 25.

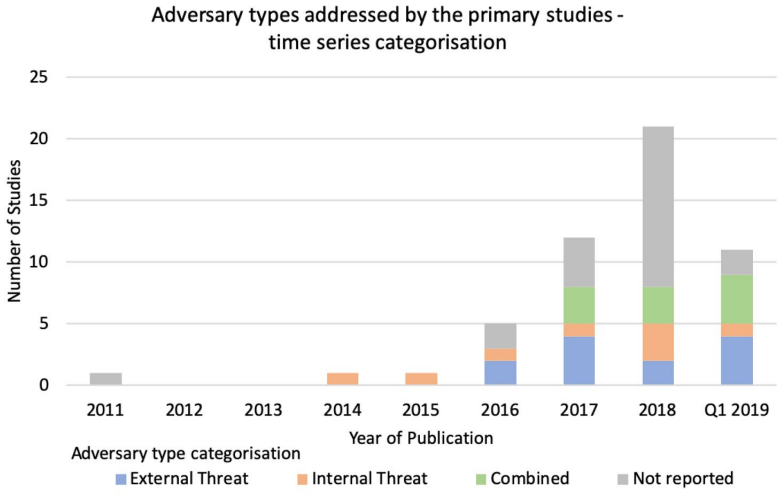


Figure 24 Time series categorization of adversary type threat factor of smart sectors addressed by the primary studies.

Some studies [73] addressed insider threats on smart devices such as smart meters, which can be compromised by an active attacker to disrupt network

communication. The study in addition to considering the compromise of the physical nodes addresses the ability of the protocol to absorb the degradation following an insider attack. In [191], the focus of the study are large-scale distributed CPS proposing a quantitative cyber-physical security assessment methodology, the following research study [218] provides an overview and discusses related risk assessment methods. Another study [179] investigated external threats and articulated that the challenges of IoT devices provide means for hackers to access such devices. Therefore, the proposed testbed aimed to facilitate the analysis of various types of IoT devices either by using the conventional penetration testing methodology or advanced security testing utilising an ML approach. Internal and external faults including malicious activity were addressed by other studies [180, 219]. In reference [212, 219], the focus of the paper is on a Multiple Characteristic Association (MCA) approach to address cyber-attacks and faults in electrical CPS and reference [212] utilised an attribute-based time-sensitive and location-centric access control model consisting of an administrative and an operational component with applicability to remote and local operations.

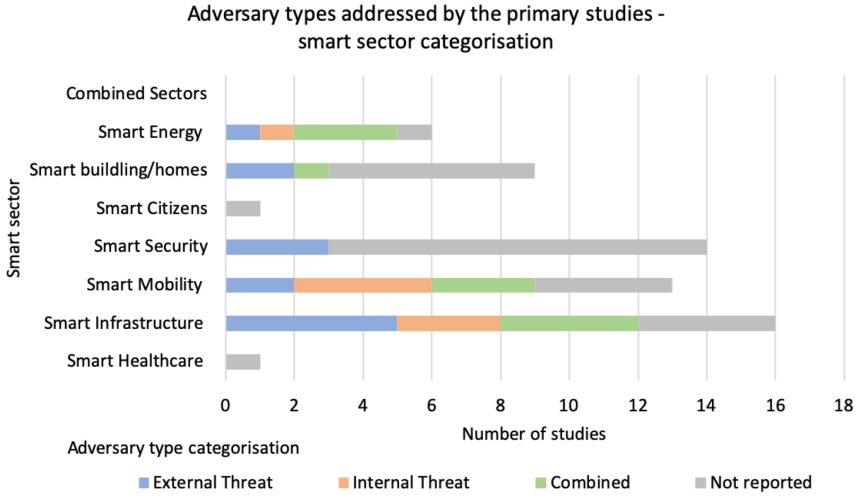


Figure 25 Adversary type threat factor categorization of smart sectors addressed by the primary studies. In this graph, multiple sectors addressed in a single study are reported individually to preserve sector visibility.

Table 6 Primary studies categorisation by adversary threat type

Threat Type	Primary Studies
Internal and External Threats	[36, 72, 125, 156, 180, 190, 194, 196, 202, 212]
External Threat	[157, 179, 184, 187, 189, 191, 193, 197, 198, 201, 203]
Internal Threat	[59, 82, 157, 182, 184, 213, 217, 220]

c. DFIR Analysis

Digital forensics forms a substantial part of IR in the cybersecurity sector; it is a recognised scientific methodology with a key focus on the process and verifiable conclusions. Although several published digital investigation models outline the steps for investigation by forensic teams, there is no single uniform IR model. The simplest lifecycle for an investigation model consists of three stages, “acquisition”, “analysis” and “reporting”. However, with the increased penetration of digital technologies into modern lives, there were several revisions to the investigation stages. The U.S. Department of Justice (DoJ) proposed a four-stage process consisting of “acquisition”, “identification”, “evaluation” and “admission as evidence” [221]; the DFRWS model consists of six phases namely “identification”, “preservation”, “collection”, “examination”, “analysis” and “presentation” [222]. The ADFM has expanded the process by three more stages: “preparation”, “approach strategy” and “returning evidence” [223]. Due to the evolving sources of DE, the digital and physical environments are closely converged where physical artefacts contain the DE, which is reflected in the Integrated Digital Investigation Process (IDIP) consisting of five stages defined as “readiness”, “deployment”, “physical crime scene”, “digital crime scene” and “review” [224]. Similar to the DFRWS model, the ISO/IEC 27050-3:2017, a general-purpose framework for Electronically Stored Information (ESI) was developed for digital investigations containing seven stages: “identification”, “preservation”, “collection”, “processing”, “analysis”, “review” and “production”. The

NIST published an IR procedure NIST 800-61 in response to the frequency of emerging incidents consisting of four stages: “preparation”, “detection and analysis”, “containment, eradication and recovery” and finally “post-incident activity”. In CPS, IR is a complex, multifaceted problem crossing the physical and cybersecurity boundaries.

The primary studies were classified by their key themes into groups according to the NIST 800-61 IR stages [178]. The studies were determined to have focused predominantly on the detection and analysis stage, as shown in Table 7.

Table 7 DFIR key stages categorisation of primary studies.

Key Stage	Primary Studies
Preparation	[69, 180, 191, 225]
Detection and Analysis	[2, 25, 36, 41, 68, 72, 125, 156-159, 179-185, 187-191, 195-202, 204-208, 210, 211, 213-215, 220]
Containment, Eradication and Recovery	[68, 70, 73, 158, 180, 183, 184, 194, 203, 209, 211, 212, 215, 217]
Post-Incident Activities	none

Preparation is an important part of the IR. Apart from compiling assets, creating a communication plan, setting metrics or creating an incident plan for each type of incident, security event simulation is also a valuable part of this stage. Simulation or modelling helps identify gaps, determine and optimise which security events and at what trigger should be investigated; therefore, they provide a controlled opportunity to strengthen weaker areas and improve cyber resilience, which was discussed in the previous section. For example, the author in reference [225] proposed a novel framework using the Fuzzy Analytic Hierarchy Process to evaluate and rank the cybersecurity challenges in smart cities. Amongst the 9 identified smart

sectors (factors) and 32 sub-factors, smart security was rated highest for being influenced by cybersecurity challenges in smart cities. The results of the study placed the sub-factors identified as part of smart security in the highest priority areas influenced by cybersecurity challenges which were identified as “surveillance and biometrics” followed by “simulation and modelling” and “intelligent threat detection”. The results show that smart security sector studies do not have a specific focus on cyber resilience aspects, see Table 4. and the research focus relates predominantly to the communication layer threats, see Figure 23. A security-by-design approach was proposed by reference [69] articulated as a framework to develop a highly secure and trustworthy smart car service and protect them from cyber-attacks. The authors argue ABA is a more suitable approach because of the sensors’ low computational power and therefore a lack of encryption techniques applicability. The sensor profiling was accomplished by using the Discrete Wavelet Transform (DWT) coefficients and the Euclidean distance was utilised for sensor classification. The presented results demonstrated up to 95% accuracy for unknown and 98% for known attacks with a low false-positive rate.

Incident Detection and Analysis (IDA) is a key phase in IR because the response cannot be manifested without accurate detection. Although incident detection is considered a reactive approach, there are detectable events that precede an incident. The results from the primary studies show that the highest distribution in the detection and analysis stage of the IR model is in the smart infrastructure sector as shown in Figure 26, and overall 67% of the sampled primary studies focus exclusively on cyber-attacks detection, as shown in Figure 27. The author in reference [68] presents a framework for smart homes and smart buildings

addressing multiple layers and threat types. The study utilised ABA-IDS to continuously monitor, detect and classify cyber-attacks against sensors with high accuracy. The study aimed to extend the methodology to other IoT security frameworks, such as smart water systems [184] and smart grid systems [189]. Both studies rely on ABA-IDS utilising JRip classification algorithm achieving up to 99.8% and 97.18% accuracy on their respective datasets. The ABA-IDS detection and the classification results for reference [68] were similar and in some instances exceeded the results of other state-of-the-art protection systems for smart grids. Different approaches were proposed to enhance the detection of cyber-attacks in ICS. For example, a secure water treatment plant often consists of distributed cyber infrastructures that control physical processes. The author in reference [201] proposed a Time Automata (TA) approach, whilst another study [25] focused on a hybrid of machine learning combined with specification-based detection. An orthogonal defence mechanism consisting of several intelligent checkers was used by the author in reference [72].

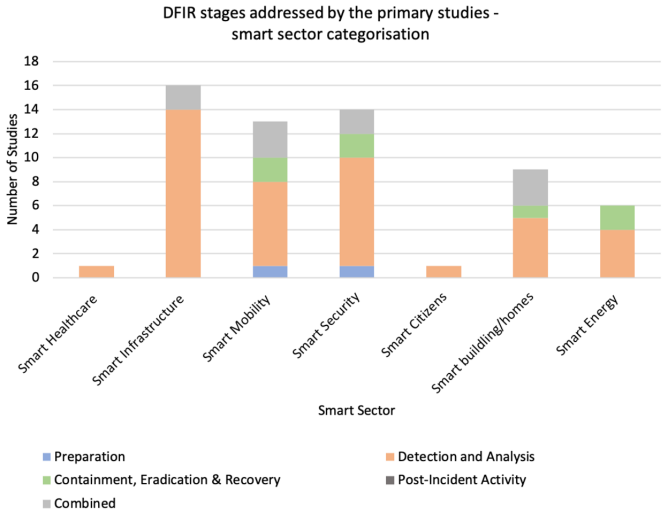


Figure 26 DFIR stages categorisation across smart sectors addressed by the primary studies. In this graph, multiple sectors addressed in a single study are reported individually to preserve sector visibility.

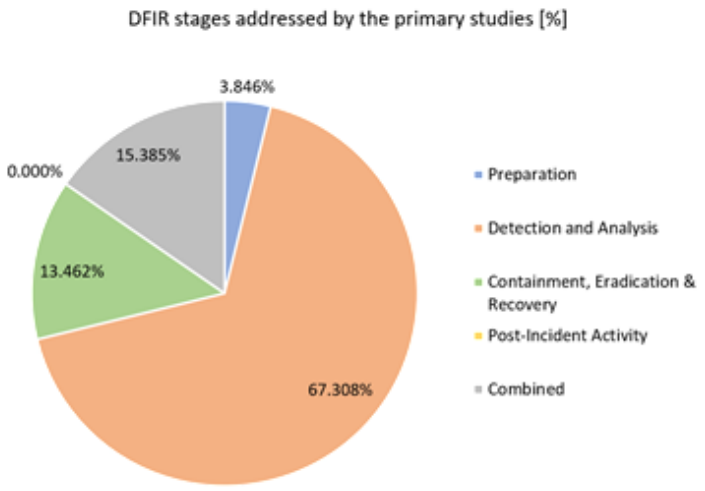


Figure 27 DFIR stages addressed by the primary studies.

Containment, Eradication and Recovery (CER) is the part of the process where models and standards differ. Whilst NIST views the CER as a single step, SANS (SysAdmin, Audit, Network, and Security), DFRWS and ISO/IEC 27050-3:2017 view them as separate segments. Furthermore, the terminology used by different frameworks and standards to identify similar steps can vary. The terminology used by NIST 800-61 refers to containment as an aim to stop the attack or threat, eradication removes it stopping cross-systems proliferation and recovery aims to get the system operation returning to business as usual. The figures show that only 13% of the primary studies investigate the CER segment of the IR procedure, as shown in Figure 27. For example, the focus of the following study [158] is on increasing resilience rather than lowering risks to demonstrate system recovery from disruption. The author argues that smart development over resilience may benefit some smart systems to achieve recovery through automation by redistributing the traffic by using alternative routes. This is part of the investigated model's algorithm. However, the limitation of the study is its consideration of large and very large urban areas; therefore, the model's applicability was not tested on smaller urban areas.

Furthermore, the modelled scenario captured only a limited set of ITS disruptions, therefore, the effect of disruptions from different cyber-attacks compared to those which were tested, and their method of recovery may vary. The author in reference [73] presents an interesting notion of extending the concept of resilience in networking to survivability, fault tolerance and security, however, acknowledges difficulties in defining quantitative metrics. Focusing on the internal threat, the reliance is on the protocol's capacity to absorb the attack under some failure behaviour and the resilient technique provides dynamicity to improve the self-healing capabilities of smart meters. Another study with a focus on resilience mechanisms [180] proposes achieving self-healing through a structural adaptation approach by substituting failed components as a method of recovery for compromised CPS. The author asserts that this is achievable provided the compromised component is redundant and can be isolated. The author in reference [68] proposed an IoT security framework and based on the detection of abnormal behaviour, recovery actions can be taken. Other studies acknowledge the elapsed period before IR starts after the attack occurs. For example, the study in reference [217] presents a hybrid solution of distributed and centralised continuously evolving trust-based intrusion detection model aggregating multiple trust data sources to enable an effective in-flight network defence. The study claims, that following an abnormal pattern's emergence, trust-value triggered IR with active defence is possible. Comparable to the results in Section 2.3.3 c, the results from the primary studies show that research often focused on one aspect of DFIR, see Figure 27.

Post-Incident Activity (PIA) is one of the most important phases of the IR process, but it is most often omitted [226]. This phase provides an opportunity to

contribute to continuous learning, an evidence-based body-of-knowledge and to form a robust CTI. The IR can be accelerated by having an effective and specific CTI context around an initial indicator [227]. Therefore, a review of what occurred and defining actionable advice that can be used to inform decisions in the IR's preparation phase is important to achieve a closure of the IR process. The PIA has not been addressed by the primary studies.

d. Data Source Analysis

Through this research, a lack of available real datasets from CPS systems was identified. Although experimentation was carried out, predominantly this was limited to software-based simulations (46%) and simulation infrastructure (42%) by the primary studies, as shown in Figure 28. The infrastructure-based simulations typically relied on testbeds to replicate real-life CPS device settings such as secure water treatment (SWaT) or water treatment plant (WTreat) testbeds [190, 196]. However, in 12% of the studies published between 2018 and early 2019, public scientific datasets like BATADAL [125] or CAIDA [207] were used either solely or in conjunction with software-based simulation. Carrying out experimentation in an isolated environment limits testing in several ways. For example, the unavailability of a current real dataset limits the reflection of the current threat types and limits the full contextualisation of the actual CPS devices' constraining factors such as resources or connectivity disruptions.

e. Analysis of Primary Studies Cross-Sector Proposals or Applications to Improve Digital Forensics

The purpose of analysing the cross-sector proposals or applications in smart cities is to explore transferable solutions that emerge from individual smart sectors to investigate possible trends and attempts to improve DF investigations. To achieve this, the primary studies were organised accordingly to the smart sector's distribution according to the scope of our research, as shown in Figure 29.

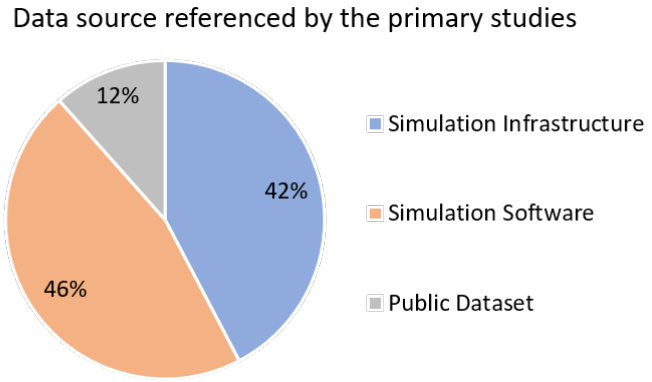


Figure 28 Data source referenced by the primary studies.

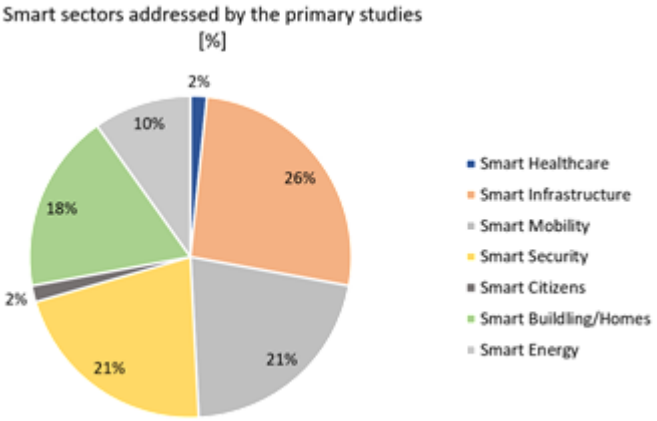


Figure 29 Smart sectors addressed by the primary studies.

The scientific community focused the research on smart infrastructure, followed by smart mobility and smart security sectors whilst smart healthcare and smart citizen were addressed only by a small number of studies, see Figure 29. Some of the studies address more than one themes, which is taken into consideration. This trend could be explained by the influences of key events such as Industry 4.0 and the

maturity of the research of the design principles and enabling technologies in these areas [1] whereas the lack of research within the smart healthcare and smart citizen sector could be impacted by regulatory restrictions, ethical challenges, lack of relevant usable datasets and the current health care models or pathways [195].

The results show that some studies address more than one smart sector [184, 195, 202, 206, 208, 215] or aim to diversify their future research [25, 72, 156, 157, 180, 182, 188, 191, 213, 216]. For example, reference [195] explores smart support for independent living of the elderly within the community to maximise their independence whilst maintaining the ability to deal with their complex medical needs across multiple smart sectors including healthcare, homes and infrastructure. Furthermore, several studies consider developing their research to generalise applicability to other smart sectors and acknowledge the need for framework adaptability as a result of the complexity and constant change of interconnected devices [215]. For example, the principles and techniques applied by reference [182] could be applied to other physical processes than the one covered by the study, whilst reference [213] suggests their methods can be applied in several CPS domains such as power networks, transportation, oil and natural gas systems.

Although the cyber threat landscape is changing from hobby-hacking to organised cybercrime, the cyber-attacks are becoming more sophisticated, organised and targeted; there is little scientific evidence of attempts for supporting modern DF, cross-organisational information security sharing or coordination [30]. Security practices remain in silos lacking collaborative cyber defences to deal with the increased sophistication and coordination of cyber-attacks including APT [4, 29].

This assertion is supported by our analysis of primary studies thus far. The transition from more traditional to IoT-enabled CPS creates highly complex ecosystems, however, the focus of research is often limited to the boundary of the individual organisation or smart sector.

f. Typology Analysis

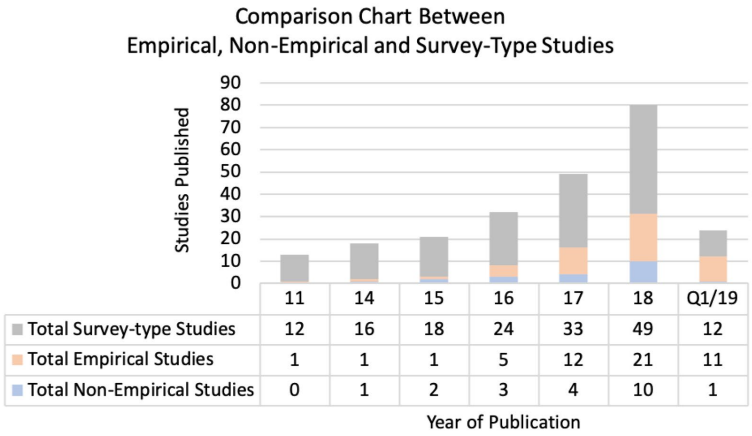
The purpose of the typology analysis is to separate the non-empirical and the empirical studies and to examine their chronological distribution. This analysis helps to better understand if the CPS frameworks and systems supporting cyber resilience or modern DFIR are predominantly academic ideas built on theory or do they emerge based on identified needs or as a result of relevant events.

Cyber-attacks are a natural progression of physical attacks; they are more economical, reduce the risk for the attacker and have fewer geographical constraints. Studies from the sample recognised the cybersecurity risk factors that the integration of connected devices, sensors and automation helped by AI has on smart ecosystems. In 2011, the focus of an [182] empirical study was attacks on sensor networks and their impact on the process control system. The research study referred to the example of targeted ICS-based attacks such as the Maroochy Shire Council sewage attacks in Queensland, Australia in 2000; Ohio's 2003 Davis-Besse Slammer worm private network attack and the 2007 Iranian nuclear plant Stuxnet worm attack. The control systems' vulnerabilities such as Stuxnet and urban migration are also referred to by reference [68]. In 2007, the disruption and economic consequences of a large-scale cyberattack on the USA power grid were studied by [108]. Several non-empirical studies investigated the theoretical concepts

or potential challenges to be addressed for different aspects of the cyber defences against targeted attacks related to the increased interconnectivity and heterogeneity of the physical and cyber convergence. In 2014, the following study [90] investigated a federated building information system as a method of preventing hostile reconnaissance, managing intellectual property and enabling operational security. The study refers to a 2013 incident in Hackney, London in which a piling rig penetrated the roof of a Network Rail tunnel.

Therefore, the proliferation of digital technologies and the integration of IoT with physical systems expands the scope of forensic science creating a need for new specialised forensic techniques to reduce the backlog, workload and cost of the forensic investigation process [228]. DF has developed as a branch of forensic science alongside the conventional forensic disciplines covering diverse digital technologies that can be exploited by criminals. The results presented in Figure 30 demonstrate the chronological trend between surveys, non-empirical and empirical studies. The focus of this SLR is on primary empirical studies. The total of the studies shows that non-empirical studies including the survey-type studies amounted to 64% compared with 36% of the empirical studies of the reviewed samples. Although the number of survey studies consistently increased, a sharp increase in empirical research is observed during 2017 and a similar surge in the non-empirical studies is observed in 2018. Depending on this evidence, it is possible to argue that this dynamic could be influenced by the key events discussed in Section 2.3.3 a. Furthermore, from the empirical studies, it emerges that the focus of the research was informed by the threats of specific events, driven by the need for defence-in-

depth mechanisms and influenced by the implementation of technological innovation and application within smart sectors.



	11	14	15	16	17	18	Q1/19
Total Survey-type Studies	12	16	18	24	33	49	12
Total Empirical Studies	1	1	1	5	12	21	11
Total Non-Empirical Studies	0	1	2	3	4	10	1

Figure 30 Comparison chart between empirical, non-empirical and survey studies. Non-empirical studies passed the systematic Phase0, Phase1 and Phase2 selection process' stages. Survey-type studies were considered, based on the original search string, in Phase0 an

g. Geographic Analysis

The purpose of the geographic analysis is to support the analysis in previous sections and gain a better understanding of where the research is concentrated, which geographical sectors have interest and opportunities for research addressing CPS-related cyber resilience and DFIR in smart cities. To achieve this, from the primary studies' authorship list, each unique country was recorded and assigned to the continent, as demonstrated in Figure 31. The colour hue represents the frequency of research carried out within the geographic region. The geographic analysis shows that the USA with 23% has the highest number of contributions of reviewed studies, followed by Singapore with 10%, the UK with 8% and Australia with 7% of contributions in the reviewed studies. In terms of continents, Figure 32 shows that Asia is the continent with the highest concentration of the relevant CPS research at 37%, closely followed by Europe at 30% and North America at 26%. Central and South Americas, Australia and Africa are the continents with the lowest number of published studies within the scope of our research.

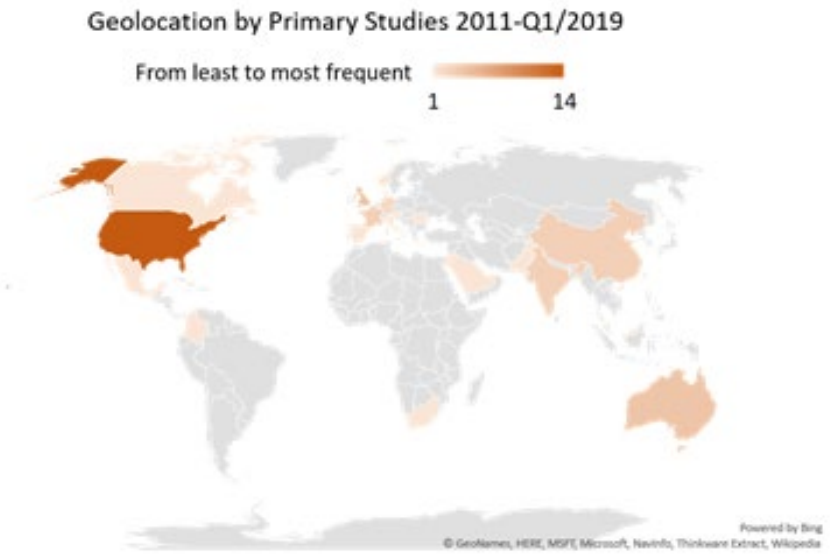


Figure 31 Geolocation by primary studies 2011-Q1/2019. (Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation. https://www.microsoft.com/en-us/maps/ product/print-rights).

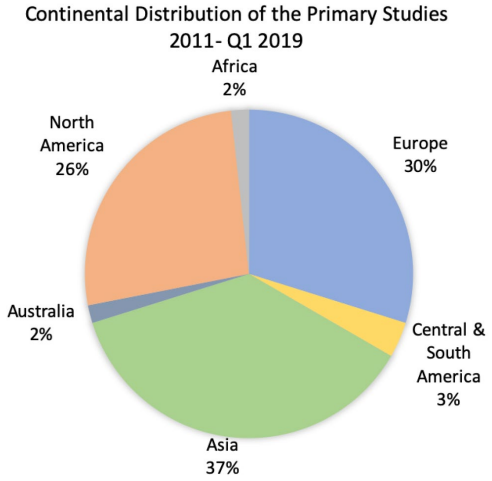


Figure 32 Continental distribution of primary studies.

2.5 Discussion

The analysis revealed that in the last decade, CPS have emerged as a new paradigm. As a result of the increased growth, complexity and heterogeneity of these infrastructures [161, 219], the volume and the variety of vulnerabilities and attacks have evolved highlighting the need for defence mechanisms [229], the need for cyber resilience and the capability to support DFIR [28, 36, 61, 62]. In this SLR,

the analysis of the primary studies supports the assertion that CPS-related cyber resilience and DFIR are active research domains. As has been noted in the analysis of the results, several empirical research studies have focused on CPS-related cyber resilience and DFIR. For example, Table 4 summarises primary studies which focused on aspects of cyber resilience across several different smart sectors, while a summary of primary studies with a focus on DFIR's key stages in smart cities is shown in Table 7. However, despite the importance of cyber resilience and support for DFIR in smart cities, these aspects have not been extensively considered by researchers in the context of CPS. As was already noted, Figure 24 demonstrates a different level of scientific interest in adversary type research while Figure 25 further analyses the phenomena and presents the gaps across specific smart sectors. Furthermore, summarised in Figure 26 and Figure 27, the analysis revealed differences in scientific interest in the DFIR stages with further variations across smart sectors. This poses an important question as to the reason for those differences. However, it is not the aim of this SLR to provide the answer but to identify the gaps and present some open challenges and findings that can be used in future research directions [68, 230].

Concerning RQ1, during the primary studies' selection process, the researcher observed the availability of studies related to CPS applications. Within those studies, aspects of security may have been mentioned but they were not the focus of the study and often cyber defence was omitted altogether [231]. Moreover, although CPS proliferate many aspects of modern lives and the demand and need for resilience in CPS increases [232], the analysis revealed a distinct lack of available empirical research focused on cyber resilience in the smart healthcare and smart

citizen sectors Figure 23. Apart from the complex and diverse ethical challenges including privacy and confidentiality concerns [233], possible reasons for disparity include the maturity of the Industry 4.0 technology compared with the smart sectors summarised in Figure 21 [163, 164]. Moreover, the scale of media coverage of attacks on CNI like the cyberattack on the Ukrainian power grid [65] or Stuxnet [64] could also contribute to the prominence of the research in those sectors. Despite this prominence, validation of the proposed solutions requires carrying out cyber-attacks or otherwise adversely impacting the infrastructure. Therefore, any validation requires a strictly controlled environment to avoid accidental disruption or damage. Particularly in ICS, building this kind of industrial capability can be economically demanding. Beyond reproducibility of research, suitably controlled environments are required to deal with the unpredictability of real-world challenges. This requires a realistic environment often involving physical infrastructure that represents prevalent adversary challenges [25, 68, 72, 190, 196, 197, 213, 220]. In their current state, mainstream systems may not be equipped with the infrastructure to facilitate such testing and would require significant change. Without such capability, it is difficult to understand new cyber risks and find effective methods to defend against modern adversaries.

In addition to challenges accessing infrastructure-based simulators or testing in a production environment, there is a lack of publicly accessible up-to-date datasets, as illustrated in Figure 28. The following study [231] stressed the need for access to public data to enable the successful adoption of technological innovations. To validate Industry 4.0-based proposals, the following study [2] relied on a combination of datasets. The limitation of the dataset used by reference [207] covering malicious

IoT devices is the use of the CAIDA darknet datasets which predominantly contain malicious material. Based on the results, the research community appears to lean on software-based simulation using established platforms, predominantly Matlab [181, 188, 200, 201, 214], but researchers also utilise UPPAAL [201] and ProModel Process [188] simulators. Therefore, software-based simulations are a frequent choice to test experimental concepts. However, using software-based simulations may not be most suitable in some cases. For example, in smart mobility scenarios involving driving where reactions could be very different in a simulated environment knowing that a simulator can be restarted with a click of a button compared to a non-simulated experiment. This may have profound consequences for the required acceleration of research on cyber defence of CPS within smart cities since there is reliance on simulators for sufficient presentation of threats compared to reliable decision-making in a real-world environment. Therefore, up-to-date real-world datasets are vital for researchers. They are valuable since they include nuances generated from real-world applications including exposure to current adversary challenges. They help researchers validate security processes and ML techniques against challenges based on current cyber threat intelligence for detection and classification of threats in CPS. Availability of real-world data, particularly concerning ICS, is challenging and a barrier for the research community to advance security research.

Infrastructure in smart cities consists of a growing number of highly integrated CPS including traditional devices or entire cities retrofitted with new technologies to facilitate IoT connectivity [4, 150, 152]. Concerning RQ2, these devices contribute very little to support a systematic DFIR process in smart cities. Therefore, there is a

need to develop a process-driven DFIR to deal with the evolving cyber threat landscape, the expanded attack surface and the attack vector introduced through IoT connectivity [30, 35]. Furthermore, as the sources of evidence evolve, DE is contained within the physical artefacts [56]. For example, image-based evidence can be gained through Closed-Circuit Television (CCTV) surveillance or from social media. Behavioural anomaly detection can be used to detect unauthorised vehicle use through driver profiling [234], detect attacks on smart water systems [184] or unauthorised access within smart workplaces [6].

DE, similar to physical evidence, seized at a crime scene or following a security incident, is relevant during DF investigations [159]. The majority of the primary studies have researched a subset of an IR process, predominantly focusing on the "detection and analysis" phase Figure 27 of an incident utilising different approaches including profile detection, behavioural anomaly, system monitoring or audit analysis [68, 69, 157, 179, 180, 183, 184, 189, 202, 205, 206, 209]. While incidents detection is a reactive activity by nature, it is a key enabler for subsequent DF processes, which cannot occur without detection and identification of an incident. However, leaning on Locard's theory, contact between items causes an exchange. Without CPS-specific support for modern DFIR, a forensic investigation from a complex interconnected cyber-physical environment may not extract DE appropriately. Therefore, the important artefacts gathered during the acquisition stage may not be admissible in a court of law because the validity and integrity of the DE are not appropriately maintained. Best practice guides are published—within the UK jurisdiction, the Association of Chief Police Officers (ACPO) [235] and, in the US, with the Best Practices for Seizing Electronic Evidence [236] to support incident practitioners.

The authors of the following study [41] argue that in some smart sectors such as smart homes, the application of digital forensics is an emerging field of study and asserts that there is a distinct lack of formal methodologies addressing the application of digital forensics in incident responses. Furthermore, recent studies show that the integration of CPS in smart cities would significantly benefit from a specific forensic methodology as part of forensic preparedness to deal with security incidents [41, 42]. However, a lack of consensus and formal process models in the DF field that can be used to determine the reliability of DE in courts is argued by reference [237]. Finally, the increasing integration of technology into modern lives and the breadth of digital technologies exploitable by criminals requires extensive research to develop appropriate frameworks.

Concerning RQ3, the significance of the primary studies investigated is that despite the transition from traditional to IoT-enabled environments, the research findings show limited evidence of cross-sector proposals or applications for improving DF. The authors of [30] claim that there is little evidence of cross-organisational information security sharing, structure and coordination. Considering this assertion within the context of CPS, although researchers recognise the lack of shared practice, efforts are made to expand and improve cyber defence often as part of their future research direction. However, the various attempts to improve the ability to withstand targeted attacks [182] remain within a smart sector; for example, discussions are initiated between groups such as the control and security practitioners but very few studies exploit the idea of cross-sector efforts to improve digital forensics. For example, the authors of [25] consider their underlying idea

applicable to multiple smart sectors which indicates recognition of more integrated approaches. The proposal of the authors of [72, 215] was to increase the flexibility and application of their system in several different environments. Generally, the explored research focused on developing and improving cyber defences within a single smart sector. Embedded throughout this thesis is a fundamental motivation to explore transferrable solutions that emerge from smart sectors and other disciplines contributing to the scientific community to improve cyber defence in CPS. Likewise, this research acknowledges the significance of generalisation and applicability of the framework and related models to contribute to advance research across other smart sectors.

In summary, this research draws on the results of the extensive SLR process, presents and discusses the outcomes of the findings. The extensive review uncovered several gaps which could provide the basis and create opportunities for future research. These include a lack of industrial capability to produce real-world datasets to develop and validate cyber defence techniques to defend against modern adversaries and a lack of up-to-date publicly accessible datasets.

3. Chapter: Cyber Resilience in Industrial Control Systems

3.1 Industrial Control Systems Architecture Components

As illustrated in Figure 33, ICS consist of collections of PLC, HMI and SCADA subsystems that control field devices or equipment through the monitoring and processing of data associated with industrial processes [238, 239]. Centralised data acquisition and control are vital for maintaining the operational efficiency of the ICS [14]. Utilities such as electric grids, power plants and water systems, chemical plants, pipelines, manufacturing, transportation, and other physical processes are supervised and controlled by SCADA systems [240]. They are commonly used in critical industrial infrastructures. While ICS encapsulates heterogeneous hardware and software tools to support industrial operations, ICS were not traditionally designed to sustain networking-enabled applications [241]. SCADA systems have evolved from first-generation large systems to dispersed systems of the second generation that are based on exclusive network technologies. This evolution guided the development of the third-generation, fully networked modern systems that leverage the use of internet technologies [242]. In modern environments, most of the ICS components are connected to telecommunications networks and interact with the internet [243]. Today's ICS are fully integrated with the fourth industrial revolution technologies such as big data analytics, IoT, cloud computing, robotics, mobile computing and AI to fulfil industry requirements.

Figure 33 shows the basic hierarchy and ICS architecture components, which are classified into three layers and five distinct levels. The Physical Control Layer

consists of the Field and Control Levels. The Field Level contains the field instruments and is the lowest level of the control stack hierarchy. This level includes sensors, pumps and actuators that are directly connected to the plant or equipment. They generate the data that will be used by the other levels to supervise and control the processes. The Control Level uses PLCs. They are adapted industrial digital computers that control the manufacturing processes. PLCs link the field instruments with the SCADA host software using a communication network. SCADA operates at the Logical Control Layer which consists of the Supervision and Management layers. SCADA monitors, maintains and engineers processes and instruments whereas the Manufacturing Execution System (MES) is responsible for process scheduling, material handling, maintenance, and inventory. The Corporate Network Layer and the corresponding Enterprise Level is the top level of industrial automation which manages the whole control or automation system. This level utilises Enterprise Resource Planning (ERP) systems for commercial activities including production planning, customer and market analysis, orders and sales. Furthermore, as illustrated in Figure 33, ICS are automated, they require expert engineering knowledge, real-time processing and dealing with deterministic data patterns. These systems are designed for safety, reliability and availability. They have not been developed around the traditional CIA cybersecurity pillars [5].

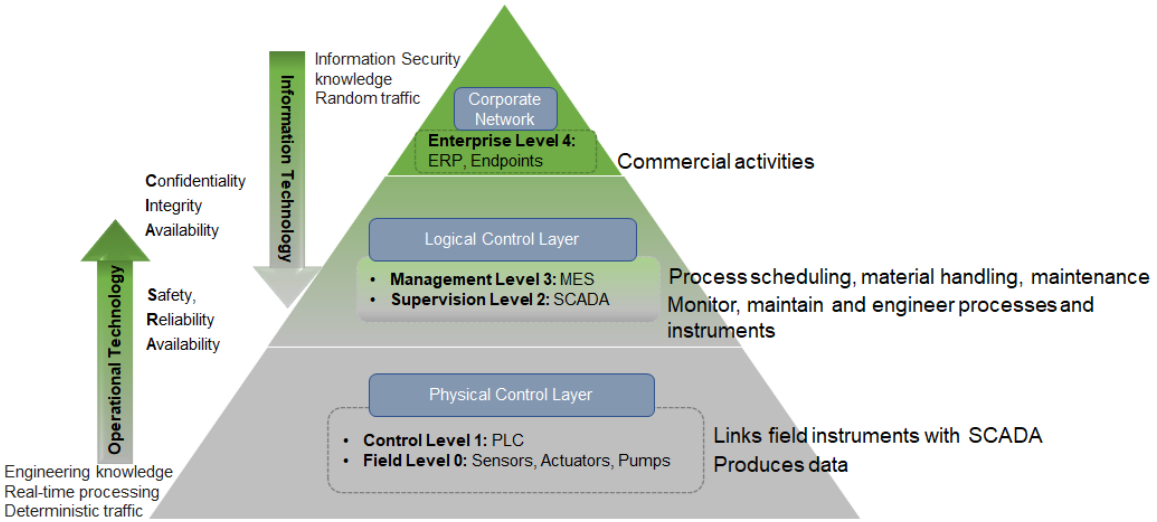


Figure 33 Components and architecture of Industrial Control Systems

3.2 Inherent and Emerging Threats

Ubiquitous sensor networks are transformational to the operations of ICS. They are integral segments of smart cities due to the level of control and intelligence gained from their sensing, processing, and communication capabilities. Broadly, CPS are subject to cyber-attacks such as targeting authentication through compromised key attacks [244], compromising the confidentiality and integrity of CPS data by targeting the CPS data storage, communication channels, actuators' controls and end-points [245]. Threats specific to ICS are often more basic such as outdated security measures. For example, in brownfield implementations where legacy systems coexist with innovative sensing technology, equipment is exposed through vulnerabilities resulting from outdated security updates. A false sense of security is provided by securing physical aspects of CPS while wireless and remote connectivity surpasses the physical boundaries. Poor configuration, lack of appropriate network segregation, compromised credentials targeting cloud-based ICS systems, backdoors, remote access channels, software vulnerabilities and smart-cyber insiders are attractive attack vectors for threat actors [7]. However,

compared with the well-established field of IS, ICS security is a less well-understood discipline and the attacks remain poorly described [18].

3.3 Critical Infrastructure and Major Attacks on ICS

Notably, the number of widely acknowledged and reported high-profile attacks on CNI is limited, see Figure 34 and related Table 37. For example, Solar Sunrise 1998 was one of the earliest multi-stage cyber-attacks against critical infrastructure which systematically exploited a vulnerability in the Sun Solaris operating systems targeting the United States (US) Department of Defence (DoD) networks [39, 246]. Another substantial incident was Stuxnet [64] where a malware attack targeted an Iranian nuclear plant. Norsk Hydro a renewable energy supplier was targeted by the LockerGoga ransomware [40]. The attack on the Ukrainian power grid compromised the SCADA system [65]. The attack on the Kemuri Water Company compromised the sensors monitoring the plant and the levels of chemicals in the water treatment plant were altered [61]. In the recent attack against Florida's Oldsmar water treatment facility, the attackers briefly increased the amount of sodium hydroxide a hundredfold. The chemical is the main ingredient in drain cleaners. The facility supplies water to commercial establishments and about fifteen thousand residents. This attack could have had profound consequences on the community [247]. Interconnected systems are subject to attacks and it may not be possible to establish the source or the motive [39, 248]. Thus, it is critical to establish an intelligence-based defence-in-depth mechanism and understand the threat models posed against ICS.

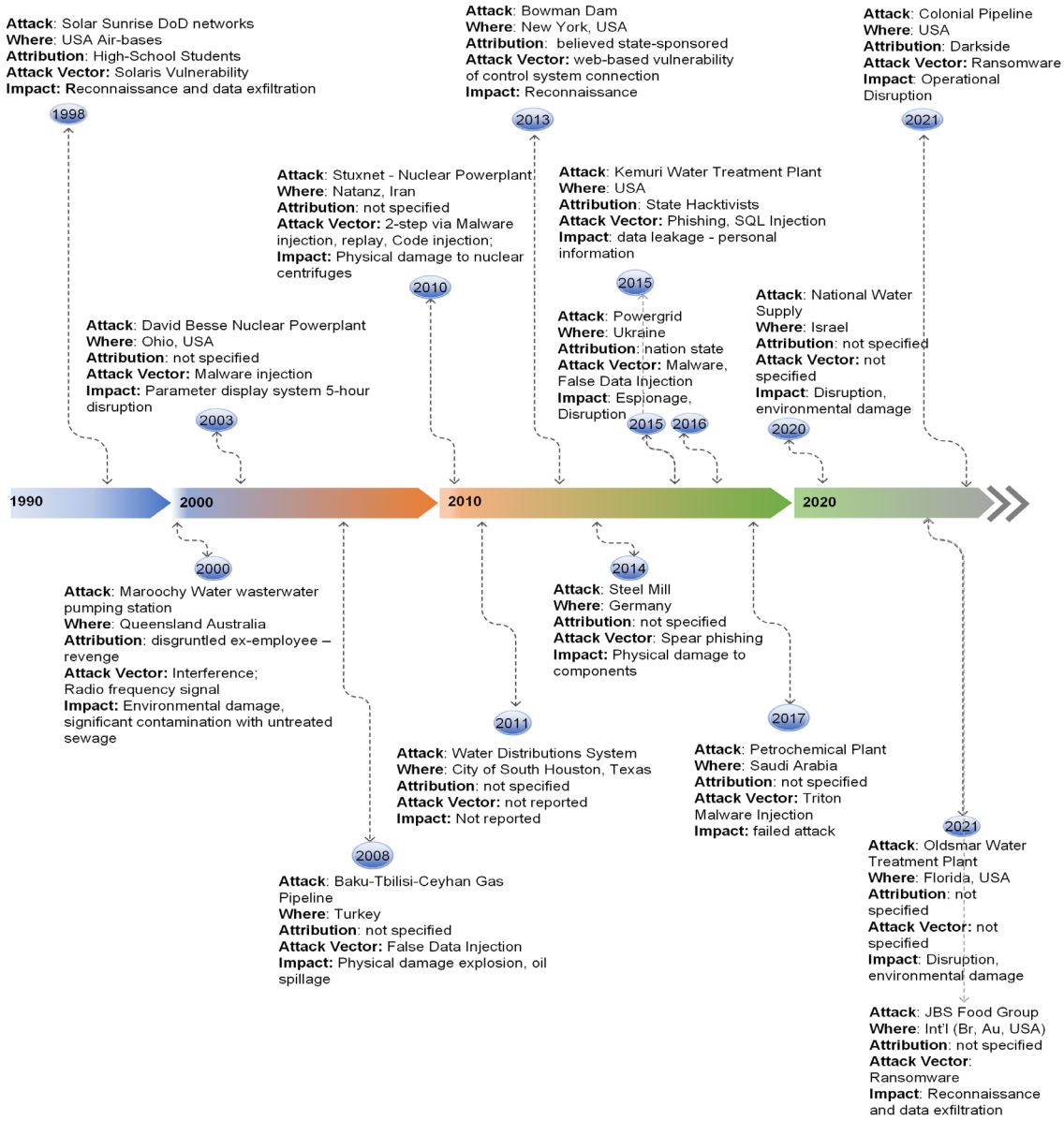


Figure 34 Timeline of reported high-profile attacks on ICS for a period between 1990 and February 2021.

3.4 Threat Modelling

This research study considers ICS-related cybersecurity attacks discussed in the literature and covered in the previous section [5, 26-31, 39, 40, 61, 64, 65, 246, 247] in addition to those listed in Table 15. This research asserts that ICS are integral components of smart cities and fundamental to the operation of industrial facilities. As acknowledged throughout this thesis that although ICS were originally designed

as isolated systems with separation of the ICT and OT environments, modern ICS are complex, distributed and interdependent [19]. The integration of disruptive technologies makes ICS an attractive target for adversaries with a substantial attack surface and attack vectors. Against this backdrop, the number of widely acknowledged attacks against ICS and CNI remains limited. Besides the prevalence of attacks, research highlights the differences between the security and operational priorities of OT and ICT systems with increasing challenges to protecting ICS [18, 19]. Furthermore, existing literature acknowledges that ICS are vulnerable to conventional IT and specific OT threats with potentially devastating consequences to the wider society [18, 19, 249]. Thus, effective countermeasures and incident response methods are required as part of layered security.

Besides threats from external adversaries, the literature indicates the prevalence of insider threats [249]. Social challenges such as accidental insiders, disgruntled employees and social engineering are underestimated and difficult to detect. These challenges fall outside of traditional cyber defence measures such as firewalls, access control, network, and host security. This research study outlined the threats and consideration for the attack vectors, for example:

- An attacker could gain access to the logic and the physical control layers. Code injection could alter the sensor values creating discrepancies between the PLC registered values and the actual state of the physical process. Likewise, command injection could compromise the actuators and create discrepancies between the expected and registered state. This thesis considers attacks on the operational infrastructure at the physical control layer.

- An employee or a supply chain contractor with legitimate and unmonitored access, knowledge of the industrial equipment, related software systems and data configurations. A scenario could evolve where access to the computer systems could result in intended or accidental manipulation leading to critical service disruption or physical damage. The attack vector is exploitable locally or remotely. Moreover, physical access to operational infrastructure and tampering with the physical process in the system could lead to physical damage and alteration to the expected functioning of the operational infrastructure.
- Emerging attack vectors due to introducing open standards and interconnectivity with enterprise and public networks in ICS create new opportunities for attackers [7, 24, 25, 244, 245]. Hostile and sophisticated threat actors such as APT will adapt their tactics, techniques and procedures to exploit these opportunities [5].
- A compromised device on the enterprise network or an ICS component exposed to the internet could be exploited as an initial attack vector by an attacker [24]. Apart from unpatched or zero-day vulnerabilities, attackers could gain access to operational infrastructure by exploiting legitimate account credentials and poor security controls [250].
- Likewise, open standards and interconnectivity with corporate and public networks in ICS create new attack vectors [7, 24, 25, 244, 245]. Resourceful attack actors such as APT will adapt their tactics, techniques and procedures to exploit these opportunities [5]. The initial attack vectors could include the attacker's ability to compromise a device on the corporate network or an internet-exposed ICS component leveraging unpatched or a zero-day

vulnerability [24]. Exploiting legitimate account credentials coupled with poorly designed or bypassed security controls, could enable the attacker to gain access to the ICS operational infrastructure [250].

Modern APT actors strategically and persistently adapt and evolve their tactics techniques and procedures to compromise technology and exploit new opportunities to achieve their aims. APT attacks follow a multi-stage and multipath attack process ranging from reconnaissance to achieving the attackers' strategic goal. This research assumes an attack vector where the attacker gains access to the logical control and the physical control layers. Thus, pertinent to this thesis is to consider attacks on the operational infrastructure at the physical control layer listed in Table 15 and Table 26. The datasets comprise operational scenarios of anomalies and malicious acts such as accident, sabotage, breakdown and cyber-attack of variable duration affecting sensors, the network and the subsystem. They highlight consequences of attacks affecting the network and the physical processes. And therefore, while APTs could exploit any of the layers illustrated in Figure 33 as an entry attack vector, the datasets utilised by this thesis represent attacks the could be exploited as part of a multi-stage attack such as by an APT adversary to impair process controls, prevent response and otherwise disrupt ICS environments.

4. Chapter: SPEAR - Super Learner Ensemble for Anomaly Detection Framework

ICS are integral parts of smart cities and critical to modern societies. Despite indisputable opportunities introduced by disruptor technologies, they proliferate the cybersecurity threat landscape, which is increasingly more hostile. The quantum of sensors utilised by ICS aided by AI enables data collection capabilities to facilitate automation, process streamlining and cost reduction. However, apart from operational use, the sensors-generated data combined with AI can be innovatively utilised to model anomalous behaviour as part of layered security to increase resilience to cyber-attacks. This chapter introduces a framework to profile anomalous behaviour in ICS and derive a cyber risk score. Firstly, in this chapter, the research focuses on anomalous behaviour detection methodology. Following that, the approach is experimentally validated by utilising an ICS liquid distribution case study. Finally, Chapter 5 is dedicated to the cyber risk quantification model.

4.1 Data Mining Methodology

The Cross-Industry Standard Process for Data Mining, generally known as the CRISP-DM process model Figure 35 inspires the approach for the framework's methodology introduced in this chapter. CRISP-DM is an open standard providing a clear model for analysis. The authors **[**251**]** highlight that CRISP-DM is the de-facto standard and an industry-independent process model applicable to data mining projects. The CRIPS-DM process model consists of 6 distinct stages, as illustrated in Figure 35.

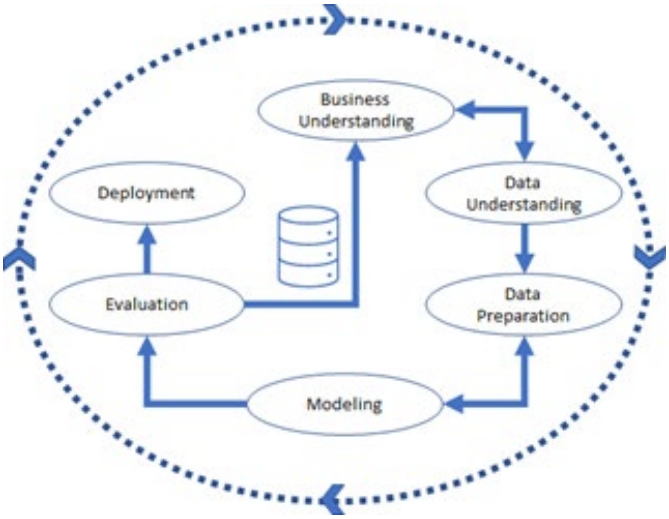


Figure 35 Schematic diagram representation of the CRISP-DM process model

4.2 The SPEAR Framework Overview

This chapter presents the SPEAR Framework shown in Figure 36, to facilitate proactive anomalous behaviour detection in ICS. The approach is motivated by a study in the field of genetics and molecular biology [252]. The authors construct a fast learner using a weighted combination of several candidates and utilise V-fold cross-validation to avoid overfitting. Their approach is aimed to generalise to any parameter. As part of the framework in this chapter, leveraging supervised learning a super learner ensemble is constructed. Using overlapping RW, this research derives the best predictor for anomalous behaviour detection for the datasets. In this study, the approach differs from other studies such as [92, 93, 252]. This study does not rely on a single classification model for the base learners [46], it uses a stack of base learners, overlapping RW and apply stratified k-fold n-repeat Cross-Validation (CV) to each base learner [45] at the time of training the model. The choice of the best learner in the stack is based on majority voting.

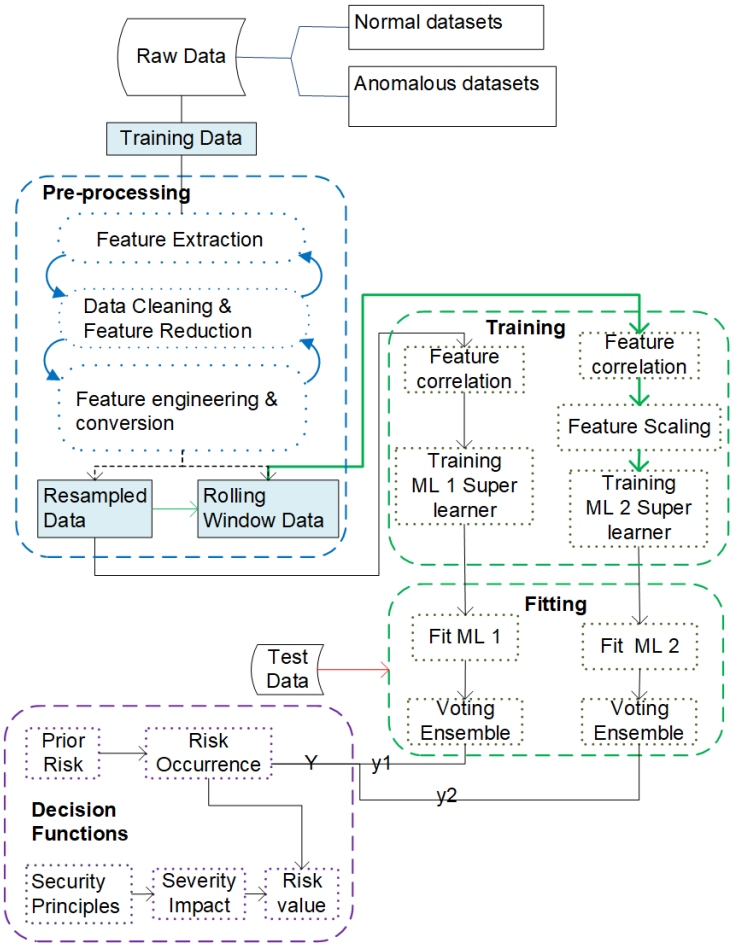


Figure 36 SPEAR Framework, which consists of the data pre-processing, model training, model fitting and decision function stages.

4.3 Procedure Design

Firstly, during the pre-processing phase, the temporal dataset is transformed, and features are extracted to solve the problem as a supervised model as shown in Figure 37. Contextual features contained within the date and timestamp are introduced including a feature to represent the elapsed time from the beginning of the event. This research thesis does not seek to establish the date and time of the events. Its interest is to uncover a behavioural anomaly as a temporal event in a sequence of events. Next, irrelevant, missing, or duplicated feature values or instances could skew the learning algorithm's performance. Such features are addressed during the data cleaning phase utilising several data cleaning techniques.

Feature engineering introduces additional features which contribute to the learning model's performance.

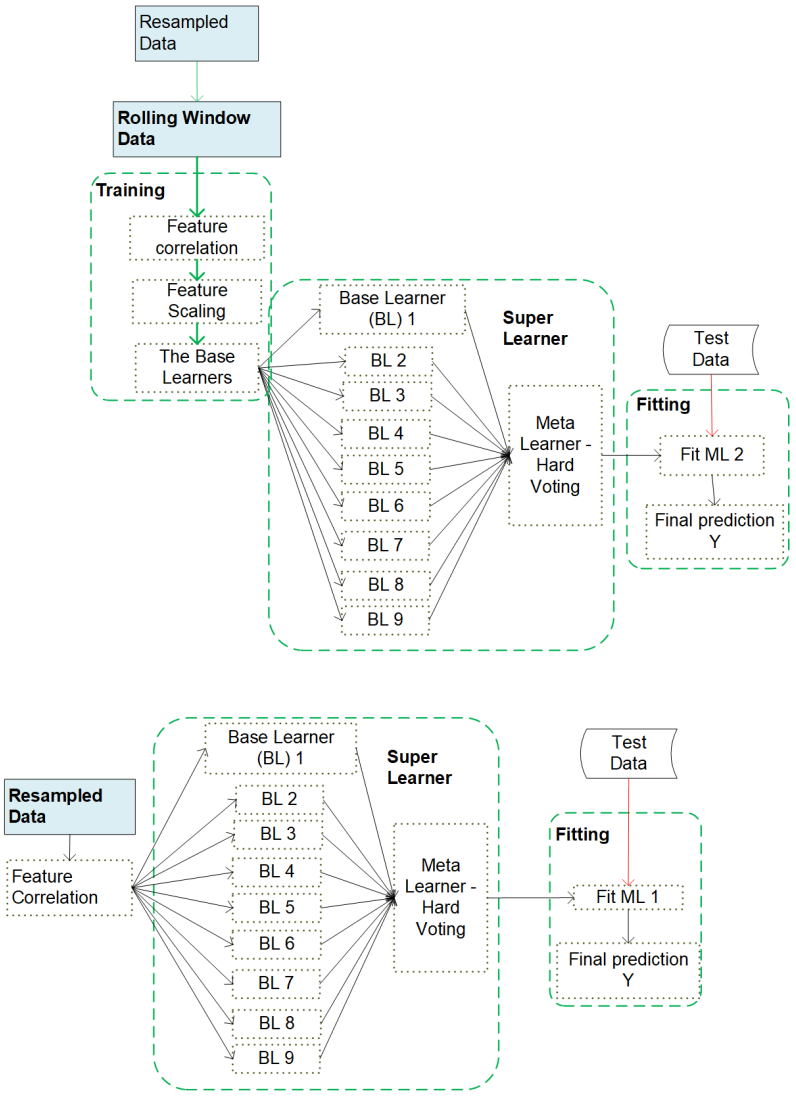


Figure 37 Supervised ML, super learner ensemble model.

To evaluate the performance of the ML model, consideration was given to the train, test and validation data subsets. The anomalous behaviour varies from a few seconds to several minutes as presented in Table 15. Therefore, consideration to handle imbalanced datasets is factored in. For the one-class binary classification for the outlier detection, the normal and anomalous datasets are combined into a single dataset. The base learners are trained on random subsets of the total training data

and are fitted with test data. The same approach is applied to individual attacks, creating a set of imbalanced datasets. The stratify parameter is used to retain the train-test split ratio for the train and test sets, setting aside 30% of the dataset for testing. Grid search is utilised for hyperparameter optimisation. To avoid overfitting or significantly reducing the number of samples in the train or test sets, repeated stratified 10-fold, 3-repeat CV is applied to the base learners during the model training. The meta-learner is trained from the outputs of the sub-models utilising a list of defined estimators from the stack as input arguments. Majority voting ensemble $\hat{y} = mode\{ \lambda_{b1}(x), \ldots \lambda_{b1n}(x)\}$ is applied before the final prediction is produced as illustrated in Figure 37 [45, 46]. The performance scores are derived from the confusion matrix, see Figure 38.

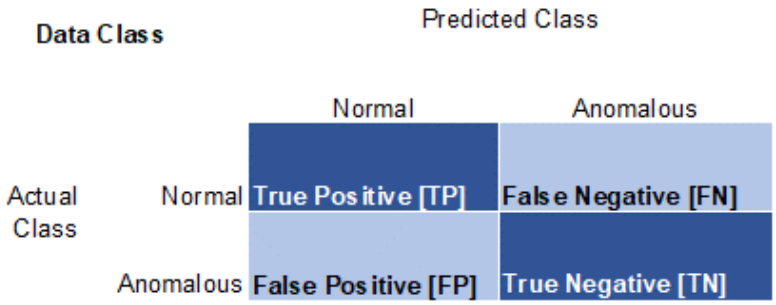


Figure 38 Confusion Matrix

4.4 Pilot Experimentation

Pilot experimental work focused on a small subset of the individual attack scenarios. Pilot experimentation helped to test and evaluate the instruments, the procedure and the formal experiment's optimal time window.

4.5 The SPEAR Framework Data Preparation

4.5.1 Feature Extraction

The timestamp feature is rearranged to the 'Date' format [dd/mm/yyyy hh:mm:ss.sss]. The single 'Register' feature has all sensor types as the feature's values. The dataset features are rearranged according to the algorithm in Table 8, such that each sensor type is represented as an individual register feature $f_{v1}...f_{vn}$ labelled *R1...Ri* in time series. Feature extraction by sensor type separates sensors by their functions while their time segments are unchanged. Furthermore, this feature extraction enables the grouping of different sensor types into learning sub-models which is an interesting approach, similar to another study [25].

Table 8 Algorithm 1: SPEAR Framework Feature Extraction Algorithm

Input raw dataset of instances $i_1...i_n$ with features $f_{r1}...f_m$, of values $[v_1....v_n]$
Output labelled dataset of instances $i_1...i_n$ and class (normal data [0], anomalous data [1]), with features $f_{v1}...f_{vn}$ of value $[v_1....v_n]$
Step 1: Load raw dataset into dataframe
for f_r identify unique values
extract v into f_v using index 'Date'
label Class for $i_1....i_n$
end for

4.5.2 Data Cleaning

Recording of sensor readings may become corrupted or erroneous for several reasons during the data collection process such as malfunctioning sensors, malicious activity, disruptions in network connectivity or the data collection infrastructure. This could result in noisy, missing, or duplicated observations within the dataset, in real-world data and large datasets the likelihood of erroneous data increases. Therefore, data cleaning is essential for a meaningful analysis of the dataset which is handled according to the algorithm in Table 9. This process identifies missing values, duplicate instances, unique feature values, single-value and low-variance features. In this dataset, features that have a single value or very few unique values, and have zero or low variance of <=0.001% are not likely to contribute to the predictive model's performance. Therefore, such features are removed from the

dataset. Missing values are often marked with a placeholder such as *'NaN'* or left blank. However, not all algorithms have the resilience to deal with missing values, particularly predictive techniques [253]. To minimise the loss of data, missing values are marked. Instances are dropped where missing values <=0.5% of the dataset, otherwise, values would be imputed using the forward-fill method propagating the last observed non-null value forward until the next non-null value is reached.

Table 9 Algorithm 2: SPEAR Framework Data Cleaning and Feature Reduction

Input raw dataset of features $f_{v1}...f_{vn}$, values $[v_1....v_n]$ and instances $i_1..i_n$
Output cleaned dataset with feature-set $f_{vc1}...f_{vcn}$ of value $[v_1....v_n]$ and instances $i_{c1}...i_{cn}$

Step 1: identify missing f_v
  for i replace missing f_v[v==NaN]
    count f_v[v==NaN]
    print summary of missing f_v
    if missing f_v[v <0.5%]
      remove missing instances
    else
      impute missing f_v[v, impute method == ffill]
       verify missing f_v[v]
    then go to step 2
  end for
Step2: identify duplicate instances i_d
  for i:
    calculate i_d
    remove i_d
    then go to step 3
  end for
Step 3: Identify features with single value, few values and near-zero variance predictors:
  for i in range f_v[v]:
    print f_v[v], len==unique
    if len==unique where (unique f_v[v]/total i*100)<=0.001%)
      drop f_v
    else
      cleaned dataset of feature-set $f_{vc1}...f_{vcn}$ of value $[v_1....v_n]$ and instances $i_{c1}...i_{cn}$
  end for

4.5.3 Feature Engineering and Visualisation

This part of the pre-processing phase introduces additional features to the dataset, according to the algorithm in Table 10. To apply an ML algorithm to train the dataset, the time-series dataset is transformed, so that it can be modelled as a supervised problem. Contextual features based on the date and timestamp are introduced. While information about business hours, public holidays, years' seasons,

part of the week could be extracted and enhance the performance of a learning algorithm, in this dataset using the date would not likely help the learning algorithm and could result in inferior performance. The dataset is resampled using seconds as the smallest time unit, and the mean values for each sensor using the default label bucket and bin interval values. Furthermore, the seasonality and the trend characteristics of the discrete, pumps and ultrasound sensors in the dataset were established. The normal dataset's repeatable patterns are shown in Figure 39.

Table 10 Algorithm 3: Feature Engineering for the SPEAR Framework

Input: cleaned dataset of features $f_{vc1}...f_{vcn}$ of value $[v_1....v_n]$ and instances $i_{c1}...i_{cn}$, index [Date]
Output: pre-processed dataset with features $f_1...f_n$, values $[v_1....v_n]$ and instances $i_1...i_n$
for i
set index f_{vc}==datetime ['%d/%m/%Y%H:%M:%S.%f']
transform f_{vc} datetime to new features where 'second'[v==%S], 'minute'[v==%M], 'hour'[v==%H]
then resample $i_{c1}....i_{cn}$, f index (v==datetime [%d/%m/%Y %H:%M:%S']), $f_1..f_n$ ([v==mean])
end for
for i of sensors f_{vc}
do kpss stationarity test
end for
Apply Algorithm 2 Step1
for i
apply rolling window (interval[s], min_periods, win_type, mean)
end for
reset index
drop f_{vc} 'datetime'

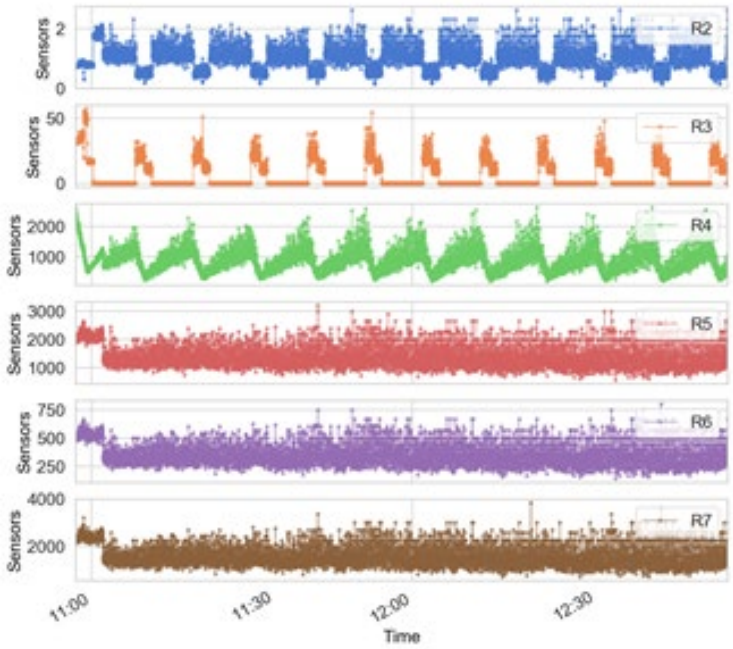


Figure 39 Sensors' temporal distribution – normal dataset.

The test for a null hypothesis whether the dataset is stationary, the statistical Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [254] for stationarity around a deterministic trend is applied to the sensors. The probability score, the p-value of the test is >0.05 (significance level), the KPSS null hypotheses not rejected, and showing the sensors' stationarity around a constant as shown in Table 11. The equation (3.20) is used to calculate the lags [254], where *'n'* represents the length of the series:

$$\text{int}\left(12 * \left(\frac{n}{100} ** \left(\frac{1}{4}\right)\right)\right) \tag{3.20}$$

No other environmental information is introduced. Further analysis of the dataset uses the seasonal data decomposition function to verify the stationarity around a deterministic trend and decomposes the data into four components: level, trend, seasonality, and noise. The components are structured as outlined in equation (3.21), where *'y(t)'* is the time series dataset over time, Level (L), Trend (T), Seasonality (S), Noise (N):

$$y(t) = L + T + S + N \tag{3.21}$$

Furthermore, 3s, 5s, and 10s size rolling mean windows and parameters including minimum period and window types are utilised to test and evaluate the model's performance.

Table 11 KPSS test output stationarity test – normal dataset

Test Output	**Sensor Output Values**		
Sensors	**Discreet**	**Pumps**	**Ultrasound**
Test Statistics:	0.142124119400179 58	0.21469603458547087	0.13686452509501715
p-value:	0.1		
Critical Values:	'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739		
num lags:	36		
Stationarity	Series is Stationary		

4.6 The SPEAR Framework Learning Algorithms Modelling

This section introduces the proposed detection scheme presenting two ML algorithms for the framework.

4.6.1 Supervised Learning Modelling

In supervised learning, the model uses the concept of a super-learner ensemble for classification algorithms for anomalous behaviour identification in CPS. This model consists of nine stacked base learners. The base learners are typically investigated independently to gain the best performance based on the optimal set of features and classifiers. The aim is to avoid selecting a suboptimal classifier to solve the problem, improve the predictive performance and increase the generalisation performance of the algorithm. The learning algorithm, as shown in Table 12, is based on the general framework of several ensemble algorithms [45]. Scientific studies accept that meta-learners may not produce better results than any of the classifiers used individually, nonetheless, their use mitigates the risk of using an inefficient classifier [46, 113].

The learning model is trained and tuned using resampling and resampling with rolling windows techniques. The stacked base learners are trained on random subsets of the total training data, they are fitted with test data and produce accuracy scores. The meta-learner is a heterogenous ensemble derived from the base learners consisting of different algorithms. The meta-learner is trained from the base learners' outputs, utilising a list of defined estimators from the stack as input arguments. The meta-learner applies the majority voting method before the final prediction is produced, see Figure 37 [45, 46]. The labelled dataset uses the stratify parameter to retain the train-test split ratio and is split into train-test sets, setting aside

30% of the dataset for testing. A dictionary of parameter values is defined for hyperparameter optimisation and uses a grid-search technique to determine the best parameter set. To avoid overfitting or significantly reducing the number of samples in the train or test sets, repeated stratified 10-fold, 3-repeat CV is applied. The two models are trained independently, applying 1s resampling and a 10s rolling window to the dataset, as shown in Figure 37.

Table 12 Algorithm 4: Supervised Learning Ensemble Super Learner for the SPEAR Framework based on the general framework of ensemble algorithms [45].

Input: Pre-processed dataset $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$,
base learners algorithm $\lambda_{t1}, \ldots \lambda_{tn}$, meta-learner algorithm λ
Output: $H(x)$
Other Definitions: h_t =base learner, T = Number of learning algorithms, h' = meta-learner

#Train the base learners by applying the base learner learning #algorithms to the pre-processed original training dataset
for t in $(t_1 \ldots t_T)$:
    $h_t = \lambda_t(D)$;
end for
#Produce a new dataset for training the meta-learner,
#The output of the base learners is the input for the meta-learner
#Original labels are retained
D' = Θ;
for i in $(i_1 \ldots i_n)$:
    for t in $(t_1 \ldots t_T)$:
        $\mathbf{z}_{it} = h_t(x_i)$;
    end for
  #The new dataset is produced from the cross-validated the total
  #number of base learners. The meta-learner is applied where h' will #become the function of $z_{it1}, \ldots z_{itT}$ for y.
    D' = D' ∪ $((z_{it1}, \ldots z_{itT}), y_i)^{in}_{i=1}$;
end for
#train the meta-learner h' by applying the meta learner algorithm λ to #the newly generated dataset D'.
  h'=λ(D');
Output: $H(x) = h'(h_{t1}(x), \ldots, h_{tT}(x))$

4.6.2 *Unsupervised Learning Modelling*

In unsupervised learning, the model covered in Figure 37 and Figure 40, uses the concept of outlier detection to identify anomalous behaviour in CPS as its main algorithm. This model uses IF which is an unsupervised ML ensemble. ML methods such as statistical, clustering or classification-based algorithms require the normal behaviour profile established first. Unlike other unsupervised ML methods, IF defines anomalies as few and different [45, 119] and uses isolation to determine anomalous

behaviour. It does not require a profile of the normal behaviour first [119] making it a fast algorithm with low demand on memory. The IF creates an ensemble of isolation trees trained on a random data subset '$d_{max\text{-}samples}$' from the main dataset '$d_{max\text{-}samples}$' $\subset$ D of the maximum number of features '$f_{max\text{-}features}$', as shown in Table 13. The IF, with several randomly created partitions, isolates the anomalies through recursive binary splitting completed by each of the created iTrees and randomly selects a split feature 'q_f' and a split value 'p_v' from the input dataset D' generating a left D_l' node and a right D_r' node until all the samples are isolated, as presented in Table 14. The splitting required for sample isolation starts at the internal root node and terminates at the external leaf node with several internal interim nodes produced if there is a possible split remaining until the maximum path depth is reached. Accepting that the anomalies are few and different, they can be isolated such that they have a shorter path. Therefore, anomalies are isolated nearer the root of the tree while normal measurements are isolated near the leaf nodes of the formed iTree. Left and right interim nodes are created at each point that a split occurs until the final external node is reached at the point which cannot create any further nodes. A density-based approach utilising Local Outlier Factor (LOF) and a distance-based approach using SVM were added to the IF algorithm to investigate performance variations of an unsupervised multi-learner ensemble model.

Table 13 Algorithm 5: IF Forest training phase of the unsupervised learning ensemble for SPEAR Framework, based on the [119]

Input: Pre-processed Dataset D = {x_1, …., x_n}, Number of tree estimators $\varepsilon_{n\text{-}estimators}$, data sub-set $d_{max\text{-}samples}$ data sub-set features $f_{max\text{-}features}$
Output: new dataset iTree D'
Initialise Forest
#for the number of trees
for i=1 to $\varepsilon_{n\text{-}estimators}$:
#The maximum number of samples which represent the data sub-set and the maximum number of features in the data sub-set to train the tree
$d_{max\text{-}samples}$ ← sample (D, $d_{max\text{-}samples}$, $f_{max\text{-}features}$)
Forest ← Forest ∪ iTree(D')
end for
return Forest

In this thesis, although setting the contamination level can be achieved by utilising subject matter expert knowledge, the labels for the dataset are known and they are used to set the contamination level *'C_o'* given the anomalous instances $i_\alpha\{1, \ldots, \alpha\}$, and normal instances $i_n\{1, \ldots, n\}$ in the dataset for the ground truth and validation of results as given by equation (3.22):

$$C_o = \frac{i_\alpha}{i_n} \tag{3.22}$$

The labels are removed and not used by the algorithm for anomalous behaviour detection.

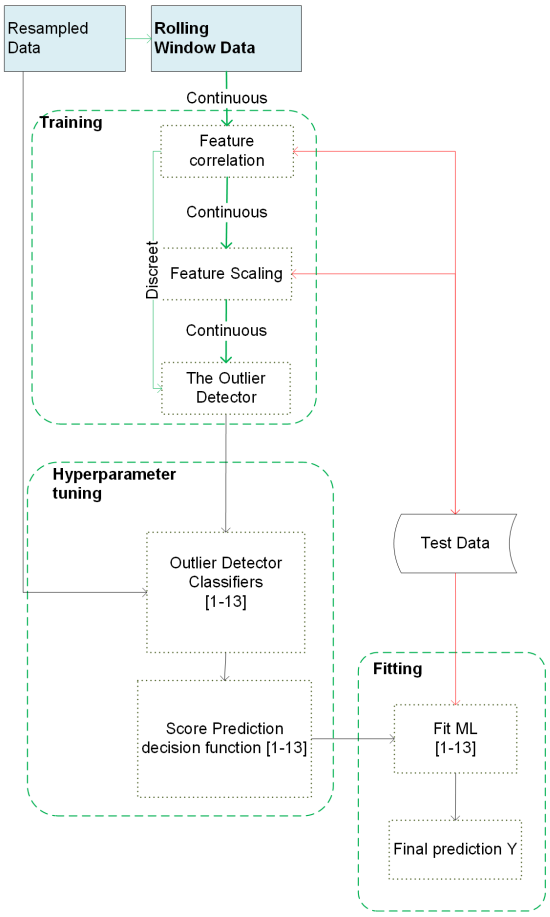


Figure 40 Unsupervised ML, multi-learner ensemble model.

Table 14 Algorithm 6: IF iTree training phase of the unsupervised learning ensemble for SPEAR Framework based on the [119]

Input: D'

Output: iTree
If D' cannot be split:
external leaf node;
elif:
let Q be D' features
randomly select $q_f \in Q$
randomly select a split p_v between the min and max of q_f in D'
$D'_l \leftarrow$ filter (D', $q_f > p_v$)
$D'_r \leftarrow$ filter (D', $q_f \leq p_v$)
return interim node {Left $\leftarrow$ iTree(D_l), Right$\leftarrow$ iTree(D_r), feature split $\leftarrow q_f$, value split $\leftarrow p_v$ }
end

4.7 Case Study: ICS Liquid Distribution Experiment Design for Piloting the SPEAR Framework

The experimental environment consisted of two liquid containers, two pumps, an ultrasound sensor, four discreet liquid level sensors, automated controls, and infrastructure for the data acquisition, as presented in Figure 41. The schematic diagram shows the main tank, the positioning of the sensors and their corresponding liquid levels. Each liquid level is coupled with the decimal representation of the value that each sensor assumes based on the PLC register's binary state. The secondary tank shows the ultrasound sensor and the depth of the liquid. The liquid depth is divided into 10,000 equal segments with 0 representing a full tank and 10,000 an empty tank. Based on the discreet and the ultrasound triggers the pumps assumed ON or OFF states alternatively or in combination. This was reflected by the values recorded in the dataset. The diagram shows the registers' Least Significant Bit (LSB), the PLC registers [R2-R7] and the dataset features allocated to the bit segment [0-15] within each PLC register. The testbed functions in manual or automated modes using a touch-screen command and remote network connectivity. The pumps were activated and deactivated depending on the liquid reaching a pre-determined level. The activation of the pump which fills the main tank depended on the ultrasound sensor values. The pumps and the registers indicate the binary state of the sensors assuming two states; an ON state represented as 1.0 and an OFF state represented

as 0.0. The dataset contains the corresponding decimal value of the sensors' binary state in the PLC register.

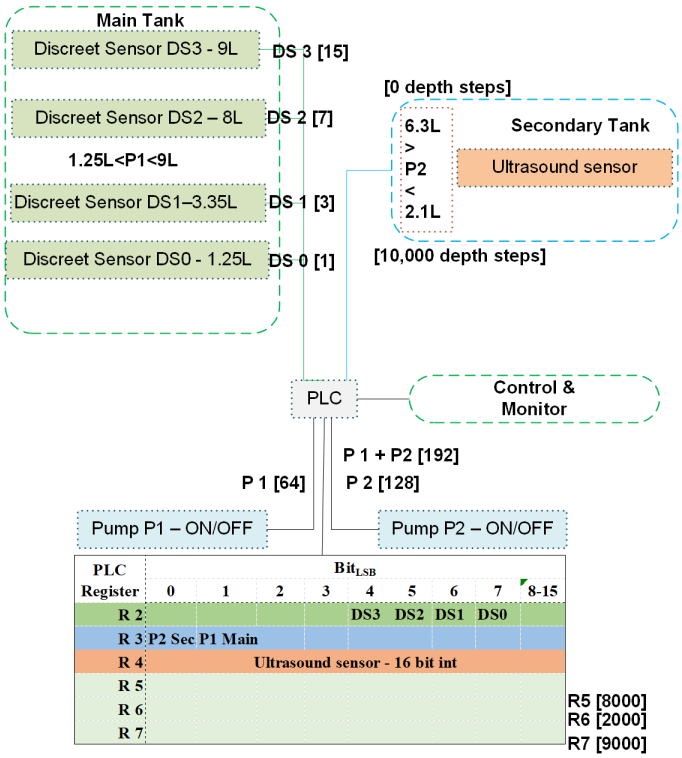


Figure 41 The 'aNomalies' testbed schematic diagram and the structure of the data log registers.

The instruments utilised in this case study for experimentation to process and analyse the collected data, and train the ML models consisted of a Jupyter Notebook scikit-learn ML library [255] and a Hewlett-Packard Envy x360 x64-based Intel® Core™ i7-8565U CPU, 4 Cores 8 logical processors 1.8Ghz, 16GB Physical and 40GB of virtual memory.

4.8 The ICS Liquid Distribution Dataset

The data used in this experiment was produced from the 'aNomalies' testbed [84]. The dataset covered five operational scenarios: normal, accident, sabotage, breakdown and cyber-attack, as shown in Table 15. The timestamp was presented

in the format of dd/mm/yyyy hh:mm:ss.sss. A read request was sent to the PLC every 100ms. The bit segment of each PLC register according to the position of the LSB held the specific sensors' values. From the total of ten registers, three PLC registers corresponded with the values recorded in the dataset. Registers one, eight, nine and ten represented no values. Register two provided the binary state of the discreet sensors, using the first four bits of the PLC register. Register three provided the binary state for the pump using the last two bits of the PLC register. Register four recorded the value of the ultrasonic sensor as a 16-bit integer. Registers 5-7 record values but it was not clear from the dataset's description what values of these registers represented.

Table 15 Files that make up the temporal dataset [84].

File	Scenario - Type	Sensors affected	Duration [hh:mm:ss]
1	Normal	None	02:01:47
2	Plastic Bag	ultrasonic	00:33:20
3	Blocked measure 1	ultrasonic	00:00:25
4	Blocked measure 2	ultrasonic	00:00:17
5	2 floating objects in the main tan	ultrasonic	00:01:35
6	7 floating objects in the main tan	ultrasonic	00:01:22
7	Humidity	ultrasonic	00:00:18
8	Failure of a discreet sensor	Discreet 1	00:13:55
9	Failure of a discreet sensor	Discreet 2	00:03:40
10	Denial of Service attack	Network	00:01:37
11	Spoofing	Network	00:34:33
12	Wrong Connection	Network	00:15:33
13	Tank hit – with low intensity	The entire system	00:00:39
14	Tank hit – with medium intensity	The entire system	00:00:32
15	Tank hit – with high intensity	The entire system	00:00:33

4.9 Experimental Results

4.9.1 Performance Metrics

To determine the best-performing classifier for anomaly detection this study derives several statistical metrics including accuracy, precision, recall and F1-score from the confusion matrix, see Figure 38 [256, 257]. These values are calculated using the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values from the classification model. The TP metric is the number of

anomalous instances in the dataset classified as anomalous. The TN metric is the number of normal instances classified as normal. These are the correctly identified instances which do not belong to the TP class. The FP are the instances that were incorrectly allocated to the class and FN are instances not recognised as part of the class, they are values of the incorrectly classified instances. Two types of errors can occur in a binary classification problem. Type I error or False Positive Rate (FPR) occurs when the model predicts a sample as anomalous when it was normal. A Type II error occurs when the model predicts anomalous instances to be normal, also known as False Negative Rate (FNR).

Accuracy is the ratio of the number of correct predictions made by the model to the total number of predictions made. Although the most instinctive measure of performance, as a ratio of correctly predicted observations to the overall observations, accuracy is not the most suitable performance measure for evaluating imbalanced datasets. Accuracy is given by the equation (4.1),

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{4.1}$$

In anomalous behaviour detection, determining the number of anomalies detected is not a sufficient measure of the performance of the classifiers due to the existence of falsely predicted values. Therefore, this study utilises Precision, also known as the Positive Predicted Value, which can be expressed as

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4.2}$$

Precision is the measure of how accurately the classifier detects an attack. This metric relates to the class agreement of the number of correctly predicted positive

instances over the number of instances labelled as positive [257]. In the case of high precision, there is a low false-positive rate.

The recall measure, also known as Sensitivity, is given by (4.3),

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{4.3}$$

Recall calculates the number of attacks detected. This metric is also referred to as the TPR and relates to the classifier's effectiveness to identify the correctly predicted positive classes over the total number of positive observations [257].

The F-measure or F1-score is a weighted average of precision and recall, providing a holistic performance evaluation measurement of the ML algorithms [257]. F1-score, given by (4.4), is an indicator of how accurately the model identifies the anomalous instances in the dataset.

$$\text{F1} = \frac{2\times(\text{Recall}\times\text{Precision})}{(\text{Recall+Precision})} \tag{4.4}$$

The Receiver Operating Characteristic Area Under the Curve (ROC AUC) metric refers to the classifier's ability to avoid misclassification, useful to evaluate the ML algorithm's performance when class distribution is unknown [257]. The AUC is defined by the area under the ROC curve and can be expressed as

$$\text{AUC} = \frac{\mathbf{1}}{\mathbf{2}}\left(\frac{\text{TP}}{\text{TP+FN}} + \frac{\text{TN}}{\text{TN+FP}}\right) \tag{4.5}$$

4.9.2 Supevised Learner Ensemble

Before training the models, Spearman's correlation coefficient was used to produce a summary of the strength between the features in the combined dataset.

Before training the model, highly correlated variables of at least 80% positive or negative correlation were removed from the dataset. Before model fitting, robust scaling standardisation was applied to tune the model. Robust scaling was a justified approach to avoid skewing the result due to the presence of instances of normal and anomalous classes in the dataset.

Despite the base learners being trained on the same training dataset, the results were produced independently. Despite a lack of a widely accepted definition of diversity [45, 257] in classifier ensembles, ensemble base learners are often complex making different assumptions about the prediction. A range of base learner classifiers was used in forming the super learner including k-NN, RF, LR, DT, Support Vector Classifier (SVC), AdaBoost Classifier (ABC), ETC, Gaussian Naïve Bayes (GNB) and Bagging Classifier (BC). The base learner algorithms were trained with the default parameters and with parameter optimisation, see Figure 42. The individual base learners do not produce a weak result, which would weaken the overall ensemble's performance. The base learners' results vary, which is likely to improve the ensemble generalisation and produce high-accuracy predictions. However, further optimisation is required, which could be an appropriate future direction to develop the model.

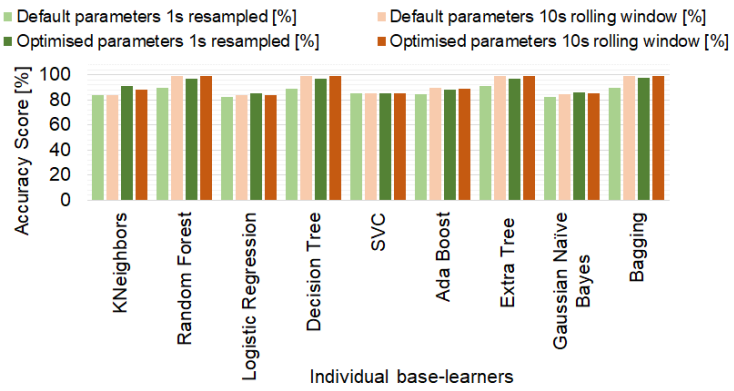


Figure 42 Individual base learners algorithms comparison at 1s intervals and 10s rolling window.

The overall performance of the models trained with a 1s resampled interval and a 10s rolling window utilising the combined dataset is covered in Figure 43. The details of the two best-performing models are presented in Figure 44.

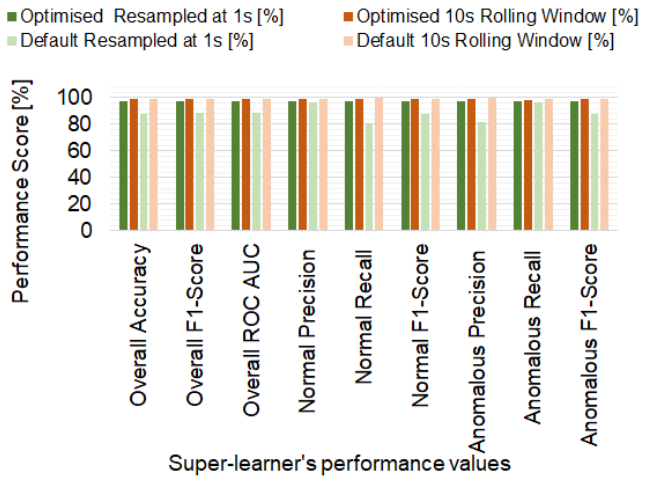


Figure 43 Overall performance of the models trained with 1s resampling and 10s rolling window.

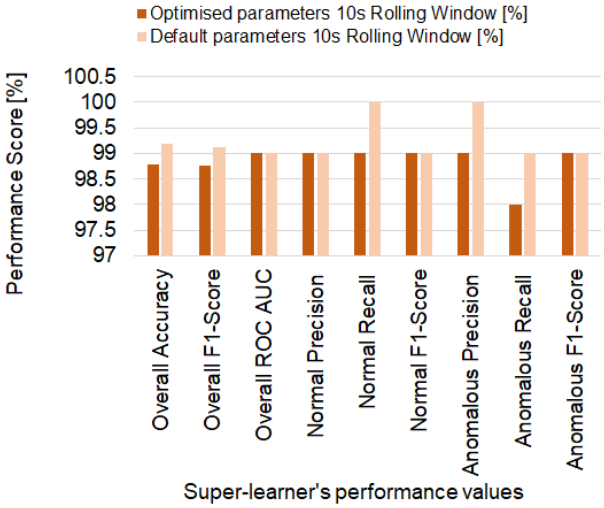


Figure 44 The overall best performing super learner models.

The difference in the performance between the two models is illustrated in Table 16. The optimisation improvement between the weakest and the best-performing model is demonstrated in Table 17. The optimisation achieved a consistent improvement in the overall F1-scores of 99.13%, an increase of 12.13% compared with the default 1s resampling rate. The most significant improvement was observed in the normal behaviour recall value by an increase of 23.46% and the anomalous behaviour precision value by an increase of 21.95% to achieve 100% in both cases,

as shown in Figure 44 The overall best performing super learner models. covers the two best-performing models. Firstly, the training time of the dataset utilising the 10s rolling window with the base learner default values was 3m 43s. Next, the training time increased to 13m 41s with additional parameters optimisation applied to the base learners. There are further notable differences in the testing time. Utilising the default base learner parameters, a testing time of 605ms, and an attack prediction time of 9.65s were achieved. The attack prediction of individual attacks ranged from 694ms to 4.19s. Whereas following optimisation of the base learner parameters, a testing time of 703ms, the attack prediction time of 12.8s was achieved.

Table 16 Overall performance details of the two best performing super learners and their percentage difference.

Super learner	Optimised parameters 10s Rolling Window [%]	Default parameters 10s Rolling Window [%]	Optimised to default parameters change [%]
Overall Performance [%]			
Accuracy	98.79	99.18	0.39
F1-score	98.77	99.13	0.36
ROC AUC	99.00	99.00	0.00
Normal behaviour performance [%]			
Precision	99.00	99.00	0.00
Recall	99.00	100.00	1.01
F1-score	99.00	99.00	0.00
Anomalous behaviour performance [%]			
Precision	99.00	100.00	1.01
Recall	98.00	99.00	1.02
F1-score	99.00	99.00	0.00

Table 17 Overall performance details of the weakest and best performing super learners and their percentage difference.

Super learner	Default parameters resampled at 1s [%]	Default parameters 10s Rolling Window [%]	Optimised to default parameters change [%]
Overall Performance [%]			
Accuracy	88.14	99.18	12.53
F1-score	88.41	**99.13**	**12.13**
ROC AUC	89.00	99.00	11.24
Normal behaviour performance [%]			
Precision	96.00	99.00	3.13
Recall	81.00	**100.00**	**23.46**
F1-score	88.00	99.00	12.50

Anomalous behaviour performance [%]			
Precision	82.00	**100.00**	**21.95**
Recall	96.00	99.00	3.13
F1-score	88.00	99.00	12.50

Further tuning was applied to the model for the individual attacks taking into consideration the imbalanced datasets. Therefore, resampling rates of 100ms, 300ms, 500ms and 1s, and 30% and 40% subsets of the normal behaviour dataset in addition to the full normal behaviour dataset were applied. The performance details of the specific attacks trained with the best-performing super learner are presented in Table 18 and Table 19. The supervised ML super learner's overall performance has been maintained consistently for a range of anomalies lasting between 17s and over 30m, as shown in Table 15 and Table 18. The model using the 10s rolling window achieved an overall F1-score of 99.13% and in the specific anomaly cases, the model's overall F1-score remained above 97.92%. However, it was noted to be below 95%, therefore a rate of >5% misclassification, in attacks 2, 3 and 6 as covered in Table 15. The corresponding results are recorded in Table 18 which presents the values of the overall ROC AUC and Table 19 shows the corresponding values of the Anomaly Recall and F1-scores.

4.9.3 Unsupervised Learners

Comparatively, the unsupervised ML model was fitted using Python's scikit-learn library [255]. As part of the dataset preparation, Spearman's correlation coefficient was applied, and highly correlated features were removed. According to the framework, before fitting the model, the features were standardised by utilising robust scaling applied to the 10s rolling window dataset. The comparison between the supervised and unsupervised models is based on the 10s rolling window. The one-

class binary classification to detect outliers in the combined dataset and the individual attacks were trained on random subsets of the dataset. A stratified 5-fold CV was applied during the model training.

A novel IF unsupervised learning approach to outlier detection was utilised. IF detects anomalies by isolating instances, as shown in Table 13 and Table 14, and not by using distance or density measures [119]. A comparison was produced by applying a density-based approach utilising LOF [258] and a distance-based approach using SVM [259, 260]. The IF algorithm is based on the characteristic that anomalies are few and different from normal observations within datasets, therefore sensitive to isolating anomalies from the typical observations [119]. The authors [119] focused on unsupervised learning and continuous values in a non-parametric approach of multivariate data detection of anomalies only. Whereas in this thesis, IF was applied to parametric discreet data values of one-class binary classification for outlier detection. According to [119], IF scales up to extremely large datasets with a high number of irrelevant features to solve high-dimensional problems. Noteworthy about this dataset is the application of the IF learning algorithm to a dataset containing a few features and short periods of recorded anomalies, see Table 15.

4.10 Discussion

4.10.1 Comparison of Learners

As the outcomes are predicted based on input data, the ML models are dependent on the quality of the datasets. This case study compared the proposed ML supervised super learner method with other ML approaches proposed in the scientific literature [45, 46, 88, 119, 252]. While compared to data-driven approaches, model-

based learning performs more effectively with lower computational overhead, particularly in larger datasets. However, in supervised learning, some of the efficiency is offset by the cost of the dataset preparation in feature labelling. This case study demonstrated that supervised learning performs comparably in training and testing to the unsupervised ML algorithms in computational complexity and performance scores based on the same dataset [88]. The experiments produced promising results which are presented in Table 18 and Table 19. The performance metrics include the anomalous precision, recall, and F1-score values, and the Confusion Matrix for the TP (the correctly identified normal behaviour instances) and TN (the correctly identified anomalous instances) values for the combined and specific anomalies datasets [45, 46, 257]. The performance of each classifier is measured by using metrics that apply to multiple classifiers. The most commonly relied upon metrics are Precision measuring the likelihood of the classifier providing the correct result, Recall indicating the detection rate and F1-score [88, 257].

Variations were observed in results between the algorithms, including their performance consistency based on the level of anomalies in the datasets as shown in Figure 45, Figure 46 and in the algorithms' total running time which is presented in Figure 47. The analysis revealed that both models have a good anomaly detection ability. The supervised super learner achieved an overall F1-score of 99.13% and an anomalous recall score of 99% compared with the IF anomalous recall score of 98%. The IF anomalous recall score values achieved above 60% in datasets 8, 9, and 13. SVM showed stronger performance where low levels of anomalous behaviour were present over a shorter period including datasets 3-7,10, 13-15 as labelled in Table 15. The respective results are presented in Table 19 and Figure 45. The lower IF precision scores compared with the supervised ML super learner could be due to the

behaviour during an attack being resemblant of normal operation hence resulting in a higher rate of FP behaviour during some of the analysed attacks.

Table 18 The Area Under ROC Curve of the individual attacks trained utilising the supervised ML super learner and the unsupervised ML algorithms with a 10s rolling window.

Dataset		AUC			
Components	Anomaly [%]	Super learner	IF	SVM	LOF
All anomalies	89	0.99	0.59	0.54	0.5
Plastic_bag	27	0.88	0.58	0.54	0.5
Spoofing	28	0.96	0.52	0.54	0.5
High_blocked	3	0.98	0.92	0.91	0.5
Second_blocked	11	0.99	0.77	0.61	0.5
Bad_connection	13	0.96	0.57	0.60	0.5
DoS_attack	1	1.00	0.74	0.98	0.5
Hits_3	0.5	1.00	1.00	0.98	0.5
Wet_sensor	2	1.00	0.50	0.98	0.5
Poly_2	1.3	0.98	0.78	0.98	0.5
Poly_7	1.1	1.00	0.69	0.98	0.5
Hits_2	4	1.00	0.50	0.98	0.5
Hits_1	5	1.00	0.50	0.98	0.5
Blocked_1	4	1.00	0.50	0.98	0.5
Blocked_2	2	1.00	0.50	0.98	0.5

Table 19 The anomalous behaviour performance metrics of the individual attacks for the supervised ML super learner and the unsupervised ML algorithms.

Main & subset dataset	Algorithm	Anomaly			Confusion Matrix	
		Precision	Recall	F1	TP	TN
All anomalies	SVM:	0.92	0.1	0.18	0.99	0.1
	IF 5:	0.52	0.98	0.68	0.20	0.98
	Super:	0.99	0.99	0.99	0.99	0.99
Plastic bag	SVM:	0.50	0.12	0.19	0.97	0.12
	IF 25:	0.31	0.39	0.34	0.76	0.39
	Super:	0.94	0.77	0.85	0.99	0.77
Spoofing	SVM:	0.50	0.11	0.18	0.97	0.11
	IF 25:	0.24	0.31	0.27	0.72	0.31
	Super:	0.97	0.92	0.95	0.99	0.92
High blocked	SVM:	0.50	0.85	0.63	0.97	0.85
	IF 25:	0.81	0.84	0.82	0.99	0.84
	Super:	0.95	0.95	0.95	1.00	0.95
Second blocked	SVM:	0.50	0.24	0.33	0.97	0.24
	IF 25:	0.55	0.61	0.57	0.94	0.61
	Super:	0.99	0.99	0.99	1.00	0.99
Bad connection	IF 5:	0.22	0.25	0.24	0.89	0.25
	Super:	0.95	0.94	0.94	0.99	0.94
DoS attack	SVM:	0.26	1.00	0.42	0.96	1.00
	IF 45:	0.48	0.49	0.49	0.99	0.49
	Super:	1.00	1.00	1.00	1.00	1.00
Hits_3	SVM:	0.09	1.00	0.16	0.95	1.00
	IF 200	0.97	1.00	0.99	1.00	1.00
	Super:	1.00	1.00	1.00	1.00	1.00

Wet sensor	SVM:	0.05	1.00	0.09	0.95	1.00
	Super:	1.00	1.00	1.00	1.00	1.00
Poly_2	SVM:	0.26	1.00	0.41	0.96	1.00
	IF 35:	0.56	0.57	0.56	0.99	0.57
	Super:	0.93	0.97	0.95	1.00	0.97
Poly_7	SVM:	0.22	1.00	0.36	0.96	1.00
	IF 15:	0.46	0.39	0.42	0.99	0.39
	Super:	1.00	1.00	1.00	1.00	1.00
Hits_2	SVM:	0.09	1.00	0.16	0.95	1.00
	Super:	1.00	1.00	1.00	1.00	1.00
Hits_1	SVM:	0.11	1.00	0.19	0.95	1.00
	Super:	1.00	1.00	1.00	1.00	1.00
Blocked_1	SVM:	0.07	1.00	0.13	0.95	1.00
	Super:	1.00	1.00	1.00	1.00	1.00
Blocked_2	SVM:	0.05	1.00	0.09	0.95	1.00
	Super:	1.00	1.00	1.00	1.00	1.00

4.10.2 Computational Complexity

The experiment results indicate that computational complexity and the cost of the supervised super learners is significantly higher in the combined dataset and in the individual datasets where the level of anomalies is above 25%. This complexity remains higher in datasets with anomalies' proportion of above 10% compared with the lower computational complexity and cost of the unsupervised multi-learners. The computational complexity was lower in supervised learning where fewer anomalies were prevalent and attacks lasted shorter. While the unsupervised multi-learners detected the attacks, IF did not detect all attacks, as illustrated in Figure 45 and Figure 46, particularly where a very low occurrence of anomalies was prevalent and over a short period. In those cases, SVM produced a consistently better performance utilising the polynomial kernel which considers the input samples, their similarity, and combinations, unlike IF. This could be explained by the behaviour in those datasets being similar to normal operations and not detected as outliers without resulting in false positives.

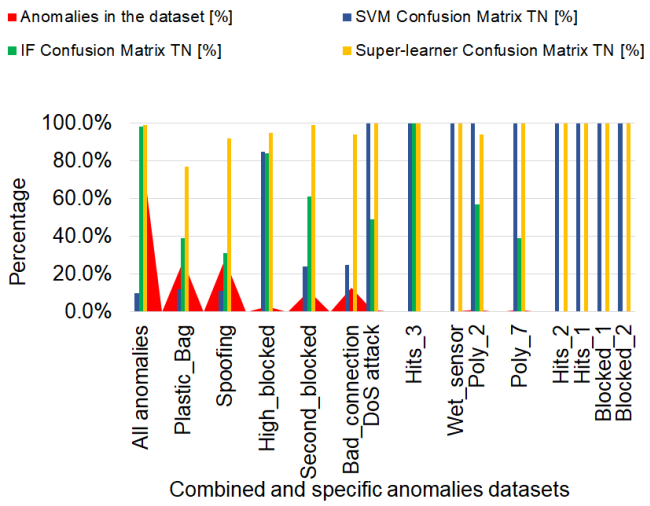


Figure 45 Comparison of the learning algorithms' confusion matrices TN values and anomalies in the combined and specific anomalies datasets.

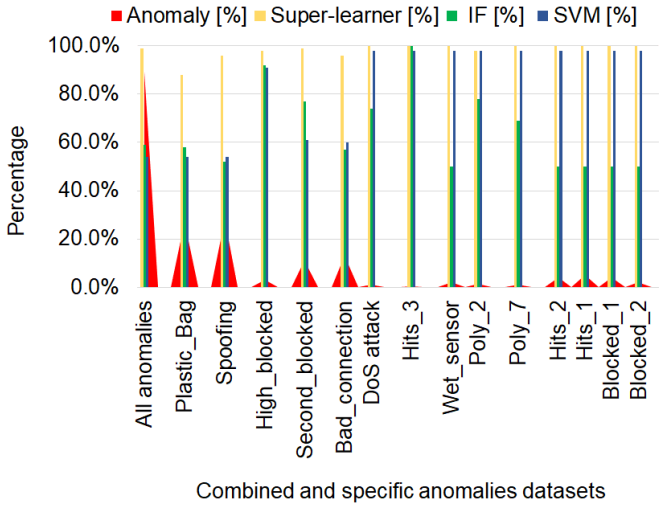


Figure 46 AUC comparison of anomalies and algorithms in the combined and specific anomalies datasets.

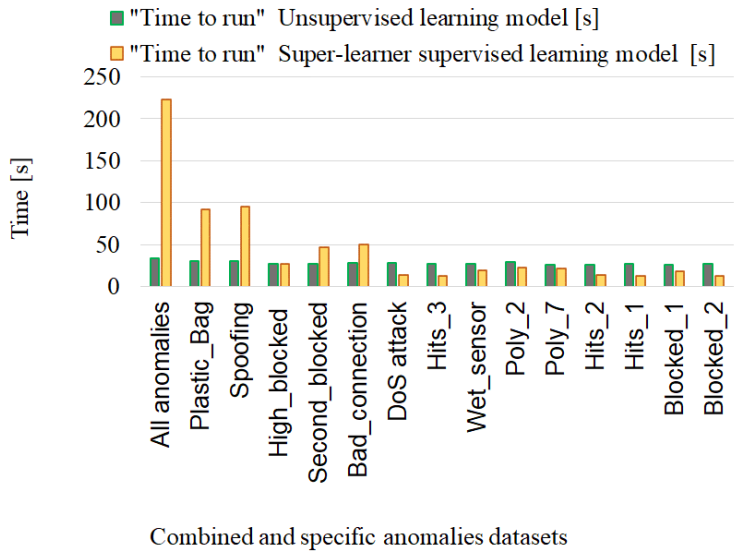


Figure 47 Comparison of algorithms performance based on the total time to run.

It is important to note that no specific data sanitisation was applied [53] such as removing part of the anomalous dataset which could be considered normal behaviour while flagged as anomalous. This could be typical during the post-attack recovery phase back to normal operation. This study asserts that the model should remain resilient to such behaviour and reflective of a typical operational pattern including a period of return to normal operation. Therefore, a future research direction could focus on the unsupervised model to further tune the hyperparameters constructing an unsupervised super learner. This could lead to simplified pre-processing, achieve lower model learning computational complexity and cost while consistently achieving performance at least similar to the super learner presented in our framework.

Our proposed approach produced encouraging attack prediction times ranging from 694ms to 4.19s for specific attacks and 9.65s for the combined dataset for the default base learner parameters. The findings indicate that several factors influence the model's performance. Such facets include the model structure, parameter tuning and the computational environment. However, another important challenge is the model's resilience when the data distribution evolves. Adapting to changes while maintaining the model efficacy in near-real-time utilising continuous data streams is critical in dealing with the time-critical nature of ICS. How to integrate effective near-real-time prediction and the trade-off in maintaining the model efficacy is a challenging problem that merits further research.

Addressing the first research question, it is noted in the analysis of the results, the models' resilience to detect anomalous behaviour in datasets increases by combining the learners. For example, as shown in Table 18 and Table 19, the

unsupervised learning model's resilience does detect anomalies improved when multiple algorithms in addition to IF hyperparameter tuning were utilised. Similar behaviour was observed in the supervised super learner model. An individual learner within a model is not likely to outperform other learners in the stack. That said, the aim of utilising a super learner is to solve the problem of selecting a suboptimal classifier, increase the model's resilience of attack detection, improve predictive ability, and generalise the performance of the algorithm. Although the authors of the following study [5] assert a lack of agreed definition or performance metrics of the term "resilience"[172, 173], the importance of cyber resilience is acknowledged by the scientific community and governments. It is acknowledged that more must be done to improve the cyber resilience of the CNI, accepting that cyber resilience is a particular challenge in IoT [66, 167, 169]. Therefore, it is argued that in this study the approach of constructing a resilient model as a key part of the framework to detect anomalous behaviour in ICS CPS is security process-driven, the model's resilience improves the CPS' defence mechanism, security situational awareness and support for DFIR.

5. Chapter: Cyber risk Quantification in ICS

Chapter 4 introduced the SPEAR framework and presented the methodology to profile anomalous behaviour in ICS. The approach was experimentally demonstrated on an ICS liquid distribution case study. This chapter presents the cyber risk quantification model.

5.1 Introduction

According to [43], cyber risk is defined as the potential that a given cyber threat will exploit the vulnerabilities of an IS and cause harm. Modern interconnected ICS are human-made environments profoundly influenced by human behaviours. Due to how ICS technologies advance and are consumed, the threat landscape evolves, trends change and diversify. Attack actors will have new opportunities with likely new attack vectors. Hence, an intelligence-driven approach toward understanding the impact of these changes on the ICS cyber risk posture is an implicit need to improve cyber resilience. This chapter considers cyber resilience as the ability to resist cyber-attacks across the physical and digital realms regardless of an external or insider attack [67, 73-75]. We can innovatively use disruptive technologies, and leverage data generated from sensors coupled with AI technologies to improve cyber resilience as part of an effective defence-in-depth strategy.

Chapter 5 investigates and addresses the problems and limitations of existing cyber risk management in ICS. The chapter presents an approach to objectively quantify cyber risk in the prevalence of anomalous behaviour in ICS by developing a

model as part of the SPEAR framework. A Bayesian Belief Network (BBN) model for Cyber risk Value Quantification (CRVQ) leveraging ICS sensor data is proposed. The model is then demonstrated to derive a CRVQ estimate. Further, following a discussion of the SPEAR Framework's applicability to objectively quantify the cyber risk value in the prevalence of anomalous behaviour, the model's support for post-incident investigations as part of DFIR is discussed. The two RQs addressed in this chapter can be expressed as:

- How can the SPEAR framework be utilised to quantify the cyber risk in CPS?
- How can the SPEAR framework support DFIR?

The reviewed scientific literature suggests that although cyber risk management is an area of scientific interest, research focusing on a quantification of cyber risk value based on anomalous behaviour detection in CPS remains limited. To the best of the researcher's knowledge, this research is the first study to combine attempts to address quantifying cyber risk value when anomalous behaviour is prevalent and to support DFIR.

5.2 Insights into cyber risk quantification in ICS

The CIA triad has been considered fundamental to good security practice. The CIA triad has been adopted and driven by the IS community; however, it does not sufficiently address the security aspects of ICS. For example, understanding control and safety facets are important in ICS due to their complexity, fragmentation and real-time interactions. Likewise, ICS can be geographically dispersed and potentially owned by multiple legal entities and jurisdictions. To overcome the limitations of the

traditional CIA approach and address the challenges in ICS this study investigates the use of the Parkerian Hexad (PH) as a forward-looking alternative to converge engineering and IS good practices [6, 90, 261]. Safety is considered a seventh dimension by the authors of the following study [90] who assert that the creation and use of the data should not be harmful. Furthermore, the authors emphasize that the safety dimension provides context for cybersecurity risk assessment and impacts situational awareness. This study does not challenge the approach and considers the safety aspect of significance in quantifying the cyber risk value in CPN ecosystems [6].

Scientific literature shows little evidence of deviation from the conventional risk formula. Only a few studies propose enhancements such as the architectural perspective of risk [262]. Studies such as those listed in Table 20 and others [263-265] focus on estimating the risk in control systems. For example, the following study [265] predicts the risk level at a particular time whereas other studies [263, 264] dynamically and timely calculate the risk. Another study [266] investigates ways to enhance the resilience of power systems against cyber-attacks. The method of assessing cyber risk remains a significant shortcoming in CPS. As illustrated in the previous chapter in Figure 33, ICS are highly automated and designed for safety, reliability and availability. Therefore, ICS have limited consideration to understanding the cyber risk value, the scale of the impact caused by cyber-attacks [20] or support for DFIR as shown in Table 20. Most empirical studies including [267-271] address static risk without necessarily quantifying the cyber risks. Throughout this thesis, the terms “cybersecurity risk” and “cyber risk” are used interchangeably.

Furthermore, this research acknowledges the scientific efforts to improve the cybersecurity posture of ICS including through frameworks such as the NIST voluntary Framework for Improving Infrastructure Cybersecurity [272]. It is not the intention of this study to challenge existing approaches. However, the study leans on the characteristics outlined in this chapter, which quantifies the cyber risk value and contextualises situational awareness comprehensively for CPS. This research seeks to address the cyber risk value quantitatively from CPS datasets in the prevalence of detected anomalies.

Table 20 Comparison of Risk Assessment (RA) approaches of similar studies in ICS and support for DFIR.

Studies	Address ICS	Dynamic RA	Support for DFIR	Empirical Study addressing RA
This Study	✓	✓	✓	✓
[263]	✓	✓	×	✓
[264]	✓	✓	×	✓
[273]	×	✓	×	✓
[270]	×	×	×	✓
[267]	✓	×	×	✓
[271]	×	×	×	✓
[269]	×	×	×	✓
[268]	×	×	×	✓
[274]	✓	✓	×	✓
[275]	×	✓	×	✓
[182]	✓	✓	×	✓
[276]	×	✓	×	✓
[250]	✓	×	×	✓
[266]	✓	✓	×	✓
[20]	✓	✓	×	✓

5.3 The Cyber risk Value Quantification (CRVQ) Model

The CRVQ model aims to objectively quantify the cyber risk value utilising intelligence learnt from CPS datasets in the prevalence of anomalous behaviour such that it is trustworthy, testable, and repeatable. Our approach to quantifying the overall Cyber risk Value (CRV_t) is inspired by the Common Vulnerability Scoring Systems (CVSS) [277]. CVSS is an open framework for communicating the severity and

attributes of vulnerabilities in software that consists of the base, temporal and environmental metrics.

In this chapter, this research introduces the concept of quantifying the CRV_t for a materialised cyber risk. This is achieved by producing initial scores for risk occurrence and risk severity impact combined with an updated score derived from the performance metrics of the detected anomalous behaviour. The metrics in the CRVQ model consist of three phases. The constant base metric group of attributes, the temporal metrics group which is expected to change over time and the environmental metrics group which is anticipated to vary between organisations and smart sectors. The attributes in the three metric groups are utilised to derive the risk occurrence score, the risk severity impact, and the safety scores.

5.3.1 Performance Scores

In phase one, the anomalous behaviour is identified using ML techniques which produce a set of performance scores based on the ML's predictions. Two confidence scores are derived, the Report Confidence Accuracy (RC_A_s) and the Report Confidence Anomalous Behaviour Detection (RC_ABD_s) which are linked to the outcome of the ML models' prediction 'Y', as shown in Figure 37 and Figure 40. Each base learner, y_1-y_n, detects anomalies independently using base learners followed by a meta-learner for the final prediction. The RC_A_s is expressed as:

Anomalies in dataset + overall accuracy - (1 – anomalies accuracy)

$$A_d + A_t - (1 - A_a) \tag{5.1}$$

The RC_ABD_s is expressed as:

Anomalies in dataset + weighted F1-score - (1- anomalies F1-score)

$$A_d + F1_w - (1 - F1_a) \quad (5.2)$$

where contamination level of anomalies A_d is derived as:

Anomalous instance in the dataset / Normal instance in the dataset

$$\frac{i_{Ad}}{i_{Nd}} \quad (5.3)$$

and the anomalies' F1-score '$F1_A$' is derived as:

2*(Anomalies' Recall * Anomalies' Precision) / (Anomalies' Recall + Anomalies' Precision)

$$2 * \frac{(R_A * P_A)}{(R_A + P_A)} \quad (5.4)$$

5.3.2 Metric Groups

Phase two utilises the concept of the CVSS framework. In addition to a subset of CVSS attributes, [277] adds traits specific to the CRVQ model in the environmental and temporal metric groups, as shown in Figure 48. New attributes are introduced to express an actual value derived quantitatively based on the ML predictions of detected anomalies. This study produces confidence scores of the occurring anomaly in combination with the prior knowledge set by the base score. The update factor is based on the actual occurrence of the cyber risk derived from anomalous behaviour detection. The safety factor is combined with the Initial Risk Occurrence (IRO) to produce the overall risk occurrence value. The two temporal metric attributes, RC_ABDs and RC_A_s are mandatory, their values are shown in Table 21. The Attack Vector (AV_b) uses Network (Ne) and Physical (Phy) annotations. The Attack Complexity (AC_b) and Privilege (Pr_b) base metrics use Low (Lo) and High (H) annotation. The Scope utilises Unchanged (Uch) and Changed (Ch) annotation. The User Interaction (UI_b) base metric uses the None (Nn) and Required (R) annotations.

The temporal metrics group replaces the 'Report Confidence' metric with the RC_A_s and the RC_ABD_s metrics, generated by the ML predictions, using the Confirmed (Co) and Unknown (U) annotation. The values could be expressed as a combined report confidence metric; however, the aim is to report their values independently. In addition, the Collateral Damage (CD_e) uses the None (Nn) and Confirmed (Co) annotation. The attributes in the base metrics are mandatory and not expected to change. Whereas the temporal metrics are expected to change over time across environments and are therefore used as an update factor. The metrics' dependencies between the variables are presented in Figure 49.

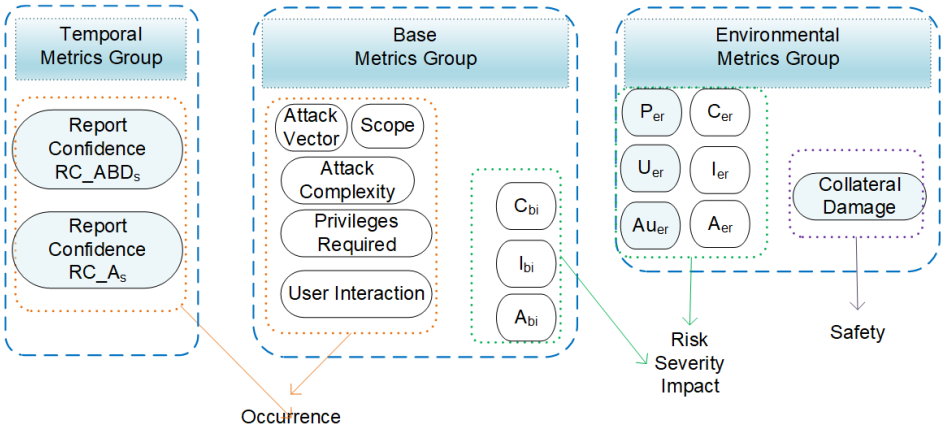


Figure 48 The metrics used in the CRVQ model.

The research in this chapter introduces the PH attributes of Possession, Utility and Authenticity in addition to the CIA as shown in Figure 48 to derive the risk severity impact. A binary state is utilised for authenticity, possession, and utility metrics are covered in Table 22 with the following considerations:

- The Base and Environmental Metric Groups in this model use the Low (Lo), High (H) binary annotation. The 'Required' column in Table 22 shows the CVSS mandatory setting for the metric and the one in this model, respectively. The setting of zero (0) means that the rating is not present. The setting is annotated to Nn when it is not mandatory and annotated to Y if it is mandatory.

- The **authenticity score** (Au_{er}) is related to the expected normal state of operation and attribution to the source of the data. As anomalies are detected, the authenticity cannot be attributed, and the operation is not considered to be in a normal state.
- The **possession score** (P_{er}) is related to the control of the CPS or their components producing the data. Literature refers to the physical disposition of media where the data is contained [278]. In CPS, possession can be lost if a CPS component producing the data is compromised in which case the control over the specific CPS component is considered lost. Therefore, possession is breached if the data does not reflect the status consistent with normal operation.
- The **utility score** (U_{er}) relates to the usefulness of the data during the normal state of CPS operations which in the presence of anomalous behaviour is considered compromised. This metric does not consider the threat actor's efforts, or the computational complexity needed to compromise the data utility. The utility score is greatest for utility compromise.

In addition, to express **safety**, this study uses the CD_e attribute introduced in the environmental metric [277]. The attribute is expected to differ between organisations, as shown in Figure 48 and Figure 49. The attribute describes the condition related to the data in the presence of detected anomalous behaviour as not considered for safe use. Therefore, having potential consequences for the organisation such as cascading effects resulting in loss of life, damage to equipment or monetary loss.

Table 21
Metrics and attributes used to derive the risk occurrence.

Required	CVSS Metric Group & Attributes	Rating	Value
	Base		
Y→Y	Attack Vector AV_b	Ne/Phy	0.85/0.62
Y→Y	Attack Complexity AC_b	Lo/H	0.77/0.44
Y→Y	Privileges Required Pr_b	Lo/H	0.62/0.27

Y→Y	Scope S_{bi}	Uch/Ch	0.06/0.23
Y→Y	User Interaction UI_b	Nn/R	0.85/0.62
	Temporal		
0→Y	Report Confidence RC_ABD_e	Co/U	96.00/85.00
0→Y	Report Confidence A_{cs}	Co/U	96.00/85.00
0→Y	Collateral Damage CD_e	Nn/Co	0.10/0.90

Table 22
Base and Environmental Metric Groups and used to derive the Risk Severity Impact using the annotation of Low (Lo) and High (H) ratings.

Required	Metric Groups & Attributes	Rating	Value
	Base		
Y→Y	Confidentiality Impact C_b	Lo/H	0.22/0.56
Y→Y	Integrity Impact I_b	Lo/H	0.22/0.56
Y→Y	Availability Impact A_b	Lo/H	0.22/0.56
	Environmental		
N→Y	Conf. Requirement C_{er}	Lo/H	0.50/1.00
N→Y	Integrity Requirement I_{er}	Lo/H	0.50/1.00
N→Y	Availability Requirement A_{er}	Lo/H	0.50/1.00
0→N	Possession Impact P_{er}	Lo/H	0.50/1.00
0→N	Utility Impact U_{er}	Lo/H	0.50/1.00
0→N	Authenticity Impact Au_{er}	Lo/H	0.50/1.00

5.3.3 Cyber risk Score

In phase three, the cyber risk score is derived from the three metric groups using a BBN. In this study, the base metric group attributes' values are used as the input for the BBN as the previous knowledge to quantify the prior distribution value deriving the IRO and the Initial Risk Severity Impact (IRSI).

The Bayes theorem is utilised to derive the Cyber Risk Value of risk occurrence (CRV_{ro}) and the Cyber Risk Value of risk severity (CRV_{rs}) values. According to [279] the Bayes theorem can be used to derive conditional probability, where in generalised terms the probability [P] of random variables 'x' given 'y' can be expressed as:

$$P(x|y) = P(y|x) * P(x) / P(y) \quad (5.5).$$

Furthermore, according to Figure 49, to derive the IRO the following applies:

- $P(AV_b)$ is not dependent.
- UI_b depends on AV_b, expressed as $P(UI_b)=P(UI_b|AV_b)$.

- AC_b depends on AV_b, expressed as $P(AC_b)=P(AC_b|AV_b)$.
- Pr_b conditionally depends on the AV_b and AC_b, and AV_b and AC_b are also internally dependent, thus $P(Pr_b) = P(Pr_b|Av_b, AC_b)$. If $P(Pr_b|Av_b, AC_b)$ is generalised as $P(x|y_1, y_2)$ the theorem is applied as:
- $P(x|y_1, y_2) = P(x) *P((y_1, y_2)|x) / (y_1, y_2)$ (5.6)

Then

- $P(x|y_1, y_2)| = P(y_1) + P(y_2) - P(y_1) * P(y_2)$ (5.7)

However, this study aims to solve the challenge of deriving a value that is based on actual detected anomalous behaviour, is accurate, trustworthy and improves cyber risk situational awareness. Therefore, the Risk Occurrence Update Factor (ROUF) and the Risk Severity Update Factor (RSUF) are produced. The ROUF is based on the ML model's performance metrics and safety factor whereas the RSUF is based on the PH Au_{er}, P_{er}, and U_{er} which are integrated into the Confidentiality (C_{er}), Integrity (I_{er}), and Availability (Av_{er}) requirement environmental metrics attributes such that combined value is used in this model for simplicity. The CRV_{ro} is derived as:

$$IRO = \int P (AC_b, AV_b, Pr_b, UI_b) \quad (5.8).$$

$$ROUF = \int P (RC\_ABD_s, RC\_A_s, CD_e) \quad (5.9)$$

$$CRV_{ro} = \int IRO \times ROUF \quad (5.10)$$

The CRV_{rs} is derived as:

$$IRSI = \int P (C_b, I_b, Av_b) \quad (5.11)$$

$$RSUF = \int P (C_{er}, I_{er}, Av_{er}) \quad (5.12)$$

$$CRV_{rs}= \int IRSI \times RSUF \quad (5.13).$$

The overall Cyber risk Value CRV_t is derived as:

$$\int CRV_{ro} \times CRV_{rs} \quad (5.14).$$

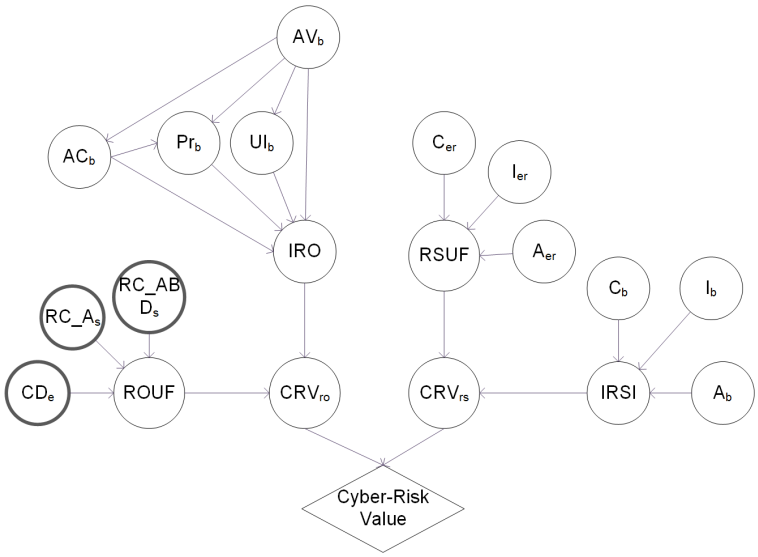


Figure 49 BBN CRVQ model

5.4 Deriving the CRVQ estimate

To demonstrate the CRVQ model, this study defines a set of input values expressed as:

- Metric Group: Rating/Value,
- Base (C_b:Lo/0.22, I_b:H/0.56,A_b:H/0.56);
- Environmental (C_{er}:Lo/0.5, I_{er}:H/1, A_{er}:Lo/0.5, P_{er}:Lo/0.5, U_{er}:H/1, Au_{er}:H/1);
- Base (AV_b:Ne/0.85, AC_b:Lo/0.77, Pr_b:Lo/0.62, S_{bi}:Uch/0.06, UI_b:R/0.62);
- Temporal (RC_ABD_e:Co/0.96, A_{cs}:Co/0.96, CD_e:Nn/0.1).

To establish the ROUF, the study derives values from the super learner ensemble performance metrics recorded in Table 23, given the equations defined in section 5.3.1. If the RC_A_s and RC_ABD_s values exceed 0.96 the maximum value of 0.96 will be retained. If the RC_As and RC_ABDs values are 0.85 or below the value of 0.85 will be recorded. For example, let the focus be on the risk occurrence part of the model, to derive the ROUF score, utilising the ROUF normalised equations

presented in section 05.3.3. The likelihood that the initially assessed cyber risk changed; based on the ML model's extracted performance metrics it is 67%, increasing the cyber risk score by 13% compared with the initial cyber risk value.

Table 23 Report Confidence metrics resampled to rolling 10s dataset individual learners accuracy score with repeated stratified 10fold 3 repeat CV 10s_R2-10s_R4.

Metric Groups & Attributes	Rating	Update Factor Value
RC_A_s		
2	Plastic Bag	1.16
10	Denial of Service attack	1.01
11	Spoofing	1.24
RC_ABD_s		
2	Plastic Bag	0.87
10	Denial of Service attack	1.01
11	Spoofing	1.12

Anomalous behaviour detection in ICS improves the cyber defence mechanisms against evolving threats and cyber-attacks from modern adversaries. Integrating this dimension as a source of information into the cyber risk quantification model to derive the cyber risk value contributes to a better understanding and contextualising of cyber risk in ICS. Data-informed cognisance contributes to the understanding of how the cyber incident evolves and articulates changes in the cyber risk posture. This helps anticipate potential consequences and take measured actions informed by data resulting in improved decision-making achieving better situational awareness of the cyber risk posture.

5.5 SPEAR Framework Discussion

5.5.1 Applicability to Cyber risk Quantification

This study addressed the research question of how the SPEAR Framework can be utilised to quantify the cyber risk in CPS. The presented CRVQ model is an integral part of the SPEAR Framework. This chapter outlined the structure of the CRVQ model, identified the algorithm and the performance metrics to objectively

quantify the cyber risk value in CPS. Further, the research study demonstrated in this chapter the model's applicability to quantifying the cyber risk score and articulated the cyber risk score change in the presence of anomalous behaviour. Existing risk assessment models are driven by the IS community and ICT systems [280-283], whereas methodologies such as the qualitative Hazard and Operability study (HAZOP) tend to focus on risks to personnel and equipment, not cybersecurity [284]. While individual maturity exists, a disconnect remains between ICT and OT, particularly in ICS. It is acknowledged that cybersecurity controls applicable in the ICT realm are not necessarily applicable to the OT realm. Nevertheless, a holistic defence-in-depth security approach with layered protective controls which consider the converged realms is needed. It is acknowledged that generalisation and scaling of the proposed CRVQ model at this stage are not possible. Further empirical studies are required to systematically investigate and report further findings from a wider pool of ICS assets to optimise the CRVQ model and its components.

That said, this research asserts that quantifying the cyber risk forms an important part to enhance a robust defence-in-depth approach. It is further asserted that CPS sensor-generated data combined with ML anomaly detection techniques can contribute to the objective evaluation of the effectiveness of the assessed cyber risk in CPS. Hence creating an opportunity for decision-making for cybersecurity protective and corrective actions proactively. Such an approach aligns with the CREST principles of intelligence often referred to as CROSSCAT (Centralised, Responsive, Objective, Systematic, Sharing, Continuous review, Accessible, Timely) [285] to improve the cyber-threat intelligence capability creating opportunities to solve real-life problems.

5.5.2 *Applicability to Support DFIR*

Finally, this chapter addressed the research question concerning how the SPEAR Framework supports DFIR. Thus far this chapter has outlined and discussed the correlation of raw data collection, information processing, generating knowledge and application of this knowledge to quantify the value of cyber risk as part of defence-in-depth capability in CPS. To illustrate the applicability of the SPEAR Framework and its components to DFIR, comparisons are drawn with the generic DFRWS and the ISO/IEC 27050:2016 standard data lifecycle phases. According to the following study [56], the stages can be broadly categorised into physical, logical and legal contexts. The physical context is concerned with capturing the data from seized physical media and maps to the identification and preservation stages, these are not the focus of this study. The logical context that is concerned with the data and information mapping to the collection, processing and analysis stages are of particular interest to this research study, see Figure 50. The initial risk assessment is a well-established field and is out of the scope of this research study. The SPEAR Framework's learning model introduces the capability of collecting and processing CPS datasets. Applying the learning algorithm to the data classifies the anomalous behaviour from the sensor-generated datasets. Performance metrics are applied to the produced results. The results are further analysed, and the cyber risk is quantified generating a score which can be applied to the effective evaluation of the cyber risk impact.

In ubiquitous CPS, particularly ICS and CNI seizing physical media is not always possible and innovative methods of gathering DE are required. Such methods could include ML-based models, as presented in the SPEAR Framework. Collection of DE

could be achieved by continuously processing and analysing data from CPS in near-real-time, applying ML algorithms to detect anomalous behaviour as an early incident indicator. Therefore, cyber-threat intelligence and the knowledge produced from sensor data utilising the SPEAR Framework's learning algorithms could assist in the reconstruction of events and identification of prior patterns. However, consideration should be given to admissibility. For example, CPS objects can be modelled as "Digital Witnesses" (DW) to support DFIR [6, 56]. In such a case chain-of-custody needs to be achieved by utilising a suitable mechanism for admissibility in the Court of Law. Therefore, the SPEAR Framework has applicability to support the logical stages of the DFIR in CPS. However, it should be recognised that more research is needed to understand the constraints and develop techniques that contribute to reducing the workload and cost of DF investigations and generating admissible and trustworthy DE.

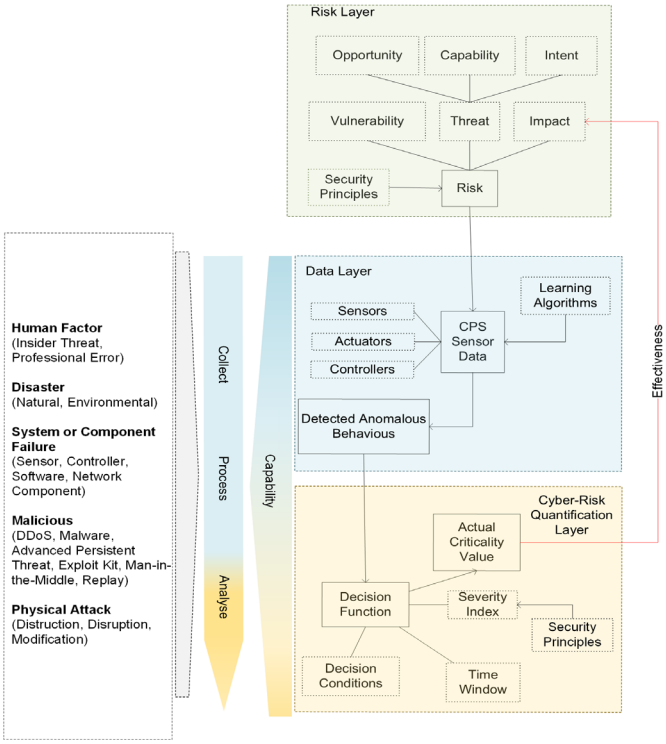


Figure 50 The SPEAR Framework applicability to DFIR based on anomalous behaviour detection from CPS sensor data and Cyber risk Quantification.

6. Chapter: A-ADC - Adaptive Learning Anomaly Detection and Classification Model

The previous chapters presented model-based learning to tackle malicious intent, accidental hazards and professional errors in ICS. The SLR findings indicate that several factors influence the learning model's performance among which are the model's structure, parameter tuning and computational environment. However, data produced from physical plant sensors in ICS is ubiquitous and the evolving distribution of the data is an important factor. Hence, in dealing with the time-critical nature of ICS, it is critical to adapt to changes while maintaining the model efficacy in near-real-time utilising continuous data streams. Due to the range of critical services that ICS provide, disruptions to operations could have devastating consequences making ICS an attractive target for sophisticated threat actors. Hence, we introduce a novel anomalous behaviour detection model for ICS data streams from physical plant sensors. A model for one-class classification is developed, leveraging stream rebalancing followed by adaptive Machine Learning (ML) algorithms coupled with drift detection methods to detect anomalies from physical plant sensor data. Our approach is demonstrated on ICS datasets. Additionally, a use case illustrates the model's applicability to post-incident investigations as part of a defence-in-depth capability in ICS. The experimental results show that the proposed model achieves an overall Matthews Correlation Coefficient (MCC) score of 0.999 and Cohen's Kappa (K) score of 0.9986 on limited variable single-type anomalous behaviour per data stream. The results on wide data streams achieve an MCC score of 0.981 and a K score of 0.9808 in the prevalence of multiple types of anomalous instances. Finally, we introduce a Performance Benchmark Criteria

framework to quantify the performance of incremental classifiers across distinct levels of cyber-physical experimental environments.

6.1 The A-ADC Model

6.1.1 Overview

Traditional ML mechanisms generate prediction models learnt on batches of data or datasets offline, see Figure 51a. In real-world applications, ICS sensors continuously produce data and process command controls with limited computational resources [19, 129]. Therefore, effective anomaly detection should utilise techniques that are capable of processing streams of data. Notably, as shown in Table 24, data stream classification algorithms should process a single data instance, and utilise a limited amount of memory and time with an ability to predict the class on demand [130, 286]. However, utilising online learning techniques does not mean that changes in data from ICS physical plant sensors would be detected. As shown in Figure 51b, they should be fused with techniques to deal with adapting to change quickly to maintain fast detection and a low rate of false alarms [142, 287].

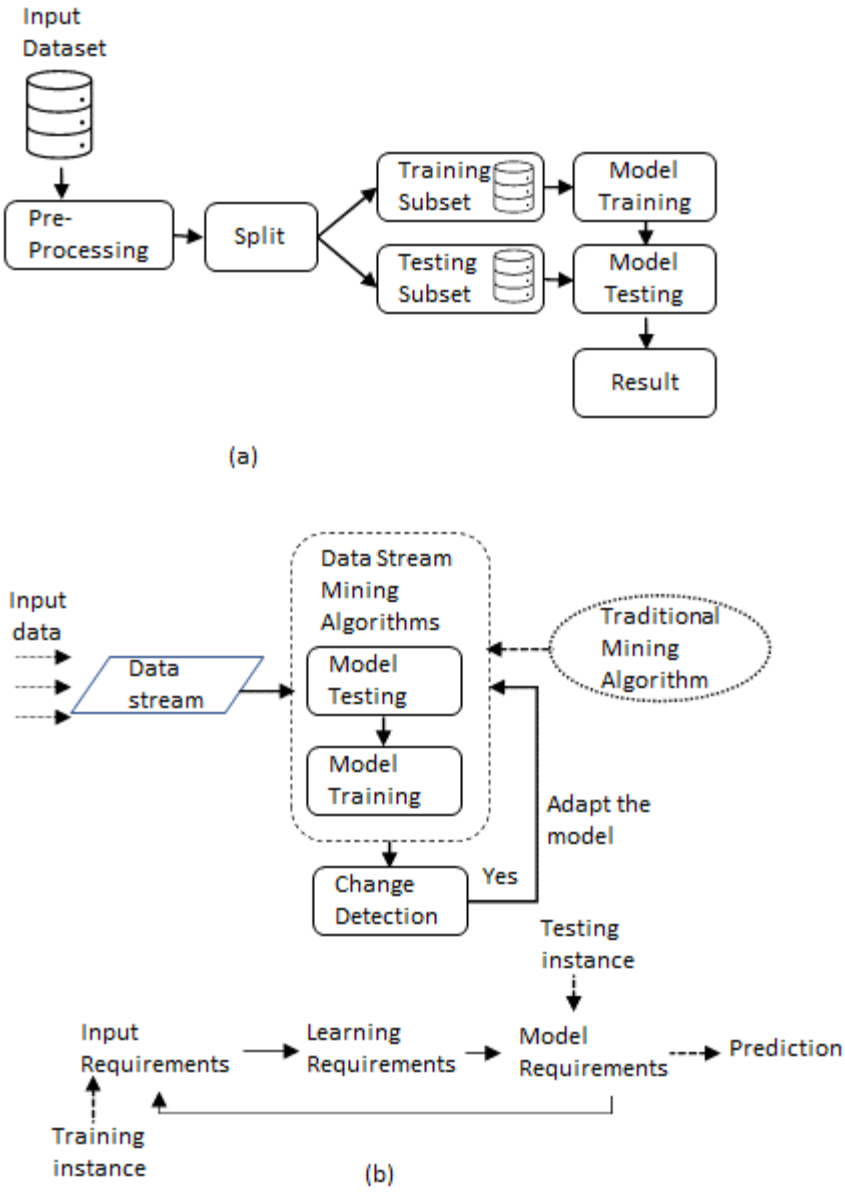


Figure 51 (a) General workflow for traditional machine learning using datasets (b) General Workflow for adaptive learning using data streams related to the MOA workflow

Table 24 Data Stream Classification Requirements

Data Stream Classification Requirements	Requirement description
Input	The initial step includes processing an instance that the algorithm is passed from the data stream one at a time. Each instance is inspected once. Once an instance is inspected, it is discarded without the ability to retrieve it.
Learning	Utilises a limited amount of memory and time. The size of a continuous data stream is larger than the memory available to the learning algorithm.
Model	Ability to predict the class on request from previously unseen data.

As part of our research towards near-real-time detection of anomalous behaviour in ICS, this research study leverages online learning techniques with change

detection ability. The study utilises the MOA open-source framework [288] and the WEKA [289]. MOA is a software environment that is designed to deal with the challenges of scaling up state-of-the-art algorithms to real-world datasets and data streams. MOA consists of algorithms for batch and online learning to gain knowledge from evolving data streams. WEKA provides learning algorithms applicable to datasets and data streams, and tools to transform, pre-process, analyse and evaluate the data. MOA supports bi-directional interactions with the WEKA environment. This research constructs the A-ADC adaptive online learning model presented in Figure 52 facilitating anomalies detection from ICS-based data streams.

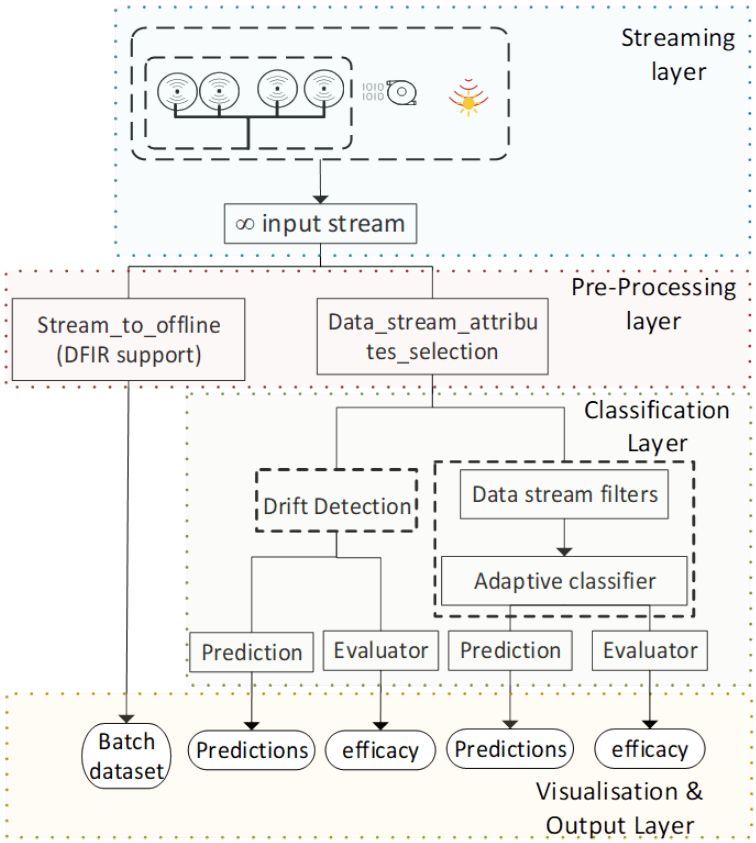


Figure 52 Components of the A-ADC anomaly detection ensemble model for data streaming from ICS

6.1.2 *Model Design*

The model's Streaming Layer utilises an arff loader to read the source data stream in an arff format one instance at a time. The data stream is read as the data flows without any allowance for random access to the instances. Although instances are discarded once read, see Table 24, the algorithm can remember the instances internally in the short term. This enables the creation of mini-batches to apply conventional ML techniques. That said, the retained instances have to be discarded to adhere to the data streaming requirements as outlined in Table 24 and to operate within the physical limits of the available working memory in near-real-time [130]. As described in Figure 52, two activities take place in the Processing Layer that are integral components of the model. Firstly, the data stream attributes selection filter is applied based on expert knowledge to retain attributes that contain actual physical plant sensor values. Additionally, the processed instances are discarded to operate within the working memory and time constraints of data stream processing. However, this approach lacks support for DF readiness. To support post-incident investigation as part of DFIR, the Stream_to_offline component produces an offline dataset of the incoming data stream. The dataset is produced by collecting the instances as they arrive in the stream from the incoming instance connection until the last instance arrives.

The Classification Layer presented in Figure 52 takes the output from the Data_stream_attributes_selection filter from the Pre-Processing Layer of the model. Initially, the data stream is passed through an arbitrary incremental multi-filter that applies successive filters. The numeric attributes are standardised to have zero mean and unit variance. Whether dealing with batches or data streams, real world-data from sensors may become corrupted or erroneous during the data collection

stages. The model handles missing data values by imputing the values of the missing numeric attributes with means and the nominal attributes with modes.

Next, the model utilises the MOA incremental (adaptive) classifiers with warning and drift detection. The change detection mechanism within the incremental predictive algorithm is utilised to maintain the model's performance in the prevalence of the detected changes in the data stream. Whereas the drift detection component detects a deviation from the expected behaviour. As already stated, in data streams the entire dataset is not available. Therefore, there is no differentiation between train and test data subsets. The instances update the classifier incrementally and the prediction is made for each instance saving the results in a confusion matrix. Then the classifier is trained on that instance. The performance is computed every 1,000 instances and piped to a graphical and textual output for visualisation.

6.1.3 Algorithms Modelling

The A-ADC model has access to several data streaming predictive algorithms to eventually utilise the best-performing algorithm. By segmenting the model's architecture, see Figure 52, the study's approach ensures that the experiment is repeatable utilising different datasets and algorithms. The study compares the anomalous behaviour detection model for data streaming with the SPEAR super learner batch learning ensemble [22].

Large parts of ICS data exhibit normal behaviour and anomalies that can impact the efficacy of the classifier are typically the minority class. The anomalous behaviour triggers the phenomenon of sudden or gradual change in the data stream. The model

utilises the data stream rebalancing to address the problem of class imbalance batch by batch while training the model instance at a time [290]. The authors [290] present the concept of rebalancing data streams online and demonstrate the impact of data stream rebalancing on classifier performance utilising synthetic data streams.

Initially, the model utilises the Temporally Augmented Classifier (TAC) with ARF and ADWIN, see Figure 53a. Next, this study utilises the RS algorithm with ARF as the base learner [127] and HDDM_A_Test [145] as the base learner's warning and drift detection method, see Figure 53b. Then, this study utilises the RS algorithm with HAT [141] combined with ADWIN as the drift detection mechanism, as shown in Figure 53c and finally, HAT coupled with Perceptron (Pe), see Figure 53d. The related configurations are shown in Table 25.

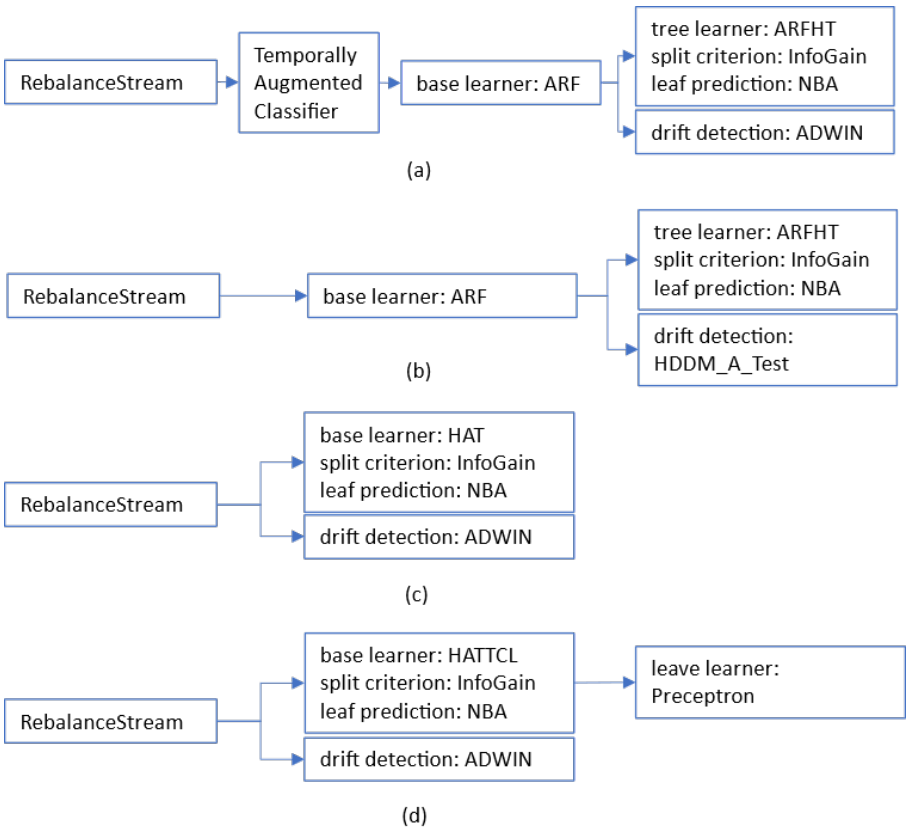


Figure 53 Algorithms evaluated as part of the A-ADC model on ICS data streams

The performance of the RS algorithm is compared with that of FeatureImportanceHoeffdingTreeEnsemble (FIHTE). The FIHTE algorithm is based on the traditional HT classifier that factors in feature importance [291], see Figure 54. The related configuration is provided in Table 25. This method utilises ensembles of incremental DT with ARF tree learner combined with HDDM_A_Test and the Mean Decrease in Impurity (MDI) to calculate the feature importance.

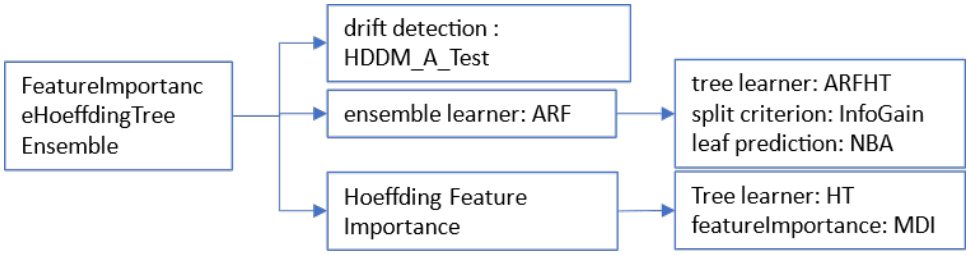


Figure 54 'Feature Importance Hoeffding Tree Ensemble' predictive algorithm as part of the A-ADC model on ICS data streams

Finally, a DDM classifier with the HDDM_A_Test is used. The DDM classifier independently detects a deviation from the expected behaviour and functions as a classifier in the ICS data stream distinguishing between normal and anomalous behaviours using NB as the base learner [145], see Figure 55 and related configuration in Table 25.

Figure 55 Drift detection algorithm as part of the A-ADC model on ICS data streams

Table 25 Configuration of the adaptive learning predictive algorithms as part of the A-ADC model on ICS data streams

Incremental Classifier	Incremental Classifier Configuration
RS_TAC_ARF_ADWIN Figure 53a	meta.imbalanced.RebalanceStream -l (meta.TemporallyAugmentedClassifier -l (meta.AdaptiveRandomForest -x (ADWINChangeDetector -a 0.001) -p (ADWINChangeDetector -a 0.01)))
RS_ARF_HDDM_A Figure 53b	meta.imbalanced.RebalanceStream -l (meta.AdaptiveRandomForest -l (ARFHoeffdingTree -e 200000 -g 50 -c 0.01 -z) -x HDDM_A_Test -p HDDM_A_Test)

RS_HAT_ADWIN Figure 53c	meta.imbalanced.RebalanceStream -l (trees.HoeffdingAdaptiveTree -e 100000 -z)
RS_HAT_P Figure 53d	meta.imbalanced.RebalanceStream -l (trees.HoeffdingAdaptiveTreeClassifLeaves -a functions.Perceptron -e 100000 -z)
FIHTE_ARF_HDDM_A Figure 54	moa.learners.featureanalysis.FeatureImportanceHoeffdingTreeEnsemble -l (meta.AdaptiveRandomForest -l (ARFHoeffdingTree -e 200000 -g 50 -c 0.01 -z) -x HDDM_A_Test -p HDDM_A_Test)
DDM_NB_HDDM_A Figure 55	drift.DriftDetectionMethodClassifier -d HDDM_A_Test

6.2 Case Studies

6.2.1 Liquid Storage and Water Distribution

In summary and as outlined in Chapter 4, section 4.7 and shown in Figure 41, the 'aNomalies' liquid distribution testbed consists of two tanks. The main tank contains four floating discreet sensors, and a secondary tank is fitted with an ultrasonic sensor. The testbed that can operate in manual or automatic modes via a touchscreen or remotely is fitted with two pumps, automated controls and infrastructure for data acquisition. The secondary tank is divided into 10,000 equal segments where 0 represents the value for a full tank and 10,000 reflects an empty tank. The pumps can be activated to an ON position represented as 1.0 and an OFF position represented as 0.0. in combination or alternatively. The pumps are triggered depending on the pre-determined level of the liquid in the tanks. The PLC consisting of ten registers utilises three registers to record the values of the testbed sensors. Register two, which records the values for the four discreet sensors utilises the first four bits of the PLC register. Register three contains the pumps' states utilising the

last two PLC register bits. Register four represents the ultrasonic sensor as a 16-bit integer value.

The dataset, outlined in Chapter 4, section 4.8 and Table 15 utilised for data stream simulation in our experiment was produced from the ‘aNomalies’ testbed [84]. The single ‘Register’ feature records all sensor types as the feature’s values. The dataset was adapted to simulate sensor-based data streams for incremental learning. The dataset features were rearranged, such that each sensor type is represented as an individual register feature in time series. The following expression shows an example measurement and a timestamp of the instances utilised; in = (tn, s1n, s2dn-x, s3pn-x, s4un, s5n, s6n, s7n, s8n, s9n, s10n,). Sensors s2d, s3p and s4u measure operational components either as a single sensor or a group of sensors. For example, s2d indicates the discreet sensors, s3p indicates the pumps sensors and s4u the ultrasonic sensors. The sensors' measurements are time-aligned. The normal behaviour of the ‘aNomalies’ dataset is stationary around a deterministic trend according to the statistical KPSS test [254], with a p-value >0.05 (significance level). That said, such characteristics may change over time or in the prevalence of anomalies introducing a concept of change in the data stream. According to statistical theory, the algorithmic error rate decreases in a stationary dataset, whereas the error rate increases as the distribution changes [143]. Therefore, this study hypothesizes that detected warning and change in a data stream with stationary characteristics is indicative of anomalous behaviour.

The Water Distribution Testbed (WDT) testbed is a hardware-in-the-loop testbed that emulates water passage between nine tanks using solenoid valves, pumps,

pressure and flow sensors. The WDT testbed uses a three-layer SCADA architecture. The WDT testbed consists of two main subsystems. The physical subsystem is made up of 5 tanks, 20 solenoid valves, and 5 pressure sensors under each of the tanks. Additionally, 8 manual valves are used for leak simulation. The simulated part of the testbed consists of 3 more tanks, 2 pumps, 4 flow sensors, 2 solenoid valves and 3 pressure sensors per tank. The first layer is the field instrumentation control layer which consists of the sensors and the actuators connected to the PLC. The second process control layer, consists of 4 PLCs and the third process control layer consists of the SCADA system [292, 293].

6.2.2 *Steam Turbine Power and Pumped-storage Hydropower Generation*

The Hardware-In-the-Loop (HIL) Augmented ICS (HAI) testbed emulates steam-turbine power generation and pumped-storage hydropower generation [294]. The testbed consists of 4 primary processes, namely the boiler process, the turbine process, the water-treatment process and the HIL simulator. The boiler process is controlled by utilising four controllers: water level, pressure, temperature and flow rate. The turbine process utilises a direct motor speed controller and the water-treatment process uses a level controller. Whereas the HIL simulation model is made up of two synchronous generator models and a power grid model with an electrical load [295].

6.2.3 *Datasets*

The details of the datasets utilised for the experimentation to simulate online learning data streams are shown in Table 26.

The ‘aNomalies’ dataset is a collection of labelled sensor values that consists of fifteen artefacts ranging from 134,225 to 189,048 instances [22, 84]. The dataset contains several differentiating factors. Besides the limited number of operational component sensors, the five real-world operational scenarios include normal, accident, sabotage, breakdown and cyber-attack behaviours that range from a few seconds to several minutes representing external attacks and insider threats. The prevalence of anomalies in the data subsets ranges between 0.5%-28%.

The characteristics of the WDT dataset represent a small-scale data stream, a large number of attributes, multiple anomalous occurrences and data imbalance. The dataset contains 9,206 instances, 40 attributes and an overall 15.85% prevalence of multiple types of anomalous instances.

The HAI dataset represents an imbalanced dataset that contains 38 attacks that include 14 primitives and 14 combination attacks. The HAI dataset is representative of a large dataset of 999,000 instances, 59 attributes and an overall prevalence of 1.83% anomalous instances.

Table 26 The datasets used to evaluate the A-ADC model

Dataset ID: Year	Instances	Anomalies	Attacks	Attributes
‘aNomalies’:2017	134,225-189,048	0.5-28%	15 situations covering 5 real-world operational scenarios,	3
WDT:2021	9,206	15.85%	28 (only 18 with an effect on the physical processes)	40
HAI:2020	999,000	1.83%	38 including 14 primitives 14 combinations	59

6.2.4 Post-incident Investigation

Thus far, this research study has addressed the applicability of the model to ICS anomalous behaviour detection as part of a defence-in-depth capability. Besides anomalous behaviour detection, the research considers the role of this model to support DF readiness. DF capabilities are needed to understand ICS events for post-incident investigation and event reconstruction [296].

This study illustrates the A-ADC model's applicability to support modern DFIR. For coherence with the discussion in chapter 5, section 5.5.2, this chapter refers to the phases of frameworks and standards such as the DFRWS and the ISO/IEC 27050:2016, see Figure 56. Additionally, it referred to a prior study [56] that proposed the categorisation of the data stages into physical, logical and legal contexts. The physical context relates to the data captured from seized physical media corresponding to the identification and preservation of data lifecycle phases. This research study asserted it was not always feasible to seize the physical media in operations-critical and dispersed infrastructures such as ICS that ubiquitously produce large quantities of data. Other mechanisms are required to gather DE such as novel ML techniques that can continuously process data streams from physical plant sensors to detect anomalous behaviour.

This chapter addresses the logical context with the related phases of the digital investigation models shown in Figure 56. As stated for data streaming anomaly detection, processed data instances are discarded to adhere to the data streaming paradigm, see Table 24 [130]. Therefore, datasets are not produced natively, limiting support for DF readiness such as event reconstruction and identification of prior patterns. The A-ADC model's support for modern DFIR is twofold. Firstly, the model

utilises adaptive learning to detect anomalous behaviour from data stream instances. Secondly, the Stream_to_offline component produces a dataset from the incoming data stream instances to support post-incident investigation and event reconstruction.

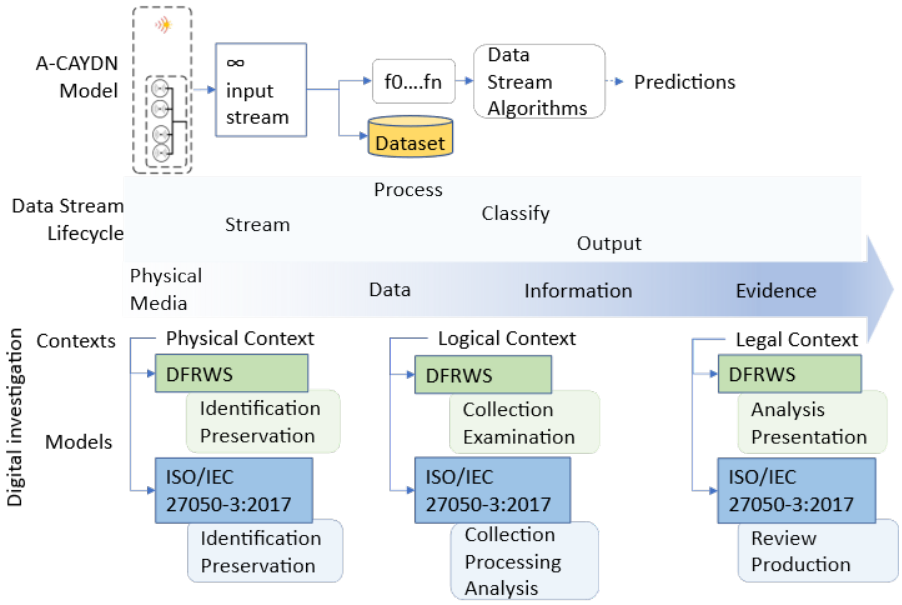


Figure 56 The A-ADC model feasibility to support modern DFIR from data streams

However, to satisfy the legal context, DE admissibility in ICS should be considered. Previous studies studied the concept of DW [297-300]. Moreover, as part of DF readiness by design, the authors [6, 56] proposed that CPO can be modelled as a DW to support DFIR. In this instance, a DW refers to a CPO that is capable to maintain admissible evidence to the Court of Law, similarly to a human witness [6, 298]. A chain-of-custody needs to be achieved and maintained throughout the investigation to address the requirements of the legal context. However, without maintaining integrity, the data could be falsified. Mechanisms such as Blockchain (BCh) are explored as a method for secure logging to maintain the integrity of the collected data [296, 301]. In [56] the authors proposed that a forensic-enabled design could be achieved by automating the identification and preservation

stages by utilising Blockchain-based systems. However, further research is needed to understand the constraints and feasibility of these techniques in ICS [296, 301-303]. While the A-ADC model is pertinent to support the logical stages of the DFIR in CPS, the admissibility of DE in ICS is a complex problem that merits further research.

6.3 Performance Evaluation

6.3.1 Metrics

Evaluation of binary classification adopts several statistical metrics depending on the experiment such as F1-scores which are computed from the confusion matrix and often used, see Figure 38 [256]. Unlike the F1-score given by equation (4.4), Precision and Recall given by equations (4.2) and (4.3) respectively, the MCC is intended to handle highly imbalanced datasets [256, 304]. The MCC score for binary classification is given by equation (6.1).

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FN} \times \mathrm{FP}}{\sqrt{(\mathrm{TP}+\mathrm{FN}) \times (\mathrm{FN}+\mathrm{TN}) \times (\mathrm{TP}+\mathrm{FP}) \times (\mathrm{FP}+\mathrm{TN})}} \tag{6.1}$$

The MCC factors in the performance of all four categories from the confusion matrix, see Figure 38, proportionate to the dataset's sizes of positive and negative elements [305, 306]. Furthermore, classifiers could exhibit high TP and low FP values. However, in the absence of reporting the number of unclassified instances, these results could be misleading about the classifier's performance. For example, if the rate of unclassified instances is high, this could still result in weak classifier performance. Therefore, the metrics of reporting the weighted MCC score, the

incorrectly classified and the unclassified instances are justified in demonstrating this classifiers' performance.

The metrics on the progress of the model's efficacy were reported every 10,000 instances. The results of the algorithms for the liquid storage and distribution case study using the 'aNomalies' dataset are provided in Table 27 and Table 28. The RS_ARF_HDDM_A and FIHTE_ARF_HDDM_A algorithms achieved the highest MCC_w scores and lowest rate of unclassified instances, see Table 28.

Table 27. Overall MCC_w scores and K statistics for the 'aNomalies' data stream

Incremental Classifier	K scores	MCC_w scores
RS_ARF_HDDM_A	0.9986	0.999
FIHTE_ARF_HDDM_A	0.9987	0.999
RS_HAT_ADWIN	0.9739	0.975
RS_HAT_P	0.2445	0.287
RS_TAC_ARF_ADWIN	0.9980	0.998
DDM_NB_HDDM_A	0.9988	0.9988

Table 28 Overall instances classification rate for the 'aNomalies' data streams.

Incremental Classifier	Unclassified [%]	Incorrectly classified [%]	Correctly classified [%]
RS_ARF_HDDM_A	0.001343	0.003371	99.995307
FIHTE_ARF_HDDM_A	0.001400	0.003000	99.995643
RS_HAT_ADWIN	0.005200	0.060836	99.933960
RS_HAT_P	0.000657	13.792786	86.206529
RS_TAC_ARF_ADWIN	0.001500	0.005914	99.992690
DDM_NB_HDDM_A	0.002810	0.002760	99.994400

The comparison of the classifiers' MCC_w scores and the percentage of the anomalous instances prevalent in the data stream is given in Table 27 and Table 29 describes the operational scenarios. The attained results from the operational scenarios are shown in Figure 57 whereas the results achieved from the two best-performing classifiers are presented in Figure 58.

Table 29 Operational Scenarios 'aNomalies' dataset

File	Scenario - Type	Sensors	Labels
1	Normal	None	All anomalies
2	Plastic Bag	Ultrasonic	Plastic_bag
3	Blocked measure 1	Ultrasonic	Spoofing
4	Blocked measure 2	Ultrasonic	High_blocked

5	2 floating objects, main tank	Ultrasonic	Second_blocked
6	7 floating objects, main tank	Ultrasonic	Bad_connection
7	Humidity	Ultrasonic	DoS_attack
8	Failure of a discreet sensor	Discreet 1	Hits_3
9	Failure of a discreet sensor	Discreet 2	Wet_sensor
10	Denial of Service attack	Network	Poly_2
11	Spoofing	Network	Poly_7
12	Wrong Connection	Network	Hits_2
13	Tank hit, low intensity	All system	Hits_1
14	Tank hit, medium intensity	All system	Blocked_1
15	Tank hit, high intensity	All system	Blocked_2

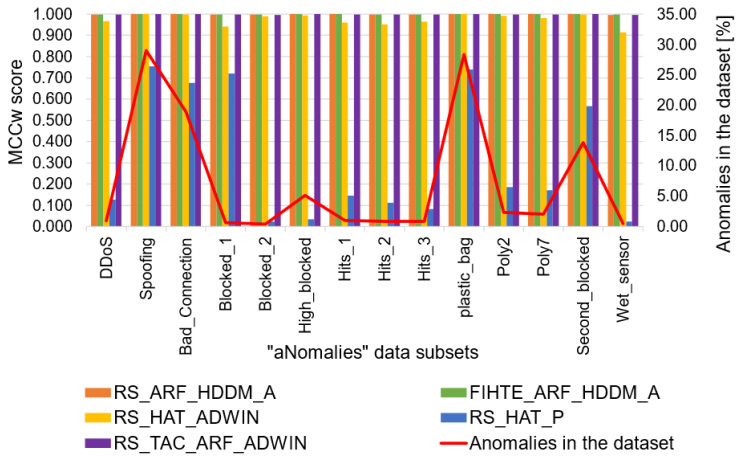


Figure 57 Comparison of the classifiers' MCC_w scores and the percentage of anomalies in the 'aNomalies' data stream.

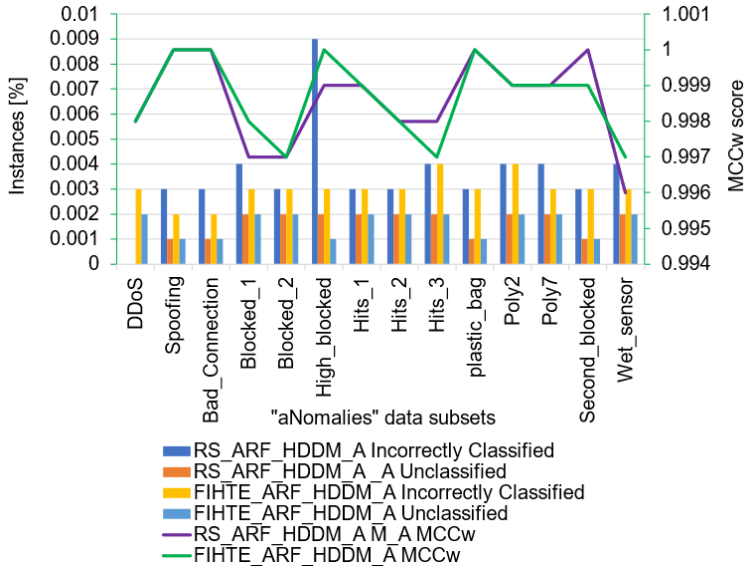


Figure 58 Comparison of the two best-performing classifiers' weighted MCC scores and the percentage in the 'aNomalies' data streams.

The experimentation reveals that the classifiers perform well in a binary classification with an occurrence of one type of anomaly, see Table 27 and Table 28 and multiple types of anomalous behaviour, see Table 30 and Table 31. According

to the literature [129, 130], classification accuracy is correlated with the number of available training instances. The training instances are expected to be inversely impacted by increases in change detections, where reducing the number of available training instances is expected to lead to lower classification accuracy. This is demonstrated according to the results recorded for the WDT data stream in Figure 59, Table 30 and Table 31.

Table 30 Performance metrics of the weighted MCC score and the Cohen Kappa (K) statistics for the WDT and HAI data streams

Incremental Classifier mFeaturesMode =60	WDT scores for		HAI scores for	
	K	MCC_w	K	MCC_w
RS_ARF_HDDM_A	0.9332	0.933	0.9809	0.981
FIHTE_ARF_HDDM_A	0.9332	0.949	0.9809	0.981
RS_TAC_ARF_ADWIN	0.9486	0.949	0.9862	0.986
DDM_NB_HDDM_A	0.9245	0.925	0.9845	0.985

Table 31 Performance metrics of the unclassified, incorrectly classified and correctly classified instances for the WDT and HAI data streams.

Incremental Classifier mFeaturesMode =60	Unclassified [%]	Incorrectly classified [%]	Correctly classified [%]
WDT			
RS_ARF_HDDM_A	0.51	1.79	97.69
FIHTE_ARF_HDDM_A	0.51	1.79	97.70
RS_TAC_ARF_ADWIN	0.54	1.38	98.08
DDM_NB_HDDM_A	1.62	1.92	96.46
HAI			
RS_ARF_HDDM_A	0.62	0.07	99.32
FIHTE_ARF_HDDM_A	0.62	0.07	99.32
RS_TAC_ARF_ADWIN	0.28	0.05	99.67
DDM_NB_HDDM_A	0.39	0.05	99.56

The results for the steam turbine power and pumped-storage hydropower generation for the HAI data stream are presented in Table 30 and Table 31. The optimisation was carried out to improve the model's performance. The mFeaturesMode parameter was tuned for the RS_ARF_HDDM_A improving the MCC_w scores from 0.981 (mFeaturesMode=60) to 0.987 (mFeaturesMode=10) while reducing the incorrectly classified instances from 0.07% to 0.0474% respectively.

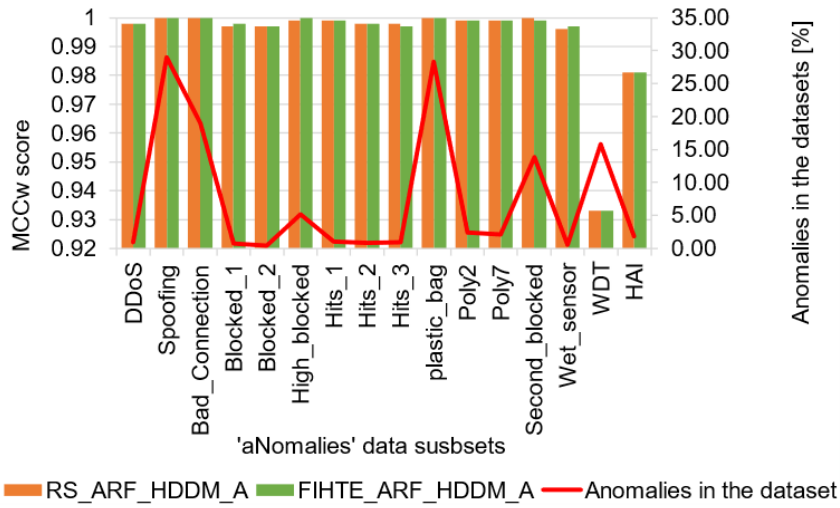


Figure 59 Comparison of the two best-performing classifiers' weighted MCC scores in the tested datasets.

6.3.2 *Performance Benchmark Criteria Framework*

A persistent lack of suitable and up-to-date benchmark datasets to evaluate the performance of data stream classifiers remains a profound disadvantage in accelerating scientific research [130, 307, 308]. Drawbacks of synthetically generated and benchmark datasets used to evaluate incremental learning, such as those listed in Table 32, include not reflecting current ICS environments and up-to-date adversary challenges [307].

The state-of-the-art addresses performance evaluation metrics as discussed in 6.3.1. However, the metrics do not factor in performance variations due to different dataset characteristics as illustrated in Table 26 and Table 32. Besides approaches to evaluate drift detectors, scientific literature uncovers the lack of consistent use of metrics and explains the drawbacks of some commonly used metrics such as classifier accuracy [307, 309-312]. In addition, the state of the art does not factor in changes due to evolving design and iterations across different types of environments

such as the research environments described by [313] in their vision for a new generation of smart machines.

Table 32 Benchmark datasets characteristics used in concept drift detection.

Benchmark datasets	Description	Instances	Year
Electricity pricing [314]	Australian New South Wales electricity market prices in 30 mins intervals	45,312	1999
Forest Covertype [315]	54 attributes that describe types of forest cover in the Roosevelt National Forest in Northern Colorado.	581,012	1998
KDD'99	23 class labels, 41 attributes that simulate military network intrusion	494,000	1999
Poker hand [316]	11 attributes that represent a 5-card poker hand of a 52-piece card deck	1,000,000	2007
Airlines [317]	13 attributes representing flight departure and arrival with related delays	120,000,000	2008

This research thesis proposes a novel Performance Benchmark Criteria (PBC) framework to address these shortcomings for ICS. This chapter introduces the concept of objectively quantifying the performance of incremental classifiers across different levels of cyber-physical experimental environments. This is achieved by producing metric criteria groups: Base, Temporal and Environment, as illustrated in Figure 60. The Base group contains performance and classification impact metrics. The Base group represents fundamental metrics that are constant within and across experimental environments and data sources. The Temporal group reflects the data characteristics that change between data sources and across experimental environments. The attributes of the Environment group are anticipated to vary across different experimental environments. The interaction of the elements guides the design of the proposed framework.

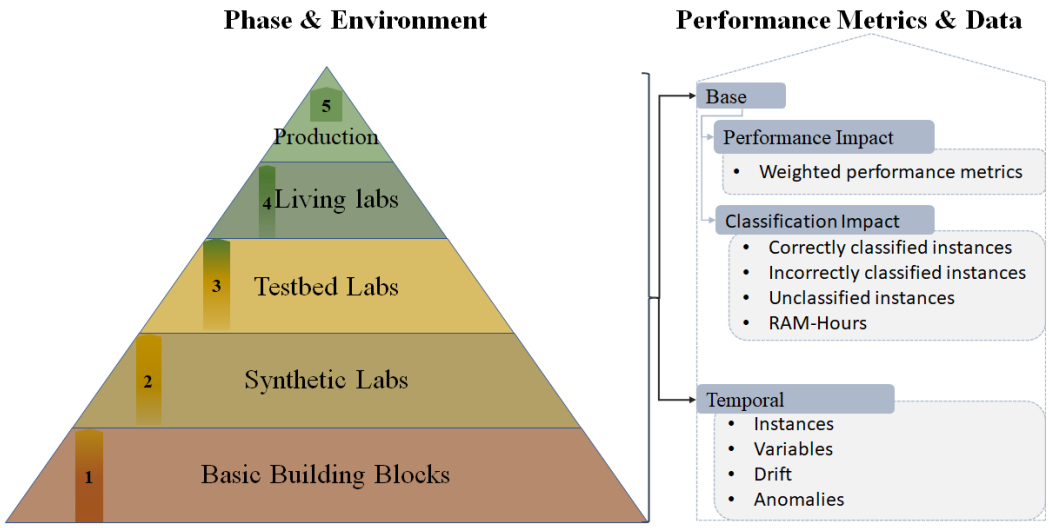


Figure 60 The Performance Benchmark Criteria (PBC) framework

This study defines the **Building Blocks** as a segment consisting of components of algorithms, libraries and modules that contribute to the design of the basic model. This segment typically uses synthetic or publicly available datasets. The **Synthetic Labs** phase includes maker and virtual environments that enable ideation and collaboration on the components of the basic building blocks to develop subsystems. This phase generates datasets from synthetic environments. Whereas, the key characteristic of the **Testbed Lab** phase is the safety of an environment to enable reproducibility, consistency and predictability. Datasets generated in these environments are often contributed to the scientific body of knowledge. While developing and iterating the idea requires attributes that characterise the first three phases of the proposed framework, **Living Labs** create a test environment to represent real-world challenges. These include physics, socio-technical influences, component noise and exposure to current adversary challenges that characterise real-world applications of unpredictability, diversity and exceptions. The datasets generated from Living Labs are valuable to researchers, policymakers and regulators. However, these datasets are likely to require curation due to additional

challenges. These include privacy and exposure of sensitive environmental information. Thus datasets are not always made available in publicly accessible repositories. The anticipated performance of the classifier within the **Production** phase should be informed by the performance benchmark as the baseline metric.

The Base metrics derive values from established evaluation techniques to capture the classifier and drift detector performance. Besides the performance metrics outlined in section 6.3.1 other performance criteria have been proposed, some with varying levels of application in published work while others with the challenge of the metrics' suitability [287, 307, 309-312]. The Temporal metrics derive the values from the characteristics of the data source. The data source characteristics such as the number of instances, attributes and computational environment impact incremental classifiers' performance. The data sources to evaluate the incremental classifiers vary across the experimentation environments, see Figure 60. Therefore, defining a common standard for metrics and reference datasets is an important challenge that merits further empirical research.

6.3.3 Model's Effectiveness

The goal of the model is to identify the minority class observations in the data stream. Although classifiers could exhibit strong performance in the majority class, such performance could be misleading due to the impact of the minority class observation in an imbalanced dataset [146]. While accuracy is considered an effective metric for balanced datasets, scientific literature highlights shortfalls in using classifier accuracy to evaluate incremental classifiers [141, 304, 307, 309-311]. For example, let's assume 1% anomalies prevalent in a dataset with a binary classifier

labelling each instance in the dataset as normal, classification achieves 99% accuracy despite not labelling any of the anomalous instances correctly. Hence, this study derived performance metrics including MCC, F1 and K scores, as produced in Table 27 and Table 30. In Table 28, Table 31 and Figure 58 the study reports the unclassified and incorrectly classified instances.

This research applied data stream rebalancing to deal with the problem of an imbalanced data stream while maintaining the data streaming paradigm [290]. Furthermore, it applied an online change detection method based on Hoeffding's Bounds moving averages to detect changes in the data streams [145]. Compared to batch learning, online learning utilises the method of concept change through moving averages where the deviation of the data distribution does not require an updated model.

The results demonstrate that the RS_ARF_HDDM_A and the FIHTE_ARF_HDDM_A algorithms produced consistently high MCC_w and K scores with fewer unclassified and incorrectly classified instances. Using the 'aNomalies' data stream produced an overall MCC_w score of 0.999 and a K score above 0.998 for the RS_ARF_HDDM_A and FIHTE_ARF_HDDM_A. Both the algorithms scaled to the HAI data stream achieved 0.981 and 0.9809 MCC_w and K scores respectively. While the RS_TAC_ARF_ADWIN algorithm achieved a slightly higher MCC_w score of 0.986 for the HAI data stream, it produced lower MCC_w scores for the 'aNomalies' data stream, see Figure 57 and Table 27. Likewise, RS_ARF_HDDM_A and FIHTE_ARF_HDDM_A algorithms produced fewer unclassified and incorrectly classified instances for the 'aNomalies' data stream, see Table 28. Finally, optimising

the RS_ARF_HDDM_A algorithm improved the MCC_w scores to 0.987 (mFeaturesMode=10) while reducing the incorrectly classified instances from 0.07% to 0.0474% respectively for the HAI data stream.

The results of our preliminary experimentation found that the k-NN SAM algorithm produced high MCC_w and K scores of 0.989 for the HAI and 0.959 for the WDT data streams. While the results partially agree with another study [126], this research found the overall instances throughput was poor. Algorithms such as RS_HAT_P and RS_HAT_ADWIN showed inferior performance and did not scale effectively.

The promising results make this research relevant in the context of addressing cyber and physical attacks within ICS as part of a defence-in-depth approach. The A-ADC model has been empirically evaluated utilising ICS data [84, 292-294]. Optimisation of parameters including mFeaturesMode and mFeatureperTreesize of the ARF base learners reveals a further decrease in the rate of unclassified instances and an increase in the flow throughput. Therefore, a direction of future work in this area could focus on additional model tuning. The results demonstrate that the A-ADC model can adapt to changes dynamically, and scale to different data stream characteristics while maintaining classification performance. Thus, the model is considered suitable for identifying normal and anomalous behaviour in ICS data streams.

6.3.4 Comparison of the Learners

Varying consistency in reporting dataset characteristics such as attack duration or anomalous instances within the dataset makes performance comparison

challenging. Apart from the variations in the computational environment, the training dataset pre-processing including feature selection and parameter utilisation could impact the effectiveness of a fair comparison between algorithms. Another issue is the split between the training and testing datasets which could differ between models thus augmenting the challenges of achieving an unbiased comparison of ML models. Likewise, the type of learning, batch learning and data stream learning evaluations fundamentally differ. The output of batch learning is a static model. The model reuses data, leveraging techniques such as cross-validation and hold-out to measure the generalisation of the model and to compare results. Whereas data streams process each instance in a limited memory allocation and time and need to be ready to predict at any time. The focus in data streams is on model efficacy over time. Incremental learning uses prequential, also known as interleaved-test-then-train and holdout evaluations. The trade-off in approaches such as holdout or prequential depends on the expected characteristics of the data stream such as the prevalence of concept drift and availability of holdout instances.

Despite using different ML approaches, to achieve a fairer comparison when training different models, this thesis considers performance metrics that can be leveraged by both ML models. Primarily, the focus of this thesis is a fair comparison of the SPEAR and A-ADC models' effectiveness in detecting anomalous behaviours using comparable metrics, datasets and computational environment. This research considered the limitations of unbiased comparison given the different ML approaches. It does not seek to compare the results of one model over the other. Instead, it presents a meaningful comparison of the effectiveness of different models to detect anomalies given comparable constraints and performance metrics. This

research thesis utilises the default processing environment memory allocation and leverages datasets with a varying number of instances, anomalies and attack types, see Table 26. A range of metrics on progress was reported every 10,000 instances. The performance of the A-ADC model was compared with the SPEAR framework super-learner ensemble supervised learning model that utilised the default parameters and 10s rolling windows [22]. This study compares the overall and anomalous behaviour F1-score and the Recall values. The results reveal that the adaptive learning model, see Table 33, achieved consistent overall $F1_A$ and $Recall_A$ scores of at least 0.997 for the anomaly class compared with the SPEAR framework's model of 0.990. The summary of the overall results is presented in Table 33 and individual attacks are in Figure 61 and Figure 62.

The results of this research thesis concerning the WDT dataset are consistent with that of [293], who evaluated the performance of several ML classifiers including RF against the WDT dataset. The RF algorithm achieved 0.97 and 0.98 F1 and Recall scores respectively for the physical dataset. Whereas these results demonstrate the ARF-based algorithms RS-ARF_HDDM_A and FIHTE_ARF_HDDM_A achieve a consistent F1 and Recall scores of 0.982 whereas the RS_TAC_ARF_ADWIN has a slightly higher score of 0.986. The summary of the F1 and Recall scores is shown in Table 34.

Table 33 Performance comparison of the SPEAR Framework Super-Learner and A-ADC Model classifiers.

Classifier	$F1_A$	$Recall_A$
Batch Learning SPEAR Framework		
Default parameters 10s rolling window	0.990	0.990
Adaptive Learning A-ADC model		
RS_ARF_HDDM_A	0.999	0.997
FIHTE_ARF_HDDM_A	0.999	0.998
DDM_NB_HDDM_A	0.998	0.996

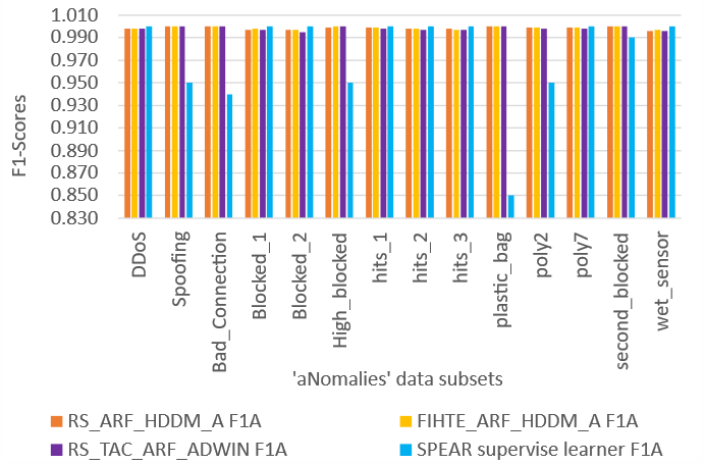


Figure 61 F-score comparison of the SPEAR Framework Super-Learner and A-ADC Model classifiers

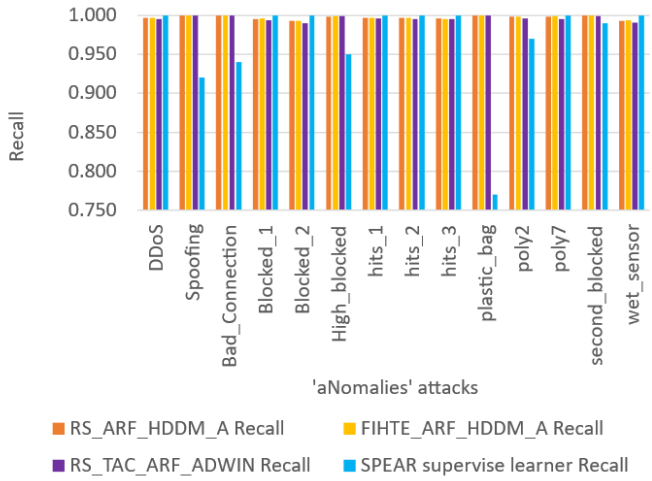


Figure 62 Recall comparison of the SPEAR Framework Super-Learner and A-ADC Model classifiers

Table 34 F1 and Recall performance metrics for the A-ADC Model classifiers for the WDT dataset.

	RS_ARF_ HDDM_A		FIHTE_ARF_ HDDM_A		RS_TAC_ARF_ ADWIN	
	Recall	F-score	Recall	F-score	Recall	F-score
Normal class	0.987	0.989	0.987	0.989	0.989	0.992
Anomaly class	0.953	0.944	0.953	0.944	0.972	0.957
Weighted Average	0.982	0.982	0.982	0.982	0.986	0.986

7. Conclusion and Future Work

7.1 Conclusion

The original contributions of this research commenced with an SLR of peer-reviewed literature with an explicit investigation of empirical primary studies that addressed cyber resilience and DFIR of CPS in smart cities. The SLR provided scientific evidence of the gaps in the literature and an evidence-based summary of key themes. This phase of the research study was published by MDPI Smart Cities [5].

This research developed an ML model-driven framework (SPEAR) to profile anomalous behaviour in ICS innovatively combining physical plant sensor-generated data with AI to address anomalous behaviour detection in ICS. Consideration was given to the threat landscape and the evolving threat model. Factoring in anomaly types and ML techniques, detection algorithms comprising a super-learner ensemble and hybrid unsupervised learning models for binary classification were developed. The learning models were underpinned by detailed procedure design and a pilot phase experimentation. The framework and the associated learning models were experimentally validated on an ICS liquid storage and distribution case study utilising a set of performance metrics. Further, a cyber risk quantification model (CRVQ) was derived demonstrating the concept of objectively producing a cyber risk score for a materialised cyber risk. This phase of the research study was published in the IEEE Internet of Things Journal [22].

Another model was then proposed for anomalous behaviour detection from ICS data streams from physical plant sensors. A novel adaptive learning model for one-class classification was developed, leveraging stream rebalancing and drift detection methods. The detailed design and piloting phase was followed by experimentally evaluating the model on ICS case studies. These included the ICS liquid storage and distribution, steam turbine power and pumped-storage hydropower generation case studies using a set of performance metrics. Further, a comparison of the performance of the proposed model with batch learning approaches was carried out. In addition, a performance benchmarking criteria framework was introduced. The framework aimed to coherently quantify the performance of classifiers across different cyber-physical experimental environments. This phase of the research study was submitted for publication and is currently under review.

Further, besides proposing a mechanism for detecting anomalous behaviour, the concept of DW as part of DF readiness in ICS was investigated. A use case presented the concept of integrating BCh technologies into the design of ICS to support modern DF readiness. BCh, a disruptor technology, due to its distributed nature is protected from integrity attacks and its immutable timestamps could offer novel approaches to achieving DCoC in ICS supporting DF readiness. This element of the research study has been published in IEEE Blockchain Technical Briefs 2022 [318].

At this stage, we evaluate the aims and objectives to evaluate this research study. The thesis has:

“two directions sharing the main goal to improve cyber resilience in ICS”.

The first aim of this research study was to:

“investigate and develop a security mechanism to improve the proactive cyber defence in ICS to support its mission objectives. The mechanism seeks to be security process-driven, integrate novel ML techniques and utilise data generated from physical plant sensors. The mechanism intends to be testable, trustworthy and repeatable”.

And because the increase in automation and interconnectedness in ICS widens the attack surface with threats ranging from external adversaries to insiders creating genuine security concerns of potentially catastrophic consequences [6, 18, 19, 249], the research study aimed secondly to

“investigate how the security mechanism addresses the reactive defence as part of DF readiness in ICS”.

The aims were advanced by setting out and addressing the following objectives:

- **Objective 1:** Conduct a Systematic Literature Review (SLR) of current cyber resilience and DFIR approaches in CPS in smart cities. This objective was addressed in Chapter 2 by conducting an SLR of scientific literature reporting on frameworks and systems that addressed CPS’ cyber resilience in smart cities to address the RQs outlined in Table 1.

The protocol utilised to achieve this objective was based on the SLR guidelines for the computer engineering discipline proposed by Kitchenham and Charters [58], which present a rigorous and credible methodology. The SLR analysis uncovered emerging themes and concluded with several key findings, among which:

- The chronological analysis of key events revealed some of the important influencing factors including Industry 4.0, government-led support and initiatives such as the National Cyber Security Strategy in the UK [66] or national infrastructure plans [66, 164, 167], innovative ideas [195] and incidents [64, 319] were among some of the key influencing factors.
- The data source analysis showed a lack of real CPS datasets with a predominant use of software-based simulations and simulation infrastructure. However, reliance on simulators may not sufficiently represent the threats compared with datasets generated from a real-world environment. For example, datasets generated from isolated or simulated environments are likely to be constrained in enabling the understanding of prevalent threat types in the context of the actual CPS.
- Further analysis of cross-sector proposals or applications to improve DF showed a distinct lack of research focusing on cyber resilience in some smart sectors such as smart healthcare and smart citizen were addressed by a small number of studies. This SLR concluded that some smart sectors have complex and diverse ethical challenges whereas, media prominence of critical infrastructures attacks may be factors that focus the spotlight on research in those sectors.

- **Objective 2:** Investigate the current ML approaches to improve the proactive cyber defence of ICS. This objective was addressed in Chapters 2, 4 and 6 by investigating the types of anomalies, followed by the application of relevant ML techniques to address the RQs outlined in Table 1.

This research investigated the ICS threat landscape including critical infrastructure and major ICS cybersecurity attacks, concluding from the findings that threats in ICS are more basic and the ICS security less well-understood compared with the well-established field of IS. Furthermore, despite the level of widely acknowledged and well-reported high-profile attacks remaining low, the threat actor motivations evolved.

Apart from the threat actors' evolving techniques, tools and procedures, this study factored in the numerous dataset-related challenges. Against this backdrop, a thorough investigation of the types of anomalies in the context of the application of ML techniques was carried out. An in-depth investigation of ML approaches leveraged for anomalous behaviour classification in ICS set the groundwork for addressing accidental and malicious activities as part of a modern defence-in-depth approach. Following that, consideration was given to the ubiquity and characteristics of data produced from sensors, actuators and controllers creating continuous data streams. Since data streams evolve and data distribution can change, online ML techniques capable of handling data streams to process data dynamically and adapt to changing and scaled data were investigated.

- **Objective 3:** Develop an approach that leverages ML techniques to improve cyber resilience in ICS. This objective was addressed in Chapters 3, 4 and 6 by developing ML model-driven approaches to address the RQs outlined in Table 1.

A novel Super learner Ensemble Anomaly detection cyber risk quantification (SPEAR) framework was proposed to profile anomalous behaviour in ICS. As part of the framework, a supervised learning super learner ensemble and hybrid unsupervised learning model were constructed. Detailed procedure design was developed factoring in the complexities of the data preparation stages. Pilot experimentation was employed as part of evaluating the instruments and the procedure design. Furthermore, a novel Adaptive learning for Anomaly Detection and Classification (A-ADC) model for data streams was introduced for near-real-time profiling of anomalous behaviour from ICS physical plant sensor data streams. The model performed novel computation, innovatively combining adaptive ML algorithms with drift detection methods. The approach of segmenting the model's architecture aims to achieve repeatability for different data streams and algorithms.

- **Objective 4:** Develop a novel ML-based anomaly detection and cyber risk quantification mechanism, evaluating and analysing the efficacy. This objective was addressed in Chapters 5. The RQs outlined in Table 1 are addressed by experimentally evaluating the models presented in Objective 3 and by proposing a BBN model as part of the SPEAR framework to quantify cyber risk in the prevalence of anomalous behaviour in ICS.

The SPEAR framework and associated learning models were experimentally validated on an ICS liquid distribution case study. The proposed approach showed promising results, an overall F1-score of 99.13%, and an anomalous recall score of 99% detecting anomalies lasting only 17 seconds ranging from 0.5% up to 89% of the dataset.

The A-ADC model's experimental validation produced an overall Matthews Correlation Coefficient (MCC) score of 0.999 and Cohen's Kappa (K) score of 0.9986 on limited variable single-type anomalous behaviour per data stream. Wide data streams achieve an MCC score of 0.981 and a K score of 0.9808 in the prevalence of multiple types of anomalous instances.

- **Objective 5:** Investigate the mechanism's support for post-incident investigations as part of DF readiness. This objective was addressed in Chapters 2, 5 and 6 by investigating and demonstrating how sensing capabilities create opportunities to protect ICS as part of defence-in-depth approaches.

The SLR investigated how the identified primary studies addressed modern DFIR and further what were the efforts to utilise interactions in CPS to improve DFIR. Apart from insights gained about the factors influencing the findings and emerging themes outlined in Objective 1, the DFIR analysis asserted that CPS-related cyber resilience and DFIR were active research domains. However, the in-depth analysis uncovered that these areas have not been extensively considered by researchers in the context of CPS and modern DF readiness. The findings pointed to differences in scientific interest in the DFIR stages with variations across smart sectors and research

diversification to other smart city sectors an ambition of future research directions. For example, the DFIR analysis revealed that 67.308% of the investigated studies focused on the detection and analysis stages of IR.

Furthermore, this research scrutinised the applicability of the SPEAR framework and the A-ADC model to support DFIR as part of DF readiness to improve defence-in-depth in ICS. The research concluded that the SPEAR framework could be applied innovatively as part of a post-incident investigation, reconstruction of events and identification of prior patterns. For example, due to the operation-critical and dispersed nature of ICS ubiquitously producing large quantities of data, it is not always possible to seize physical media. Innovative methods of gathering DE are needed to exploit the opportunities created by the pervasive nature of data. This could be achieved by modelling cyber-physical objects as DW coupled with the concept of DCoC to support modern DFIR.

7.2 Future Work

This research developed a novel ML model-driven framework to detect and classify anomalous behaviours and derive a cyber risk score from ICS physical plant sensor data. Although the laboratory studies produced promising results, future research is still required to address additional optimisation and generalisation issues. For example, the data produced from physical plant sensors are ubiquitous and the data distribution evolves. While the fusion of OT and ICT increases process control, monitoring and automation, the increased integration of connected technologies into daily lives creates an expanding and dynamic attack surface with numerous attack

vectors. Besides targeted attacks from external adversaries including APT and insider threats, ICS are predisposed to challenges resulting from the disparity between security and operational priorities of ICT and OT, which further complicates the protection of ICS from cybersecurity threats [18, 19]. Hence, ICS have to adapt to a complex and evolving threat landscape. The integration of near-real-time prediction and the trade-off to maintain the model efficacy is a challenging problem that merits further research. Additionally, further research is vital to integrate CTI into the modelling of cyber-attacks against critical functions to support modern defence-in-depth strategies for the smart cities of the future.

Moreover, this research scrutinised the applicability of the framework as part of reactive cyber defence. For example, logs systematically record events in digital systems that help understand ongoing and occurred events. Secure logs protection is a well-understood technique in computer security to maintain the integrity of the logs to support incident responders. Apart from investigating digital crimes, logs can be leveraged to deal with insider threats such as accidental hazards and professional errors. Hence, secure logs form an important part of DF readiness. In a digital investigation, any piece of data is a potential DE artefact. However, the method in which DE is handled influences whether that digital artefact is admissible in the Court of Law [5]. For example, DE requires a DCoC, which maintains traceability of the digital artefact to ensure attribution, specifically referencing the chronology of ownership, custody and location of the DE.

Against this backdrop, due to its security capabilities, Blockchain (BCh) technology can be integrated into ICS as an enabler to achieve modern DF

readiness. Apart from operational activities such as process control and automation, physical plant sensors could collaborate on providing specific cyber threat intelligence from the physical plant's sensing capabilities to support incident investigations. However, in pervasive systems particularly critical infrastructures, it is not always viable to seize physical media to gather DE. While the surge of disruptive technologies integrated into ICS introduces complexities and increases their attack surface, disruptive technologies such as BCh present an opportunity to leverage the concept of DW to support investigations. The concept of DW in ICS is illustrated in Figure 64 in section 9.6. DW is referred to as cyber-physical objects with functional sensing capability to confirm a crime-related event [6, 56, 298], in this instance extended to accidental and malicious anomalous activities detected from ICS physical plant sensors. DF readiness should be carefully embedded into the ICS design and architecture as well as industrial business practice, illustrated in Figure 65 in section 9.6 and Figure 66 in section 9.7. Therefore, future research is vital to focus on modelling cyber-physical objects as DW to support DF readiness. That being said, the integration of forward-looking innovative enablers such as BC technology to achieve DCoC as part of DF readiness in ICS remains limited. The complexity of ICS environments by utilising DW to achieve DF readiness requires extensive empirical research.

The results of evaluating the ML models developed throughout the research are based on laboratory conditions and theoretical discussions. Countermeasures were taken to maintain the characteristics of a real-world environment by utilising a dataset from a purpose-built testbed. However, a further appropriate field study will create an environment representing real-world challenges giving more accurate data to

investigate how the components of the framework would operate under such conditions in practice. An actual field study can last for many months allowing the capture of a range of anomalous activities. However, utilising actual environments in ICS many of which support critical infrastructures could have catastrophic consequences. Therefore, defining a common standard for metrics and reference datasets across different types of experimental environments is an important challenge for future research.

8. References

[1] B. Vogel-Heuser, and D. Hess, “Guest Editorial Industry 4.0–Prerequisites and Visions”, IEEE Transactions on Automation Science and Engineering, vol. 13, no. 2, pp. 411-413, 2016, [Online], https://doi.org/10.1109/TASE.2016.2523639

[2] N. Moustafa, E. Adi, B. Turnbull, and J. K. Hu, “A New Threat Intelligence Scheme for Safeguarding Industry 4.0 Systems”, IEEE Access, vol. 6, pp. 33910-33924, 2018, [Online], https://doi.org/10.1109/access.2018.2844794

[3] M. Lom, O. Pribyl, and M. Svitek, "Industry 4.0 as a part of smart cities", in 2016 Smart Cities Symposium Prague (SCSP), pp. 1-6, 26-27 May 2016, IEEE, https://doi.org/10.1109/SCSP.2016.7501015

[4] M. Postránecký, and M. Svítek, "Smart city near to 4.0 — an adoption of industry 4.0 conceptual model", in 2017 Smart City Symposium Prague (SCSP), pp. 1-5, 25-26 May 2017, https://doi.org/10.1109/SCSP.2017.7973870

[5] G. Ahmadi-Assalemi, H. Al-Khateeb, G. Epiphaniou, and C. Maple, “Cyber Resilience and Incident Response in Smart Cities: A Systematic Literature Review”, MDPI Smart Cities, vol. 3, no. 3, pp. 894-927, Aug 2020, [Online], https://doi.org/10.3390/smartcities3030046

[6] G. Ahmadi-Assalemi, H. M. Al-Khateeb, G. Epiphaniou, J. Cosson, H. Jahankhani, and P. Pillai, "Federated Blockchain-Based Tracking and Liability Attribution Framework for Employees and Cyber-Physical Objects in a Smart Workplace", in 2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3). London, UK, pp. 1-9, 16-18 Jan. 2019, https://doi.org/10.1109/ICGS3.2019.8688297

[7] S. Schrecker, H. Soroush, J. Molina, J. LeBlanc, F. Hirsch, M. Buchheit, A. Ginter, R. Martin, H. Banavara, and S. Eswarahally, “Industrial internet of things volume G4: security framework”, Ind. Internet Consort, pp. 1-173, 2016, [Online], Accessed: 28 Dec 2020, Available: https://www.iiconsortium.org/pdf/IIC_PUB_G4_V1.00_PB.pdf

[8] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, “The industrial internet of things (IIoT): An analysis framework”, Computers in Industry, vol. 101, pp. 1-12, 2018, [Online], https://doi.org/10.1016/j.compind.2018.04.015

[9] W. Z. Khan, M. H. Rehman, H. M. Zangoti, M. K. Afzal, N. Armi, and K. Salah, “Industrial internet of things: Recent advances, enabling technologies and open challenges”, Computers & Electrical Engineering, vol. 81, pp. 106522, 2020, [Online], https://doi.org/10.1016/j.compeleceng.2019.106522

[10] B. Dorsemaine, J. Gaulier, J. Wary, N. Kheir, and P. Urien, "Internet of Things: A Definition & Taxonomy", in 2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies, pp. 72-77, 9-11 Sept. 2015, https://doi.org/10.1109/NGMAST.2015.71

[11] K. A. Rose, S. Sable, D. L. DeAngelis, S. Yurek, J. C. Trexler, W. Graf, and D. J. Reed, “Proposed best modeling practices for assessing the effects of ecosystem restoration on fish”, Ecological Modelling, vol. 300, pp. 12-29, 2015, [Online], https://doi.org/10.1016/j.ecolmodel.2014.12.020

[12] T. Lu, X. Guo, Y. Li, Y. Peng, X. Zhang, F. Xie, and Y. Gao, “Cyberphysical Security for Industrial Control Systems Based on Wireless Sensor Networks”, International Journal of Distributed Sensor Networks, vol. 10, no. 6, pp. 438350, 2014, [Online], https://doi.org/10.1155/2014/438350

[13] Y. Peng, Y. Wang, C. Xiang, X. Liu, Z. Wen, D. Chen, and C. Zhang, "Cyber-Physical Attack-Oriented Industrial Control Systems (ICS) Modeling, Analysis and Experiment Environment", in 2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), pp. 322-326, 23-25 Sept. 2015, https://doi.org/10.1109/IIH-MSP.2015.110

[14] S. V. B. Rakas, M. D. Stojanović, and J. D. Marković-Petrović, “A Review of Research Work on Network-Based SCADA Intrusion Detection Systems”, IEEE Access, vol. 8, pp. 93083-93108, 2020, [Online], https://doi.org/10.1109/ACCESS.2020.2994961

[15] L. L. Dhirani, E. Armstrong, and T. Newe, “Industrial IoT, Cyber Threats, and Standards Landscape: Evaluation and Roadmap”, Sensors, vol. 21, no. 11, pp. 3901, 2021, [Online], https://doi.org/10.3390/s21113901

[16] N. Bakalos, A. Voulodimos, N. Doulamis, A. Doulamis, A. Ostfeld, E. Salomons, J. Caubet, V. Jimenez, and P. Li, “Protecting water infrastructure from cyber and physical threats: Using multimodal data fusion and adaptive deep learning to monitor critical systems”, IEEE

Signal Processing Magazine, vol. 36, no. 2, pp. 36-48, 2019, [Online], https://doi.org/10.1109/MSP.2018.2885359
[17] I. Kiss, B. Genge, P. Haller, and G. Sebestyén, "Data clustering-based anomaly detection in industrial control systems", in 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP), pp. 275-281, 4-6 Sept. 2014, https://doi.org/10.1109/ICCP.2014.6937009
[18] S. Berger, O. Bürger, and M. Röglinger, "Attacks on the Industrial Internet of Things – Development of a multi-layer Taxonomy", Computers & Security, vol. 93, pp. 101790, June 2020, [Online], https://doi.org/10.1016/j.cose.2020.101790
[19] G. Li, Y. Shen, P. Zhao, X. Lu, J. Liu, Y. Liu, and S. C. Hoi, "Detecting cyberattacks in industrial control systems using online learning algorithms", Neurocomputing, vol. 364, pp. 338-348, 2019, [Online], https://doi.org/10.1016/j.neucom.2019.07.031
[20] W. P. Mardyaningsih, P. H. Rusmin, and B. Rahardjo, "Anomaly Detection and Data Recovery on Mini Batch Distillation Column based Cyber Physical System", in 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). Bandung, Indonesia, pp. 454-458, 18-20 Sept. 2019, IEEE, https://doi.org/10.23919/EECSI48112.2019.8977070
[21] F. Kammüller, J. R. Nurse, and C. W. Probst, "Attack tree analysis for insider threats on the IoT using Isabelle", in International Conference on Human Aspects of Information Security, Privacy, and Trust, pp. 234-246, Springer, https://doi.org/10.1007/978-3-319-39381-0_21
[22] G. Ahmadi-Assalemi, H. Al-Khateeb, G. Epiphaniou, and A. Aggoun, "Super Learner Ensemble for Anomaly Detection and Cyber-risk Quantification in Industrial Control Systems", IEEE Internet of Things Journal, pp. 1-1, 2022, [Online], https://doi.org/10.1109/JIOT.2022.3144127
[23] D. F. Hsu, and D. Marinucci,"Advances in cyber security: technology, operations, and experiences", pp. 272, Oxford: Fordham University Press, ISBN: 9780823244584, 2012. [Online], https://doi.org/10.5422/fordham/9780823244560.001.0001
[24] M. R. Asghar, Q. Hu, and S. Zeadally, "Cybersecurity in industrial control systems: Issues, technologies, and challenges", Computer Networks, vol. 165, pp. 106946, Dec 2019, [Online], https://doi.org/10.1016/j.comnet.2019.106946
[25] Q. Lin, S. Adepu, S. Verwer, and A. Mathur, "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, Incheon, Republic of Korea, ACM, pp. 525-536, 2018, https://doi.org/10.1145/3196494.3196546
[26] World Economic Forum, "The Global Risks Report 2020", 2020, [Online], Accessed: 29 Dec 2020, Available: http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf
[27] F-Secure, "Attack Landscape H12019", 2019, [Online], Accessed: 29 Dec 2020, Available: https://blog-assets.f-secure.com/wp-content/uploads/2019/09/12093807/2019_attack_landscape_report.pdf
[28] ENISA, "ENISA Threat Landscape Report 2018 15 Top Cyberthreats and Trends", 2019, Accessed: 20 Oct 2019, [Online], Accessed: 20 Oct 2019, Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2018
[29] C. Tankard, "Advanced Persistent threats and how to monitor and deter them", Network Security, vol. 2011, no. 8, pp. 16-19, Aug 2011, [Online], https://doi.org/10.1016/S1353-4858(11)70086-1
[30] F. Skopik, G. Settanni, and R. Fiedler, "A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing", Computers & Security, vol. 60, pp. 154-176, July 2016, [Online], https://doi.org/10.1016/j.cose.2016.04.003
[31] G. Settanni, Y. Shovgenya, F. Skopik, R. Graf, M. Wurzenberger, and R. Fiedler, "Correlating cyber incident information to establish situational awareness in Critical Infrastructures", in 2016 14th Annual Conference on Privacy, Security and Trust (PST). Auckland, New Zealand, pp. 78-81, 14 Dec 2016, IEEE, https://doi.org/10.1109/PST.2016.7906940
[32] D. C. Le, and N. Zincir-Heywood, "Anomaly detection for insider threats using unsupervised ensembles", IEEE Transactions on Network and Service Management, 2021, [Online], https://doi.org/10.1109/TNSM.2021.3071928
[33] M. Collins, "Common sense guide to mitigating insider threats", CARNEGIE-MELLON UNIV PITTSBURGH PA PITTSBURGH United States, 2016, Accessed: 15/06/2021, [Online], Accessed: 15/06/2021, Available: https://apps.dtic.mil/sti/pdfs/AD1044922.pdf

[34] M. Kalech, “Cyber-attack detection in SCADA systems using temporal pattern recognition techniques”, Computers & Security, vol. 84, pp. 225-238, 2019, [Online], https://doi.org/10.1016/j.cose.2019.03.007

[35] E. Bajramovic, K. Waedt, A. Ciriello, D. Gupta, and Ieee,"Forensic Readiness of Smart Buildings Preconditions for Subsequent Cybersecurity Tests", New York, Ieee, ISBN: 978-1-5090-1845-1, 2016. [Online], https://doi.org/10.1109/ISC2.2016.7580754

[36] Y. Wang, and G. Yan, “A new model approach of electrical cyber physical systems considering cyber security”, IEEJ Transactions on Electrical and Electronic Engineering, vol. 14, no. 2, pp. 201-213, 2019, [Online], https://doi.org/10.1002/tee.22798

[37] N. Martindale, M. Ismail, and D. A. Talbert, “Ensemble-Based Online Machine Learning Algorithms for Network Intrusion Detection Systems Using Streaming Data”, Information, vol. 11, no. 6, pp. 315, 2020, [Online], https://doi.org/10.3390/info11060315

[38] S. McLaughlin, C. Konstantinou, X. Wang, L. Davi, A. Sadeghi, M. Maniatakos, and R. Karri, “The Cybersecurity Landscape in Industrial Control Systems”, Proceedings of the IEEE, vol. 104, no. 5, pp. 1039-1057, March 2016, [Online], https://doi.org/10.1109/JPROC.2015.2512235

[39] B. v. Lier, "The industrial internet of things and cyber security: An ecological and systemic perspective on security in digital industrial ecosystems", in 2017 21st International Conference on System Theory, Control and Computing (ICSTCC). Sinaia, Romania, pp. 641-647, 19-21 Oct. 2017, https://doi.org/10.1109/ICSTCC.2017.8107108

[40] Europol, “Internet Organised Crime Threat Assessment”, pp. 63, 2019, [Online], Accessed: 22 Oct 2019, Available: https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2019

[41] Q. Do, B. Martini, and K.-K. R. Choo, “Cyber-physical systems information gathering: A smart home case study”, Computer Networks, vol. 138, pp. 1-12, 2018, [Online], https://doi.org/10.1016/j.comnet.2018.03.024

[42] E. Oriwoh, D. Jazani, G. Epiphaniou, and P. Sant, "Internet of things forensics: Challenges and approaches", in 9th IEEE International Conference on Collaborative computing: networking, Applications and Worksharing, pp. 608-615, IEEE, https://doi.org/10.4108/icst.collaboratecom.2013.254159

[43] H. Government, "National Cyber Strategy 2022 Pioneering a cyber future with the whole of the UK", Cabinet Office, 2022, [Online], Accessed: 30/01/2022, Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1040805/National_Cyber_Strategy_-_FINAL_VERSION.pdf

[44] N. Elmrabit, F. Zhou, F. Li, and H. Zhou, "Evaluation of Machine Learning Algorithms for Anomaly Detection", in 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), pp. 1-8, 15-19 June 2020, https://doi.org/10.1109/CyberSecurity49315.2020.9138871

[45] Z.-H. Zhou,"Ensemble methods: foundations and algorithms", pp. 236, New York, CRC press, ISBN: 9780429151095, 2012. [Online], https://doi.org/10.1201/b12207

[46] L. I. Kuncheva,"Combining pattern classifiers: methods and algorithms", pp. 384, Hoboken, New Jersey, USA, John Wiley & Sons, ISBN: 9781118914564, 2014. [Online], https://doi.org/10.1002/9781118914564

[47] O. A. Alimi, K. Ouahada, A. M. Abu-Mahfouz, S. Rimer, and K. O. A. Alimi, “A Review of Research Works on Supervised Learning Algorithms for SCADA Intrusion Detection and Classification”, Sustainability, vol. 13, no. 17, pp. 9597, 2021, [Online], https://doi.org/10.3390/su13179597

[48] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey”, ACM Comput. Surv., vol. 41, no. 3, pp. 1-58, 2009, [Online], https://doi.org/10.1145/1541880.1541882

[49] S. C. Tan, K. M. Ting, and T. F. Liu, “Fast anomaly detection for streaming data”, in Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Two, Barcelona, Catalonia, Spain, AAAI Press, pp. 1511–1516, 2011, https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-254

[50] K. Demertzis, L. Iliadis, P. Kikiras, and N. Tziritas, "Cyber-typhon: an online multi-task anomaly detection framework", in IFIP International Conference on Artificial Intelligence Applications and Innovations, pp. 19-36, Springer, https://doi.org/10.1007/978-3-030-19823-7_2

[51] P. Zehnder, and D. Riemer, "Representing Industrial Data Streams in Digital Twins using Semantic Labeling", in 2018 IEEE International Conference on Big Data (Big Data), pp. 4223-4226, 10-13 Dec. 2018, https://doi.org/10.1109/BigData.2018.8622400

[52] L. Xing, K. Demertzis, and J. Yang, “Identifying data streams anomalies by evolving spiking restricted Boltzmann machines”, Neural Computing and Applications, pp. 1-15, 2019, [Online], https://doi.org/10.1007/s00521-019-04288-5

[53] M. Elnour, N. Meskin, K. Khan, and R. Jain, “A Dual-Isolation-Forests-Based Attack Detection Framework for Industrial Control Systems”, IEEE Access, vol. 8, pp. 36639-36651, Feb 2020, [Online], https://doi.org/10.1109/ACCESS.2020.2975066

[54] K. Stouffer, J. Falco, and K. Scarfone, “Guide to industrial control systems (ICS) security (NIST SP 800-82)”, NIST special publication, vol. 800, no. 82, pp. 16-16, 2011, [Online], Accessed: 02/01/2021, Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r2.pdf

[55] A. A. Cardenas, S. Amin, and S. Sastry, "Secure Control: Towards Survivable Cyber-Physical Systems", in 2008 The 28th International Conference on Distributed Computing Systems Workshops, pp. 495-500, 17-20 June 2008, IEEE, https://doi.org/10.1109/ICDCS.Workshops.2008.40

[56] H. M. Al-Khateeb, G. Epiphaniou, and H. Daly, "Blockchain for Modern Digital Forensics: The Chain-of-Custody as a Distributed Ledger", Blockchain and Clinical Trial: Securing Patient Data, pp. 149-168, Cham: Springer International Publishing, 2019, https://doi.org/10.1007/978-3-030-11289-9_7

[57] A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, “IoT in Smart Cities: A Survey of Technologies, Practices and Challenges”, Smart Cities, vol. 4, no. 2, pp. 429-475, 2021, [Online], https://doi.org/10.3390/smartcities4020024

[58] B. A. Kitchenham, and S. Charters. "Guidelines for performing Systematic Literature Reviews in Software Engineering 2.3", Accessed: 10/08/2020; Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.154.1446&rep=rep1&type=pdf

[59] E. Negri, L. Fumagalli, and M. Macchi, “A review of the roles of digital twin in cps-based production systems”, Procedia Manufacturing, vol. 11, pp. 939-948, 2017, [Online], https://doi.org/10.1016/j.promfg.2017.07.198

[60] Z. DeSmit, A. E. Elhabashy, L. J. Wells, and J. A. Camelio, “An approach to cyber-physical vulnerability assessment for intelligent manufacturing systems”, Journal of Manufacturing Systems, vol. 43, pp. 339-351, 2017, [Online], https://doi.org/10.1016/j.jmsy.2017.03.004

[61] Verizon, "Data Breach Digest", 2016, Accessed: 22 Oct 2019, [Online], Accessed: 22 Oct 2019, Available: https://enterprise.verizon.com/resources/reports/2016/data-breach-digest.pdf

[62] Verizon, “2016 Data Breach Investigations Report”, pp. 85, 2016, [Online], Accessed, Available: https://regmedia.co.uk/2016/05/12/dbir_2016.pdf

[63] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, “Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains”, Leading Issues in Information Warfare & Security Research, vol. 1, no. 1, pp. 80, 2011, [Online], Accessed: 07/01/2020, Available: https://www.nationalcyberwatch.org/resource/intelligence-driven-computer-network-defense-informed-by-analysis-of-adversary-campaigns-and-intrusion-kill-chains-2/

[64] R. Langner, “Stuxnet: Dissecting a Cyberwarfare Weapon”, IEEE Security & Privacy, vol. 9, no. 3, pp. 49-51, May 2011, [Online], https://doi.org/10.1109/MSP.2011.67

[65] SANS ICS, “Analysis of the cyber attack on the Ukrainian power grid”, Electricity Information Sharing and Analysis Center (E-ISAC), 2016, [Online], Accessed: 15 Oct 2019, Available: https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf

[66] HM Government, "National Cyber Security Strategy 2016-2021", 2016, Accessed: 30 Aug 2019, [Online], Accessed: 30 Aug 2019, Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567242/national_cyber_security_strategy_2016.pdf

[67] J. P. G. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, “Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines”, Computer Networks, vol. 54, no. 8, pp. 1245-1265, 2010, [Online], https://doi.org/10.1016/j.comnet.2010.03.005

[68] J. Pacheco, and S. Hariri,"IoT Security Framework for Smart Cyber Infrastructures", New York, IEEE, ISBN: 978-1-5090-3651-6, 2016. [Online], https://doi.org/10.1109/fas-w.2016.58

[69] J. Pacheco, S. Satam, S. Hariri, C. Grijalva, and H. Berkenbrock, "IoT Security Development Framework for Building Trustworthy Smart Car Services", in Ieee International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data. New York, pp. 237-242, IEEE, https://doi.org/10.1109/ISI.2016.7745481

[70] M. A. Rahman, M. M. Rashid, M. S. Hossain, E. Hassanain, M. F. Alhamid, and M. Guizani, “Blockchain and IoT-Based Cognitive Edge Framework for Sharing Economy

Services in a Smart City", IEEE Access, vol. 7, pp. 18611-18621, 2019, [Online], https://doi.org/10.1109/ACCESS.2019.2896065
[71] J. Lee, B. Bagheri, and H.-A. Kao, "A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems", Manufacturing Letters, vol. 3, pp. 18-23, 2015, [Online], https://doi.org/10.1016/j.mfglet.2014.12.001
[72] S. Shrivastava, S. Adepu, and A. Mathur, "Design and assessment of an Orthogonal Defense Mechanism for a water treatment facility", Robotics and Autonomous Systems, vol. 101, pp. 114-125, 2018, [Online], https://doi.org/10.1016/j.robot.2017.12.005
[73] O. Erdene-Ochir, M. Abdallah, K. Qaraqe, M. Minier, and F. Valois, "Routing resilience evaluation for smart metering: Definition, metric and techniques", in 2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), pp. 1867-1871, 2-5 Sept. 2014, https://doi.org/10.1109/PIMRC.2014.7136474
[74] T. R. Farley, and C. J. Colbourn, "Multiterminal resilience for series-parallel networks", Networks, vol. 50, no. 2, pp. 164-172, 2007, [Online], https://doi.org/10.1002/net.20186
[75] P. Cholda, A. Mykkeltveit, B. E. Helvik, O. J. Wittner, and A. Jajszczyk, "A survey of resilience differentiation frameworks in communication networks", IEEE Communications Surveys & Tutorials, vol. 9, no. 4, pp. 32-55, 2007, [Online], https://doi.org/10.1109/COMST.2007.4444749
[76] R. Leszczyna, "A review of standards with cybersecurity requirements for smart grid", Computers & Security, vol. 77, pp. 262-276, 2018, [Online], https://doi.org/10.1016/j.cose.2018.03.011
[77] M. Lezzi, M. Lazoi, and A. Corallo, "Cybersecurity for Industry 4.0 in the current literature: A reference framework", Computers in Industry, vol. 103, pp. 97-110, 2018, [Online], https://doi.org/10.1016/j.compind.2018.09.004
[78] M. Daneva, and B. Lazarov, "Requirements for smart cities: Results from a systematic review of literature", in 2018 12th International Conference on Research Challenges in Information Science (RCIS), pp. 1-6, 29-31 May 2018, https://doi.org/10.1109/RCIS.2018.8406655
[79] J. P. G. Sterbenz,"Smart City and IoT Resilience, Survivability, and Disruption Tolerance: Challenges, Modelling, and a Survey of Research Opportunities", New York, IEEE, ISBN: 978-1-5386-0671-1, 2017. [Online], https://doi.org/10.1109/RNDM.2017.8093025
[80] D. W. McKee, S. J. Clement, J. Almutairi, and J. Xu, "Survey of advances and challenges in intelligent autonomy for distributed cyber-physical systems", CAAI Transactions on Intelligence Technology, vol. 3, no. 2, pp. 75-82, 2018, [Online], https://doi.org/10.1049/trit.2018.0010
[81] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, and X. Yang, "A Survey on the Edge Computing for the Internet of Things", IEEE Access, vol. 6, pp. 6900-6919, 2018, [Online], https://doi.org/10.1109/ACCESS.2017.2778504
[82] J. Lin, W. Yu, N. Zhang, X. Y. Yang, H. L. Zhang, and W. Zhao, "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications", IEEE Internet of Things Journal, vol. 4, no. 5, pp. 1125-1142, 2017, [Online], https://doi.org/10.1109/jiot.2017.2683200
[83] G. Ahmadi-Assalemi, H. M. al-Khateeb, C. Maple, G. Epiphaniou, M. Hammoudeh, H. Jahankhani, and P. Pillai, "Optimising driver profiling through behaviour modelling of in-car sensor and global positioning system data", Computers & Electrical Engineering, vol. 91, pp. 107047, 2021, [Online], https://doi.org/10.1016/j.compeleceng.2021.107047
[84] P. M. Laso, D. Brosset, and J. Puentes, "Dataset of anomalies and malicious acts in a cyber-physical subsystem", Data in Brief, vol. 14, pp. 186-191, Oct 2017, [Online], https://doi.org/10.1016/j.dib.2017.07.038
[85] M. I. Jordan, and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects", Science, vol. 349, no. 6245, pp. 255-260, July 2015, [Online], https://doi.org/10.1126/science.aaa8415
[86] Z. Zhang, and E. Sejdić, "Radiological images and machine learning: Trends, perspectives, and prospects", Computers in Biology and Medicine, vol. 108, pp. 354-370, May 2019, [Online], https://doi.org/10.1016/j.compbiomed.2019.02.017
[87] N. A. Jalil, H. J. Hwang, and N. M. Dawi, "Machines Learning Trends, Perspectives and Prospects in Education Sector", in Proceedings of the 2019 3rd International Conference on Education and Multimedia Technology. Nagoya Japan, pp. 201-205, https://doi.org/10.1145/3345120.3345147

[88] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido, and M. Marchetti, "On the effectiveness of machine and deep learning for cyber security", in 2018 10th International Conference on Cyber Conflict (CyCon). Tallinn, Estonia, pp. 371-390, 29 May-1 June 2018, IEEE, https://doi.org/10.23919/CYCON.2018.8405026

[89] J. M. Torres, C. I. Comesaña, and P. J. García-Nieto, "Machine learning techniques applied to cybersecurity", International Journal of Machine Learning and Cybernetics, vol. 10, no. 10, pp. 2823-2836, Jan 2019, [Online], https://doi.org/10.1007/s13042-018-00906-1

[90] H. A. Boyes, R. Isbell, P. Norris, and T. Watson, "Enabling intelligent cities through cyber security of building information and building systems", in IET Conference on Future Intelligent Cities. London, UK, pp. 1-6, 4-5 Dec. 2014, https://doi.org/10.1049/ic.2014.0046

[91] T. G. Dietterich, "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms", Neural Computation, vol. 10, no. 7, pp. 1895-1923, Oct 1998, [Online], https://doi.org/10.1162/089976698300017197

[92] D. B. Araya, K. Grolinger, H. F. ElYamany, M. A. Capretz, and G. Bitsuamlak, "An ensemble learning framework for anomaly detection in building energy consumption", Energy and Buildings, vol. 144, pp. 191-206, June 2017, [Online], https://doi.org/10.1016/j.enbuild.2017.02.058

[93] N. Moustafa, B. Turnbull, and K. R. Choo, "An Ensemble Intrusion Detection Technique based on proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things", IEEE Internet of Things Journal, pp. 1-1, Sep 2018, [Online], https://doi.org/10.1109/JIOT.2018.2871719

[94] L. H. N. Lorena, A. C. P. L. F. Carvalho, and A. C. Lorena, "Filter Feature Selection for One-Class Classification", Journal of Intelligent & Robotic Systems, vol. 80, no. 1, pp. 227-243, Sep 2014, [Online], https://doi.org/10.1007/s10846-014-0101-2

[95] F. Schuster, A. Paul, R. Rietz, and H. Koenig, "Potentials of Using One-Class SVM for Detecting Protocol-Specific Anomalies in Industrial Networks", in 2015 IEEE Symposium Series on Computational Intelligence, pp. 83-90, 7-10 Dec. 2015, https://doi.org/10.1109/SSCI.2015.22

[96] A. Terai, S. Abe, S. Kojima, Y. Takano, and I. Koshijima, "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, 26-28 April 2017, https://doi.org/10.1109/EuroSPW.2017.62

[97] M. Turkoz, S. Kim, Y. Son, M. K. Jeong, and E. A. Elsayed, "Generalized support vector data description for anomaly detection", Pattern Recognition, vol. 100, pp. 107119, 2020, [Online], https://doi.org/10.1016/j.patcog.2019.107119

[98] R. C. B. Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, "Machine learning for power system disturbance and cyber-attack discrimination", in 2014 7th International Symposium on Resilient Control Systems (ISRCS), pp. 1-8, 19-21 Aug. 2014, https://doi.org/10.1109/ISRCS.2014.6900095

[99] F. A. Alhaidari, and E. M. A.-. Dahasi, "New Approach to Determine DDoS Attack Patterns on SCADA System Using Machine Learning", in 2019 International Conference on Computer and Information Sciences (ICCIS), pp. 1-6, 3-4 April 2019, https://doi.org/10.1109/ICCISci.2019.8716432

[100] S. Tamy, H. Belhadaoui, M. A. Rabbah, N. Rabbah, and M. Rifi, "An Evaluation of Machine Learning Algorithms To Detect Attacks in Scada Network", in 2019 7th Mediterranean Congress of Telecommunications (CMT), pp. 1-5, 24-25 Oct. 2019, https://doi.org/10.1109/CMT.2019.8931327

[101] R. L. Perez, F. Adamsky, R. Soua, and T. Engel, "Machine Learning for Reliable Network Attack Detection in SCADA Systems", in 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), pp. 633-638, 1-3 Aug. 2018, https://doi.org/10.1109/TrustCom/BigDataSE.2018.00094

[102] A. Choubineh, D. A. Wood, and Z. Choubineh, "Applying separately cost-sensitive learning and Fisher's discriminant analysis to address the class imbalance problem: A case study involving a virtual gas pipeline SCADA system", International Journal of Critical Infrastructure Protection, vol. 29, pp. 100357, 2020, [Online], https://doi.org/10.1016/j.ijcip.2020.100357

[103] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Gradient Boosting Feature Selection With Machine Learning Classifiers for Intrusion Detection on Power Grids", IEEE

Transactions on Network and Service Management, vol. 18, no. 1, pp. 1104-1116, 2021, [Online], https://doi.org/10.1109/TNSM.2020.3032618
[104] M. Al-Asiri, and E.-S. M. El-Alfy, “On Using Physical Based Intrusion Detection in SCADA Systems”, Procedia Computer Science, vol. 170, pp. 34-42, 2020, [Online], https://doi.org/10.1016/j.procs.2020.03.007
[105] I. A. Siddavatam, S. Satish, W. Mahesh, and F. Kazi, "An ensemble learning for anomaly identification in SCADA system", in 2017 7th International Conference on Power Systems (ICPS), pp. 457-462, 21-23 Dec. 2017, https://doi.org/10.1109/ICPES.2017.8387337
[106] S. Ahmed, Y. Lee, S. Hyun, and I. Koo, “Unsupervised Machine Learning-Based Detection of Covert Data Integrity Assault in Smart Grid Networks Utilizing Isolation Forest”, IEEE Transactions on Information Forensics and Security, vol. 14, no. 10, pp. 2765-2777, March 2019, [Online], https://doi.org/10.1109/TIFS.2019.2902822
[107] V. Vapnik, "The Support Vector Method of Function Estimation", Nonlinear Modeling: Advanced Black-Box Techniques, pp. 55-85, Boston, MA: Springer US, 1998, https://doi.org/10.1007/978-1-4615-5703-6_3
[108] W. Shang, J. Cui, C. Song, J. Zhao, and P. Zeng, "Research on Industrial Control Anomaly Detection Based on FCM and SVM", in 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), pp. 218-222, 1-3 Aug. 2018, https://doi.org/10.1109/TrustCom/BigDataSE.2018.00042
[109] P. Arora, B. Kaur, and M. A. Teixeira, “Evaluation of Machine Learning Algorithms Used on Attacks Detection in Industrial Control Systems”, Journal of The Institution of Engineers (India): Series B, vol. 102, no. 3, pp. 605-616, 2021, [Online], https://doi.org/10.1007/s40031-021-00563-z
[110] A. Gumaei, M. M. Hassan, S. Huda, M. R. Hassan, D. Camacho, J. Del Ser, and G. Fortino, “A robust cyberattack detection approach using optimal features of SCADA power systems in smart grids”, Applied Soft Computing, vol. 96, pp. 106658, 2020, [Online], https://doi.org/10.1016/j.asoc.2020.106658
[111] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, “Scikit-learn: Machine learning in Python”, the Journal of machine Learning research, vol. 12, pp. 2825-2830, 2011, [Online], Accessed: 30/11/2021, Available: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?source=post_page
[112] O. A. Alimi, K. Ouahada, A. M. Abu-Mahfouz, and K. O. A. Alimi, "Empirical Comparison of Machine Learning Algorithms for Mitigating Power Systems Intrusion Attacks", in 2020 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1-5, 20-22 Oct. 2020, https://doi.org/10.1109/ISNCC49221.2020.9297340
[113] T. G. Dietterich, "Ensemble methods in machine learning", in International workshop on multiple classifier systems. Cagliari, Italy, pp. 1-15, 1 Dec 2000, Springer, https://doi.org/10.1007/3-540-45014-9_1
[114] L. Breiman, “Random forests”, Machine learning, vol. 45, no. 1, pp. 5-32, 2001, [Online], https://doi.org/10.1023/A:1010933404324
[115] L. E. B. Ferreira, H. M. Gomes, A. Bifet, and L. S. Oliveira, "Adaptive random forests with resampling for imbalanced data streams", in 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1-6, IEEE, https://doi.org/10.1109/IJCNN.2019.8852027
[116] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees”, Machine Learning, vol. 63, no. 1, pp. 3-42, 2006, [Online], https://doi.org/10.1007/s10994-006-6226-1
[117] Y. Freund, and R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”, Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119-139, 1997, [Online], https://doi.org/10.1006/jcss.1997.1504
[118] L. Breiman, “Bagging predictors”, Machine learning, vol. 24, no. 2, pp. 123-140, 1996, [Online], https://doi.org/10.1007/BF00058655
[119] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation Forest", in 2008 Eighth IEEE International Conference on Data Mining. Pisa, Italy, pp. 413-422, 15-19 Dec. 2008, IEEE, https://doi.org/10.1109/ICDM.2008.17
[120] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, “State-of-the-art in artificial neural network applications: A survey”, Heliyon, vol. 4, no. 11, pp. e00938, 2018, [Online], https://doi.org/10.1016/j.heliyon.2018.e00938
[121] S. Alkaabi, S. Yussof, H. M. Al-Khateeb, G. Ahmadi-Assalemi, and G. Epiphaniou, “Deep Convolutional Neural Networks for Forensic Age Estimation: A Review”, Cyber Defence in

the Age of AI, Smart Societies and Augmented Humanity, pp. 375, 2020, [Online], https://doi.org/10.1007/978-3-030-35746-7_17

[122] A. M. Kosek, and O. Gehrke, "Ensemble regression model-based anomaly detection for cyber-physical intrusion detection in smart grids", in 2016 IEEE Electrical Power and Energy Conference (EPEC), pp. 1-7, 12-14 Oct. 2016, https://doi.org/10.1109/EPEC.2016.7771704

[123] K. Demertzis, L. Iliadis, and S. Spartalis, "A spiking one-class anomaly detection framework for cyber-security on industrial control systems", in International Conference on Engineering Applications of Neural Networks, pp. 122-134, Springer, https://doi.org/10.1007/978-3-319-65172-9_11

[124] D. Shalyga, P. Filonov, and A. Lavrentyev, "Anomaly detection for water treatment system based on neural network with automatic architecture optimization", ICML Workshop for Deep Learning for Safety-Critical in Engineering Systems, 2018, [Online], Accessed: 03/09/2021, Available: https://arxiv.org/abs/1807.07282

[125] D. T. Ramotsoela, G. P. Hancke, and A. M. Abu-Mahfouz, "Attack detection in water distribution systems using machine learning", Human-centric Computing and Information Sciences, vol. 9, no. 1, pp. 13, 2019, [Online], https://doi.org/10.1186/s13673-019-0175-8

[126] K. Demertzis, L. Iliadis, and V.-D. Anezakis, "MOLESTRA: a multi-task learning approach for real-time big data analytics", in 2018 Innovations in Intelligent Systems and Applications (INISTA), pp. 1-8, IEEE, https://doi.org/10.1109/INISTA.2018.8466306

[127] H. M. Gomes, A. Bifet, J. Read, J. P. Barddal, F. Enembreck, B. Pfharinger, G. Holmes, and T. Abdessalem, "Adaptive random forests for evolving data stream classification", Machine Learning, vol. 106, no. 9, pp. 1469-1495, 2017, [Online], https://doi.org/10.1007/s10994-017-5642-8

[128] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble learning for data stream analysis: A survey", Information Fusion, vol. 37, pp. 132-156, 2017, [Online], https://doi.org/10.1016/j.inffus.2017.02.004

[129] U. Adhikari, T. H. Morris, and S. Pan, "Applying hoeffding adaptive trees for real-time cyber-power event and intrusion classification", IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 4049-4060, 2017, [Online], https://doi.org/10.1109/TSG.2017.2647778

[130] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer, "MOA data stream mining-A practical approach", COSI (Centre for Open Software Innovation), 2011, Accessed: 15/08/2021, Available: http://jwijffels.github.io/RMOA/MOA_2014_04/doc/pdf/StreamMining.pdf

[131] Gartner, "Gartner Top 10 Data and Analytics Trends for 2021", Accessed: 01/07/2021, [Online]. Available: https://www.gartner.com/smarterwithgartner/gartner-top-10-data-and-analytics-trends-for-2021/

[132] M. U. Togbe, M. Barry, A. Boly, Y. Chabchoub, R. Chiky, J. Montiel, and V.-T. Tran, "Anomaly Detection for Data Streams Based on Isolation Forest Using Scikit-Multiflow", in International Conference on Computational Science and Its Applications, pp. 15-30, Springer, https://doi.org/10.1007/978-3-030-58811-3_2

[133] C. Nixon, M. Sedky, and M. Hassan, "Autoencoders: A Low Cost Anomaly Detection Method for Computer Network Data Streams", in Proceedings of the 2020 4th International Conference on Cloud and Big Data Computing, pp. 58-62, https://doi.org/10.1145/3416921.3416937

[134] Z. Ding, and M. Fei, "An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window", IFAC Proceedings Volumes, vol. 46, no. 20, pp. 12-17, 2013, [Online], https://doi.org/10.3182/20130902-3-CN-3020.00044

[135] K. Demertzis, L. Iliadis, and V.-D. Anezakis, "A dynamic ensemble learning framework for data stream analysis and real-time threat detection", in International Conference on Artificial Neural Networks, pp. 669-681, Springer, https://doi.org/10.1007/978-3-030-01418-6_66

[136] V. Losing, B. Hammer, and H. Wersing, "KNN Classifier with Self Adjusting Memory for Heterogeneous Concept Drift", in 2016 IEEE 16th International Conference on Data Mining (ICDM), pp. 291-300, 12-15 Dec. 2016, https://doi.org/10.1109/ICDM.2016.0040

[137] R. O. Duda, and P. E. Hart,"Pattern classification", John Wiley & Sons, ISBN: 8126511168, 2006. [Online], Accessed: 03/02/2021, Available: https://cds.cern.ch/record/683166/files/0471056693_TOC.pdf

[138] S. Shalev-Shwartz, and S. Ben-David,"Understanding machine learning: From theory to algorithms", Cambridge university press, ISBN: 1139952749, 2014. [Online], Accessed: 20/02/2021, Available: https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf

[139] I. Ghafir, M. Hammoudeh, V. Prenosil, L. Han, R. Hegarty, K. Rabie, and F. J. Aparicio-Navarro, “Detection of advanced persistent threat using machine-learning correlation analysis”, Future Generation Computer Systems, vol. 89, pp. 349-359, 2018, [Online], https://doi.org/10.1016/j.future.2018.06.055
[140] G. J. Ross, N. M. Adams, D. K. Tasoulis, and D. J. Hand, “Exponentially weighted moving average charts for detecting concept drift”, Pattern Recognition Letters, vol. 33, no. 2, pp. 191-198, 2012, [Online], https://doi.org/10.1016/j.patrec.2011.08.019
[141] N. Dahal, O. Abuomar, R. King, and V. Madani, “Event stream processing for improved situational awareness in the smart grid”, Expert Systems with Applications, vol. 42, no. 20, pp. 6853-6863, 2015, [Online], https://doi.org/10.1016/j.eswa.2015.05.003
[142] A. Bifet, and R. Gavaldà, "Adaptive Learning from Evolving Data Streams", in. Berlin, Heidelberg, pp. 249-260, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-642-03915-7_22
[143] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with Drift Detection", in. Berlin, Heidelberg, pp. 286-295, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-540-28645-5_29
[144] M. Baena-Garcıa, J. del Campo-Ávila, R. Fidalgo, and A. Bifet, "Early drift detection method", in, Accessed: 05/10/2021, Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.6101&rep=rep1&type=pdf
[145] I. Frías-Blanco, J. d. Campo-Ávila, G. Ramos-Jiménez, R. Morales-Bueno, A. Ortiz-Díaz, and Y. Caballero-Mota, “Online and Non-Parametric Drift Detection Methods Based on Hoeffding’s Bounds”, IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 3, pp. 810-823, 2015, [Online], https://doi.org/10.1109/TKDE.2014.2345382
[146] P. Branco, L. Torgo, and R. P. Ribeiro, “A survey of predictive modeling on imbalanced domains”, ACM Computing Surveys (CSUR), vol. 49, no. 2, pp. 1-50, 2016, [Online], https://doi.org/10.1145/2907070
[147] C. Elkan, “The foundations of cost-sensitive learning”, in Proceedings of the 17th international joint conference on Artificial intelligence - Volume 2, Seattle, WA, USA, Morgan Kaufmann Publishers Inc., pp. 973–978, 2001, https://doi.org/10.5555/1642194.1642224
[148] H. Abdulsalam, D. B. Skillicorn, and P. Martin, "Streaming Random Forests", in 11th International Database Engineering and Applications Symposium (IDEAS 2007), pp. 225-232, 6-8 Sept. 2007, https://doi.org/10.1109/IDEAS.2007.4318108
[149] H. Abdulsalam, D. B. Skillicorn, and P. Martin, "Classifying Evolving Data Streams Using Dynamic Streaming Random Forests", in. Berlin, Heidelberg, pp. 643-651, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-540-85654-2_54
[150] E. Parliament, “Mapping Smart Cities in the EU”, pp. 200, 2014, [Online], Accessed: 26/04/2019, Available: http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/507480/IPOL-ITRE_ET%282014%29507480_EN.pdf
[151] V. Albino, U. Berardi, and R. M. Dangelico, “Smart Cities: Definitions, Dimensions, Performance, and Initiatives”, Journal of Urban Technology, vol. 22, no. 1, pp. 3-21, 2015, [Online], https://doi.org/10.1080/10630732.2014.942092
[152] C. Harrison, B. Eckman, R. Hamilton, P. Hartswick, J. Kalagnanam, J. Paraszczak, and P. Williams, “Foundations for Smarter Cities”, IBM Journal of Research and Development, vol. 54, no. 4, pp. 1-16, 2010, [Online], https://doi.org/10.1147/JRD.2010.2048257
[153] A. Caragliu, C. Del Bo, and P. Nijkamp, “Smart Cities in Europe”, Journal of Urban Technology, vol. 18, no. 2, pp. 65-82, 2011, [Online], https://doi.org/10.1080/10630732.2011.601117
[154] G. C. Lazaroiu, and M. Roscia, “Definition methodology for the smart cities model”, Energy, vol. 47, no. 1, pp. 326-332, 2012, [Online], https://doi.org/10.1016/j.energy.2012.09.028
[155] J. Manuel Barrionuevo, P. Berrone, and J. Ricart,"Smart Cities, Sustainable Progress: Opportunities for Urban Development", ISBN: 2012. [Online], https://doi.org/10.15581/002.ART-2152
[156] C. Cheh, K. Keefe, B. Feddersen, B. Chen, W. G. Temple, and W. H. Sanders, “Developing Models for Physical Attacks in Cyber-Physical Systems”, in Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy, Dallas, Texas, USA, ACM, pp. 49-55, 2017, https://doi.org/10.1145/3140241.3140249
[157] F. Li, Y. Shi, A. Shinde, J. Ye, and W. Z. Song, “Enhanced Cyber-Physical Security in Internet of Things through Energy Auditing”, IEEE Internet of Things Journal, pp. 1-1, 2019, [Online], https://doi.org/10.1109/JIOT.2019.2899492

[158] A. A. Ganin, A. C. Mersky, A. S. Jin, M. Kitsak, J. M. Keisler, and I. Linkov, “Resilience in Intelligent Transportation Systems (ITS)”, Transportation Research Part C: Emerging Technologies, vol. 100, pp. 318-329, 2019, [Online], https://doi.org/10.1016/j.trc.2019.01.014
[159] X. H. Feng, E. S. Dawam, and S. Amin,"A New Digital Forensics Model of Smart City Automated Vehicles", New York, Ieee, ISBN: 978-1-5386-3066-2, June 2017. [Online], https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.47
[160] A. Clarke, and A. Crane, “Cross-Sector Partnerships for Systemic Change: Systematized Literature Review and Agenda for Further Research”, Journal of Business Ethics, vol. 150, no. 2, pp. 303-313, 2018, [Online], https://doi.org/10.1007/s10551-018-3922-2
[161] S. Pfeiffer, “The Vision of “Industrie 4.0” in the Making—a Case of Future Told, Tamed, and Traded”, NanoEthics, vol. 11, no. 1, pp. 107-121, 2017, [Online], https://doi.org/10.1007/s11569-016-0280-3
[162] L. Elliott, and J. Kollewe, "Germany's smaller firms emerge intact from the recession", Theguardian, 2011, Accessed: 04/07/2019, Available: https://www.theguardian.com/world/2011/mar/15/new-europe-germany-manufacturing
[163] D. B. Hancké, and D. S. Coulter, "The German manufacturing sector unpacked: institutions, policies and future trajectories", London School of Economics and Political Science, Foresight, Government Office for Science, 2013, Accessed: 04/07/2019, [Online], Accessed: 04/07/2019, Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/283889/ep13-german-manufacturing.pdf
[164] Infrastructure and Projects Authority, "National Infrastructure Delivery Plan 2016-2021", UK, 2016, [Online], Accessed: 06/07/2019, Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/520086/2904569_nidp_deliveryplan.pdf
[165] The White House, "Fact Sheet: Cybersecurity National Action Plan", 2016, Accessed: 06/07/2019, Available: https://obamawhitehouse.archives.gov/the-press-office/2016/02/09/fact-sheet-cybersecurity-national-action-plan
[166] World Economic Forum, "World Economic Forum Annual Meeting 2016 Mastering the Fourth Industrial Revolution", REF 300116, Davos - Klosters, 2016, Accessed: 23/04/2020, [Online], Accessed: 23/04/2020, Available: http://www3.weforum.org/docs/WEF_AM16_Report.pdf
[167] Australian Cyber Security Growth Network, "Australia’s Cyber Security Sector Competitiveness Plan", Australia, 2018, [Online], Accessed: 06 July 2019, Available: https://www.austcyber.com/file-download/download/public/415
[168] G. National Cyber Security Centre, "The cyber threat to UK business", UK government, UK, 2018, [Online], Accessed: 06/07/2019, Available: https://www.ncsc.gov.uk/information/the-cyber-threat-to-uk-business-2017-2018-report
[169] House of Lords House of Commons Joint Committee on the National Security Strategy, "Cyber Security of the UK’s Critical National Infrastructure", UK government, 2018, [Online], Accessed: 06 July 2019, Available: https://publications.parliament.uk/pa/jt201719/jtselect/jtnatsec/1708/1708.pdf
[170] D. Buhr, and T. Stehnken, "INDUSTRY 4.0 AND EUROPEAN INNOVATION POLICY Big plans, small steps", ISBN 978-3-96250-109-9, The Friedrich-Ebert-Stiftung - Economic and Social Policy Department, 2018, Accessed: 04/07/2019, [Online], Accessed: 04/07/2019, Available: http://library.fes.de/pdf-files/wiso/14455.pdf
[171] P. Maresova, I. Soukal, L. Svobodova, M. Hedvicakova, E. Javanmardi, A. Selamat, and O. Krejcar, “Consequences of Industry 4.0 in Business and Economics”, Economies, vol. 6, no. 3, pp. 46, 2018, [Online], https://doi.org/10.3390/economies6030046
[172] I. Friedberg, K. McLaughlin, P. Smith, and M. Wurzenberger, "Towards a Resilience Metric Framework for Cyber-Physical Systems", in ICS-CSR. Belfast, UK, 23 - 25 August 2016, https://doi.org/10.14236/ewic/ICS2016.3
[173] R. Arghandeh, A. Von Meier, L. Mehrmanesh, and L. Mili, “On the definition of cyber-physical resilience in power systems”, Renewable and Sustainable Energy Reviews, vol. 58, pp. 1060-1069, May 2016, [Online], https://doi.org/10.1016/j.rser.2015.12.193
[174] National Institute of Standard and Technology, "Glossary of Key Information Security Terms", Accessed, [Online]. Available:
[175] I. Linkov, D. A. Eisenberg, K. Plourde, T. P. Seager, J. Allen, and A. Kott, “Resilience metrics for cyber systems”, Environment Systems and Decisions, vol. 33, no. 4, pp. 471-476, 2013, [Online], https://doi.org/10.1007/s10669-013-9485-y

[176] J.-P. Watson, R. Guttromson, C. Silva-Monroy, R. Jeffers, K. Jones, J. Ellison, C. Rath, J. Gearhart, D. Jones, and T. Corbet, “Conceptual framework for developing resilience metrics for the electricity oil and gas sectors in the United States”, Sandia National Laboratories, Albuquerque, NM (United States), Tech. Rep, 2014, [Online], Accessed: 01/09/2019, Available: https://www.energy.gov/sites/prod/files/2015/09/f26/EnergyResilienceReport_(Final)_SAND2015-18019.pdf

[177] Internet Engineering Task Force, "Requirements for Internet Hosts -- Communication Layers", 1989, p. 116, Accessed: 20/09/2019, Available: https://history-computer.com/Library/rfc1122.pdf

[178] National Institute of Standards and Technology NIST, "Computer Security Incident Handling Guide", Accessed, [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf

[179] S. Siboni, V. Sachidananda, Y. Meidan, M. Bohadana, Y. Mathov, S. Bhairav, A. Shabtai, and Y. Elovici, “Security Testbed for Internet-of-Things Devices”, IEEE Transactions on Reliability, vol. 68, no. 1, pp. 23-44, 2019, [Online], https://doi.org/10.1109/TR.2018.2864536

[180] D. Ratasich, F. Khalid, F. Geissler, R. Grosu, M. Shafique, and E. Bartocci, “A Roadmap Toward the Resilient Internet of Things for Cyber-Physical Systems”, IEEE Access, vol. 7, pp. 13260-13283, 2019, [Online], https://doi.org/10.1109/ACCESS.2019.2891969

[181] B. Mohandes, R. A. Hammadi, W. Sanusi, T. Mezher, and S. E. Khatib, “Advancing cyber–physical sustainability through integrated analysis of smart power systems: A case study on electric vehicles”, International Journal of Critical Infrastructure Protection, vol. 23, pp. 33-48, 2018, [Online], https://doi.org/10.1016/j.ijcip.2018.10.002

[182] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response", in Proceedings of the 6th ACM symposium on information, computer and communications security. Hong Kong China, pp. 355-366, March 2011, https://doi.org/10.1145/1966913.1966959

[183] S. Marrone, R. J. Rodríguez, R. Nardone, F. Flammini, and V. Vittorini, “On synergies of cyber and physical security modelling in vulnerability assessment of railway systems”, Computers & Electrical Engineering, vol. 47, pp. 275-285, 2015, [Online], https://doi.org/10.1016/j.compeleceng.2015.07.011

[184] J. Pacheco, D. Ibarra, A. Vijay, and S. Hariri, "IoT Security Framework for Smart Water System", in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), pp. 1285-1292, 30 Oct.-3 Nov. 2017, https://doi.org/10.1109/AICCSA.2017.85

[185] S. Lakshminarayana, T. Z. Teng, R. Tan, and D. K. Y. Yau, “Modeling and Detecting False Data Injection Attacks against Railway Traction Power Systems”, ACM Trans. Cyber-Phys. Syst., vol. 2, no. 4, pp. 1-29, 2018, [Online], https://doi.org/10.1145/3226030

[186] A. Bathelt, N. L. Ricker, and M. Jelali, “Revision of the tennessee eastman process model”, IFAC-PapersOnLine, vol. 48, no. 8, pp. 309-314, 2015, [Online], https://doi.org/10.1016/j.ifacol.2015.08.199

[187] X. Jia, X. Li, and Y. Gao, “A Novel Semi-Automatic Vulnerability Detection System for Smart Home”, in Proceedings of the International Conference on Big Data and Internet of Thing, London, United Kingdom, ACM, pp. 195-199, 2017, https://doi.org/10.1145/3175684.3175718

[188] G. Comert, J. Pollard, D. M. Nicol, K. Palani, and B. Vignesh, “Modeling cyber attacks at intelligent traffic signals”, Transportation research record, vol. 2672, no. 1, pp. 76-89, 2018, [Online], https://doi.org/10.1177/0361198118784378

[189] A. Orozco, J. Pacheco, and S. Hariri, "Anomaly behavior analysis for smart grid automation system", in 2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), pp. 1-7, 8-10 Nov. 2017, https://doi.org/10.1109/ROPEC.2017.8261614

[190] C. M. Ahmed, J. Zhou, and A. P. Mathur, “Noise Matters: Using Sensor and Process Noise Fingerprint to Detect Stealthy Cyber Attacks and Authenticate sensors in CPS”, in Proceedings of the 34th Annual Computer Security Applications Conference, San Juan, PR, USA, ACM, pp. 566-581, 2018, https://doi.org/10.1145/3274694.3274748

[191] L. Xiaoxue, Z. Jiexin, and Z. Peidong, "Dependence analysis based cyber-physical security assessment for critical infrastructure networks", in 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 1-7, 13-15 Oct. 2016, https://doi.org/10.1109/IEMCON.2016.7746296

[192] I. Abeykoon, and X. Feng, "A Forensic Investigation of the Robot Operating System", in 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 851-857, 21-23 June 2017, https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.131

[193] V. R. Palleti, Y. C. Tan, and L. Samavedham, "A mechanistic fault detection and isolation approach using Kalman filter to improve the security of cyber physical systems", Journal of Process Control, vol. 68, pp. 160-170, 2018, [Online], https://doi.org/10.1016/j.jprocont.2018.05.005

[194] A. Tundis, R. Egert, and M. Mühlhäuser, "Attack Scenario Modeling for Smart Grids Assessment through Simulation", in Proceedings of the 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy, ACM, pp. 1-10, 2017, https://doi.org/10.1145/3098954.3098966

[195] N. Bryant, N. Spencer, A. King, P. Crooks, J. Deakin, and S. Young, "IoT and Smart City Services to Support Independence and Wellbeing of Older People", 2017 25th International Conference on Software, Telecommunications and Computer Networks, International Conference on Software in Telecommunications and Computer Networks, 2017, https://doi.org/10.23919/SOFTCOM.2017.8115553

[196] C. M. Ahmed, M. Ochoa, J. Zhou, A. P. Mathur, R. Qadeer, C. Murguia, and J. Ruths, "NoisePrint: Attack Detection Using Sensor and Process Noise Fingerprint in Cyber Physical Systems", in Proceedings of the 2018 on Asia Conference on Computer and Communications Security, Incheon, Republic of Korea, ACM, pp. 483-497, 2018, https://doi.org/10.1145/3196494.3196532

[197] E. Kang, S. Adepu, D. Jackson, and A. P. Mathur, "Model-based security analysis of a water treatment system", in Proceedings of the 2nd International Workshop on Software Engineering for Smart Cyber-Physical Systems, Austin, Texas, ACM, pp. 22-28, 2016, https://doi.org/10.1145/2897035.2897041

[198] M. Aloqaily, S. Otoum, I. A. Ridhawi, and Y. Jararweh, "An intrusion detection system for connected vehicles in smart cities", Ad Hoc Networks, 2019, [Online], https://doi.org/10.1016/j.adhoc.2019.02.001

[199] A. Elsaeidy, K. S. Munasinghe, D. Sharma, and A. Jamalipour, "Intrusion detection in smart cities using Restricted Boltzmann Machines", Journal of Network and Computer Applications, vol. 135, pp. 76-83, 2019, [Online], https://doi.org/10.1016/j.jnca.2019.02.026

[200] F. Firoozi, V. I. Zadorozhny, and F. Y. Li, "Subjective Logic-Based In-Network Data Processing for Trust Management in Collocated and Distributed Wireless Sensor Networks", IEEE Sensors Journal, vol. 18, no. 15, pp. 6446-6460, 2018, [Online], https://doi.org/10.1109/JSEN.2018.2848205

[201] G. Sugumar, and A. Mathur, "Testing the Effectiveness of Attack Detection Mechanisms in Industrial Control Systems", in 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), pp. 138-145, 25-29 July 2017, https://doi.org/10.1109/QRS-C.2017.29

[202] A. Elsaeidy, I. Elgendi, K. S. Munasinghe, D. Sharma, A. Jamalipour, and Ieee,"A Smart City Cyber Security Platform for Narrowband Networks", New York, IEEE, ISBN: 978-1-5090-6796-1, 2017. [Online], https://doi.org/10.1109/ATNAC.2017.8215388

[203] J. Liu, C. Zhang, and Y. Fang, "EPIC: A Differential Privacy Framework to Defend Smart Homes Against Internet Traffic Analysis", IEEE Internet of Things Journal, vol. 5, no. 2, pp. 1206-1217, 2018, [Online], https://doi.org/10.1109/JIOT.2018.2799820

[204] S. Garg, A. Singh, S. Batra, N. Kumar, and L. T. Yang, "UAV-Empowered Edge Computing Environment for Cyber-Threat Detection in Smart Vehicles", IEEE Network, vol. 32, no. 3, pp. 42-51, 2018, [Online], https://doi.org/10.1109/MNET.2018.1700286

[205] J. Pacheco, X. Y. Zhu, Y. Badr, S. Hariri, and Ieee,"Enabling Risk Management for Smart Infrastructures with an Anomaly Behavior Analysis Intrusion Detection System", New York, IEEE, ISBN: 978-1-5090-6558-5, 2017. [Online], https://doi.org/10.1109/fas-w.2017.167

[206] X. Zhu, Y. Badr, J. Pacheco, and S. Hariri, "Autonomic Identity Framework for the Internet of Things", in 2017 International Conference on Cloud and Autonomic Computing (ICCAC), pp. 69-79, 18-22 Sept. 2017, https://doi.org/10.1109/ICCAC.2017.14

[207] F. Shaikh, E. Bou-Harb, J. Crichigno, and N. Ghani, "A Machine Learning Model for Classifying Unsolicited IoT Devices by Observing Network Telescopes", in 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 938-943, 25-29 June 2018, https://doi.org/10.1109/IWCMC.2018.8450404

[208] B. A. Mozzaquatro, C. Agostinho, D. Goncalves, J. Martins, and R. Jardim-Goncalves, “An Ontology-Based Cybersecurity Framework for the Internet of Things”, Sensors, vol. 18, no. 9, pp. 3053, 2018, [Online], https://doi.org/10.3390/s18093053

[209] A. Hamza, D. Ranathunga, H. H. Gharakheili, M. Roughan, and V. Sivaraman, “Clear as MUD: Generating, Validating and Applying IoT Behavioral Profiles”, in Proceedings of the 2018 Workshop on IoT Security and Privacy, Budapest, Hungary, ACM, pp. 8-14, 2018, https://doi.org/10.1145/3229565.3229566

[210] Z. A. Khan, “Using energy-efficient trust management to protect IoT networks for smart cities”, Sustainable Cities and Society, vol. 40, pp. 1-15, 2018, [Online], https://doi.org/10.1016/j.scs.2018.03.026

[211] E. Anthi, S. Ahmad, O. Rana, G. Theodorakopoulos, and P. Burnap, “EclipseIoT: A secure and adaptive hub for the Internet of Things”, Computers & Security, vol. 78, pp. 477-490, 2018, [Online], https://doi.org/10.1016/j.cose.2018.07.016

[212] M. Gupta, J. Benson, F. Patwa, and R. Sandhu, “Dynamic Groups and Attribute-Based Access Control for Next-Generation Smart Cars”, in Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy, Richardson, Texas, USA, ACM, pp. 61-72, 2019, https://doi.org/10.1145/3292006.3300048

[213] S. Adepu, and A. Mathur, “Distributed Detection of Single-Stage Multipoint Cyber Attacks in a Water Treatment Plant”, in Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security, Xi'an, China, ACM, pp. 449-460, 2016, https://doi.org/10.1145/2897845.2897855

[214] S. Garg, A. Singh, K. Kaur, G. S. Aujla, S. Batra, N. Kumar, and M. S. Obaidat, “Edge Computing-Based Security Framework for Big Data Analytics in VANETs”, IEEE Network, vol. 33, no. 2, pp. 72-81, 2019, [Online], https://doi.org/10.1109/MNET.2019.1800239

[215] L. Vegh, "Cyber-physical systems security through multi-factor authentication and data analytics", in 2018 IEEE International Conference on Industrial Technology (ICIT), pp. 1369-1374, 20-22 Feb. 2018, https://doi.org/10.1109/ICIT.2018.8352379

[216] Z. Alansari, N. B. Anuar, A. Kamsin, M. R. Belgaum, J. Alshaer, S. Soomro, and M. H. Miraz, "Internet of Things: Infrastructure, Architecture, Security and Privacy", in 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), pp. 150-155, 16-17 Aug. 2018, https://doi.org/10.1109/iCCECOME.2018.8658516

[217] P. Seymer, D. Wijesekera, and Ieee, "In-Flight Aircraft Smart Space Security using Multi-Entity Trust Evaluations", 2018 Ieee/Aiaa 37th Digital Avionics Systems Conference, IEEE-AIAA Digital Avionics Systems Conference, pp. 53-62, New York: Ieee, 2018, https://doi.org/10.1109/DASC.2018.8569865

[218] P. A. Ralston, J. H. Graham, and J. L. Hieb, “Cyber security risk assessment for SCADA and DCS networks”, ISA transactions, vol. 46, no. 4, pp. 583-594, 2007, [Online], https://doi.org/10.1016/j.isatra.2007.04.003

[219] A. S. Elmaghraby, and M. M. Losavio, “Cyber security challenges in Smart Cities: Safety, security and privacy”, Journal of Advanced Research, vol. 5, no. 4, pp. 491-497, 2014, [Online], https://doi.org/10.1016/j.jare.2014.02.006

[220] R. Qadeer, C. Murguia, C. M. Ahmed, and J. Ruths, "Multistage Downstream Attack Detection in a Cyber Physical System", in Computer Security. Cham, pp. 177-185, 2018//, Springer International Publishing, https://doi.org/10.1007/978-3-319-72817-9_12

[221] J. Navarro-Ortiz, S. Sendra, P. Ameigeiras, and J. M. Lopez-Soler, “Integration of LoRaWAN and 4G/5G for the Industrial Internet of Things”, IEEE Communications Magazine, vol. 56, no. 2, pp. 60-67, 2018, [Online], https://doi.org/10.1109/MCOM.2018.1700625

[222] M. Pollitt, "A history of digital forensics", in IFIP International Conference on Digital Forensics, pp. 3-15, Springer, https://doi.org/10.1007/978-3-642-15506-2_1

[223] M. Reith, C. Carr, and G. Gunsch, “An examination of digital forensic models”, International Journal of Digital Evidence, vol. 1, no. 3, pp. 1-12, 2002, [Online], Accessed: 15/08/2019, Available: https://pdfs.semanticscholar.org/c73f/47d8385f452dfd25bbaab754874b65594ccd.pdf

[224] A. Agarwal, M. Gupta, S. Gupta, and S. C. Gupta, “Systematic digital forensic investigation model”, International Journal of Computer Science and Security (IJCSS), vol. 5, no. 1, pp. 118-131, 2011, [Online], Accessed: 15/08/2019, Available: https://www.researchgate.net/profile/Yatendra_Gupta/publication/228410430_Systematic_Digital_Forensic_Investigation_Model/links/56ea8cd208ae95bddc2bcc6b/Systematic-Digital-Forensic-Investigation-Model.pdf

[225] M. R. Belgaum, Z. Alansari, R. Jain, and J. Alshaer, “A framework for evaluation of cyber security challenges in smart cities”, pp. 1-6, 2018, [Online], https://doi.org/10.1049/cp.2018.1372

[226] National Institute of Standards and Technology (NIST), "NIST Special Publication 800-183 Nentworks of ‘Things’", Accessed: July 2016, [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-183.pdf

[227] J. Friedman, and M. Bouchard,"Definitive Guide to Cyber Threat Intelligence: Using Knowledge about Adversaries to Win the War against Targeted Attacks", CyberEdge Group, ISBN: 0996182705, 2015. [Online], Accessed: 04 May 2019, Available: https://cryptome.org/2015/09/cti-guide.pdf

[228] R. Boddington,"Practical Digital Forensics", UK, Packt Publishing Ltd, ISBN: 978-1-78588-710-9, 2016. [Online], Accessed: 07/01/2020, Available: https://www.packtpub.com/product/practical-digital-forensics/9781785887109

[229] G. Ahmadi-Assalemi, H. M. Al-Khateeb, C. Maple, G. Epiphaniou, Z. A. Alhaboby, S. Alkaabi, and D. Alhaboby, “Digital Twins for Precision Healthcare”, Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity, pp. 133, 2020, [Online], https://doi.org/10.1007/978-3-030-35746-7_8

[230] M. Mackintosh, G. Epiphaniou, H. Al-Khateeb, K. Burnham, P. Pillai, and M. Hammoudeh, “Preliminaries of Orthogonal Layered Defence Using Functional and Assurance Controls in Industrial Control Systems”, Journal of Sensor and Actuator Networks, vol. 8, no. 1, pp. 14, 2019, [Online], https://doi.org/10.3390/jsan8010014

[231] R. G. Hollands, “Will the real smart city please stand up?”, City, vol. 12, no. 3, pp. 303-320, 2008, [Online], https://doi.org/10.1080/13604810802479126

[232] H. M. Al-Khateeb, G. Epiphaniou, A. Reviczky, P. Karadimas, and H. Heidari, “Proactive Threat Detection for Connected Cars Using Recursive Bayesian Estimation”, IEEE Sensors Journal, vol. 18, no. 12, pp. 4822-4831, 2018, [Online], https://doi.org/10.1109/JSEN.2017.2782751

[233] P. Paolini, N. D. Blas, S. Copelli, and F. Mercalli, "City4Age: Smart cities for health prevention", in 2016 IEEE International Smart Cities Conference (ISC2), pp. 1-4, 12-15 Sept. 2016, https://doi.org/10.1109/ISC2.2016.7580804

[234] B. I. Kwak, J. Woo, and H. K. Kim, "Know your master: Driver profiling-based anti-theft method", in 2016 14th Annual Conference on Privacy, Security and Trust (PST), pp. 211-218, 12-14 Dec. 2016, https://doi.org/10.1109/PST.2016.7906929

[235] Association of Chief Police Officers, "ACPO Good PRactice Guide for Digital Evidence", 2012, p. 43, Accessed: 07/08/2020, Available: https://www.digital-detective.net/digital-forensics-documents/ACPO_Good_Practice_Guide_for_Digital_Evidence_v5.pdf

[236] US Department of Homeland Security, "Best Practice for Seizing Electronic Evidence v4.2", p. 27, Accessed: 07/08/2020, Available: https://www.cwagweb.org/wp-content/uploads/2018/05/BestPracticesforSeizingElectronicEvidence.pdf

[237] R. Montasari, R. Hill, V. Carpenter, and A. Hosseinian-Far, "The Standardised Digital Forensic Investigation Process Model (SDFIPM)", Blockchain and Clinical Trial: Securing Patient Data, pp. 169-209, Cham: Springer International Publishing, 2019, https://doi.org/10.1007/978-3-030-11289-9_8

[238] D. Hadžiosmanović, R. Sommer, E. Zambon, and P. H. Hartel, “Through the eye of the PLC: semantic security monitoring for industrial processes”, in Proceedings of the 30th Annual Computer Security Applications Conference, New Orleans, Louisiana, USA, Association for Computing Machinery, pp. 126–135, 2014, https://doi.org/10.1145/2664243.2664277

[239] C. Feng, V. R. Palleti, A. Mathur, and D. Chana, "A Systematic Framework to Generate Invariants for Anomaly Detection in Industrial Control Systems", in Network and Distributed Systems Security (NDSS) Symposium 2019. San Diego, CA, USA, 24-27 Feb 2019, https://doi.org/10.14722/ndss.2019.23265

[240] J. Weiss, “Industrial Control System Cyber Security And The Critical Infrastructures”, INSIGHT, vol. 19, no. 4, pp. 33-36, 2016, [Online], https://doi.org/10.1002/inst.12124

[241] J. Yeckle, and S. Abdelwahed, “An Evaluation of Selection Method in the Classification of Scada Datasets Based on the Characteristics of the Data and Priority of Performance”, in Proceedings of the International Conference on Compute and Data Analysis, Lakeland, FL, USA, Association for Computing Machinery, pp. 98–103, 2017, https://doi.org/10.1145/3093241.3093271

[242] N. Tariq, M. Asim, and F. A. Khan, “Securing SCADA-based Critical Infrastructures: Challenges and Open Issues”, Procedia Computer Science, vol. 155, pp. 612-617, 2019, [Online], https://doi.org/10.1016/j.procs.2019.08.086
[243] I. Kaspersky, “Threat Landscape for Industrial Automation Systems in The Second Half of 2016”, 2017, [Online], Accessed: 01/12/2021, Available: https://ics-cert.kaspersky.com/reports/2017/03/28/threat-landscape-for-industrial-automation-systems-in-the-second-half-of-2016/
[244] Q. Shafi, "Cyber Physical Systems Security: A Brief Survey", in 2012 12th International Conference on Computational Science and Its Applications. Salvador, Brazil, pp. 146-150, 18-21 June 2012, IEEE, https://doi.org/10.1109/ICCSA.2012.36
[245] D. Gollmann, and M. Krotofil, "Cyber-Physical Systems Security", The New Codebreakers: Essays Dedicated to David Kahn on the Occasion of His 85th Birthday, pp. 195-204, Berlin, Heidelberg: Springer Berlin Heidelberg, Mar 2016, https://doi.org/10.1007/978-3-662-49301-4_14
[246] S. A. Hildreth, “Congressional Research Service, Report for Congress. Cyberwarfare. ”, 19 June 2001, [Online], Accessed: 28 Dec 2020, Available: https://fas.org/sgp/crs/intel/RL30735.pdf
[247] BBC, “Hacker tries to poison water supply of Florida city”, 2021, [Online], Accessed: 13 Feb 2021, Available: https://www.bbc.co.uk/news/world-us-canada-55989843
[248] K. Schwab,"The fourth industrial revolution", pp. 192, New York, USA, Crown Currency, ISBN: 1524758876, 2017. [Online], Accessed: 30 Feb 2020, Available: https://books.google.co.uk/books?id=OetrDQAAQBAJ&printsec=frontcover
[249] A. Hassanzadeh, A. Rasekh, S. Galelli, M. Aghashahi, R. Taormina, A. Ostfeld, and M. K. Banks, “A review of cybersecurity incidents in the water sector”, Journal of Environmental Engineering, vol. 146, no. 5, pp. 03120003, 2020, [Online], https://doi.org/10.1061/(ASCE)EE.1943-7870.0001686
[250] W. Matsuda, M. Fujimoto, T. Aoyama, and T. Mitsunaga, "Cyber Security Risk Assessment on Industry 4.0 using ICS testbed with AI and Cloud", in 2019 IEEE Conference on Application, Information and Network Security (AINS). Pulau Pinang, Malaysia, pp. 54-59, 19-21 Nov. 2019, IEEE, https://doi.org/10.1109/AINS47559.2019.8968698
[251] C. Schröer, F. Kruse, and J. M. Gómez, “A Systematic Literature Review on Applying CRISP-DM Process Model”, Procedia Computer Science, vol. 181, pp. 526-534, 2021, [Online], https://doi.org/10.1016/j.procs.2021.01.199
[252] M. J. van der Laan, E. C. Polley, and A. E. Hubbard, “Super Learner”, Statistical Applications in Genetics and Molecular Biology, vol. 6, no. 1, Sep 2007, [Online], https://doi.org/10.2202/1544-6115.1309
[253] M. Kuhn, and K. Johnson,"Feature engineering and selection: A practical approach for predictive models", pp. 310, FL, USA, CRC Press, ISBN: 9781315108230, 2019. [Online], https://doi.org/10.1201/9781315108230
[254] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, “Testing the null hypothesis of stationarity against the alternative of a unit root”, Journal of econometrics, vol. 54, no. 1-3, pp. 159-178, Dec 1992, [Online], https://doi.org/10.1016/0304-4076(92)90104-Y
[255] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, “Scikit-learn: Machine Learning Without Learning the Machinery”, GetMobile: Mobile Comp. and Comm., vol. 19, no. 1, pp. 29–33, 2015, [Online], https://doi.org/10.1145/2786984.2786995
[256] D. Chicco, and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation”, BMC Genomics, vol. 21, no. 1, pp. 6, 2020, [Online], https://doi.org/10.1186/s12864-019-6413-7
[257] M. Sokolova, and G. Lapalme, “A systematic analysis of performance measures for classification tasks”, Information processing & management, vol. 45, no. 4, pp. 427-437, July 2009, [Online], https://doi.org/10.1016/j.ipm.2009.03.002
[258] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: identifying density-based local outliers”, in Proceedings of the 2000 ACM SIGMOD Intl. Conf. on Mnmgt of data, Dallas, Texas, USA, Association for Computing Machinery, pp. 93–104, 2000, https://doi.org/10.1145/342009.335388
[259] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the Support of a High-Dimensional Distribution”, Neural Computation, vol. 13, no. 7, pp. 1443-1471, July 2001, [Online], https://doi.org/10.1162/089976601750264965
[260] D. M. J. Tax, and R. P. W. Duin, “Support Vector Data Description”, Machine Learning, vol. 54, no. 1, pp. 45-66, Jan 2004, [Online], https://doi.org/10.1023/B:MACH.0000008084.60811.49

[261] D. B. Parker, "Toward a New Framework for Information Security?", Computer Security Handbook: John Wiley and Sons, 2015, https://doi.org/doi:10.1002/9781118851678.ch3
[262] L. Bass, R. Nord, W. Wood, and D. Zubrow, "Risk Themes Discovered through Architecture Evaluations", in 2007 Working IEEE/IFIP Conference on Software Architecture (WICSA'07), pp. 1-1, 6-9 Jan. 2007, IEEE, https://doi.org/10.1109/WICSA.2007.37
[263] Q. Zhang, C. Zhou, Y. Tian, N. Xiong, Y. Qin, and B. Hu, "A Fuzzy Probability Bayesian Network Approach for Dynamic Cybersecurity Risk Assessment in Industrial Control Systems", IEEE Transactions on Industrial Informatics, vol. 14, no. 6, pp. 2497-2506, Nov 2017, [Online], https://doi.org/10.1109/TII.2017.2768998
[264] Q. Zhang, C. Zhou, N. Xiong, Y. Qin, X. Li, and S. Huang, "Multimodel-Based Incident Prediction and Risk Assessment in Dynamic Cybersecurity Protection for Industrial Control Systems", IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 46, no. 10, pp. 1429-1444, Dec 2015, [Online], https://doi.org/10.1109/TSMC.2015.2503399
[265] S. H. Houmb, and V. N. L. Franqueira, "Estimating ToE Risk Level Using CVSS", in 2009 International Conference on Availability, Reliability and Security. Fukuoka, Japan, pp. 718-725, 16-19 March 2009, IEEE, https://doi.org/10.1109/ARES.2009.151
[266] J. Yan, M. Govindarasu, C. Liu, and U. Vaidya, "A PMU-based risk assessment framework for power control systems", in 2013 IEEE Power & Energy Society General Meeting. Vancouver, BC, Canada, pp. 1-5, 21-25 July 2013, IEEE, https://doi.org/10.1109/PESMG.2013.6672731
[267] M. Ni, J. D. McCalley, V. Vittal, and T. Tayyib, "Online risk-based security assessment", IEEE Transactions on Power Systems, vol. 18, no. 1, pp. 258-265, Feb 2003, [Online], https://doi.org/10.1109/TPWRS.2002.807091
[268] J. D. McCalley, A. Fouad, V. Vittal, A. Irizarry-Rivera, B. Agrawal, and R. G. Farmer, "A risk-based security index for determining operating limits in stability-limited electric power systems", IEEE Transactions on Power Systems, vol. 12, no. 3, pp. 1210-1219, Aug 1997, [Online], https://doi.org/10.1109/59.630463
[269] I. Molloy, L. Dickens, C. Morisset, P.-C. Cheng, J. Lobo, and A. Russo, "Risk-based security decisions under uncertainty", in Proceedings of the second ACM conference on Data and Application Security and Privacy. San Antonio Texas USA, pp. 157-168, Feb 2012, https://doi.org/10.1145/2133601.2133622
[270] H. Tsai, and Y. Huang, "An Analytic Hierarchy Process-Based Risk Assessment Method for Wireless Networks", IEEE Transactions on Reliability, vol. 60, no. 4, pp. 801-816, Oct 2011, [Online], https://doi.org/10.1109/TR.2011.2170117
[271] W. Fu, S. Zhao, J. D. McCalley, V. Vittal, and N. Abi-Samra, "Risk assessment for special protection systems", IEEE Transactions on Power Systems, vol. 17, no. 1, pp. 63-72, Aug 2002, [Online], https://doi.org/10.1109/59.982194
[272] K. Stine, K. Quill, and G. Witte, "Framework for improving critical infrastructure cybersecurity", National Institute of Standards and Technology, 2014, Accessed: 20/01/2021, [Online], Accessed: 20/01/2021, Available: https://www.nist.gov/system/files/documents/cyberframework/cybersecurity-framework-021214.pdf
[273] N. Poolsappasit, R. Dewri, and I. Ray, "Dynamic Security Risk Management Using Bayesian Attack Graphs", IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 1, pp. 61-74, June 2011, [Online], https://doi.org/10.1109/TDSC.2011.34
[274] R. Gowland, "The accidental risk assessment methodology for industries (ARAMIS)/layer of protection analysis (LOPA) methodology: A step forward towards convergent practices in risk assessment?", Journal of Hazardous Materials, vol. 130, no. 3, pp. 307-310, March 2006, [Online], https://doi.org/10.1016/j.jhazmat.2005.07.007
[275] J. Ren, I. Jenkinson, J. Wang, D. Xu, and J. Yang, "An offshore risk analysis method using fuzzy Bayesian network", Journal of Offshore Mechanics and Arctic Engineering, vol. 131, no. 4, Sep 2009, [Online], https://doi.org/10.1115/1.3124123
[276] K. Wrona, and G. Hallingstad, "Real-time automated risk assessment in protected core networking", Telecommunication Systems, vol. 45, no. 2-3, pp. 205-214, Jan 2010, [Online], https://doi.org/10.1007/s11235-009-9242-1
[277] P. Mell, K. Scarfone, and S. Romanosky, "Common Vulnerability Scoring System", IEEE Security & Privacy, vol. 4, no. 6, pp. 85-89, Dec 2006, [Online], https://doi.org/10.1109/MSP.2006.145
[278] J. Andress,"The basics of information security: understanding the fundamentals of InfoSec in theory and practice", pp. 240, Syngress, Elsevier, ISBN: 978-0-12-800744-0, 2014. [Online], https://doi.org/10.1016/C2013-0-18642-4

[279] T. D. Nielsen, and F. V. Jensen,"Bayesian networks and decision graphs", pp. 268, New york, USA, Springer Science & Business Media, ISBN: 978-1-4757-3502-4, 2001. [Online], https://doi.org/10.1007/978-1-4757-3502-4

[280] K. Sultan, A. En-Nouaary, and A. Hamou-Lhadj, "Catalog of Metrics for Assessing Security Risks of Software throughout the Software Development Life Cycle", in 2008 International Conference on Information Security and Assurance (isa 2008). Busan, Korea (South), pp. 461-465, 24-26 April 2008, IEEE, https://doi.org/10.1109/ISA.2008.104

[281] B. Romero, M. Villegas, and M. Meza, "Simon's Intelligence Phase for Security Risk Assessment in Web Applications", in Fifth International Conference on Information Technology: New Generations (itng 2008). Las Vegas, NV, USA, pp. 622-627, 7-9 April 2008, IEEE, https://doi.org/10.1109/ITNG.2008.163

[282] L. Xiao, Y. Qi, and Q. Li, "Information Security Risk Assessment Based on Analytic Hierarchy Process and Fuzzy Comprehensive", in 2008 International Conference on Risk Management & Engineering Management. Beijing, China, pp. 404-409, 4-6 Nov. 2008, IEEE, https://doi.org/10.1109/ICRMEM.2008.71

[283] K. Clark, E. Singleton, S. Tyree, and J. Hale, “Strata-Gem: risk assessment through mission modeling”, in Proceedings of the 4th ACM workshop on Quality of protection, Alexandria, Virginia, USA, Association for Computing Machinery, pp. 51–58, 2008, https://doi.org/10.1145/1456362.1456374

[284] H. Pasman, and W. Rogers, “How can we improve HAZOP, our old work horse, and do more with its results? An overview of recent developments”, Chemical Engineering Transactions, vol. 48, pp. 829-834, 2016, [Online], https://doi.org/10.3303/CET1648139

[285] CREST, “What is Cyber Threat Intelligence and how is it used?”, 2019, [Online], Accessed: 15 Nov 2020, Available: https://www.crest-approved.org/wp-content/uploads/CREST-Cyber-Threat-Intelligence.pdf

[286] A. Bifet, G. Holmes, B. Pfahringer, P. Kranen, H. Kremer, T. Jansen, and T. Seidl, “MOA: Massive Online Analysis, a framework for stream classification and clustering”, in, Conference held at Windsor, UK, JMLR, pp. 44-50, 2010, Accessed: 24/10/2021, Available: https://hdl.handle.net/10289/4934

[287] A. Bifet, J. Read, B. Pfahringer, G. Holmes, and I. Žliobaitė, "CD-MOA: Change Detection Framework for Massive Online Analysis", in. Berlin, Heidelberg, pp. 92-103, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-642-41398-8_9

[288] A. Bifet, G. Holmes, B. Pfahringer, P. Kranen, H. Kremer, T. Jansen, and T. Seidl, “MOA: Massive Online Analysis, a Framework for Stream Classification and Clustering”, in Proceedings of the First Workshop on Applications of Pattern Analysis, Proceedings of Machine Learning Research, PMLR, pp. 44--50, 2010, Accessed: 23/08/2021, Available: https://proceedings.mlr.press/v11/bifet10a.html

[289] I. H. Witten, E. Frank, M. Hall, and C. Pal, "The WEKA workbench. Online appendix for “Data mining: practical machine learning tools and techniques”", Morgan Kaufmann, 2016, Accessed: 15/07/2021, Available: https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf

[290] A. Bernardo, E. Della Valle, and A. Bifet, "Incremental Rebalancing Learning on Evolving Data Streams", in 2020 International Conference on Data Mining Workshops (ICDMW), pp. 844-850, IEEE, https://doi.org/10.1109/ICDMW51313.2020.00121

[291] H. M. Gomes, R. F. d. Mello, B. Pfahringer, and A. Bifet, "Feature Scoring using Tree-Based Ensembles for Evolving Data Streams", in 2019 IEEE International Conference on Big Data (Big Data), pp. 761-769, 9-12 Dec. 2019, https://doi.org/10.1109/BigData47090.2019.9006366

[292] S. F. Guarino, L; Setola, R; Flammini, F;, "A hardware-in-the-loop water distribution testbed (WDT) dataset for cyber-physical security testing", 2021, https://doi.org/10.21227/rbvf-2h90

[293] L. Faramondi, F. Flammini, S. Guarino, and R. Setola, “A Hardware-in-the-Loop Water Distribution Testbed Dataset for Cyber-Physical Security Testing”, IEEE Access, vol. 9, pp. 122385-122396, 2021, [Online], https://doi.org/10.1109/ACCESS.2021.3109465

[294] H.-K. Shin, W. Lee, J.-H. Yun, and H. Kim, "Implementation of programmable {CPS} testbed for anomaly detection", in 12th {USENIX} Workshop on Cyber Security Experimentation and Test ({CSET} 19), Accessed: 15/08/2021, Available: https://www.usenix.org/conference/cset19/presentation/shin

[295] H.-K. Shin, W. Lee, J.-H. Yun, and H. Kim, "{HAI} 1.0: HIL-based Augmented {ICS} Security Dataset", in 13th {USENIX} Workshop on Cyber Security Experimentation and Test ({CSET} 20), Accessed: 15/03/2021, Available: https://github.com/icsdataset/hai

[296] S. Schorradt, E. Bajramovic, and F. Freiling, "On the feasibility of secure logging for industrial control systems using blockchain", in Proceedings of the Third Central European Cybersecurity Conference, pp. 1-6, https://doi.org/10.1145/3360664.3360668
[297] A. Nieto, R. Rios, and J. Lopez, "Digital Witness and Privacy in IoT: Anonymous Witnessing Approach", in 2017 IEEE Trustcom/BigDataSE/ICESS, pp. 642-649, 1-4 Aug. 2017, https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.295
[298] A. Nieto, R. Roman, and J. Lopez, "Digital Witness: Safeguarding Digital Evidence by Using Secure Architectures in Personal Devices", IEEE Network, vol. 30, no. 6, pp. 34-41, 2016, [Online], https://doi.org/10.1109/MNET.2016.1600087NM
[299] A. Nieto, R. Rios, and J. Lopez,"IoT-Forensics Meets Privacy: Towards Cooperative Digital Investigations", ISBN: 2018. [Online], https://doi.org/10.3390/s18020492
[300] M. Asante, G. Epiphaniou, C. Maple, H. Al-Khateeb, M. Bottarelli, and K. Z. Ghafoor, "Distributed Ledger Technologies in Supply Chain Security Management: A Comprehensive Survey", IEEE Transactions on Engineering Management, 2021, [Online], https://doi.org/10.1109/TEM.2021.3053655
[301] A. Maw, S. Adepu, and A. Mathur, "ICS-BlockOpS: Blockchain for operational data security in industrial control system", Pervasive and Mobile Computing, vol. 59, pp. 101048, 2019, [Online], https://doi.org/10.1016/j.pmcj.2019.101048
[302] T. Spyridopoulos, T. Tryfonas, and J. May, "Incident analysis & digital forensics in SCADA and industrial control systems", 2013, [Online], https://doi.org/10.1049/cp.2013.1720
[303] T. R. Vance, and A. Vance, "Cybersecurity in the Blockchain Era : A Survey on Examining Critical Infrastructure Protection with Blockchain-Based Technology", in 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), pp. 107-112, 8-11 Oct. 2019, https://doi.org/10.1109/PICST47496.2019.9061242
[304] R. Delgado, and X.-A. Tibau, "Why Cohen's Kappa should be avoided as performance measure in classification", PlosOne, vol. 14, no. 9, pp. e0222916, 2019, [Online], https://doi.org/10.1371/journal.pone.0222916
[305] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme", Biochimica et Biophysica Acta (BBA)-Protein Structure, vol. 405, no. 2, pp. 442-451, 1975, [Online], https://doi.org/10.1016/0005-2795(75)90109-9
[306] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview", Bioinformatics, vol. 16, no. 5, pp. 412-424, 2000, [Online], https://doi.org/10.1093/bioinformatics/16.5.412
[307] S. Wares, J. Isaacs, and E. Elyan, "Data stream mining: methods and challenges for handling concept drift", SN Applied Sciences, vol. 1, no. 11, pp. 1412, 2019, [Online], https://doi.org/10.1007/s42452-019-1433-0
[308] G. Krempl, I. Žliobaite, D. Brzeziński, E. Hüllermeier, M. Last, V. Lemaire, T. Noack, A. Shaker, S. Sievi, M. Spiliopoulou, and J. Stefanowski, "Open challenges for data stream mining research", SIGKDD Explor. Newsl., vol. 16, no. 1, pp. 1–10, 2014, [Online], https://doi.org/10.1145/2674026.2674028
[309] A. Bifet, "Classifier Concept Drift Detection and the Illusion of Progress", in. Cham, pp. 715-725, Springer International Publishing, https://doi.org/10.1007/978-3-319-59060-8_64
[310] A. Bifet, J. Read, I. Žliobaitė, B. Pfahringer, and G. Holmes, "Pitfalls in Benchmarking Data Stream Classification and How to Avoid Them", in. Prague, Czech Republic, pp. 465-479, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-642-40988-2_30
[311] I. Žliobaitė, A. Bifet, J. Read, B. Pfahringer, and G. Holmes, "Evaluation methods and decision theory for classification of streaming data with temporal dependence", Machine Learning, vol. 98, no. 3, pp. 455-482, 2015, [Online], https://doi.org/10.1007/s10994-014-5441-4
[312] F. Gustafsson,"Adaptive filtering and change detection", Citeseer, ISBN: 9780471492870 2000. [Online], https://doi.org/10.1002/0470841613
[313] The UK Robotics Growth Partnership, "The Cyber-Physical Infrastructure Report", 2022, Accessed: 15/02/2022, [Online], Accessed: 15/02/2022, Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1053955/cyber-physical-infrastructure-vision.pdf
[314] M. Harries, and N. S. Wales, "Splice-2 comparative evaluation: Electricity pricing", 1999, [Online], Accessed: 04/03/2022, Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.43.9013
[315] J. A. Blackard, and D. J. Dean, "Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables",

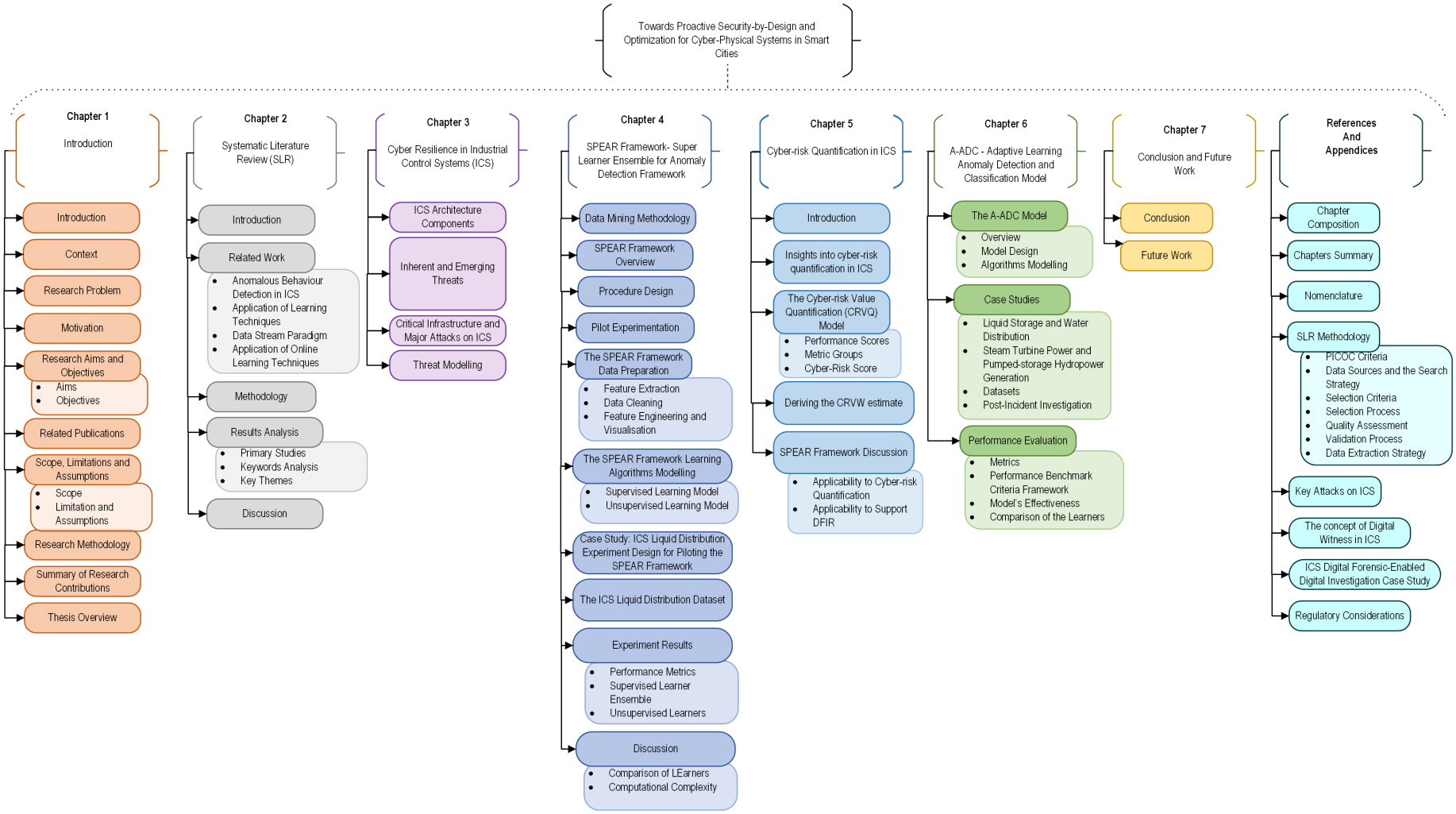
Computers and Electronics in Agriculture, vol. 24, no. 3, pp. 131-151, 1999, [Online], https://doi.org/10.1016/S0168-1699(99)00046-0
[316] D. G. Dua, C. , "UCI Machine Learning Repository", C. U. o. C. Irvine, School of Information and Computer Science., ed., 2019, Accessed: 04/03/2022, Available: http://archive.ics.uci.edu/ml]
[317] H. Dataverse, "Data Expo 2009: Airline on time data", Harvard Dataverse, 2008, https://doi.org/doi:10.7910/DVN/HG7NV7
[318] G. Ahmadi-Assalemi, and H. Al-Khateeb, "Blockchain technologies in the design of Industrial Control Systems for Smart Cities", IEEE Blockchain Technical Briefs, vol. Q2 2022, [Online], Accessed: 04/09/2022, Available: https://blockchain.ieee.org/images/files/pdf/techbriefs-2022-q2/blockchain-technologies-in-the-design-of-industrial-control-systems-for-smart-cities.pdf
[319] C. Miller, and C. Valasek, "Remote exploitation of an unaltered passenger vehicle", 2015, [Online], Accessed: 12/03/2020, Available: https://ericberthomier.fr/IMG/pdf/remote_car_hacking.pdf
[320] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering", in Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, England, United Kingdom, ACM, pp. 1-10, 2014, https://doi.org/10.1145/2601248.2601268
[321] Z. Li, and M. Shahidehpour, "Deployment of cybersecurity for managing traffic efficiency and safety in smart cities", The Electricity Journal, vol. 30, no. 4, pp. 52-61, 2017, [Online], https://doi.org/10.1016/j.tej.2017.04.003
[322] A. S. Sani, D. Yuan, J. Jin, L. Gao, S. Yu, and Z. Y. Dong, "Cyber security framework for Internet of Things-based Energy Internet", Future Generation Computer Systems, vol. 93, pp. 849-859, 2019, [Online], https://doi.org/10.1016/j.future.2018.01.029
[323] M. Salimitari, S. Bhattacharjee, and M. Chatterjee, "Prospect Theoretic Approach for Data Integrity in IoT Networks under Manipulation Attacks", arXiv preprint arXiv:1809.07928, 2018, [Online], Accessed: 07/01/2020, Available: https://arxiv.org/abs/1809.07928
[324] E. Arnautovic, "Consolidated state-of-the-art report", Computer Networks, vol. 54, no. 15, pp. 2787-2805, 2010, [Online], Accessed: 07/01/2020, Available: https://iot4cps.at/wp-content/uploads/2019/03/IoT4CPS_D2.1_V1.2b.pdf
[325] C. Schmittner, D. Ratasich, and M. Matschnig, "Design & Methods Concept", Transactions on Emerging Telecommunications Technologies, vol. 29, pp. e3308, 2018, [Online], Accessed: 07/01/2020, Available: https://iot4cps.at/wp-content/uploads/2019/03/IoT4CPS_D3.1_V1.0.pdf
[326] L. W. Xia Zhuoqun, Jiang Lalin, Xu Ming. Electric power CPS attack prediction method based on path analysis. Journal of Tsinghua University (Natural Science Edition), 2018, 58(2): 157-163., "Electric power CPS attack prediction method based on path analysis", 2018, [Online], Accessed: 11/07/2020, Available: http://jst.tsinghuajournals.com/CN/rhhtml/20180207.htm
[327] D. Püllen, N. A. Anagnostopoulos, T. Arul, and S. Katzenbeisser, "Poster: Hierarchical Integrity Checking in Heterogeneous Vehicular Networks", in 2018 IEEE Vehicular Networking Conference (VNC), pp. 1-2, IEEE, https://doi.org/10.1109/VNC.2018.8628375
[328] M. Suárez-Albela, P. Fraga-Lamas, and T. Fernández-Caramés, "A practical evaluation on RSA and ECC-based cipher suites for IoT high-security energy-efficient fog and mist computing devices", Sensors, vol. 18, no. 11, pp. 3868, 2018, [Online], https://doi.org/10.3390/s18113868
[329] V. Clincy, and H. Shahriar, "Detection of Anomaly in Firewall Rule-Sets", in International Conference on Applications and Techniques in Cyber Security and Intelligence, pp. 422-431, Springer, https://doi.org/10.1007/978-3-319-98776-7_46
[330] S. P. Singh, A. Nayyar, R. Kumar, and A. Sharma, "Fog computing: from architecture to edge computing and big data processing", The Journal of Supercomputing, vol. 75, no. 4, pp. 2070-2105, 2019, [Online], https://doi.org/10.1007/s11227-018-2701-2
[331] S. Hosseini, B. Turhan, and D. Gunarathna, "A Systematic Literature Review and Meta-Analysis on Cross Project Defect Prediction", IEEE Transactions on Software Engineering, vol. 45, no. 2, pp. 111-147, 2019, [Online], https://doi.org/10.1109/TSE.2017.2770124
[332] T. Hall, S. Beecham, D. Bowes, D. Gray, and S. Counsell, "A Systematic Literature Review on Fault Prediction Performance in Software Engineering", IEEE Transactions on Software Engineering, vol. 38, no. 6, pp. 1276-1304, 2012, [Online], https://doi.org/10.1109/TSE.2011.103
[333] T. Register, "Hacker jaled for revenge sewage attacks", 2001, [Online], Accessed: 25/02/2021, Available: https://www.theregister.com/2001/10/31/hacker_jailed_for_revenge_sewage/
[334] J. Slay, and M. Miller, "Lessons learned from the maroochy water breach", in International conference on critical infrastructure protection, pp. 73-82, Springer, https://doi.org/10.1007/978-0-387-75462-8_6

[335] A. S. Musleh, G. Chen, and Z. Y. Dong, “A Survey on the Detection Algorithms for False Data Injection Attacks in Smart Grids”, IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2218-2234, 2020, [Online], https://doi.org/10.1109/TSG.2019.2949998
[336] T. Alladi, V. Chamola, and S. Zeadally, “Industrial control systems: Cyberattack trends and countermeasures”, Computer Communications, vol. 155, pp. 1-8, 2020, [Online], https://doi.org/10.1016/j.comcom.2020.03.007
[337] Bloomberg, “Mysterious '08 Turkey Pipeline Blast Opened New Cyberwar”, 2008, [Online], Accessed: 25/02/2021, Available: https://www.bloomberg.com/news/articles/2014-12-10/mysterious-08-turkey-pipeline-blast-opened-new-cyberwar
[338] SANS, “Media report of the Baku-Tbilisi-Ceyhan (BTC) pipeline Cyber Attack”, 2014, [Online], Accessed: 25/02/2021, Available: https://ics.sans.org/media/Media-report-of-the-BTC-pipeline-Cyber-Attack.pdf
[339] D. Kushner, "The Real Story of Stuxnet: How Kaspersky Lab Tracked Down the Malware that Stymied Iran’s Nuclear-Fuel Enrichment Program,(I. Specturm, Producer)", 2013, Accessed: 28/02/2021, Available: https://courses.cs.duke.edu/spring20/compsci342/netid/readings/cyber/stuxnet-ieee-spectrum.pdf
[340] K. Kimani, V. Oduol, and K. Langat, “Cyber security challenges for IoT-based smart grid networks”, International Journal of Critical Infrastructure Protection, vol. 25, pp. 36-49, 2019, [Online], https://doi.org/10.1016/j.ijcip.2019.01.001
[341] K. E. Hemsley, and E. Fisher, "History of industrial control system cyber incidents", Idaho National Lab.(INL), Idaho Falls, ID (United States), Idaho, United States, 2018, Accessed: 15/10/2021, [Online], https://doi.org/10.2172/1505628
[342] United States Department of Justice U.S. Attorney's Office Southern District of New York, “Sealed Indictment”, 2016, [Online], Accessed: 28/02/2021, Available: https://www.justice.gov/usao-sdny/pr/manhattan-us-attorney-announces-charges-against-seven-iranians-conducting-coordinated
[343] R. M. Lee, M. J. Assante, and T. Conway, “German steel mill cyber attack”, Industrial Control Systems, vol. 30, pp. 62, 2014, [Online], Accessed: 27/02/2021, Available: https://ics.sans.org/media/ICS-CPPE-case-Study-2-German-Steelworks_Facility.pdf
[344] C. Lin, S. Wu, and M. Lee, "Cyber attack and defense on industry control systems", in 2017 IEEE Conference on Dependable and Secure Computing, pp. 524-526, 7-10 Aug. 2017, https://doi.org/10.1109/DESEC.2017.8073874
[345] J. E. Sullivan, and D. Kamensky, “How cyber-attacks in Ukraine show the vulnerability of the U.S. power grid”, The Electricity Journal, vol. 30, no. 3, pp. 30-35, 2017, [Online], https://doi.org/10.1016/j.tej.2017.02.006
[346] O. Andreeva, S. Gordeychik, G. Gritsai, O. Kochetova, E. Potseluevskaya, S. I. Sidorov, and A. A. Timorin, “Industrial control systems vulnerabilities statistics”, [Online], Accessed: 27/02/2021, Available: https://www.researchgate.net/profile/Sergey-Gordeychik/publication/337732465_INDUSTRIAL_CONTROL_SYSTEMS_VULNERABILITIES_STATISTICS/links/5de7842e92851c8364600e7e/INDUSTRIAL-CONTROL-SYSTEMS-VULNERABILITIES-STATISTICS.pdf
[347] N. H. Carreras Guzman, M. Wied, I. Kozine, and M. A. Lundteigen, “Conceptualizing the key features of cyber-physical systems in a multi-layered representation for safety and security analysis”, Systems Engineering, vol. 23, no. 2, pp. 189-210, 2020, [Online], https://doi.org/10.1002/sys.21509
[348] FireEye, “Attackers Deploy New ICS Attack Framework “TRITON” and Cause Operational Disruption to Critical Infrastructure”, 2017, [Online], Accessed: 28/02/2021, Available: https://www.fireeye.com/blog/threat-research/2017/12/attackers-deploy-new-ics-attack-framework-triton.html
[349] A. Di Pinto, Y. Dragoni, and A. Carcano, "TRITON: The first ICS cyber attack on safety instrument systems", in, Accessed: 28/02/2021, Available: https://scadahacker.com/library/Documents/Cyber_Events/Nozomi%20-%20TRITON%20-%20The%20First%20SIS%20Cyberattack.pdf
[350] S. Leppänen, S. Ahmed, and R. Granqvist, “Cyber Security Incident Report—Norsk Hydro”, 2019, [Online], Accessed: 27/02/2021, Available: https://mycourses.aalto.fi/pluginfile.php/923542/mod_folder/content/0/Group%20CSS%20Norsk%20Hydro%202019.pdf
[351] S. Falas, C. Konstantinou, and M. K. Michael, "Special Session: Physics- Informed Neural Networks for Securing Water Distribution Systems", in 2020 IEEE 38th International Conference on Computer Design (ICCD), pp. 37-40, 18-21 Oct. 2020, https://doi.org/10.1109/ICCD50377.2020.00022
[352] Cybersecurity and Infrastructure Security Agency, "Compromise of U.S. Water Treatment Facility", US, 2021, Accessed: 19/07/2022, [Online], Accessed: 19/07/2022, Available:

https://www.cisa.gov/uscert/sites/default/files/publications/AA21-042A_Joint_Cybersecurity_Advisory_Compromise_of_U.S._Drinking_Treatment_Facility.pdf
[353] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system”, Decentralized Business Review, pp. 21260, 2008, [Online], Accessed: 17/05/2022, Available: https://www.debr.io/article/21260.pdf
[354] I. Makhdoom, M. Abolhasan, H. Abbas, and W. Ni, “Blockchain's adoption in IoT: The challenges, and a way forward”, Journal of Network and Computer Applications, vol. 125, pp. 251-279, 2019, [Online], https://doi.org/10.1016/j.jnca.2018.10.019
[355] M. Banerjee, J. Lee, and K.-K. R. Choo, “A blockchain future for internet of things security: a position paper”, Digital Communications and Networks, vol. 4, no. 3, pp. 149-160, 2018, [Online], https://doi.org/10.1016/j.dcan.2017.10.006
[356] G. Epiphaniou, P. Pillai, M. Bottarelli, H. Al-Khateeb, M. Hammoudesh, and C. Maple, “Electronic Regulation of Data Sharing and Processing Using Smart Ledger Technologies for Supply-Chain Security”, IEEE Transactions on Engineering Management, vol. 67, no. 4, pp. 1059-1073, 2020, [Online], https://doi.org/10.1109/TEM.2020.2965991
[357] S.-W. Lin, B. Miller, J. Durand, G. Bleakley, A. Chigani, R. Martin, B. Murphy, and M. Crawford, “The industrial internet of things volume G1: reference architecture”, Industrial Internet Consortium, pp. 10-46, 2017, [Online], Accessed: 30/12/2020, Available: https://www.iiconsortium.org/IIC_PUB_G1_V1.80_2017-01-31.pdf
[358] B. Leander, A. Čaušević, and H. Hansson, “Applicability of the IEC 62443 standard in Industry 4.0 / IIoT”, in Proceedings of the 14th International Conference on Availability, Reliability and Security, Canterbury, CA, United Kingdom, Association for Computing Machinery, pp. Article 101, 2019, https://doi.org/10.1145/3339252.3341481
[359] National Institute of Standards and Technology, "Guidelines for smart grid cyber security", 2014, Accessed, [Online], https://doi.org/10.6028/NIST.IR.7628r1
[360] IEEE, “IEEE Standard for Intelligent Electronic Devices Cyber Security Capabilities”, IEEE Std 1686-2013 (Revision of IEEE Std 1686-2007), pp. 1-29, 2014, [Online], https://doi.org/10.1109/IEEESTD.2014.6704702
[361] National Institute of Standards and Technology Force Joint Task, "Security and Privacy Controls for Information Systems and Organizations", National Institute of Standards and Technology, 2020, Accessed, [Online], https://doi.org/10.6028/NIST.SP.800-53r5
[362] P. Cihon, “Standards for AI governance: international standards to enable global coordination in AI research & development”, Future of Humanity Institute. University of Oxford, 2019, [Online], Accessed: 02/01/2021, Available: https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf

9. Appendices

9.1 Chapters Composition



9.2 Nomenclature

Acronym	Description
A-ADC	Adaptive learning for Anomaly Detection and Classification
ABA	Anomaly Behaviour Analysis
ABC	AdaBoost Classifier
AC	Autoencoder
AC_b	Attack Complexity
ACM DL	Association of Computing Machinery Digital Library
ACPO	Association of Chief Police Officers
ADFM	Abstract Digital Forensic Model
ADWIN	Adaptive Windowing
AI	Artificial Intelligence
ANN	Artificial Neural Network
APT	Advanced Persistent Threat
Au_{er}	Authenticity Score
AV_b	Attack Vector
BBN	Bayesian Belief Network
BC	Bagging Classifier
BCh	Blockchain
C	Conference
Ch	Changed
Co	Confirmed
CART	Classification and Regression Trees
CCTV	Closed-Circuit Television
CD_e	Collateral Damage
CER	Containment, Eradication and Recovery
CIA	Confidentiality, Integrity and Availability
CISA	Cybersecurity and Infrastructure Security Agency
CNI	Critical National Infrastructure
CPN	Cyber-Physical-Natural
CPO	Cyber-Physical-Objects
CPPSs	Cyber–Physical Production Systems
CPS	Cyber Physical Systems
CRISP-DM	Cross-Industry Standard Process for Data Mining
CRISP-DM	Cross-Industry Standard Process for Data Mining
CROSSCAT	Centralised, Responsive, Objective, Systematic, Sharing, Continuous review, Accessible, Timely
CRVQ	Cyber-risk Value Quantification
CRV_{ro}	Cyber-risk Value of risk occurrence
CRV_{rs}	Cyber-risk Value of risk severity
CRV_t	Cyber-Risk Value
CTI	Cyber Threat Intelligence
CUSUM	Cumulative Sum
CV	Cross-Validation
CVSS	Common Vulnerability Scoring Systems
DCoC	Digital Chain-of-Custody

DDM	Drift Detection Method
DE	Digital Evidence
DF	Digital Forensics
DFIR	Digital Forensic and Incident Response
DFRWS	Digital Forensic Research Workshop
DL	Digital Library
DoD	Department of Defence
DoJ	Department of Justice
DoS	Denial of Service
DT	Decision Tree
DW	Digital Witnesses
DWT	Discrete Wavelet Transform
EC	Exclusion Criteria
ECDD	Exponentially weighted moving average Concept Drift Detector
EDDM	Early Drift Detection Method
ENISA	European Union Agency for Network and Information Security
EO	Executive Order
ERP	Enterprise Resource Planning
ESI	Electronically Stored Information
ETC	Extra Tree Classifier
EWMA	Exponential Weight Moving Average
FDI	False Data Injection
FIHTE	FeatureImportanceHoeffdingTreeEnsemble
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
GDPR	General Data Protection Regulation
GNB	Gaussian Naïve Bayes
H	High
HAT	Hoeffding Adaptive Tree
HAZOP	Hazard and Operability (study)
HIL	Hardware-In-the-Loop
HIPPA	Health Insurance Portability and Accountability Act
HMI	Human-Machine-Interface
HT	Hoeffding Trees
HWT	Hoeffding Window Tree
IACS	Industrial Automation and Control Systems
IC	Inclusion Criteria
ICS	Industrial Control Systems
ICT	Information Communication Technologies
IDA	Incident Detection and Analysis
IDIP	Integrated Digital Investigation Process
IDS	Intrusion Detection System
IEC	International Electro-technical Commission
IECEE	IEC System of Conformity Assessment Schemes for Electrotechnical Equipment and Components
IEEE	Institute of Electrical and Electronics Engineers
IF	Isolation Forest

IIoT	Industrial Internet of Things
IoT	Internet of Things
IP	Intellectual Property
IR	Incident Response
IRO	Initial Risk Occurrence
IRSI	Initial Risk Severity Impact
IS	Information Systems
ISO/IEC	International Organization for Standardization and International Electrotechnical Commission
IT	Information Technologies
ITS	Intelligent Transportation Systems
J	Journal
K	Cohen's Kappa
k-NN	k-Nearest Neighbour
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
L	Level
Lo	Low
LIE	Linear Incremental Estimator
LOF	Local Outlier Factor
LR	Logistic Regression
LRD	Local Reachability Density
LSB	Least Significant Bit
MCA	Multiple Characteristic Association
MCC	Matthews Correlation Coefficient
MDI	Mean Decrease in Impurity
MES	Manufacturing Execution System
MiM	Man-in-the-Middle
ML	Machine Learning
MOA	Massive Online Analysis
Ne	Network
Nn	None
N	Noise
NB	Naïve Bayes
NCSC	National Cyber Security Centre
NHS	National Health Service
NIS	Network and Information Security
NIST	National Institute of Standards and Technology
OSELM	Online Sequential Extreme Learning Machines
OT	Operational Technologies
Pe	Perceptron
Phy	Physical
P	Probability
PBC	Performance Benchmark Criteria
PCA	Principal Component Analysis
P_{er}	Possession Score
PESGS	Primal Estimated SubGradient Solver
PH	Parkerian Hexad
PHT	Page-Hinkley Test
PIA	Post-Incident Activity

PICOC	Population, Intervention, Comparison, Outcomes and Context
PLC	Programmable Logic Controllers
Pr_b	Privilege
Q	Quarter
QA	Quality Assessment
R	Required
RBM	Restricted Boltzmann Machines
RC_ABD_s	Report Confidence Anomalous Behaviour Detection
RC_A_s	Report Confidence Accuracy
ReLu	Rectified Linear Unit
RF	Random Forest
ROC AUC	Receiver Operating Characteristic Area Under the Curve
ROUF	Risk Occurrence Update Factor
RQ	Research Questions
RS	Rebalance Stream
RSUF	Risk Severity Update Factor
RW	Rolling Windows
S	Seasonality
SAM	Self Adjusting Memory
SANS	SysAdmin, Audit, Network, and Security
SCADA	Supervisory Control and Data Acquisition
SLR	Systematic Literature Review
SPEAR	Super learner Ensemble Anomaly detection cyber-Risk quantification (framework)
SVC	Support Vector Classifier
SVM	Support Vector Machine
SWS	Smart Water Systems
T	Trend
TA	Time Automata
TAC	Temporally Augmented Classifier
TCP/IP	Transmission Control Protocol/Internet Protocol
TDL-CNN	Tapped Delay Line Convolutional Neural Network
TE_PCS	Tennessee-Eastman process control system
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
Uch	Unchanged
U	Unknown
U_{er}	Utility Score
UI_b	User Interaction
US	United States
WEF	World Economic Forum
WEKA	Waikato Environment for Knowledge Analysis
WTD	Water Distribution Testbed

9.3 SLR Methodology

9.3.1 *PICOC Criteria*

The PICOC (population, intervention, comparison, outcomes, context) criteria as demonstrated in Table 35 are used from an engineering point of view, as proposed by Kitchenham and Charters [58] to frame the RQs effectively.

Table 35 Application of PICOC criteria [58] to the Research Questions (RQs).

PICOC Criteria	Criteria Description
Population	Frameworks addressing smart cities
Intervention	Digital forensic incident response (DFIR) frameworks that support cyber resilience
Comparison	Frameworks addressing cyber resilience
Outcomes	Scope, technique, security application and sector of the studies analysed
Context	Academic research

9.3.2 *Data Sources and the Search Strategy*

Digital library (DL) sources for computer science research publications were used. To help answer the RQs, keywords representative of the research topic were pre-defined and a search string was constructed using Boolean operators, key terms and synonyms to fetch all relevant studies. The Boolean operators were limited to AND and OR. The following search string was used:

('Cyber Physical Systems' OR 'Cyber-Physical Systems' OR 'CPS' OR 'Cyber Physical Object' OR 'CPO' OR 'smart device' OR 'IoT device') AND ('cybersecurity' OR 'cybersecurity' OR 'cyber-resilience' OR 'resilience') AND ('smart cities' OR 'smart city') AND ('model' OR 'modeling' OR 'technique' OR 'framework' OR 'information modeling' OR 'modeling technique' OR 'analytical modeling' OR 'reference architecture' OR 'reference model' OR 'Security Solutions' OR 'IoT Architecture')

The DLs used in this SLR were the Institute of Electrical and Electronics Engineers (IEEE), Association of Computing Machinery Digital Library (ACM DL), Science Direct, Web of Knowledge and Scopus. The search string was aligned to the built-in options within the DLs' search engines to filter the results. Where possible, searches were performed to match the search string from the title, abstract, keywords, and full text. The search of the specified DLs concluded by 5 April 2019 taking into consideration all studies returned by the defined search string published to that date. In addition to the set of studies produced through the search of the DLs, a snowballing approach search strategy was applied, as outlined by Wohlin [320], which produced a further set of relevant studies. This was a manual process applied to the studies collected by the pre-identified search criteria until no further studies met the inclusion criteria. After identifying studies from the specific data sources using the defined search string, the rest of the protocol outlined in Sections 9.4.3 - 9.4.7 was applied to the studies identified by the initial search.

9.3.3 Selection Criteria

Rigorous inclusion and exclusion criteria, as defined in Table 36, were applied to the produced set of studies from the DLs to ascertain that only relevant studies are retained in response to the RQs.

Table 36 Inclusion and exclusion criteria for the primary studies.

Inclusion Criteria (IC)	Exclusion Criteria (EC)
IC1: Must be a peer-reviewed, English-language primary study.	**EC1:** Duplicate studies.
IC2: Must contain CPS-specific information related to cyber resilience, modern DFIR or frameworks.	**EC2:** Study is not a framework that supports cyber resilience or DFIR.
IC3: Must include empirical evidence related to the cyber resilience security application and use of CPSs.	

Included studies must satisfy all IC. i.e., they must be primary, peer-reviewed, written in English and contain appropriate information on new applications or development of an existing mechanism for cyber resilience, modern DFIR or framework in CPS, providing empirical findings.

9.3.4 Selection Process

The selection process consisted of three key phases as demonstrated in Figure 63. The selection process was critically reviewed.

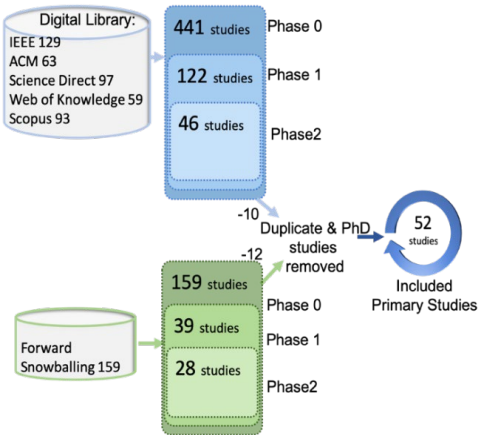


Figure 63 Primary studies selection process. IEEE—Institute of Electrical and Electronics Engineers; ACM—Association of Computing Machinery.

Phase 0—Keyword Filtering. During this phase, the identified search string was applied to each of the DLs utilised returning a combined result of 441 research studies. These studies were passed through to the next phase.

Phase 1—Title, Indexing Keywords, Abstract, and Conclusion Filtering. Following the initial keyword filtering, in phase 0, the titles, indexing keywords, abstracts and conclusion were scrutinised against the IC. Studies showing relevance to the research topic were included in the next phase. In this phase, 319 studies were excluded and 122 were put through to the final phase.

Phase 2—Full-Text Filtering. The full texts of the 122 studies were read. After applying the selection criteria in this final phase, some studies were excluded for

several reasons. For example, references [41, 321, 322] did not include an empirical study, references [323-325] at the time of review were not peer-reviewed publications, reference [326] is not an English language study, reference [327] is a poster, the focuses of references [328-330] were not specific to CPS cyber resilience or modern DFIR. Additionally, 10 studies were identified as duplicates and excluded from the final selection list. Snowballing identified an additional 159 studies. After applying the selection process, these studies were reduced to 19 after excluding nine duplicate studies and three PhD theses.

The final list of primary studies included in this SLR resulted in 52 articles, as shown in Figure 63.

9.3.5 Quality Assessment

Motivated by the guidance in reference [58], a checklist was developed according to references [331, 332] to make sure all included studies satisfy Quality Assessment (QA) criteria. This evidence-based approach assesses the validity of experimental data and reduces bias. The following QA criteria were applied:

Phase 1: CPS. The study must be focused predominantly on CPS security or the application of the CPS framework to a specific cyber resilience problem and appropriately documented.

Phase 2: Context. The context of the study must be provided in sufficient detail to accurately interpret the research.

Phase 3: Detail. The framework details are critical to answering RQ1 and RQ2. Sufficient detail about the approach to build the framework and comparison with other approaches must be presented clearly in assisting to answer RQ3.

Phase 4: Data. Sufficient detail about the type of training and test data identified and how the data was acquired, measured and reported must be provided clearly to determine the accuracy of the results reported.

9.3.6 Validation Process

A random set of 30 primary studies from the pool of studies were selected and had the inclusion/exclusion criteria re-applied to validate the effectiveness and objectivity of the process application. Further 30 random primary studies were selected from the pool of studies and had the QA criteria applied to validate the effectiveness and the application of the QA process.

9.3.7 Data Extraction Strategy

The data extraction was applied to the final 52 primary studies. Initially, the process and format were trialled on a subset of studies before extending the process to all included studies. The data was categorized, stored in a spreadsheet and tabulated using the following characteristics:

Context: year of publication, type of article, application of the study, sector, model type and security approach.

Qualitative data: was recorded including the conclusion and future research directions provided by the authors

Quantitative data: experiment observations were noted including the technique and dataset source.

9.4 Key Attacks on ICS

Table 37 Key attacks on ICS timeline of reported high profile-attacks on ICS for a period between 1990 and February 2021

Attack	Year	Description
Solar Sunrise	1998	Solar Sunrise is referred to as one of the earliest multi-stage cyber-attacks with international response against critical infrastructure. The threat actors targeted the US Department of Defence networks, breaching the US Navy Naval Sea Systems Command in Maryland [39]. The attack vector was a systematic exploitation of a vulnerability in the Sun Solaris operating systems that embedded a program to gather and later exfiltrated the data [39, 246]. Although this attack was considered the most organised and systematic by the Department of Defence, it was attributed to two Californian High School students [246].
Maroochy Water	2000	The Maroochy wastewater system in Queensland Australia suffered problems with the radio frequency signals that controlled the wastewater pumping stations. The relevant pump alarms did not activate. The threat actor, a disgruntled ex-employee, was found with SCADA equipment and software on their laptop to control the sewage management control system [333]. The attack vector in this case was a radio transmitter used to deliberately interfere with the radio signals of approximately 150 sewage pumping stations for about three months [334]. Over two months the attacker made 46 attempts to take control of the sewage system. This attack resulted in the release of millions of gallons of untreated sewage contaminating waterways and local parks [333, 334]. The attack is an example of an insider threat attack against ICS which resulted in physical damage and cascading effect on the environment and the public.
David-Besse	2003	A Denial Service Attack (DoS) Slammer worm compromised the David-Besse nuclear powerplant in Ohio, United States [335, 336]. The target of this attack was not the powerplant and no specific threat actor attribution was made. The attack vector used a backdoor in its ICS network from the Internet provider. This attack was caused by bypassing a firewall for one of the consultants' applications and an unpatched vulnerability for which a patch was available for at least six months leading up to the attack. The Slammer worm entered the ICS network through the bypassed firewall and exploited a buffer overflow in Microsoft's SQL engine slowing the servers down [336]. The consequence of the compromise was a five-hour outage of the parameter display system. The reactor was offline for major repairs at the time of the attack and the analogue reads from the sensors were not affected. However, the consequence of this attack could have been catastrophic.
Baku-Tbilisi-Ceyhan	2008	The Turkey Baku-Tbilisi-Ceyhan gas pipeline explosion was initially attributed to a cyber-attack [337]. However, no threat actor attribution was made. Moreover, according to the following reference [338], the initial attribution's credibility was doubted

		with a possibility of a physical attack. That said, a possible attack vector outlined in the SANS ICS report was the penetration of the ICS network using misconfigured IP-based cameras and vulnerable camera communication software gaining persistent access by further exploiting a Windows-based alarms server [338].
Stuxnet	2010	The Stuxnet malware attack targeted an Iranian nuclear plant [64]. The attack infected the control system networks of fourteen industrial sites in Iran including causing significant physical damage to a fifth of the Iranian nuclear centrifuges in the Natanz uranium-enrichment facility [5, 339]. The 500-kilobyte worm, one of the most advanced malware at the time, is believed to have targeted specific equipment within the Natanz nuclear plant. Stuxnet used two attack vectors. The first attack vector was delivered by a worker believed to have used a USB stick and malware wormed its way onto it, thus creating a simple propagation vector onto another machine, without a need for an internet connection. The second attack vector compromised the Centrifuges Drive Systems with malware posing as a legitimate windows driver impacting the centrifuges' rotor speeds leading to damage [336]. The attack was executed in three phases. Initially, it exploited four unpatched Microsoft vulnerabilities, two of which resulted in self-replication and the other two in privilege escalation which was unknown before this point, possibly zero-day vulnerabilities. In the second phase, it looked for a very specific Siemens Step7 Windows-based software which was used to program ICS which operated the centrifuges. Lastly, the worm compromised the PLC modifying the data such that the HMI displayed the wrong information which went undetected and enabled the attackers to cause physical damage [64, 68, 182, 339]. This is a shift from the conventional paradigm of the CIA [64]. Despite the initial threat actor, the paradigm shift to a well-financed and precisely executed sophistication of the attack is observed. This indicates a nation-state sponsor, with a political and not a financial gain motive [64, 339].
Bowman Avenue	2013	The Bowman Avenue Dam is a small hydraulic infrastructure to control storm surges [340-342]. The threat actors gained access to the floodgates [340-342]. The Bowman Dam's SCADA system used a cellular modem for internet connectivity. The attack vector leveraged unprotected infrastructure accessible from the Internet with no firewall protection or access control implementation. The attack took place during the SCADA maintenance during which the gate was manually disconnected with no control and only status monitoring capability in place. A Federal indictment attributed the attack to nation-state sponsored [341, 342].
German Steel Mill	2014	In December 2014 the German government's Bundesamt für Sicherheit in der Informationstechnik (BSI) (translated as Federal Office of Information Security) reported on an unspecified attack on a steel facility. While no attribution was asserted and the attack vector was not disclosed, according to SANS ICS [50], the threat actors had detailed knowledge of the ICS network. The attack failed multiple physical

		components. The attackers used spear phishing to gain access to the enterprise network, followed by lateral movement into the ICS network [343, 344].
Ukrainian Power Grid	2015	This attack on the Ukrainian power grid [65] seized control of the power grid's connected control system by compromising the SCADA system. The threat actors leveraged BlackEnergy malware exploiting Excel macros delivered using a spear-phishing technique [65]. The attackers used open-source information about the infrastructure from the ICS vendors. Additionally, the attackers leveraged a lack of two-factor authentication between the enterprise and the ICS network. Finally, the firewall configuration permitting remote administrative level connection from within the ICS utilising native system capabilities was also exploited. The consequence of the attack was a significant blackout leaving over 200,000 customers without electricity [5, 340]. Overall this attack resulted in physical damage to a number of the substations' physical components [65]. The attack impacted thirty substations and rendered the SCADA equipment inoperable. The electricity had to be manually restored. The Ukrainian power grid attack was the first to publicly acknowledge an incident which resulted in a power outage [65]. The threat actors showed expert knowledge of the network-connected infrastructure, operating ICS through a HMI [65]. The tactics, techniques and procedures employed at the Ukrainian power grid could be reused against other infrastructure across the world and a repeat of an attack is a real possibility [65, 345].
Ukrainian Power Grid	2016	The second attack launched against the Ukrainian power grid resulted in a power outage in its capital city, Kyiv. The attack affected 30 substations. The impact of the population scale was similar to the 2015 attack. However, additionally, a telephone denial of service attack was also launched. Although the recovery from the attack was within 3 hours, the recovery was manual due to the management systems affected by the attack. The attack was attributed to nation-state actors [341]. The fundamental difference from the first attack was the increase in sophistication of the threat actor's tactics, techniques and procedures. The second attack used a sophisticated malware "Crashoverride" which targeted the SCADA system, not a manual trip of the circuit breakers as was the case in the 2015 attack [341].
Kemuri Water Company	2015	The cyber-attack on an undisclosed water company in the United States uses the pseudonym 'Kemuri' to protect its identity [61]. The threat actors gained access to the applications controlling the PLCs for valve and flow control. Additionally, the sensors monitoring the plant were compromised and the levels of chemicals in the water treatment plant were altered [61, 346]. In this case, the ICS was internet facing and the attack could have resulted in serious damage to the public. Despite this attack exposing customer personal data, there is no reported subsequent evidence that this was misused [336]. The attack was attributed to a Syrian hacktivist group [336, 346].

Saudi Petrochemical Plant	2017	In the Saudi petrochemical plant, the threat actors targeted the industrial control unit's safety system to cause physical damage. However, a malfunctioning code resulted in the shutdown of the operation instead [336, 340]. This could be an indicator that the threat actors did not have detailed knowledge to successfully execute the attack in its entirety, however, they may use the developing knowledge to launch future larger-scale attacks. The attack vector leveraged the Triton malware to gain remote access. According to [336], the threat actors aimed to alter the codebase of the Safety Instrumentation System responsible for the plant's operational safety to cause significant physical harm. When the attack occurred, the failsafe mode was triggered resulting in an operational shutdown [336, 347]. No attribution has been made, and despite some indicators of Triton deployed attack suggesting state sponsorship other research does not draw the same conclusion [348, 349].
Norsk Hydro	2019	Manufacturing sector such as the Norwegian Norsk Hydro a renewable energy supplier that was targeted by the LockerGoga ransomware infecting forty sites and leading to production stoppage across Europe and the United States [40, 350]. Switching their operation to manual helped Norsk Hydro minimise the impact of the attack. In this attack, the threat actors exploited a weakness in the corporate Active Directory system for the attack propagation despite considerable weaknesses in their infrastructure related to endpoint security, security monitoring and system design [350].
National Water Supply	2020	The attack on Israel's Water Authority water treatment station's command and control targeted the PLCs operating the valves at several locations [351]. The threat actors aimed to disrupt the supply of water and alter the levels of chlorine. According to [351] the attackers succeeded in taking over the operations system at one of the stations.
Oldsmar Water Treatment Facility	2021	The attackers gained access to the SCADA system at Florida's Oldsmar water treatment facility and briefly increased the amount of sodium hydroxide a hundred-fold [352]. Poor password hygiene and outdated operating system were the likely attack vector used by the attackers. The facility provides water to about fifteen thousand residents and commercial establishments. The chemical being the main ingredient in drain cleaners, this attack could have had profound consequences [247].

9.5 The Concept of Digital Witness in ICS

Blockchain (BCh) emerged from the underpinning technology behind the Bitcoin cryptocurrency [353, 354]. Leveraged for financial transaction recording, BCh has unique characteristics. It is composed of cryptographically chained immutable blocks

that form a trusted, shared and distributed ledger of transactions. The blocks in the BCh are kept by peer-to-peer distributed management adopting consensus algorithms without needing a central authority or another intermediary [6, 354, 355]. Notably, nodes managing a BCh are constrained by the use of the same consensus algorithm. BCh is intrinsically incremental with each block being append-only, linked using secure hashes to the previous and subsequent blocks. The block comprises the hash, random nonce, root hash, timestamp and the metadata of all transactions immutably recorded with the ability to trace back to the Genesis block. Genesis is the name given to the first block in a given BCh. Due to the security capabilities including the decentralised architecture of trusted sources, cryptographic security, authenticity, responsibility for integrity and fault tolerance, BCh has the potential to support securing the IoT [70, 354, 355]. Scientific literature suggests that leveraging BCh to address the security of CPS and critical infrastructures to support modern DFIR is an active research area [6, 56, 356].

The role of the DW is to identify and preserve DE artefacts that can be stored on the device or transferred to other devices in the cloud, also known as Hearsay DW. To describe the DE data lifecycle from creation to destruction, reference is made to the Digital Forensic Research Workshop (DFRWS), the ISO/IEC 27050 general frameworks alongside adopting the categorisation of the data lifecycle context proposed by [56], see Figure 64. Additionally, to achieve admissibility, BCh has a unique advantage to initiate and maintain a DCoC. For example, the following study [6] introduced a tracking and liability attribution framework leveraging DW to enable the tracking of objects' behaviour within smart controlled business environments to detect insider threats. The authors proposed a framework leveraging BCh technology

to achieve DF readiness by establishing a DCoC and introduced the concept of DW to support post-incident investigations in smart controlled work environments. Another study [298] investigated the use of DW in personal devices, where personal devices can acquire, store and transmit DE to an authorised entity reliably and securely.

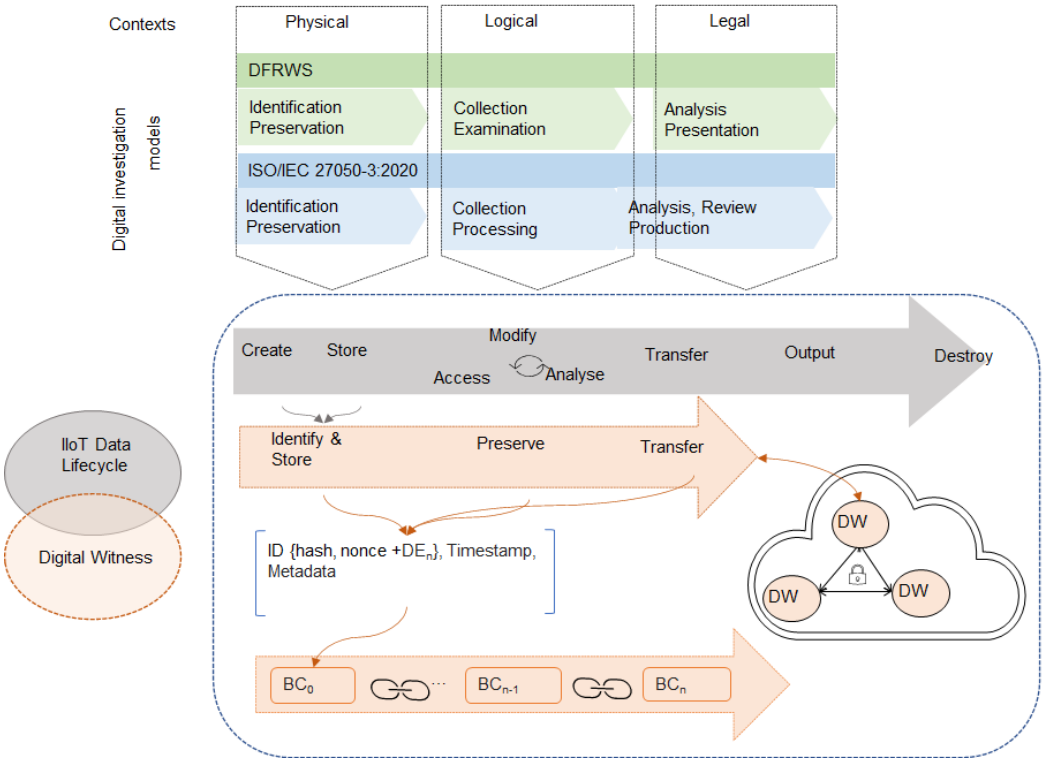


Figure 64 Concept of Digital Witness in ICS

A drawback of using BCh technologies is the computational complexity of the public BCh model. Despite the public BCh being based on Proof-of-Work and while it can withstand up to 50% compromised nodes, the implementation of the consensus protocol is capable of fewer transactions per second. However, the Proof-of-Authority leverages pre-authorised validators suited for a permissioned network. The related private BCh Practical Byzantine Fault Tolerance or Stellar Consensus is less computationally demanding, thus functionally capable of higher throughputs. That being said, they require a higher number of trustworthy nodes. Nonetheless, a

key characteristic of anomaly detection in ICS is that all objects must be known and pre-registered due to digital identity and access control. Hence, a permissioned BCh is more likely to provide the throughput, manageability, traceability, trust and integrity across a complex interdependent and distributed ICS within a common framework. BCh technology should be engineered and built into the ICS' design to achieve a verifiable audit trail, as shown in Figure 65, to support incident responders and facilitate DF readiness as part of proactive defence-in-depth.

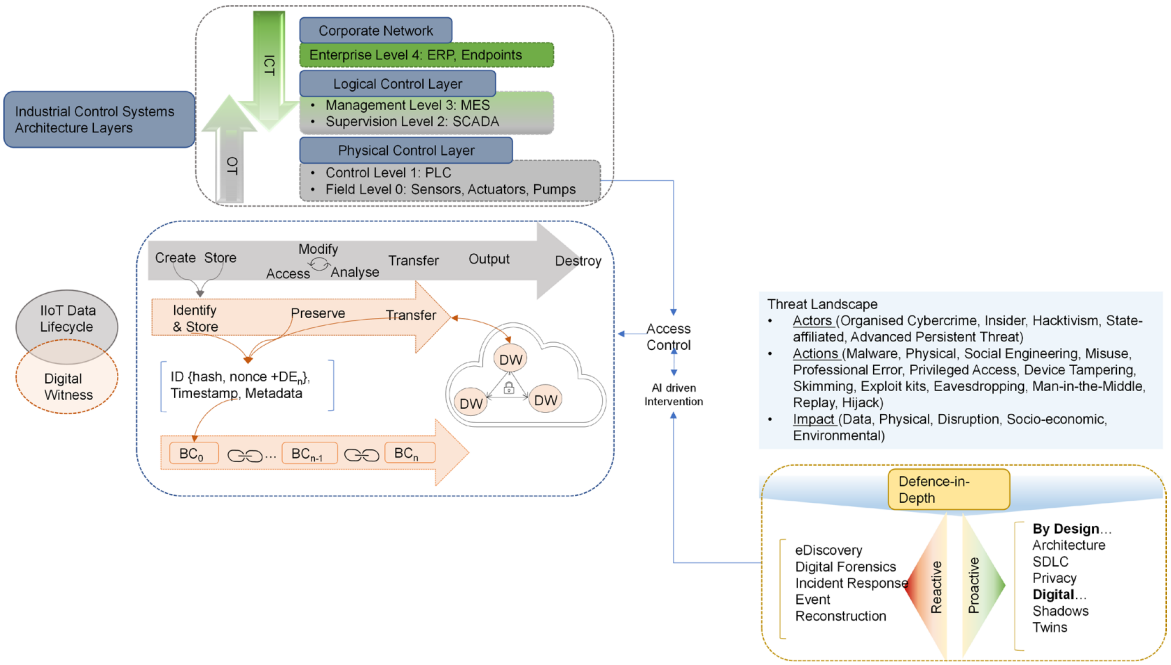


Figure 65 Integrating Blockchain technology into ICS design

9.6 ICS Digital Forensic-Enabled Digital Investigation Case Study

Alongside the value gained from the operational use of sensors-generated data, these data sources create multi-layered opportunities including support for developing intelligence capabilities and DF readiness. Firstly, Figure 66 demonstrates typical participants and how their interactions relate to a BCh-enabled ICS architecture for the physical sensors data pipeline. DW are considered cyber-physical objects which are functionally capable to maintain admissible DE including

receiving, storing and transferring DE following a pre-determined ruleset. *Employees and CPS objects* include authorised persons and smart physical or virtual objects with tamper-proof storage capable of performing tasks. *The supply chain* is considered to have a role in the development, operation and maintenance of ICT and OT concerning products, components, environmental and system parameters. *The incident investigation* includes the internal physical and virtual entities required for anomalous behaviour analysis and gatekeeping coordination of DE to legal authorities such as law enforcement agencies and the Courts of Law. Next, besides the participants and their interactions, the architectural approach requires the integration of distinct composable elements comprising data sources, innovative use of ML techniques and BC technologies, as referenced in Figure 66. On the premise of a digital investigation, all defined data sources are potentially DE analysed by leveraging AI-driven predictive modelling. Finally, depending on the threat model, permissioned BCh with smart contracts can be applied to control the ownership transfer at authorised hand-off gates to the DF investigation pipeline.

The threat landscape affecting ICS includes conventional IT and specific OT threats that range from external adversaries including APT to the prevalence of insider threats including social challenges such as accidental hazards, social engineering and disgruntled employees. In summary, the threat model for this use case could include:

- ICS interconnectivity with public networks could result in resourceful adversaries exploiting an attack vector and gaining access to the logic and the physical control layers, see Figure 33. This could result in the alteration of values resulting in an inconsistency between the actual and expected state of the physical process

resulting in anomalous behaviour captured within the physical sensors produced data.

- A use case could also suppose internal factors such as social engineering, disgruntled employee and supply chain exposure. Such actors have authorised unmonitored access to the operational infrastructure. This could result in an alteration of the software and environmental parameters configuration, which could cause deviation from the expected data patterns. Internal threats are often underestimated and challenging to detect [6, 22].

Either use case could potentially alter the physical processes resulting in physical damage with an impact on the wider society.

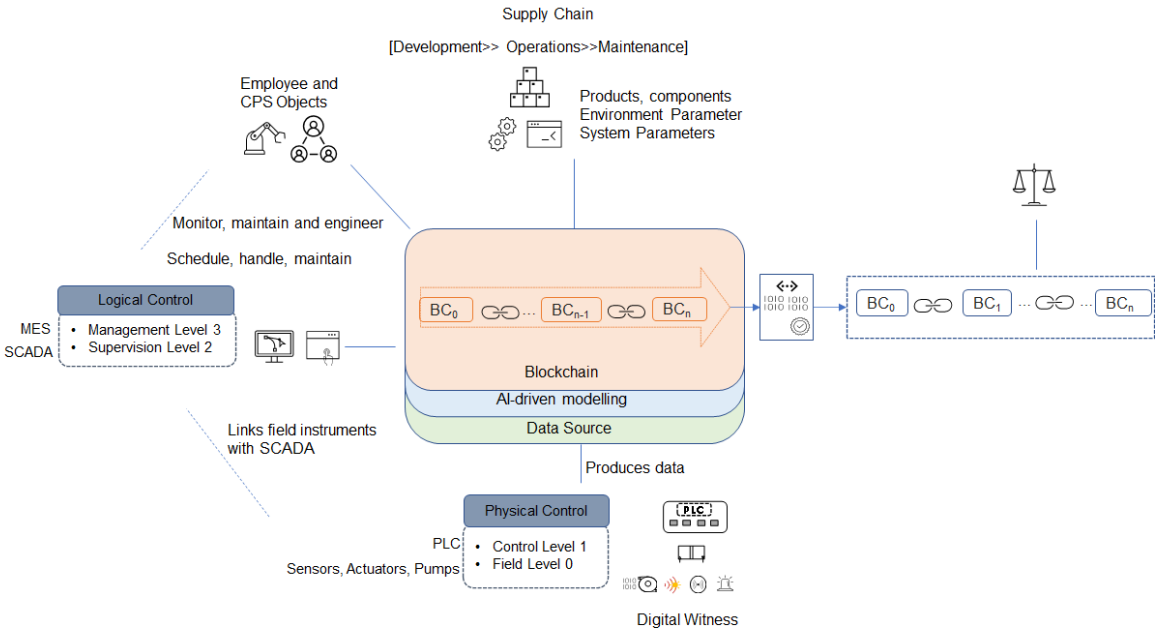


Figure 66 BC enabled ICS general framework participants' interactions

9.7 Regulatory Considerations

The regulatory landscape is multifaceted. Apart from the IIoT Reference Architecture [7, 357], IIoT systems have consumer-centric industry-specific standards and regulatory compliance requirements for information handling such as

the Health Insurance Portability and Accountability Act (HIPPA) [7] or the General Data Protection Regulation (GDPR). Specific to the security in IACS, a catalogue of standards is published by the International Electro-technical Commission (IEC) such as the IEC 62443 covering electronic security of control systems across several industry sectors [7, 358]. Specifically, IEC62443-3-3 which relates to the details of system security requirements and security levels has been accredited by the IEC System of Conformity Assessment Schemes for Electrotechnical Equipment and Components (IECEE). The NIST's NISTIR7628 revision 1 standard relates to smart grid cybersecurity [359] whereas the IEEE published standards on cybersecurity for intelligent electronic devices to accommodate critical infrastructure protection [360].

Additionally, NIST produced guidelines for the Network-of-Things in the NIST SP800-183 [226], and specifically a guide to ICS security in NIST SP 800-82 revision 2 [54] while mapping to further detailed security recommendations in NIST SP 800-53 revision 5 guidelines [361]. The Department of Homeland Security in the United States issued strategic principles for the security of IoT, whilst the European Union Network and Information Security (NIS) Directive guide to securely manage the connectivity between operational environments such as ICS or SCADA and the respective enterprises. In addition, the ISO published a security catalogue covering such as ISO/IEC 30141:2018 which focuses on IoT Reference Architecture. The ISO/IEC 27001:2013 covers Security Techniques, ISO 27002:2013 is aimed at information security management practice, and the ISO/IEC 27017:2015 specifically focuses on cloud services.

Although the regulatory landscape specific to DFIR aims to establish an international baseline it is also equally diverse, covered by several standards discussed by authors in [6, 56]. For example, the ISO/IEC 27043 focuses on incident investigation principles and processes, while ISO/IEC 27037 details DE acquisition. Methods for assuring the suitability and adequacy of incident investigations are covered by ISO/IEC 27041, ISO/IEC 27042 provides clarity on the analysis and interpretation phases of DE. The ISO/IEC 27050-1:2019, ISO/IEC 27050-3:2020 and ISO27050-2:2018 focus on the electronic discovery of DE.

Despite ongoing efforts to develop AI standards in the realm of AI-enabled systems, standards' proliferation and fragmented approach to threats make the convergence of standards challenging with currently over 80 frameworks in AI and related ethics [26, 362]. The regulations, standards and guidelines presented in this section are not exhaustive and are aimed at demonstrating the multi-disciplined complexity of the related regulatory and standards landscape.