



International Journal of Electronic Security and Digital Forensics

ISSN online: 1751-9128 - ISSN print: 1751-911X
<https://www.inderscience.com/ijesdf>

Feature-driven anomalous behaviour detection and incident classification model for ICS in water treatment plants

Gabriela Ahmadi-Assalemi, Haider Al-Khateeb, Tanaka Laura Makonese, Vladlena Benson, Samiya Khan, Usman Javed Butt

DOI: [10.1504/IJESDF.2025.10058572](https://doi.org/10.1504/IJESDF.2025.10058572)

Article History:

Received:	18 April 2023
Last revised:	16 June 2023
Accepted:	22 June 2023
Published online:	23 December 2024

Feature-driven anomalous behaviour detection and incident classification model for ICS in water treatment plants

Gabriela Ahmadi-Assalemi

University of Cambridge,
Cambridge, UK
Email: ga429@cam.ac.uk

Haider Al-Khateeb*

Cyber Security Innovation Centre,
Aston University,
Birmingham, UK
Email: h.al-khateeb@aston.ac.uk
*Corresponding author

Tanaka Laura Makonese

Faculty of Science and Engineering,
University of Wolverhampton,
Wolverhampton, UK
Email: t.l.makonese@wlv.ac.uk

Vladlena Benson

Cyber Security Innovation Centre,
Aston University,
Birmingham, UK
Email: v.benson@aston.ac.uk

Samiya Khan

Faculty of Science and Engineering,
University of Wolverhampton,
Wolverhampton, UK
Email: samiya.khan@wlv.ac.uk

Usman Javed Butt

Northumbria University,
London, UK
Email: usman.butt@northumbria.ac.uk

Abstract: Industry 5.0 envisions humans working alongside emerging technologies and enabled by the fusion of devices and sensors using information and communication technologies (ICT) to facilitate process automation, monitoring and distributed control in industrial control systems (ICS). However, the application of disruptor technologies and exposure of insecure devices broadens the attack surface making ICS an attractive target for sophisticated threat actors. Furthermore, ICS deliver a range of critical services hence disruption of industrial operations and services could have serious consequences. This study proposes an anomaly-based intrusion detection system for a water treatment plant based on a new model to determine variable significance for improved detection accuracy using machine learning (ML) algorithms coupled with incident classification based on functional impact. Determining statistical significance for independent ICS variables was addressed using logistic regression. Overall, 39 variables are deemed relevant in diagnosing the system state of the ICS operation to be expected or under attack. Our approach is validated using the secure water treatment (SWaT) testbed. Experimental results reveal that anomaly detection was effective using k-NN, ANN and SVM achieving an F1-score of 0.99, 0.98 and 0.97 respectively.

Keywords: critical national infrastructure; fifth industrial revolution; operational technology; smart city; advanced persistent threats; APT; artificial intelligence.

Reference to this paper should be made as follows: Ahmadi-Assalemi, G., Al-Khateeb, H., Makonese, T.L., Benson, V., Khan, S. and Butt, U.J. (2025) 'Feature-driven anomalous behaviour detection and incident classification model for ICS in water treatment plants', *Int. J. Electronic Security and Digital Forensics*, Vol. 17, Nos. 1/2, pp.1–29.

Biographical notes: Gabriela Ahmadi-Assalemi received her MSc degree in Cyber Security from the University of Northumbria, UK and her PhD degree in Cyber Security from the University of Wolverhampton, UK in 2018 and 2023 respectively. She is a Deputy Chief Information Security Officer at the University of Cambridge with over 20 years of experience in IT technical leadership. She has contributed research papers in high impact publications of international repute and presented her research at international conferences. Her research interests focus on defence-in-depth, attribution and anomalous behaviour detection in cyber-physical systems leveraging machine learning. She is a Fellow of the Higher Education Academy (FHEA) and a member of the Institution of Engineering and Technology (MIET).

Haider Al-Khateeb is an Associate Professor in Cyber Security at the Cyber Security Innovation (CSI) Research Centre and Aston Business School, Aston University, UK. He is a Senior Fellow of the Higher Education Academy (SFHEA), with a first-class BSc (Hons) in Computer Science, and a PhD in Cyber Security. He is a project lead for R&D and Knowledge Transfer activities with over 15 years of experience in the UK's HE sector. He is actively engaged with the latest trends in technical and operational cyber security knowledge areas such as adversarial behaviours, security operations and incident management, and privacy and online rights for vulnerable groups. He works with businesses to launch new products and delivered a range of hands-on technical training.

Tanaka Laura Makonese earned her MSc in Cyber Security from the University of Wolverhampton. Her research interests combine cyber security with digital technology advancement and data/information system analysis. Additionally, her experience in the financial sector has helped her to appreciate the importance of effective information and resource management.

Vladlena Benson is a Professor of Cyber Security Management, and director of the Cyber Security Innovation Centre at Aston University. Her research focuses primarily on secure technology exploitation, cyber security risk management and privacy enhancing technologies, which is published by world-leading journals. Alongside her academic role, she is working with multiple organisations and government moulding future national policy within cyber security, standards and partnerships. This includes, membership of the UK Cyber Security Council (UK CSC) working groups, Information Systems Audit and Control Association (ISACA) Board, and serving on the European Union Agency for Cybersecurity (ENISA) task force defining the criteria for European Cybersecurity Framework at the EU level.

Samiya Khan is a Research Fellow at University of Wolverhampton, UK. She is an alumna of University of Delhi, India and did her PhD in Computer Science from Jamia Millia Islamia, India. She has contributed several research papers and her publications extend across journal articles, conference papers, book chapters, magazine articles and edited books in high impact publications of international repute. His research interests include the multi-faceted use of artificial intelligence and edge computing for IoT applications.

Usman Javed Butt is a Cyber Security practitioner and an academic with over 19 years of experience. He is currently Associate Dean and Head of Department Computing at Northumbria University London. Over the years he has gain experience in strategic leadership, cyber security training and have delivered international talks on contemporary cyber security challenges. He is also external examiner to various UK universities for their cyber security and artificial intelligence programmes and is an active researcher and cyber security consultant.

1 Introduction

Industrial control systems (ICS) integrate with disruptive technologies and innovative solutions to support modern critical networked utility infrastructures such as water treatment systems, electricity generation plants and gas pipelines (Rakas et al., 2020; Dhirani et al., 2021). However, with disruptive technologies including 5G, edge computing, and artificial intelligence (AI) increase the ICS' attack surface by introducing new attack vectors (Dhirani et al., 2021; Bakalos et al., 2019; Kiss et al., 2014k; Ahmadi-Assalemi et al., 2020). Additionally, the proliferating availability of cyber tools and service lowers the entry point for threat actors to gain adversarial capabilities. These threats usually degenerate into real-time adversarial attacks that cause disruption. Consequently, cyber-attacks against ICS could result in real-world damage and have profound consequences for entire communities (Lin et al., 2018; Ahmadi-Assalemi et al., 2019). Therefore, ICS must continually adapt to deal with the current and manage the risk against future increasingly sophisticated and hostile threat landscape within their complex threat environment. Additionally, organised cybercrime, including emerging risks from state-aligned adversaries, requires timely detection to isolate and eradicate attack methods (Bakalos et al., 2019; Bristow, 2021). The motivation for this research study is rooted in the opportunity offered by machine learning (ML) techniques to address the prevalent and emerging threats in ICS and to advance knowledge in this field.

In recent years, ML approaches have been adopted in numerous science domains to solve problems relating to threat detection (Abdulhammed et al., 2019). Meanwhile, today's industry-standard ICS lack inbuilt capabilities to effectively detect anomalous behaviour and forecast system malfunctions (Liu et al., 2018). Statistical, specification, signature and rules-based approaches cannot deal with encrypted traffic and unknown attacks (Elmrabit et al., 2020). Besides requiring high processing time to check and update signatures, in the absence of an identified vulnerability, an attack signature is not available (Moustafa et al., 2018). Likewise, specification-based detection requires detailed system definitions and an accurate understanding of the physical process which is often challenging due to the complexity of the ICS environment (Junejo and Goh, 2016). Therefore, in several scenarios, these methods are outperformed by behaviour-based ML techniques that detect previously unseen attacks. This is accomplished by training a predictive ML model to classify any behaviour that deviates from normal events (Elmrabit et al., 2020; Junejo and Goh, 2016; Apruzzese et al., 2018; Ahmadi-Assalemi et al., 2022). As part of a defence mechanism, ADS can be developed and optimised to be fit-for-purpose to address anomalies manifesting in ICS that include large-scale and high-dimensional datasets (Alimi et al., 2021; Feng et al., 2019; Mitchell and Chen, 2014). ML methods that consist of supervised, semi-supervised and unsupervised algorithms, can achieve a binary or multiclass classification of normal and anomalous behaviour from physical plant sensor data in ICS (Robles-Durazno et al., 2018). Furthermore, ML analytical capabilities require no or limited human intervention to improve industrial processes (Elmrabit et al., 2020). While the range of ML techniques shares the same objectives, the structure of their models including their strengths and weaknesses vary. In this study, we apply supervised ML algorithms for binary classification to realise an indication of anomaly (IoA), then assess the indication of compromise (IoC) as part of an incident classification model for water treatment plants (IWaP). Prior research informs the selection of the algorithms, based on their characteristics, use and performance to improve and automate the detection of anomalies as part of resilient ICS.

This study addresses the following research questions: How can we determine the right sensors/actuators data variables to facilitate anomaly detection for a given ICS scenario? How can anomalies in ICS be detected using ML leveraging support vector machines (SVM), k-nearest neighbour (k-NN) and artificial neural network (ANN)? Which of the selected ML algorithms accurately detect anomalies in a given ICS scenario? And how can we classify detected incidents in this context? This research aims to develop and evaluate a new model to improve the detection and classification of anomalous behaviour in ICS in the context of a water treatment system and present a direction for future research. Our key contributions are as follows:

- a physical process variables selection method to facilitate anomaly detection in ICS
- ICS physical control process-based anomalous behaviour detection method modelled to leverage SVM, k-NN and ANN learning algorithms
- incident classification process for ICS based on the functional impact.

In the remainder of the paper, we discuss the background and related work in Section 2. Introduce the IWaP model in Section 3 and the research methodology in Section 4. Then,

we present our results and related evaluation in Section 5, the discussion in Section 6, and finally the conclusion in Section 7.

2 Background and related works

2.1 ICS and water treatment systems

ICS consist of collections of programmable logic controllers (PLC), human machine interfaces (HMI), supervisory control and data acquisition (SCADA) subsystems that control field devices or equipment through the monitoring and processing of data associated with industrial processes (Feng et al., 2019; Hadžiosmanović et al., 2014). Centralised data acquisition and control are vital for maintaining the operational efficiency of the ICS (Rakas et al., 2020). SCADA systems are commonly used in critical industrial infrastructures such as water treatment systems. According to the following study (Yang et al., 2017), ICS embody a variety of heterogeneous hardware and software tools that synergistically support industrial operations. However, Yeckle and Abdelwahed (2017) highlighted that ICS were not traditionally designed to sustain networking-enabled applications. SCADA systems have evolved from large first-generation systems to dispersed systems of the second generation that were based on exclusive network technologies. This evolution guided the development of the third-generation, fully networked modern systems that leveraged the use of internet technologies (Tariq et al., 2019). In modern environments, most of the ICS components are connected to telecommunications networks and interact with the internet (Kaspersky, 2017). Today's ICS are fully integrated with the fourth industrial revolution technologies such as big data analytics, internet of things (IoT), cloud computing, robotics, mobile computing and AI to fulfil industry requirements.

2.2 Inherent and emerging threats in ICS

ICS components that are connected to the internet expose them to sophisticated adversaries who target and disrupt the steady functioning of these systems with destructive consequences, often exploiting poorly protected elements that support control processes (Lin et al., 2018; Ahmadi-Assalemi et al., 2019; Lier, 2017). While literature indicates the prevalence of attacks against ICS and CNI (Hemsley and Fisher, 2018; Bristow, 2021), research also highlights vulnerabilities affecting the supply chain (Claroty Team 82, 2021). Furthermore, the report by Kaspersky (2021) noted that the internet, removable media and email are the most prominent threat sources in industrial infrastructures. Of the 33.8% who attacked ICS machines, 18.2% utilised the internet as an attack vector.

Although the frequency of announced attacks against ICS is limited, several substantial high-profile incidents have been reported and state-aligned adversaries including with ideological motivation are emerging. For example, in 1982, a Trojan was implanted in SCADA software which triggered a gas burst along the Trans-Siberian pipeline (Robles-Durazno et al., 2018). In 2000, a sewerage control system in Australia's Queensland was attacked causing massive spillage of raw sewage into the surrounding town (Cárdenas et al., 2011). Besides these, the nuclear network infrastructure in Iran was targeted by the Stuxnet worm between 2005 and 2010 (Robles-Durazno et al., 2018).

Other noteworthy attacks include the Bowman Avenue Dam in 2013, where hackers took control of floodgates that monitor water levels and control flow rates (Hemsley and Fisher, 2018), and the German Steel Mill security breach in 2014, which was perpetrated by a credential theft via a spear-phishing email announcement, causing damage to the plant's blast furnace (Lee et al., 2014). In addition, the 2015 attack on the power grid in Ukraine's capital city, Kyiv, caused a power outage affecting 200,000 Ukrainians (SANS ICS, 2016).

Besides threats from external actors, ICS are susceptible to internal factors such as malicious intent and accidental hazards including professional error. Anomalies that arise from internal or external attacks require effective detection near real-time to ensure stable systems operation. We define *anomalies* as random variations in the state or condition of a system that causes measurable effects, and it is typically an outlier event (Cook et al., 2020). Devices within an ICS process use operating actuators and reading sensors to handle variables such as temperature, pressure, and flow (Wang et al., 2014). According to Goh et al. (2017) any anomalous behaviour due to erroneous sensor measurements or readings, malfunctioning sensors and actuators, incorrect states of devices and mismatches with defined ICS standard thresholds may correspond to a detection event. Faulty sensor and actuator measurements or states could directly violate the operational safety bounds and conditions set for the water treatment ICS and potentially endanger public health and safety. Thus, developing methods that detect anomalous system states is critical to improving the cyber resilience in ICS as part of a layered defence-in-depth approach.

2.3 Anomaly detection

Anomaly or outlier detection is a class of techniques that identifies unusual activities, events or elements concerning large amounts of data being generated or ingested into a system (Cook et al., 2020). The development of robust anomaly detection techniques is an ever-growing challenge as attackers' appetite, motivation and capabilities of attackers to launch attacks on critical national infrastructures grow by the day (Elmrabit et al., 2020). Various threats including botnets, spoofing, session-hijacking, Denial-of-Service (DoS) attacks, botnets and SQL injections are some of the threats endangering ICS stability industry-wide.

Traditionally, signature-based techniques are used for detecting attacks on a network or system. These techniques make use of pre-identified indicators for detection. However, there has been a significant rise in the number of zero-day attacks in modern environments. These high-risk attacks are capable of causing severe ICS disturbances because of their ability to exploit software vulnerabilities before the developers become aware of them (Moustafa et al., 2018). Attacks such as these cannot be detected using traditional signature-based techniques, which leads to the need to develop advanced techniques for this purpose. Moreover, existing anomaly detection techniques use a reference point of the ICS, which is inadequate evidence for ascertaining if the system abnormality results from an attack or a fault (Kaouk et al., 2019). With this backdrop, Elmrabit et al. (2020) emphasises the urgent need to develop robust systems, specifically for industries where sensitive information is held and critical operations occur.

Incorporating new technologies has created numerous paths and vectors for cyber-criminals to execute malicious acts (Givehchi et al., 2013). For example, water treatment and distribution systems are usually large, dispersed and comprise several

exposed elements in public areas (Abokifa et al., 2019). Given such high exposure, the water supply systems encounter a significantly large number of anomalies that are either caused by general system failures or intended system attacks. Unfortunately, industry-standard ICSs have limited security capabilities for detecting anomalous behaviour and forecasting system malfunctions (Liu et al., 2018). Consequently, the need for the development of models that fetch and classify ICS anomalous behaviour is of high priority, and the state-of-the-art suggests the use of ML algorithms for effective anomaly detection (Feng et al., 2019).

2.4 Learning techniques for anomaly detection

ML algorithms are the most used techniques for detecting, classifying, and predicting anomalies (Elmrabit et al., 2020). These algorithms have initiated many technological breakthroughs in the cybersecurity space. Furthermore, their ability to develop systems that learn from past experiences can effectively be used to classify the behaviour of industrial systems (Feng et al., 2019). ML algorithms are categorised into supervised, semi-supervised and unsupervised learning algorithms. Supervised learning algorithms use labelled data to train the model, making predictions for an unknown dataset (Dalal, 2020). On the other hand, unsupervised ML algorithms use unlabelled data for training the model. They identify input data points by comprehending the underlying structure of the dataset. Semi-supervised ML algorithms draw inspiration from both supervised and unsupervised ML. In the case of SCADA systems, supervised ML algorithms are arguably the most popular choice for implementing security models and solutions (Alimi et al., 2021).

SVMs are extensively applied for classification and regression (Shang et al., 2018). Most SCADA system operations can be reduced and represented as a binary classification problem thus making the use of SVM algorithms suitable for these tasks. Existing literature attests to the effective use of SVM for such problems (Schuster et al., 2015; Terai et al., 2017; Turkoz et al., 2020). K-NN is a distance-based supervised ML algorithm, which is simple to design, easy to interpret and incurs low computation time (Alimi et al., 2021). K-NN classifies unknown data points based on their distance from the neighbouring, known data points (Arora et al., 2021). This technique has been used for modelling intrusion detection on the SCADA dataset in Alimi et al. (2021) and Gumaie et al. (2020). ANN is similar to the SVM in handling multi-dimensional data with good generalisation. In its simplest form, ANN uses perceptrons for identifying linearly separable patterns (Alimi et al., 2021). Classification models with proven effectiveness for SCADA systems have been developed and validated (Kosek and Gehrke, 2016; Demertzis et al., 2017; Shalyga et al., 2018).

A previous study (Junejo and Goh, 2016) applied nine supervised ML algorithms for detecting and classifying behaviour-based attacks on the raw water stage 1 (P1) of the secure water treatment (SWaT) testbed. The data was exposed to 18 different attacks based on sensor and actuator values. SVM, J48 decision tree, multinomial logistic regression (LR), Bayesian network (BayesNet), neural networks (NN), naive Bayes (NB), random forest (RF), instance-based learning (IBK) and best-fit (BF) tree algorithms were used to develop corresponding models. The results of the study demonstrated the best accuracy and precision for a BF tree algorithm. However, the proposed system is susceptible to new attacks as it cannot detect attacks that are not included in pre-selected categories. Another study utilising this dataset (Goh et al., 2017) targeted the

identification of attacks on P1 proposed using an unsupervised learning algorithm, recurrent neural networks (RNN), as the predictor model. After normalising the features of all the P1 sensor and actuator values, the RNN model is then trained and tested. A cumulative sum technique is applied to ascertain irregularities between the actual sensor data and the outputs of the trained model. The created RNN model detects most of the anomalies in the control network with low false-positive rates (FPR) in addition to determining which of the sensors was attacked.

Both previously discussed studies (Junejo and Goh, 2016; Goh et al., 2017) investigated intrusions against the SWaT testbed. Inoue et al. (2017) study used the complete dataset to investigate the performance of a deep neural network (DNN) and one-class support vector machine (OCSVM) for detecting irregularities in a water treatment plant. Inoue et al. (2017) applied unsupervised learning for model development. The advantage of using this approach comes from the fact that this type of learning permits the derivation of models solely based on data presented and does not require an understanding of the target ICS complexities. In comparison to injecting part of the SWaT dataset with 18 attack scenarios in Junejo and Goh (2016), Inoue et al. (2017) made use of the 36 attack scenarios applicable to the entire dataset. The latter study included actuator, sensor and network traffic data gathered over 11 days of non-stop operations. The water treatment dataset contained benign data produced under days of the regular operation of the testbed and malicious data comprising 36 anomaly scenarios. The results of the study proved that DNN performs better than OCSVM.

Detection of integrity attacks on SCADA systems was studied by Almalawi et al. (2014), in which k-NN and clustering techniques were employed to facilitate automatic identification and extraction. During the identification stage, the density factor for the k-NN of a data point was used for computing an inconsistency score. An optimal inconsistency threshold was then determined to differentiate between anomalous and expected behaviour. During the extraction stage, the prominent fixed-width clustering method was extended to extract proximity-detection rules, which were then used to form a small dataset that best represented both normal and anomalous behaviours visible in the training dataset. Extensive experiments were carried out on one real dataset from an urban wastewater treatment plant and two simulated water distribution system datasets. The real dataset consisted of daily measuring of sensors from an urban wastewater plant that consisted of 38 data nodes, approximately 527 observations with 14 labelled inconsistent. The simulated datasets were produced from a virtual SCADA lab for water distribution systems using visualisation features to represent the SCADA components and the EPANET library of hydraulic and water quality models. These datasets consisted of 23 nodes with approximately 105,000 observations with 100 inconsistent observations per dataset. The anomaly detection accuracy of the proposed approach was determined to be high and comparable to other state-of-the-art approaches. However, since this approach requires the computation of an inconsistency score for each observation, it proves to be computationally expensive, which is one of its most profound disadvantages.

Other approaches include a hybrid model for intrusion detection system (IDS) that detects anomalies using a combination of Bloom filter and instance-based k-NN algorithm (Khan et al., 2019) and combined use of convolutional neural networks (CNN) and principle component analysis (PCA) for detection of anomalous behaviour (Priyanga S et al., 2020). Both these methods have exhibited effectiveness in identifying a high number of attacks within minimal time. Besides these, a deep learning-based approach for anomaly detection system (ADS) was also proposed, which uses CNN for updating

the model of normality according to changes observed through continuous monitoring of sensor states in the industrial environment (Abdelaty et al., 2021). In another study, an effective ML-based approach was proposed for real-time anomaly detection in a clean water supply system (CWSS) (Robles-Durazno et al., 2018). One of the sensors in the system monitors energy consumption. Data collected from this sensor is marked benign or malicious, which is then used to test the performance of SVM, k-NN, and RF algorithms. A testbed implemented by Festo MPA Control Process Rig at Edinburgh Napier University was used for validating the proposed approach. The results of this study indicate that RF performs better than SVM and k-NN in terms of accuracy and F-measure. However, k-NN incurred lesser computation time with compromised accuracy and F-measure.

Along similar lines, Arora et al. (2021) also validated that RF and k-NN are better-performing algorithms for anomaly detection and classification in SCADA systems than SVM, NB and Decision Tree (DT). Existing research (Phillips et al., 2020) has also examined the viability of ML algorithms in detecting new security threats specific to SCADA systems and the Modbus protocol. This work demonstrated the use of ML algorithms such as SVM, DT, k-NN and k-means clustering, to detect unsafe traffic in a dataset produced via a remote terminal unit (RTU), which leverages the Modbus protocol. This communication protocol is typically used by industrial devices in SCADA systems and does not have an established authentication or authorisation system. As a result, data integrity is not verified. Results conclude that SVM is the best-performing algorithm among the four algorithms used in the study. In another study (Robles-Durazno et al., 2020) on the detection of attacks on the Edinburgh Napier University CWSS generated from a Festo MPA compact workstation, k-NN and SVM performed better than NB, DT and multilayer perceptron (MLP) in terms of accuracy, false positive and false negative rates. In contrast to the results presented by previous studies, Robles-Durazno et al. (2018) identified RF as the best-performing algorithm, highlighting the relatively lower performance of SVM for its use case. Thus, it can be inferred that SVM and k-NN are robust single classifiers, but they prove to be weaker when compared to ensemble classifiers such as RF.

A forecasting approach using NN for early anomaly detection was provided by Shalyga et al. (2018). This approach introduces a genetic algorithm-based technique for optimising network architecture selection and was tested on the SWaT ICS testbed dataset. NAB-score and F-measure are used as evaluation metrics. The best-performing NN model (MLP) could accurately detect 25 out of 34 anomalies, translating to seven false positives. This model had an F-score of 0.812, with a precision of 0.976 and a recall of 0.696. Furthermore, the proposed model had only an average anomaly detection delay of 11% compared to the complete anomaly's duration. This approach can also interpret anomalies concerning the sensor's location, which can be incorporated into future systems.

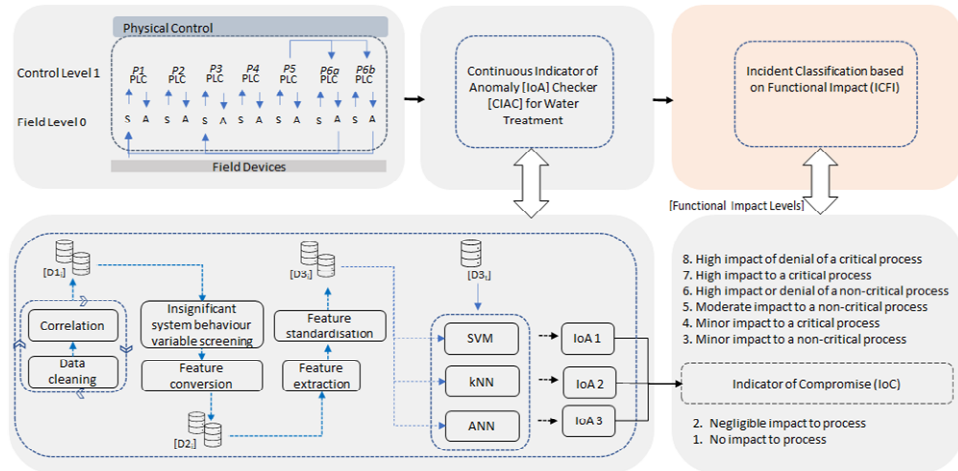
Although existing literature suggests using supervised and unsupervised learning methods for intrusion detection in ICS, supervised learning algorithms have reduced false alarm rates and better detection performance (M.R. et al., 2021). The use of standalone supervised algorithms for intrusion detection in SCADA systems has been proposed by many authors, but ensemble methods have demonstrated improved performance (Robles-Durazno et al., 2020; Tamy et al., 2019). There is a wide range of supervised algorithms that have been proposed for this purpose, including SVM, k-NN, ANN, Bayesian networks (BN), decision trees (DT), and random forests (RF). Each of these

models is suited for different classification scenarios and has different strengths and weaknesses, as highlighted by Alimi et al. (2021). Intrusion detection in ICS is a binary classification problem, which makes SVM well-suited for this purpose. Other advantages of using SVM for this purpose include robustness to high dimensional data and good generalisation. This algorithm performs exceptionally well even in scenarios where the data is non-linearly separable in feature space. ANN has similar benefits of robustness and generalisation. Many authors have proposed k-NN as the algorithm of choice for SCADA systems because of its low computation time and its ability to be easily interpreted (Alimi et al., 2020). Hence, in this study, we employed supervised ML leveraging distance-based and shallow neural networks algorithms utilising SVM, k-NN and ANN to develop the models for anomaly detection and classification.

Other supervised methods such as DT, RF and BN were eliminated because of specific disadvantages reported about them. DTs require the creation of trees and tend to struggle if the data is high-dimensional. Moreover, overfitting is a well-known problem in DTs, and proper pruning is necessary for its management. RF extends DT in a way that requires the creation of many trees. As a result, this algorithm is computationally intensive. Finally, even though some authors have proposed it, BN also underperforms with high-dimensional data. Additionally, SCADA/ICS systems contain multiple nodes, and BN implementations are expected to involve high computational requirements in this scenario (Borujeni et al., 2021). Moreover, our study proposes a variable selection method that uses the logistic regression model (LRM) for feature selection to improve system performance because of its interpretability in the form of probabilities instead of binary outcomes.

3 Incident classification model for water treatment plants

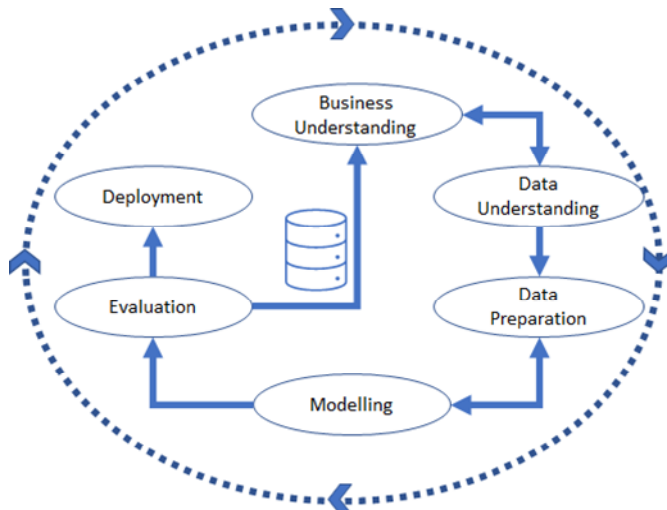
ICS provide a range of critical services, which if disrupted, could result in substantial real-world damage. Furthermore, data-driven methods to protect ICS require different conditions to enable testing of individual events to determine incidents. The IWaP model aims to fetch systems events from the physical control layer of water treatment plants and perform process-based detection analyses to identify indicators of anomalies (IoA). Not all anomalies are incidents with an impact on the ability to provide all services. IWaP leverages the National Institute of Standards and Technology (NIST) Special Publication (SP) 800-61 Rev 2 (NIST, n.d.) and adapts to include the functional impact categorisation where the potential IoC is mapped to eight levels of functional impact where Level 1 has *no impact* and Level 8 has a high impact of a *denial of a critical process* to the water treatment plant. In this context, an IWaP event is any change, error or recorded interruption. However, since not all events qualify as an imminent threat of violation of the system's security principles and resilience, IWaP considers events classified as Level 3 (minor impact to a non-critical process) or above as incidents and logs them accordingly for further analysis by the incident response team. The IWaP model is shown in Figure 1.

Figure 1 IWaP model (see online version for colours)

Notes: IWaP includes an automated process to check indicators of anomalies (CIAC) which are then tested against the bespoke functional impact levels on the water treatment plant. The impact that is recorded as level 3 and above will be logged and classified.

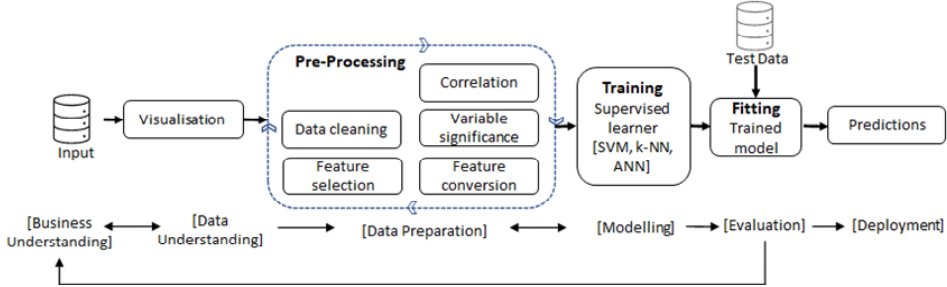
3.1 Data mining for supervised learners training

IWaP adapts the cross-industry standard process for data mining, generally known as CRISP-DM (CRISP-DM) framework (illustrated in Figure 2) to guide the data mining process. CRISP-DM was created as an open standard to provide a clear model for analysis. CRISP-DM is the de-facto standard and an industry-independent process model applicable to data mining projects (Schröder et al., 2021).

Figure 2 Schematic diagram representation of the CRISP-DM framework (see online version for colours)

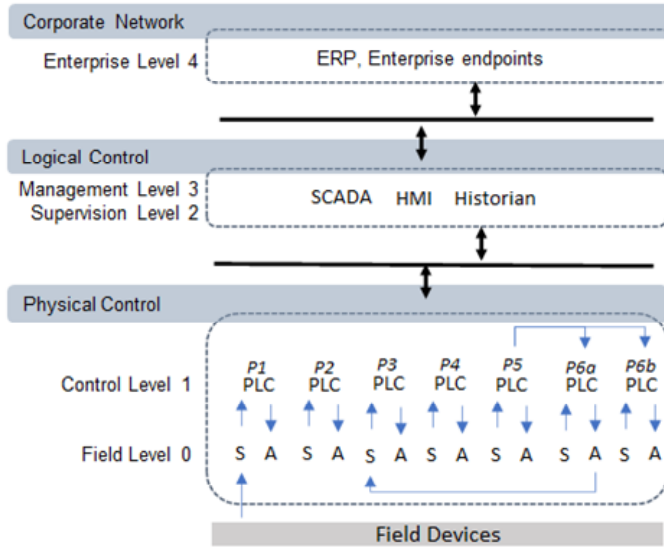
The IoC process is deployed using supervised ML algorithms, to facilitate training on the current and future attack samples Figure 3 demonstrates a training procedure mapped to the CRISP-DM framework phases. Pre-processing includes *data cleaning* and other fundamental phases such as *feature selection*. Feature engineering approaches such as feature selection, weighting, extraction and reduction have an impact on the performance of the predictive classifiers and the predictive performance optimisation. They can also reduce the computational complexity of imbalanced datasets used for training purposes (Alimi et al., 2021).

Figure 3 IWaP datamining and supervised learning procedure (see online version for colours)



3.2 Process-based anomaly detection

Figure 4 illustrates simplified ICS architecture components consisting of devices and subsystems categorised into three layers. IWaP is a physical control process-based anomaly detection model. The physical and control components of a water treatment system comprise sensors (S) and actuators (A) connected to the plant equipment that provide readings from and modify the physical plant field instruments' processes (P1, P2 ... Pn) via received and transmitted PLC signals. For example, water level meters, pressure indicators, pH monitors, flow meters and conductivity are some of these sensors while actuators comprise several motorised valves and pumps. The PLC link the plant instruments with SCADA, HMI, Historian and other engineering workstations at the logical control layer leveraging communication networks. The corporate network forms the top layer of industrial automation comprising information technologies that manage the system control and automation. This study does not focus on network traffic nor zone segmentation between the ICS layers, and they are not shown in Figure 4. The first process (P1) typically represents the intake and storage of raw water in the water treatment systems undergoing various phases of treatment to reach the last process (6b) outputting filtered water. The remaining processes are logically assigned in order such as the second process (P2), where water flows through a chemical pre-treatment process involving water quality assessment and chemical dosing of low-quality water. The plant in each of the process phases contains sensors and actuators that report and modify the state of the physical process.

Figure 4 Simplified architecture components of ICS (see online version for colours)

3.3 Data preparation for modelling

The first step towards developing ML models for anomaly detection in ICS is to determine the statistical significance of variables and their impact on anomaly detection so that insignificant variables can be screened out. The *data preparation* phase of the model determines which variables to include. For example, the 'timestamp' variable representing the sensors' and actuators' data collection time and the date are not necessary for the model-building to classify anomalies. Data cleaning follows, where missing values and possible records with outliers are typically removed (Elmrabit et al., 2020). The LRM is an example of an appropriate feature selection method to screen out insignificant variables but also to avoid overfitting. The Logit model takes the format represented in (1),

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 FIT_101 + \beta_2 LIT_101 + \dots + \beta_{51} P_603 + \mu_i, \quad (1)$$

where P_i is the probability of having an attacked state, β_i are the model slope coefficients, the independent variables in the data and $\ln$ is the natural logarithm. In (1), $\ln(P_i/1 - P_i)$ is the Logit transformation defined as the log of the odds given in equation (2),

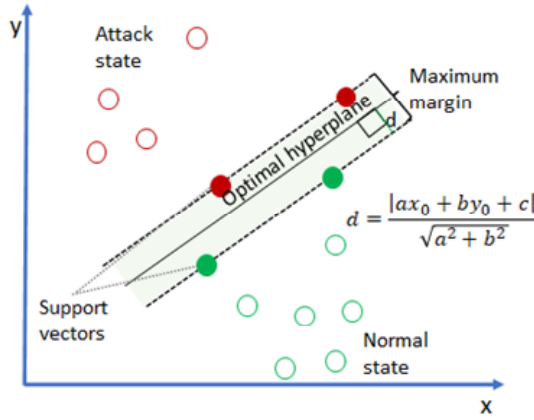
$$odds_i = \frac{P_i}{1-P_i} \quad (2)$$

After screening the variables for significance, ML models are developed. Our model is based on SVM, ANN and k-NN algorithms. Each of the classifiers was applied separately and the attained results were compared.

3.4 Model building with SVM algorithm

Proposed by Vapnik (1998), the SVM algorithm is based on a concept of decision boundaries that are commonly used in SCADA binary classification problems. A robust classification method that is suitable to deal with high dimensional datasets with good generalisation ability (Alimi et al., 2021; Xin et al., 2018). In this study, SVM acts as a classifier to determine whether a combination of sensor and actuator readings denote an attacked mode or a normal one. The SVM operates by choosing the optimal hyperplane, see Figure 5. The hyperplane is then placed at the centre of the support vectors (the data points) where the distance is the maximum from the closest points of either class giving the least test errors.

Figure 5 SVM structure in two dimensions and distance measure calculation (see online version for colours)



We map a new system behavioural mode on the hyperplane with complex feature space x , represented in (3):

$$\Phi(x) \in R^M, \text{ where } x \in R^D. \quad (3)$$

The transformed feature space for each input feature is mapped to a transformed basis vector $\Phi(x)$, as defined in equation (4):

$$\Phi(x): R^D \rightarrow R^M. \quad (4)$$

After the transformation, we separate these points to get a decision boundary that divides the points into their respective classes. The equation of the main separator line is called a hyperplane equation which is given as:

$$H: w^T(x) + b = 0 \quad (5)$$

where b is the intercept and bias term of the hyperplane equation in D dimensional space. SVM segregates the data points to minimise the misclassification errors by computing the distance between data points and the separating line as shown in Figure 5. The distance of a hyperplane equation $w^T\Phi(x) + b = 0$ from a given point vector $\Phi(x_0)$ is given as:

$$d_H(\Phi(x_0)) = \frac{|w^T(\Phi(x_0)) + b|}{\|w\|_2} \quad (6)$$

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

$$\|w\|_2 = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2} \quad (7)$$

where $n = 1, 2, \dots, 51$.

We utilised an 80:20 dataset ratio to train and test the model. Data manipulation was performed where log transformations were done so that the data points from both sensor and actuator readings fall within the same range.

3.5 Model building using ANN algorithm

The second ML model was constructed using the ANN algorithm to classify the system's behavioural mode as normal or attacked, see Figure 6. The ANN algorithm is robust to high-dimensional data with good generalisation ability (Apruzzese et al., 2018; Alimi et al., 2021). It imitates biological neurons by using interconnected basic processing units and nodes (Abiodun et al., 2018). The network neurons fire above a certain threshold; hence a function that behaves similarly was employed. We utilised the step function in Python defined by (8),

$$H(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

The input to the neuron must exceed the threshold to trigger firing due to various neural connections. However, the neurons must be merged uniquely. A weighted sum of the incoming input was used to permit firing to achieve this. Let x denote the entering neuron's output vector and $w \cdot x$ denotes the dot product, where w is the weight vector. To improve the neuron's capacity, a threshold value was set arbitrarily that was calculated by multiplying the weighted total of the inputs by a scalar. This calculation is performed because the new step function $H(x + (-b))$ at $x = b$ equals 0, which is the step function's threshold. Adding a scalar of $-b$ forced the neuron's activation threshold to be set to b . The bias b is the amount by which the step function is shifted away from the natural threshold at $x = 0$. Python was used to create the neuron output model for the following two processes. The activation function given in (9) was applied to the first-stage outcome

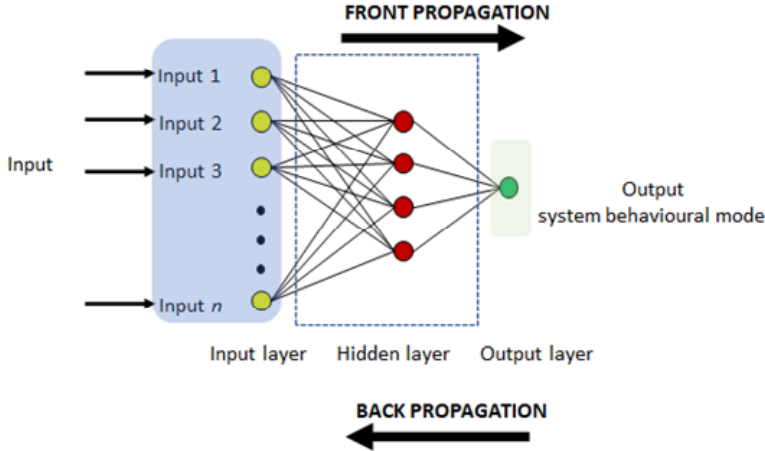
$$H(\vec{w} \cdot \vec{x} + b) = \begin{cases} 1, & \vec{w} \cdot \vec{x} + b \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

The final step in the ANN model-building process is a choice of an appropriate activation function. This step is essential because the neural network consists of linearly connected neurons, resulting in a non-linear function. This fraction cannot be modelled by the step function discussed. Hence, we utilised a more complex continuous sigmoid function (Zhou, 2012). The sigmoid function given in (10) was computed in R Studio, with an output range of 0 to 1.

$$\partial(x) = \frac{1}{1 + e^{-x}}. \quad (10)$$

The ANN's multi-layer perceptron (MLP) architecture was utilised with an 80:20 dataset ratio to train and test the model. Data manipulation was performed where log transformations were done so that the data points from both sensor and actuator readings fall within the same range.

Figure 6 ANN structure algorithm (see online version for colours)



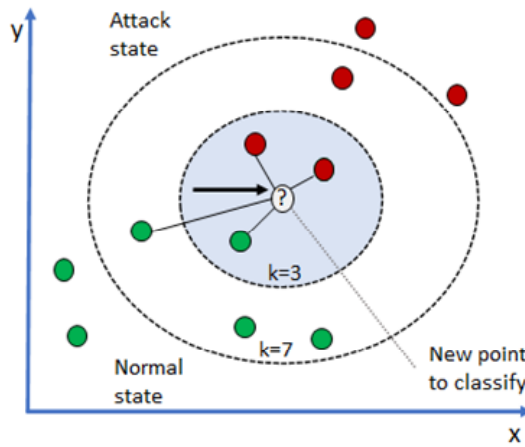
3.6 Model building using the *k*-NN algorithm

The third ML model was constructed using *k*-NN with an 80:20 dataset ratio to train and test the model. The *k*-NN algorithm is a universal function approximator under certain conditions with a simple underlying concept illustrated in Figure 7. It is known as a lazy learning algorithm, *k*-NN is a popular classification method to model anomaly detection in SCADA systems (Elmrabit et al., 2020; Xin et al., 2018). *K*-NN is a simple non-parametric classification algorithm based on a distance function that classifies data points and computes the class label based on the *k* nearest points where *k* is always a positive integer. Instead of approximating the target function $f(x) = y$ globally, during each prediction, the *k*-NN algorithm approximates the target function locally. It uses the datasets closest in proximity as it is easier to learn to approximate a function locally than globally. The *k*-NN model is a classification algorithm often described as a majority voting where *k* data points are referenced to a class with >50% data points nearer to it.

The distances are calculated using the computation in (11):

$$d(x^{[a]}, x^{[b]}) = \sqrt{\sum_{j=1}^m (x_j^{[a]} - x_j^{[b]})^2} \quad (11)$$

where d is the Euclidean Distance (Zhou, 2012; Kuncheva, 2014). The Euclidean norm is used to calculate the distances between the datasets and the neighbours, which is used to classify the data into either normal or attacked mode. The distances are stored in Python, sorted in ascending order and the first *k* entries are listed to assign a class to the new sample based on the majority of the classes present in the neighbour points.

Figure 7 k-NN structure algorithm in two dimensions (see online version for colours)

4 Case study: piloting the IWaP model

4.1 Dataset description

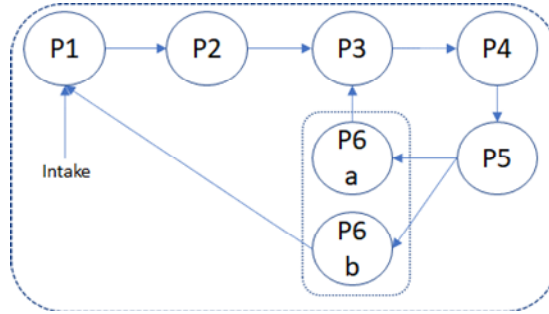
The experiment utilises the SWaT testbed built by iTrust, Centre for Research in Cyber Security, Singapore University of Technology and Design (Mathur and Tippenhauer, 2016). The testbed represents a purpose-built scaled-down version of a water treatment plant designed for cyber security research, with the ability to produce 5 gallons of drinking water per minute (Goh et al., 2017). The physical plant consists of industrial hardware to simulate the real behaviour of the operational components allowing data gathering in a controlled environment. The dataset contains fifty-one independent variables recorded every second (Goh et al., 2017). Data collection took eleven days. The plant generated data under normal operations during the first seven days and attacks were launched against the system during the remaining four days. The data was gathered under two behavioural modes: normal and attacks. Some attacks were launched consecutively whereas other scenarios permitted the system to reach a normal operating state before the launch of the next attack. The attacks lasted from a few minutes to an hour. The dataset to be used contains sensors and actuator data related to the physical properties of the plant. The monitoring sensors include level and flow indication transmitters, analyser indicator transmitters for conductivity, pH and oxidation reduction potential, pressure and differential pressure transmitters.

4.2 Process analysis

The *process analysis* phase in our model concluded six main processes (P1, P2 ... P6) corresponding to the physical and control components of the water treatment system. The system houses a total of fifty-one actuators and sensors. The first process (P1) involves the intake and storage of raw water to be treated in the water treatment systems. In the second process (P2), water flows through a pre-treatment process involving water quality

assessment and chemical dosing of low-quality water. Unwanted materials in the water are filtered and removed via the Ultra Filtration (UF) system during the third process (P3). In the fourth process (P4), the water is further dechlorinated using Ultraviolet lamps. Any inorganic impurities are reduced from the water using a Reverse Osmosis system during the fifth process (P5). Lastly, following the Reverse Osmosis the sixth process (P6b) entails water storage awaiting distribution in a water distribution system, which for the purpose of the testbed is recycled. However, the treated water can be transferred back using the UF backwash pumps (P6a) for re-processing (P3). The process summary is shown in Figure 8.

Figure 8 Testbed process summary (see online version for colours)



4.3 Experiment procedure and tools

The experiment procedure corresponds to the IWaP phases illustrated in Figure 3. During the *data preparation* phase, we determined which of the fifty-one dataset explanatory variables can be employed in the model. Then, a conversion of the categorical system response represented by the ‘attack state’, which indicates anomalous instances, was encoded to the numeric value 1. The ‘normal state’ which is indicative of the normal instances was encoded to 0. The numeric values help represent the states of the system as binary properties. In this study, we use the terms *attacked system mode* and *attacked state* to denote anomalous behaviour instances in the dataset. Therefore, enabling mathematical computations relating to performing ML to occur. In total, 449,919 records composed of fifty-one columns representing variables, independent variables and one column representing the dependent variable, which is also referred to as the system behavioural mode, were used in the *data preparation* phase.

Data pre-processing and feature extraction from the dataset were performed using R-studio. First, to predict the properties of unknown data, we defined the features that will be used to classify the corresponding system state, represented as the x values, and behavioural mode, represented as the y values. Then, given the reduced dataset deduced from the LRM phase, we carried out feature selection which involves grouping the data into 39 features represented by variable x and its corresponding system behaviour mode represented as y. The data collection, pre-processing and feature extraction delineated as D1_i, D2_i and D3_i in Figure 1, are replicated for the three employed algorithms, SVM, k-NN and ANN applied in the *modelling* phase. Once the variables have been defined, the dataset uses an 80:20 dataset split ratio to train and test the model. The pre-processing function named Standard Scalar is called to standardise the data features by removing

means and scaling to unit variance. As supervised learning is employed, the implemented model is required to learn the behaviours of normal and attack data. Then, a supervised ML model is constructed implementing the SVM, k-NN and ANN learning techniques and trained using the training data subset.

Software, platforms and libraries needed for this experiment include Anaconda Individual Edition; Python and Jupyter Integrated Development Environment (IDE); the Scikit-learn library in Python to implement the SVM, k-NN and ANN algorithms; and Pandas and NumPy to visualise and analyse the dataset.

5 Results

5.1 Determination of significant variables for anomaly detection in ICSs

Before training the model, LRM was used to screen the variables with a statistically significant impact on anomaly detection. Determining the significant variables is crucial in the classification process to determine the variables the system's behavioural mode is dependent on. Independent variables with less than two modes or levels violate regression analysis conditions, which results in a contrasts error reflecting that contrasts can be applied only to factors with two or more levels. A subset of categorical variables representing actuators had only one state or mode. Hence, several variables failed to meet the conditions for the logistic regression variable screening. Therefore, seven variables were excluded: P401, P202, P301, P404, P601, P603 and P502. Further screening of the remaining 44 independent variables for the statistical significance of the exogenous variables on the outcome variable resulted in excluding further five variables. The remaining 39 variables are statistically significant with $p < 0.1$ in determining the system's behavioural mode. The ORP Meter (AIT402) has the least significance ($p = 0.08859$), meaning that the variable ORP Meter reading has the least effect on detecting the system anomalies. This variable is associated with a 4.419e-03 estimate. Thus, a unit change in the readings of the ORP Meter is associated with a 0.004419 change in the probability of having a normal system mode. At the same time, several variables have very small p-values ($p < 2e-16$) such as the RO reject flow meter sensor (FIT503) that are highly likely to have a significant effect on determining the system behavioural mode based on the highest estimate of 1.770e+02. In other words, a unit change in the readings from this sensor is associated with 177 points shift in the log odds of the normal system behavioural mode. Therefore, 39 statistically significant variables shown in Figure 9 were used in the *modelling phase*.

5.2 Classification-based anomaly detection leveraging SVM, k-NN and ANN

First, a model employed the SVM algorithm for anomaly detection. The SVM achieved an accuracy of 98.72% and an F1-score of 0.97, as shown in Table 1. These results correspond to a precision score of 0.99 and a recall score of 0.95. As shown in Table 2, SVM correctly classified a total of 78,973 modes as normal and 9,868 modes as attacked. However, 125 attacked instances were misclassified as normal and 1,026 normal system modes were misclassified as attacked. The SVM model performed well with a slight misclassification error rate of 1.28% of both Type I and Type II errors, which is the model's shortfall in classifying a system behavioural mode as either normal or attacked.

Figure 9 Statistically significant variables included in the modelling phase (see online version for colours)

Deviance Residuals:					
	Min	1Q	Median	3Q	Max
	-4.9128	-0.2715	-0.1893	-0.0915	5.8602
Coefficients:					
	Estimate	Std. Error	Z value	Pr(> z)	
(Intercept)	-6.996e+03	4.009e+02	-17.449	< 2e-16 ***	
FIT101	3.171e-01	2.807e-02	11.300	< 2e-16 ***	AIT401 4.734e+01 2.635e+00 17.963 < 2e-16 ***
LIT101	2.351e-03	1.128e-04	20.836	< 2e-16 ***	AIT402 4.419e-03 2.595e-03 1.703 0.08859 .
MV101	5.185e-01	6.799e-02	7.626	2.42e-14 ***	FIT401 -5.325e+00 1.179e-01 -45.183 < 2e-16 ***
P101	-3.799e+00	1.248e-01	-30.444	< 2e-16 ***	LIT401 -7.088e-04 1.217e-04 -5.823 5.79e-09 ***
P102	7.405e+00	1.207e-01	61.375	< 2e-16 ***	P402 8.758e+00 1.342e+00 6.528 6.67e-11 ***
AIT201	2.012e-03	1.795e-03	1.121	0.26244	P403 -1.886e+01 4.520e+01 -0.417 0.67647
AIT202	4.795e+00	6.627e-01	7.236	4.63e-13 ***	UV401 -6.606e+00 2.033e+00 -3.249 0.00116 **
AIT203	-4.095e-02	2.227e-03	-18.384	< 2e-16 ***	AIT501 -6.179e+00 5.082e-01 -12.158 < 2e-16 ***
FIT201	1.495e+00	8.362e-02	17.874	< 2e-16 ***	AIT502 4.328e-02 2.808e-03 15.410 < 2e-16 ***
MV201	3.620e+00	7.570e-02	47.825	< 2e-16 ***	AIT503 -8.946e-02 2.097e-03 -42.660 < 2e-16 ***
P201	-3.373e+00	9.440e-02	-35.733	< 2e-16 ***	AIT504 2.628e-02 9.290e-04 28.284 < 2e-16 ***
P203	-1.597e+01	1.593e+00	-10.023	< 2e-16 ***	FIT501 -2.567e+01 1.210e+00 -21.208 < 2e-16 ***
P204	5.693e-01	5.354e+02	0.001	0.99915	FIT502 -1.482e+00 5.947e-01 -2.492 0.01270 *
P205	1.317e+01	1.625e+00	8.107	5.20e-16 ***	FIT503 1.770e+02 6.695e+00 26.430 < 2e-16 ***
P206	1.348e+01	5.402e+02	0.025	0.98009	FIT504 6.564e+00 6.956e+00 0.944 0.34538
DPT301	2.291e-01	5.991e-03	38.231	< 2e-16 ***	PE01 6.291e+00 2.085e+00 3.017 0.00255 ***
FIT301	-1.394e+00	5.638e-02	-24.720	< 2e-16 ***	PIT501 -2.525e+00 6.656e-02 -37.934 < 2e-16 ***
LIT301	-4.986e-03	1.144e-04	-43.568	< 2e-16 ***	PIT502 6.980e-01 4.740e-02 14.725 < 2e-16 ***
MV301	2.758e+00	8.136e-02	33.896	< 2e-16 ***	PIT503 2.795e+00 7.746e-02 36.085 < 2e-16 ***
MV302	1.022e+00	7.088e-02	14.423	< 2e-16 ***	FIT601 2.298e-01 9.828e-02 2.338 0.01939 *
MV303	-1.956e+00	4.975e-02	-39.318	< 2e-16 ***	PE02 -1.583e+00 1.737e-01 -9.113 < 2e-16 ***
MV304	3.144e+00	4.925e-02	63.852	< 2e-16 ***	---
P302	-2.587e+00	9.055e-02	-28.573	< 2e-16 ***	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ''

The performance of SVM models developed in previous research on varying datasets is illustrated in Table 3. Our results are competitive with the research findings of reviewed literature. Moreover, they are similar to Junejo and Goh (2016) who also used the SWaT dataset and achieved an accuracy of 98.71%, a precision score of 0.980, and a recall score of 0.987. Likewise, the developed model outperforms that of Shalyga et al. (2018) which achieved an F1-score of 0.796 and (Bernieri et al., 2019) whose F1-score was 0.8053.

Table 1 Performance of the IWaP model using SVM, k-NN and ANN for anomaly detection (N = 89,984)

	Precision	Recall	F1-score	Total
<i>SVM</i>				
0	0.99	1.00	0.99	79,098
1	0.99	0.91	0.94	10,886
Accuracy			0.99	89,984
Macro avg	0.99	0.95	0.97	89,984
Weighted Avg	0.99	0.99	0.99	89,984
<i>k-NN</i>				
0	1.00	1.00	1.00	78,977
1	0.99	0.98	0.99	11,007
Accuracy			1.00	89,984
Macro avg	0.99	0.99	0.99	89,984
Weighted Avg	1.00	1.00	1.00	89,984
<i>ANN</i>				
0	0.99	1.00	1.00	79,141
1	0.99	0.96		10,843
Accuracy			0.99	89,984
Macro avg	0.99	0.98	0.98	89,984
Weighted Avg	0.99	0.99	0.99	89,984

Table 2 Confusion matrix attributes of the IWaP model using SVM, k-NN and ANN for anomaly detection (N = 89,984)

<i>Predicted</i>	<i>Actual</i>		
	<i>Normal</i>	<i>Attacked</i>	<i>Total</i>
<i>SVM</i>			
Normal	78,973	125	79,098
Attacked	1,026	9,868	10,886
Total	79,999	9,985	89,984
<i>k-NN</i>			
Normal	78,878	99	78,977
Attacked	187	10,820	11,007
Total	79,065	10,919	89,984
<i>ANN</i>			
Normal	78,998	143	79,141
Attacked	263	10,580	10,843
Total	79,261	10,723	89,984

Second, the model employed the k-NN algorithm for anomaly detection. The 80:20 training and testing split ratio was maintained to evaluate the model's performance. The optimal number of neighbours producing the best result is $k = 21$. As shown in Table 2, the k-NN model correctly classified 78,878 modes as normal and 10,820 modes as attacked. There were 99 attack mode records were classified as normal mode, while 187 records were incorrectly classified as attacked mode. Overall, the k-NN model achieved a classification accuracy of 99.68% with 0.32% misclassification of Type I and Type II errors combined. Thus, by adopting the k-NN model, there will be a <1% chance that some of the classifications will be incorrect. The k-NN model achieved a 0.99 score on precision, recall and F1-score, see Table 1.

The performance of k-NN models developed in previous research on varying datasets is illustrated in Table 3. Our results outperformed that of Bernieri et al. (2019) which achieved an F1-score of 0.839. Whereas, our model underperformed in comparison to Phillips et al. (2020) which obtained an F1-score of 0.99664 and Sebestyen et al. (2021) which obtained an F1-score of 0.9947.

Third, the model used the ANN algorithm for anomaly detection employing MLP architecture to classify the system's behavioural mode. The MLP-ANN is comprised of two hidden layers. The first layer had 10 neurons and the second had 20 neurons. The output layer comprises 39 neurons corresponding to the number of independent variables used in classifying the system's behavioural mode as either normal or attacked. The 80:20 training and testing split ratio was retained to evaluate the model. The output of the confusion matrix for this model is shown in Table 2. The ANN model correctly classified 78,998 modes as normal and 10,580 modes as attacked. However, it misclassified 263 normal instances as being attacked and 143 attack instances as being normal. The ANN model detected the system under attack with 99.35% accuracy meaning a misclassification error rate of 0.65%. Results in Table 1 show that the ANN model is robust as evidenced by a precision score of 0.99 and recall and F1-score of 0.98. The

output of the confusion matrix for this model is shown in Table 2 indicating that the rate of false-positive and false-negative classifications remained minimal.

Table 3 Summary of classification results of past studies and this study leveraging SVM, k-NN, ANN, NN, MLP algorithms

<i>Reference</i>	<i>Dataset</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Junejo and Goh (2016)	SWaT	NN: 0.9824 SVM: 0.9871	NN: 0.984 SVM: 0.98	NN: 0.982 SVM: 0.987	
Phillips et al. (2020)	Morris, Vaughn and Dandass's gas pipeline SCADA system dataset	SVM: 0.99994 k-NN: 0.99982			SVM: 0.99991 k-NN: 0.99664
Shalyga et al. (2018)	SWaT		MLP: 0.967 SVM: 0.925	MLP: 0.696 SVM: 0.699	MLP: 0.812 SVM: 0.796
Ramotsoela et al. (2020)	Battle of the Attack Detection algorithms (BATADAL) dataset	ANN: 0.7838			ANN: 0.5219
Sebestyen et al. (2021)	GECCO drinking water quality 2017 dataset				MLP: 0.9949 k-NN: 0.9947 SVM: 0.9891 NN> 0.5786
Bernieri et al. (2019)	SWaT	SVM: 0.9603 k-NN: 0.9658			SVM: 0.8053 k-NN: 0.839
This study	SWaT	SVM: 0.9872 k-NN: 0.9968 ANN: 0.9935	SVM: 0.99 k-NN: 0.99 ANN: 0.99	SVM: 0.95 k-NN: 0.99 ANN: 0.98	SVM: 0.97 k-NN: 0.99 ANN: 0.98

The performance of ANN models developed in previous research on varying datasets is illustrated in Table 3. The proposed ANN model outperformed several studies that used the ANN-based models on a range of ICS datasets. For example, other studies that leveraged the SWaT dataset include Junejo and Goh (2016) who developed an ANN model achieving a precision score of 0.984 and a recall score of 0.982 while Shalyga et al. (2018) developed an MLP ANN model achieving a precision score of 0.967, recall score of 0.696 and an F1-score of 0.812. Based on the Battle of the Attack Detection Algorithms (BATADAL) dataset, Ramotsoela et al. (2020) constructed the ANN model and achieved an F1-score of 0.5219 while Sebestyen et al. (2021) developed an MLP ANN model using the Genetic and Evolutionary Computation Conference (GECCO) 2017 drinking water quality dataset achieving an F1-score of 0.9949.

6 Discussion

Direct comparison of an algorithm is challenging due to several factors including the choice of datasets, the experimentation environment and the experiment procedure (Apruzzese et al., 2018). The experiment procedure in this study performs an unbiased

comparison of the selected ML models' effectiveness. The experimentation in this study consistently follows the same data-mining framework, see Error! Reference source not found. and Figure 3, utilises the same dataset maintaining the train-test data subset split, considers the same features and systematically applies the data pre-processing to meaningfully compare the performance of the selected algorithms.

ICS datasets are typically imbalanced where the majority class belongs to the normal behaviour and the goal of the model is to correctly classify the minority class (Branco et al., 2016). Against this backdrop, disruptive experimentation to develop new defensive methods, anomaly detection, classification and optimisation to defend against evolving ICS attack vectors is not achievable in operational ICS (Junejo and Goh, 2016; Alimi et al., 2021). Attacks against real-world ICS are unsystematic. Hence, the lack of publicly available suitable real-life datasets requires approaches where real-world activities can be simulated in controlled environments (Pospisil et al., 2021).

Existing literature supported the use of supervised learning techniques for anomaly detection in ICS because of higher detection rates and lower false alarms. Our study uses SVM, k-NN and ANN algorithms for modelling because they are deemed appropriate for high dimensional data and do not suffer from limitations such as high computational complexity. It is evident from the results that k-NN performed better than its counterparts, with consistent precision, recall and an F1-score of 0.99. Furthermore, the accuracy for k-NN was found to be 99.68% compared with 98.72% and 99.35% for SVM and ANN, respectively. Besides this, for SVM, 1.28% of the system behavioural modes were misclassified. On the other hand, the misclassification errors were lower than 1% for both k-NN and ANN. A summary of the results and their comparison with state-of-the-art works is presented in Table 3.

The results of this study confirm the better performance of k-NN with respect to SVM. These results are consistent with previous studies (Bernieri et al., 2019), which used the SWaT dataset and Almalawi et al. (2014), which proposed a system for detecting integrity attacks on a water distribution system. Moreover, the findings are congruent with Sebestyen et al. (2021) who used GECCO drinking water quality 2017 dataset to detect anomalies. This evaluation included MLP, k-NN, SVM and a general NN concluded that the best model for detecting attacks on ICSs is the MLP model followed by the k-NN and SVM. The inferior performance of SVM was also reported by Shalyga et al. (2018) who compared the MLP-ANN and SVM using the SWaT dataset.

In contrast to these reports, Phillips et al. (2020) who used Morris, Vaughn and Dandass' gas pipeline SCADA system dataset, compared the performance of SVM and k-NN. The authors concluded that the SVM was marginally superior with an F1-score of 0.99991 as compared to 0.99664 for k-NN. Notably, the findings of our study are partially in agreement with those of Robles-Durazno et al. (2020) and Arora et al. (2021) who examined the viability of ML algorithms in detecting new security threats in various cyber-physical systems and reported the strengths of k-NN in comparison to SVM in classifying attacks for these systems.

Although SVM performed slightly lower than k-NN and ANN, our findings show that given an optimal set of features from physical plant sensor readings the classifiers' performance is effective for anomaly detection in ICS' water treatment plants with the error range consistently falling under 2%. However, k-NN and ANN can achieve an accuracy as high as 99% with the incorporation of 39 statistically significant variables. Therefore, as the IWaP model leverages multiple classifiers combined with an optimal feature set mitigating the risk of using a single sub-optimal classifier. The promising

results of this study make it relevant in the context of addressing the rising number of attacks on modern ICSs interconnected with information and communication technologies (ICT) and agree with the recommendations by Kaspersky Labs ICS CERT and Claroty (Claroty Team 82, 2021; Kaspersky, 2021; PR Newswire, 2021). Thus, this research can play a crucial role in directing future research in this domain.

7 Conclusions

The recent past has witnessed a rapidly increasing number of attacks on ICS. The proliferation of connected devices in ICS coupled with the commoditisation of intrusion cyber capabilities and services increase the threats of successful cyber attacks against operational technologies. Data produced from the physical sensors can be innovatively used leveraging ML techniques as part of defence-in-depth in ICS. The performance results of the SVM, k-NN and ANN ML approaches utilised by the IWaP model were analysed. An accuracy of 99.72%, 98.72% and 99.35% was achieved respectively, demonstrating the feasibility of this approach for anomaly detection in ICS in water treatment systems. Furthermore, while experimental work was conducted for water treatment systems to validate the indicators of anomaly (IoA) capability within our model using the SWaT dataset (Goh et al., 2017) with 26 actuator states and 25 sensor readings, results suggest the potential to generalise the capability to other ICS. However, this will require a process to automate the identification of sensor data variables to provide the best detection rates. As such, this study determined the statistical significance of independent variables using logistic regression, and the process concluded 39 variables of statistical significance in ascertaining the system behavioural mode. Additionally, the proposed ICS anomaly classification approach using k-NN outperformed existing state-of-the-art approaches. We attribute that to the inclusion of fit-for-purpose pre-processing and variable screening phases. The results of this study attest that the ML approaches comprising the IWaP model can effectively classify anomalies within a 2% error range. However, this system could still be susceptible to zero-day attacks because the SWaT dataset was deliberately injected to induce anomalies. A typical limitation to acknowledge considering the challenge faced by the scientific community is to construct datasets with real-life advanced persistent threats (APT) data. That being said, conclusions from this study can be utilised to address emerging APT threats in two ways. Firstly, the proposed technique has limited capabilities with regard to handling zero-day attacks. However, to enable effective handling of attacks with unknown behaviour, the use of online ML concepts where algorithms learn pattern behaviours one instance at a time shall be explored in future work. Secondly, by profiling the characteristics of a normal system operation as part of the overall intrusion detection capabilities. Therefore, this research effort can be further extended to develop a full-fledged interactive application for anomaly detection in ICS.

Acknowledgements

The authors want to thank iTrust, Centre for Research in Cyber Security, Singapore University of Technology and Design for presenting the SWaT dataset.

References

- Abdelaty, M.F., Corin, R.D. and Siracusa, D. (2021) 'DAICS: a deep learning solution for anomaly detection in industrial control systems', *IEEE Transactions on Emerging Topics in Computing*, p.1, <https://doi.org/10.1109/TETC.2021.3073017>.
- Abdulhammed, R., Musafer, H., Alessa, A., Faezipour, M. and Abuzneid, A. (2019) 'Features dimensionality reduction approaches for machine learning based network intrusion detection', *Electronics*, Vol. 8, No. 3, p.322, <https://doi.org/10.3390/electronics8030322>.
- Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. and Arshad, H. (2018) 'State-of-the-art in artificial neural network applications: a survey', *Heliyon*, Vol. 4, No. 11, p.e00938, <https://doi.org/10.1016/j.heliyon.2018.e00938>.
- Abokifa, A.A., Haddad, K., Lo, C. and Biswas, P. (2019) 'Real-time identification of cyber-physical attacks on water distribution systems via machine learning-based anomaly detection techniques', *Journal of Water Resources Planning and Management*, Vol. 145, No. 1, p.04018089, [https://doi.org/doi:10.1061/\(ASCE\)WR.1943-5452.0001023](https://doi.org/doi:10.1061/(ASCE)WR.1943-5452.0001023).
- Ahmadi-Assalemi, G., Al-Khateeb, H., Epiphaniou, G. and Aggoun, A. (2022) 'Super learner ensemble for anomaly detection and cyber-risk quantification in industrial control systems', *IEEE Internet of Things Journal*, p.1, <https://doi.org/10.1109/JIOT.2022.3144127>.
- Ahmadi-Assalemi, G., Al-Khateeb, H., Epiphaniou, G. and Maple, C. (2020) 'Cyber resilience and incident response in smart cities: a systematic literature review', *MDPI Smart Cities*, August, Vol. 3, No. 3, pp.894–927, <https://doi.org/10.3390/smartcities3030046>.
- Ahmadi-Assalemi, G., Al-Khateeb, H.M., Epiphaniou, G., Cosson, J., Jahankhani, H. and Pillai, P. (2019) 'Federated blockchain-based tracking and liability attribution framework for employees and cyber-physical objects in a smart workplace', *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*, London, UK, pp.1–9, 16–18 January, <https://doi.org/10.1109/ICGS3.2019.8688297>.
- Alimi, O.A., Ouahada, K., Abu-Mahfouz, A.M. and Alimi, K.O.A. (2020) 'Empirical comparison of machine learning algorithms for mitigating power systems intrusion attacks', *2020 International Symposium on Networks, Computers and Communications (ISNCC)*, 20–22 October, pp.1–5, <https://doi.org/10.1109/ISNCC49221.2020.9297340>.
- Alimi, O.A., Ouahada, K., Abu-Mahfouz, A.M., Rimer, S. and Alimi, K.O.A. (2021) 'A review of research works on supervised learning algorithms for SCADA intrusion detection and classification', *Sustainability*, Vol. 13, No. 17, p.9597, <https://doi.org/10.3390/su13179597>.
- Almalawi, A., Yu, X., Tari, Z., Fahad, A. and Khalil, I. (2014) 'An unsupervised anomaly-based detection approach for integrity attacks on SCADA systems', *Computers & Security*, Vol. 46, pp.94–110, <https://doi.org/10.1016/j.cose.2014.07.005>.
- Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A. and Marchetti, M. (2018) 'On the effectiveness of machine and deep learning for cyber security', *2018 10th International Conference on Cyber Conflict (CyCon)*, IEEE, Tallinn, Estonia, 29 May to 1 June, pp.371–390, <https://doi.org/10.23919/CYCON.2018.8405026>.
- Arora, P., Kaur, B. and Teixeira, M.A. (2021) '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, <https://doi.org/10.1007/s40031-021-00563-z>.
- Bakalos, N., Voulodimos, A., Doulamis, N., Doulamis, A., Ostfeld, A., Salomons, E., Caubet, J., Jimenez, V. and Li, P. (2019) '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, <https://doi.org/10.1109/MSP.2018.2885359>.
- Bernieri, G., Conti, M. and Turrin, F. (2019) 'Evaluation of machine learning algorithms for anomaly detection in industrial networks', *2019 IEEE International Symposium on Measurements & Networking (M&N)*, 8–10 July, pp.1–6, <https://doi.org/10.1109/IWMN.2019.8805036>.

- Borujeni, S.E., Nannapaneni, S., Nguyen, N.H., Behrman, E.C. and Steck, J.E. (2021) 'Quantum circuit representation of Bayesian networks', *Expert Systems with Applications*, Vol. 176, p.114768, <https://doi.org/10.1016/j.eswa.2021.114768>.
- Branco, P., Torgo, L. and Ribeiro, R.P. (2016) 'A survey of predictive modeling on imbalanced domains', *ACM Computing Surveys (CSUR)*, Vol. 49, No. 2, pp.1–50, <https://doi.org/10.1145/2907070>.
- Bristow, M. (2021) *A SANS 2021 Survey: OT/ICS Cybersecurity* [online] <https://sansorg.egnyte.com/dl/q52PNTByLB> (accessed 4 December 2021).
- Cárdenas, A. A., Amin, S., Lin, Z.-S., Huang, Y.-L., Huang, C.-Y. and Sastry, S. (2011) 'Attacks against process control systems: risk assessment, detection, and response', *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, Hong Kong China, March, pp.355–366, <https://doi.org/10.1145/1966913.1966959>.
- Claroty Team 82 (2021) *Claroty Biannual ICS Risk & Vulnerability Report: 1H 2021*, Claroty https://claroty.com/wp-content/uploads/2021/08/Claroty_Biannual_ICS_Risk_Vulnerability_Report_1H_2021.pdf (accessed 4 December 2021).
- Cook, A.A., Mısırlı, G. and Fan, Z. (2020) 'Anomaly detection for IoT time-series data: a survey', *IEEE Internet of Things Journal*, Vol. 7, No. 7, pp.6481–6494, <https://doi.org/10.1109/JIOT.2019.2958185>.
- Dalal, K.R. (2020) 'Analysing the role of supervised and unsupervised machine learning in IoT', *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2–4 July, pp.75–79, <https://doi.org/10.1109/ICESC48915.2020.9155761>.
- Demertzis, K., Iliadis, L. and Spartalis, S. (2017) 'A spiking one-class anomaly detection framework for cyber-security on industrial control systems', *International Conference on Engineering Applications of Neural Networks*, pp.122–134, Springer, https://doi.org/10.1007/978-3-319-65172-9_11.
- Dhirani, L.L., Armstrong, E. and Newe, T. (2021) 'Industrial IoT, cyber threats, and standards landscape: evaluation and roadmap', *Sensors*, Vol. 21, No. 11, p.3901 <https://doi.org/10.3390/s21113901>.
- Elmrabit, N., Zhou, F., Li, F. and Zhou, H. (2020) 'Evaluation of machine learning algorithms for anomaly detection', *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 15–19 June, pp.1–8, <https://doi.org/10.1109/CyberSecurity49315.2020.9138871>.
- Feng, C., Palleti, V.R., Mathur, A. and Chana, D. (2019) 'A systematic framework to generate invariants for anomaly detection in industrial control systems', *Network and Distributed Systems Security (NDSS) Symposium 2019*, San Diego, CA, USA, 24–27 February, <https://doi.org/10.14722/ndss.2019.23265>.
- Givehchi, O., Trsek, H. and Jasperneite, J. (2013) 'Cloud computing for industrial automation systems – a comprehensive overview', *2013 IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA)*, 10–13 September, pp.1–4, <https://doi.org/10.1109/ETFA.2013.6648080>.
- Goh, J., Adepu, S., Junejo, K.N. and Mathur, A. (2017) *A Dataset to Support Research in the Design of Secure Water Treatment Systems*, pp.88–99, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-71368-7_8.
- Gumaci, A., Hassan, M.M., Huda, S., Hassan, M.R., Camacho, D., Del Ser, J. and Fortino, G. (2020) 'A robust cyberattack detection approach using optimal features of SCADA power systems in smart grids', *Applied Soft Computing*, Vol. 96, p.106658, <https://doi.org/10.1016/j.asoc.2020.106658>.
- Hadžiosmanović, D., Sommer, R., Zambon, E. and Hartel, P.H. (2014) 'Through the eye of the PLC: semantic security monitoring for industrial processes', *Proceedings of the 30th Annual Computer Security Applications Conference*, New Orleans, Louisiana, USA, Association for Computing Machinery, pp.126–135, <https://doi.org/10.1145/2664243.2664277>.

- Hemsley, K.E. and Fisher, E. (2018) *History of Industrial Control System Cyber Incidents*, Idaho National Lab. (INL), Idaho Falls, ID, Idaho, USA, 15 October 2021 <https://doi.org/10.2172/1505628>.
- Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C.M. and Sun, J. (2017) 'Anomaly detection for a water treatment system using unsupervised machine learning', *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 18–21 November, pp.1058–1065, <https://doi.org/10.1109/ICDMW.2017.149>.
- Junejo, K.N. and Goh, J. (2016) 'Behaviour-based attack detection and classification in cyber physical systems using machine learning', *Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security*, Xi'an, China, Association for Computing Machinery, pp.34–43, <https://doi.org/10.1145/2899015.2899016>.
- Kaouk, M., Flaus, J., Potet, M. and Groz, R. (2019) 'A review of intrusion detection systems for industrial control systems', *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, 23–26 April, pp.1699–1704, <https://doi.org/10.1109/CoDIT.2019.8820602>.
- Kaspersky, I. (2017) *Threat Landscape for Industrial Automation Systems in The Second Half of 2016* [online] <https://ics-cert.kaspersky.com/reports/2017/03/28/threat-landscape-for-industrial-automation-systems-in-the-second-half-of-2016/> (accessed 1 December 2021).
- Kaspersky, I.C. (2021) *Threat Landscape for Industrial Automation Systems Statistics for H1 2021*, Kaspersky Lab [online] <https://ics-cert.kaspersky.com/media/Kaspersky-ICS-CERT-Threat-landscape-for-industrial-automation-systems-statistics-for-H1-2021-En.pdf> (accessed 4 December 2021).
- Khan, I.A., Pi, D., Khan, Z.U., Hussain, Y. and Nawaz, A. (2019) 'HML-IDS: a hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems', *IEEE Access*, Vol. 7, pp.89507–89521, 2019 <https://doi.org/10.1109/ACCESS.2019.2925838>.
- Kiss, I., Genge, B., Haller, P. and Sebestyén, G. (2014) 'Data clustering-based anomaly detection in industrial control systems', *2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 4–6 September, pp.275–281, <https://doi.org/10.1109/ICCP.2014.6937009>.
- Kosek, A.M. and Gehrke, O. (2016) 'Ensemble regression model-based anomaly detection for cyber-physical intrusion detection in smart grids', *2016 IEEE Electrical Power and Energy Conference (EPEC)*, 12–14 October, pp.1–7, <https://doi.org/10.1109/EPEC.2016.7771704>.
- Kuncheva, L.I. (2014) *Combining Pattern Classifiers: Methods and Algorithms*, p.384, Hoboken, New Jersey, USA, John Wiley & Sons, ISBN: 9781118914564, <https://doi.org/10.1002/9781118914564>.
- Lee, R.M., Assante, M.J. and Conway, T. (2014) 'German steel mill cyber attack', *Industrial Control Systems*, Vol. 30, p.62 [online] https://ics.sans.org/media/ICS-CPPE-case-Study-2-German-Steelworks_Facility.pdf (accessed 27 February 2021).
- Lier, B.v. (2017) 'The industrial internet of things and cyber security: an ecological and systemic perspective on security in digital industrial ecosystems', *2017 21st International Conference on System Theory, Control and Computing (ICSTCC)*, Sinaia, Romania, 19–21 October, pp.641–647, <https://doi.org/10.1109/ICSTCC.2017.8107108>.
- Lin, Q., Adep, S., Verwer, S. and Mathur, A. (2018) 'TABOR: a graphical model-based approach for anomaly detection in industrial control systems', *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, Incheon, Republic of Korea, ACM, pp.525–536, <https://doi.org/10.1145/3196494.3196546>.
- Liu, J., Yin, L., Hu, Y., Lv, S. and Sun, L. (2018) 'A novel intrusion detection algorithm for industrial control systems based on CNN and process state transition', *2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC)*, 17–19 November, pp.1–8, <https://doi.org/10.1109/PCCC.2018.8710993>.
- M.R., G.R., Ahmed, C.M. and Mathur, A. (2021) 'Machine learning for intrusion detection in industrial control systems: challenges and lessons from experimental evaluation', *Cybersecurity*, Vol. 4, No. 1, p.27, <https://doi.org/10.1186/s42400-021-00095-5>.

- Mathur, A.P. and Tippenhauer, N.O. (2016) 'SWaT: a water treatment testbed for research and training on ICS security', *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*, 11 April, pp.31–36, <https://doi.org/10.1109/CySWater.2016.7469060>.
- Mitchell, R. and Chen, I-R. (2014) 'A survey of intrusion detection techniques for cyber-physical systems', *ACM Comput. Surv.*, Vol. 46, No. 4, Article 55, <https://doi.org/10.1145/2542049>.
- Moustafa, N., Adi, E., Turnbull, B. and Hu, J.K. (2018) 'A new threat intelligence scheme for safeguarding Industry 4.0 systems', *IEEE Access*, Vol. 6, pp.33910–33924, <https://doi.org/10.1109/access.2018.2844794>.
- National Institute of Standards and Technology (NIST) (n.d.) *Computer Security Incident Handling Guide* [online] <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf> (accessed 15 March 2023).
- Phillips, B., Gamess, E. and Krishnaprasad, S. (2020) 'An evaluation of machine learning-based anomaly detection in a SCADA system using the modbus protocol', *Proceedings of the 2020 ACM Southeast Conference*, Tampa, FL, USA, Association for Computing Machinery, pp.188–196, <https://doi.org/10.1145/3374135.3385282>.
- Pospisil, O., Blazek, P., Kuchar, K., Fujdiak, R. and Misurec, J. (2021) 'Application perspective on cybersecurity testbed for industrial control systems', *Sensors*, Vol. 21, No. 23, p.8119, <https://doi.org/10.3390/s21238119>.
- PR Newswire (2021) *Industrial Control Systems (ICS) Security Market Worth \$23.5 Billion by 2026 – Exclusive Report by MarketsandMarkets™*, Bloomberg [online] <https://www.bloomberg.com/press-releases/2021-11-25/industrial-control-systems-ics-security-market-worth-23-5-billion-by-2026-exclusive-report-by-marketsandmarkets> (accessed 15 December 2021).
- Priyanga S, P., Krithivasan, K., P, S. and Sriram, V.S.S. (2020) 'Detection of cyberattacks in industrial control systems using enhanced principal component analysis and hypergraph-based convolution neural network (EPCA-HG-CNN)', *IEEE Transactions on Industry Applications*, Vol. 56, No. 4, pp.4394–4404, <https://doi.org/10.1109/TIA.2020.2977872>.
- Rakas, S.V.B., Stojanović, M.D. and Marković-Petrović, J.D. (2020) 'A review of research work on network-based SCADA intrusion detection systems', *IEEE Access*, Vol. 8, pp.93083–93108, <https://doi.org/10.1109/ACCESS.2020.2994961>.
- Ramotsoela, T.D., Hancke, G.P. and Abu-Mahfouz, A.M. (2020) 'Behavioural intrusion detection in water distribution systems using neural networks', *IEEE Access*, Vol. 8, pp.190403–190416, <https://doi.org/10.1109/ACCESS.2020.3032251>.
- Robles-Durazo, A., Moradpoor, N., McWhinnie, J. and Russell, G. (2018) 'A supervised energy monitoring-based machine learning approach for anomaly detection in a clean water supply system', *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 11–12 June, pp.1–8, <https://doi.org/10.1109/CyberSecPODS.2018.8560683>.
- Robles-Durazo, A., Moradpoor, N., McWhinnie, J. and Russell, G. (2020) 'Real-time anomaly intrusion detection for a clean water supply system, utilising machine learning with novel energy-based features', *2020 International Joint Conference on Neural Networks (IJCNN)*, 19–24 July, pp.1–8, <https://doi.org/10.1109/IJCNN48605.2020.9207462>.
- SANS ICS (2016) *Analysis of the Cyber Attack on the Ukrainian Power Grid*, Electricity Information Sharing and Analysis Center (E-ISAC) [online] https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf (accessed 15 October 2019).
- Schröer, C., Kruse, F. and Gómez, J.M. (2021) 'A systematic literature review on applying CRISP-DM process model', *Procedia Computer Science*, Vol. 181, pp.526–534, <https://doi.org/10.1016/j.procs.2021.01.199>.
- Schuster, F., Paul, A., Rietz, R. and Koenig, H. (2015) 'Potentials of using one-class SVM for detecting protocol-specific anomalies in industrial networks', *2015 IEEE Symposium Series on Computational Intelligence*, 7–10 December, pp.83–90, <https://doi.org/10.1109/SSCI.2015.22>.

- Sebestyen, G., Hangan, A. and Czako, Z. (2021) 'Anomaly detection in water supply infrastructure systems', *2021 23rd International Conference on Control Systems and Computer Science (CSCS)*, 26–28 May, pp.349–355, <https://doi.org/10.1109/CSCS52396.2021.00064>.
- Shalyga, D., Filonov, P. and Lavrentyev, A. (2018) '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* [online] <https://arxiv.org/abs/1807.07282> (accessed 3 September 2021).
- Shang, W., Cui, J., Song, C., Zhao, J. and Zeng, P. (2018) 'Research on industrial control anomaly detection based on FCM and SVM', *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)*, 1–3 August, pp.218–222, <https://doi.org/10.1109/TrustCom/BigDataSE.2018.00042>.
- Tamy, S., Belhadaoui, H., Rabbah, M.A., Rabbah, N. and Rifi, M. (2019) 'An evaluation of machine learning algorithms to detect attacks in SCADA network', *2019 7th Mediterranean Congress of Telecommunications (CMT)*, 24–25 October, pp.1–5, <https://doi.org/10.1109/CMT.2019.8931327>.
- Tariq, N., Asim, M. and Khan, F.A. (2019) 'Securing SCADA-based critical infrastructures: challenges and open issues', *Procedia Computer Science*, Vol. 155, pp.612–617, <https://doi.org/10.1016/j.procs.2019.08.086>.
- Terai, A., Abe, S., Kojima, S., Takano, Y. and Koshijima, I. (2017) 'Cyber-attack detection for industrial control system monitoring with support vector machine based on communication profile', *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 26–28 April, pp.132–138, <https://doi.org/10.1109/EuroSPW.2017.62>.
- Turkoz, M., Kim, S., Son, Y., Jeong, M.K. and Elsayed, E.A. (2020) 'Generalized support vector data description for anomaly detection', *Pattern Recognition*, Vol. 100, p.107119, <https://doi.org/10.1016/j.patcog.2019.107119>.
- Vapnik, V. (1998) 'The support vector method of function estimation', *Nonlinear Modeling: Advanced Black-Box Techniques*, pp.55–85, Springer US, Boston, MA. https://doi.org/10.1007/978-1-4615-5703-6_3.
- Wang, Y., Xu, Z., Zhang, J., Xu, L., Wang, H. and Gu, G. (2014) *SRID: State Relation Based Intrusion Detection for False Data Injection Attacks in SCADA*, pp.401–418, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-11212-1_23.
- Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H. and Wang, C. (2018) 'Machine learning and deep learning methods for cybersecurity', *IEEE Access*, Vol. 6, pp.35365–35381, <https://doi.org/10.1109/ACCESS.2018.2836950>.
- Yang, Y., Xu, H., Gao, L., Yuan, Y., McLaughlin, K. and Sezer, S. (2017) 'Multidimensional intrusion detection system for IEC 61850-based SCADA networks', *IEEE Transactions on Power Delivery*, Vol. 32, No. 2, pp.1068–1078, <https://doi.org/10.1109/TPWRD.2016.2603339>.
- Yeckle, J. and Abdelwahed, S. (2017) 'An evaluation of selection method in the classification of SCADA datasets based on the characteristics of the data and priority of performance', *Proceedings of the International Conference on Compute and Data Analysis*, Lakeland, FL, USA, Association for Computing Machinery, pp.98–103, <https://doi.org/10.1145/3093241.3093271>.
- Zhou, Z.-H. (2012) *Ensemble Methods: Foundations and Algorithms*, p.236, CRC Press, New York, ISBN: 9780429151095, <https://doi.org/10.1201/b12207>.