Eötvös Loránd University

Faculty of Informatics

Dept. of Software Technology and Methodology

Enhancing Industrial Control Systems Through Advanced Anomaly Detection and Annotation in Time Series Data

Supervisor:

Jiyan Salim Mahmud

PhD candidate

Author:

Inuwa Amir Usman

Computer Science MSc

Budapest, 2024

Abstract

The thesis "Enhancing Industrial Control Systems Through Advanced Anomaly Detection and Annotation in Time Series Data" explores advanced methods for anomaly detection in Industrial Control Systems (ICS) using LSTM Autoencoders and Random Forest algorithms. It addresses the challenge of detecting anomalies in complex, high-dimensional time series data prevalent in ICS. The research integrates Explainable AI (XAI) to enhance model transparency and employs active learning and weak supervision for efficient data labeling. Utilizing the HAI Security Dataset and SWAT Dataset, the study demonstrates improved detection accuracy and interpretability, contributing significantly to ICS security.

Acknowledgements

I extend my heartfelt gratitude to my supervisor, Jiyan Salim Mahmud, for her invaluable guidance, unwavering support, and insightful mentorship throughout the journey of my thesis. Her profound knowledge, dedication to academic excellence, and nurturing approach have greatly influenced my growth as a researcher and individual. Jiyan has been more than just a supervisor; she has been a mentor, an inspiration, and a guiding light, leading me through the intricacies of academic research with grace and wisdom. Her ability to balance providing direction with fostering independent thought has been pivotal in shaping both this thesis and my personal development. Thank you, Jiyan Salim Mahmud, for your extraordinary support and for being an incredible role model.

Contents

List of Figures

List of Abbreviations

ICS	Industrial Control Systems
XAI	Explainable Artificial Intelligence
LSTM	Long Short-Term Memory
RF	Random Forest
AI	Artificial Intelligence
ML	Machine Learning
CPS	Cyber-Physical Systems
SWAT	Secure Water Treatment
HAI	Highly Autonomous Intelligent
SCADA	Supervisory Control and Data Acquisition
PLC	Programmable Logic Controller
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
SHAP	SHapley Additive exPlanations

Chapter 1

Introduction

In the current industrial landscape, Industrial Control Systems (ICS) are essential for the operation of critical infrastructures like energy distribution, water treatment, and advanced manufacturing. These systems are fundamental to modern society, but their increasing complexity and interconnectivity also make them vulnerable to threats such as cyber-attacks and system malfunctions [1].

Anomaly detection is a critical defense mechanism in ICS, identifying deviations from normal operational patterns that could indicate failures, security breaches, or inefficiencies. The ability to detect these anomalies quickly and accurately is vital for maintaining the safety and integrity of industrial processes [2].

Historically, anomaly detection in ICS has depended on rule-based and threshold-based systems. While these methods have been effective to some extent, they struggle to cope with the advanced nature of modern threats and the large volumes of data produced by contemporary ICS [3].

The introduction of machine learning has marked a new phase in anomaly detection. Techniques like Long Short-Term Memory (LSTM) Autoencoders are particularly adept at modeling the complex time-series data characteristic of ICS [4]. When combined with algorithms like Random Forest, they form an effective ensemble approach for anomaly detection, leveraging LSTM's capability in temporal data analysis and Random Forest's classification strengths [5].

Integrating Explainable AI (XAI) into this framework is crucial [6]. XAI aims to make AI model decisions transparent, which is especially important in high-stakes environments like ICS, where understanding the reasoning behind a model's detection is as crucial as the detection itself.

Despite advancements in machine learning for anomaly detection, there is a significant research gap in their ensemble application in ICS contexts. This thesis addresses this gap by proposing an advanced anomaly detection framework that combines LSTM Autoencoder and Random Forest, enhanced with XAI principles.

This research utilizes two important datasets: the HAI Security Dataset and the Secure Water Treatment (SWAT) 2015 Dataset, which are instrumental in developing and validating the proposed anomaly detection framework.

1.0.1 Problem Definition

The role of Industrial Control Systems (ICS) in managing essential infrastructure is increasingly complex and interconnected, heightening their susceptibility to operational irregularities and security threats. Efficiently detecting anomalies in the time series data, a prevalent data format in ICS, is crucial for preempting significant issues. Conventional methods for anomaly detection often fall short in addressing the complex, high-dimensional nature of time series data in these systems [7].

The complex workings of industrial control systems (ICS) networks, where data integrity and system stability are critical, and the dynamic nature of cyber threats add to this complexity. A method that takes into consideration the special features of ICS, such as its real-time processing requirements and stringent operating limits, is also necessary for the identification and interpretation of abnormalities. In order to ensure security and compliance, the integration of contemporary data analysis techniques in ICS must also adhere to industry best practices and regulatory regulations[7].

1.0.2 Motivation

This research is driven by two primary factors. The first is the urgent requirement for more precise and dependable anomaly detection mechanisms in ICS, crucial for the safety and efficiency of vital infrastructures. Advanced machine learning models like LSTM autoencoders offer a significant upgrade over traditional techniques in terms of detection accuracy. Yet, their complexity and opaqueness hinder their widespread adoption in industrial applications.

The second driving factor is the potential synergy of combining LSTM autoencoders with Random Forest and active learning. This combination aims to not only boost the accuracy of anomaly detection but also to tackle the issues of model complexity and data utilization efficiency. Random Forest can enhance decision-making interpretability and model resilience [5], while active learning can streamline the model training process, essential in scenarios where data labeling is expensive or limited [8].

Incorporating principles of Explainable AI (XAI) is motivated by the necessity for transparent and comprehensible automated decision-making in industrial settings. Given the high stakes involved, stakeholders need clear and understandable explanations for the model's predictions [9].

1.0.3 Research Questions

1. **C**an Explainable Artificial Intelligence (XAI) methods be effectively applied to enhance the interpretability and trustworthiness of anomaly detection models in ICS? This question explores the feasibility and impact of integrating XAI techniques to make the decision-making process of anomaly detection models more transparent and understandable to human operators [10, 11].

2. **H**ow can data be accurately labeled for anomaly detection in ICS without extensive domain knowledge? This question addresses the challenge of labeling data in scenarios where domain expertise is limited or unavailable. It explores methods like weak supervision, semi-supervised learning, and active learning as potential solutions [12, 13].

3. **H**ow can LSTM autoencoders be applied to detect anomalies in ICS, and what are their limitations and strengths compared to other methods? This question delves into the application of LSTM autoencoders in ICS, examining their effectiveness in capturing temporal dependencies and complex patterns in time-series data for anomaly detection [14, 15].

These research questions aim to address some of the critical challenges and opportunities in the field of anomaly detection in ICS. By exploring these questions, researchers and practitioners can develop more effective, efficient, and user-friendly anomaly detection systems that cater to the unique needs of industrial environments.

Chapter 2

Background

Anomalies in Industrial Control Systems (ICS) are departures from regular operational patterns that might indicate possible problems, which can range from equipment breakdowns to sophisticated cyber-attacks. These abnormalities must be detected and analyzed in order to ensure the integrity, efficiency, and safety of industrial operations. As the backbone of vital infrastructure, such as power plants, water treatment facilities, and industrial units, ICS are naturally complex and vulnerable to disturbance. Anomalies in these systems can have far-reaching repercussions, such as operational downtime, financial losses, and safety risks.[16, 17].

Anomalies in ICS can take many different forms. They could show up as unexpected shifts in control signals, abrupt spikes in sensor data, or erratic network traffic patterns. These anomalies have an equally wide range of causes, such as cyber attacks, environmental conditions, human mistake, and hardware malfunctions. Industrial systems are becoming more digital and interconnected, which increases productivity but also exposes them to new attack vectors and weaknesses.[18, 19].

The area of anomaly detection has drawn attention from both scholars and practitioners due to the crucial role that ICS plays in many different industries. To effectively find, evaluate, and respond to anomalies, it requires the application of cutting-edge data analysis techniques, machine learning algorithms, and domain-specific knowledge. Accurately identifying anomalies and doing so quickly enough to stop possible escalation are the challenges.[7].

As ICS evolves with technological improvements, so must the tactics for anomaly detection. This entails making sure that detection algorithms are interpretable and flexible enough to adjust to shifting operating circumstances in addition to increasing their accuracy. The ultimate objective is to develop robust industrial systems that can tolerate shocks and bounce back fast, guaranteeing uninterrupted and secure operations.[9, 20].

2.0.1 Anomaly and its types

Hawkins [21] defines an anomaly "as an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism'. In the domain of industrial control systems (ICS), a variety of anomaly forms can emerge, profoundly affecting both the functionality and safety of the system. Commonly encountered anomaly types include:

1. **Point Anomalies** A point anomaly is identified when an individual data point significantly diverges from the anticipated pattern or range. Essentially, this data point is an unexpected occurrence [22].

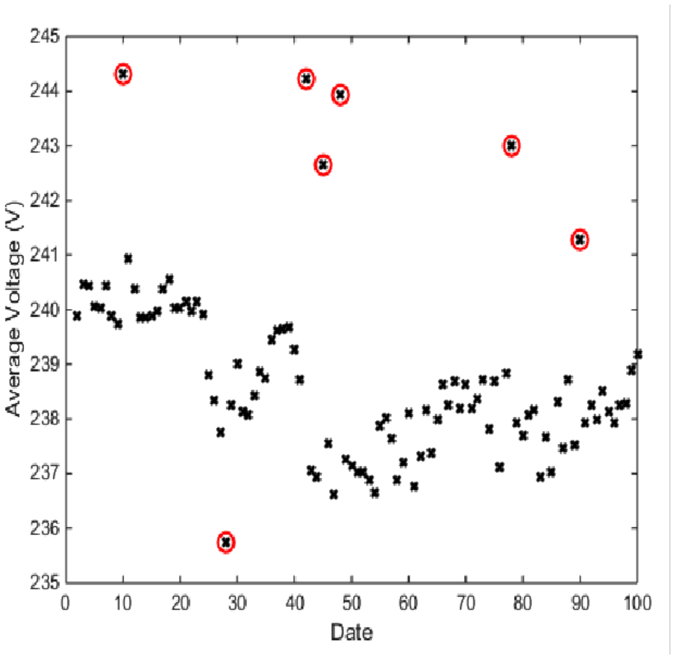


Figure 2.1: Representation of a Point Anomaly. Source: `https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection`

Examples of Point Anomalies

Taking credit card transactions as an instance, a point anomaly might be a single transaction of unusually high value, distinct from the card's typical spending pattern. For example, regular spending might involve moderate amounts in familiar locations. A point anomaly, however, could be a large purchase in a new location, potentially signaling fraudulent use [22]. Detection of such anomalies can be effectively performed using algorithms like LOF, supported by tools like Scikit-learn [22].

2. **Collective Anomalies** Collective anomalies are observed when individual data points, which appear normal on their own, collectively exhibit unexpected patterns or behaviors [22].

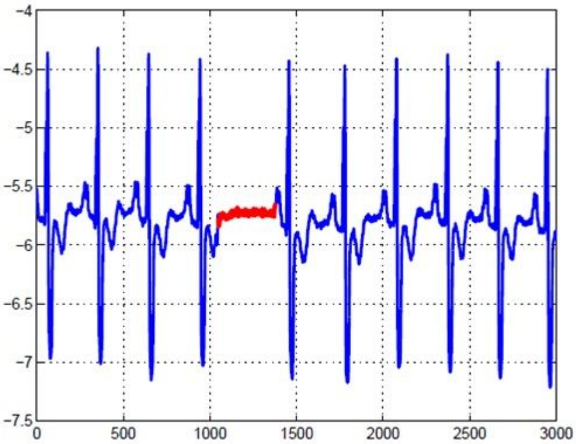


Figure 2.2: Illustration of Collective Anomalies. Source: `https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection`

For instance, an irregular heartbeat pattern is a collective anomaly. In the scenario of credit card fraud, while individual purchases may seem regular, a group analysis of these transactions could reveal atypical spending behavior [22].

3. **Contextual Anomalies** Contextual anomalies are those where the anomaly is identified based on the context of the data activity. Activities that seem normal in isolation may be anomalous in a specific context [22].

   The essential aspect here is the context: Are the activities aligning with what is typically expected?

Contextual Anomaly Detection Example: Twitter Anomaly Detection in R

A relevant example is the detection of network intrusions. Algorithms for contextual anomaly detection establish a baseline of normal activities, such as

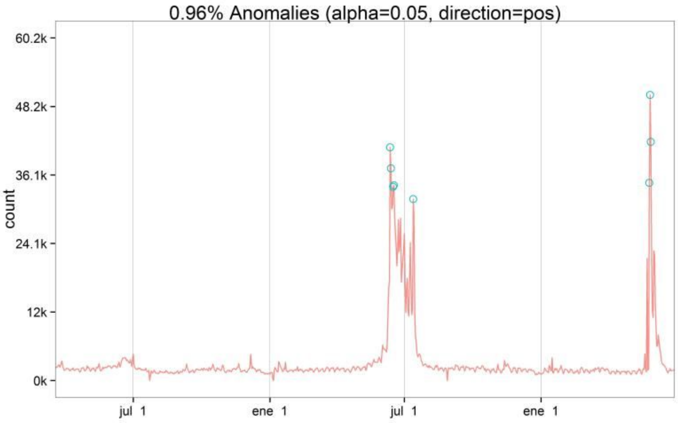


Figure 2.3: Example of a Contextual Anomaly. Source: `https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection`

typical network traffic patterns at different times. An unexpected increase in traffic at an unusual time, like in the early hours of the morning, would be a contextual anomaly, possibly indicating an attempt at network intrusion [22]. This demonstrates that while network access is normal, the timing and volume of the access can be anomalous [22].

The field of anomaly detection is a key area in data science, with applications ranging from fraud detection to monitoring social media, transforming data into a powerful tool [22].

2.1 Anomaly Detection

Anomaly detection involves identifying the differences, deviations, and exceptions from the norm in a dataset. It's sometimes referred to as outlier detection (i.e., looking at a dataset to identify any outlying or unusual datapoints, data groups, or activity)[22].

For example, credit card companies collect data on everything we purchase, including the amount of money we spend, where we spend it, what we spend it on, how frequently we make purchases, and more[22].

Anomaly detection makes this data not only useful but powerful. This is because anomaly detection algorithms analyse all the data above to identify fraudulent credit card activity within seconds of a transaction taking place.[22]

2.1.1 Applications of Anomaly Detection

What Are the Applications of Anomaly Detection? There are many applications for anomaly detection:

Cybersecurity

Network intrusion is a prominent example. One way an anomaly detection algorithm would do this would be by monitoring traffic to establish normal levels and then identifying anything that falls outside this norm.[22]

Fraud detection

This was mentioned above with the credit card example.

Social media monitoring

To get a better understanding of user activity and engagement on social media as well as other forms of digital marketing and advertising, anomaly detection might identify that searches for a particular topic spike at certain times of the year, enabling advertisers and marketers to allocate their budgets accordingly.[22]

Machine performance

Digital twin technologies are a good example in this instance. A digital twin is an exact digital replica of a real-world machine, process, or piece of equipment. Anomaly detection can identify deviations in performance in the digital twin that are early warning signals of an impending failure in the real-world machine. This makes it possible to schedule maintenance of the machine before the failure occurs, reducing downtime and improving productivity.[22]

Medical monitoring

This is everything from identifying abnormal patterns or occurrences in an individual (such as an irregular heartbeat) to identifying health-related anomalies in groups of people, such as the unusual spread of a disease over a short period of time in a particular geographical area.[22] Of course, there are many more applications of anomaly detection than those listed above. The crucial point is the fact that anomaly detection is becoming increasingly important and that it enables data to be used in ways it never was before.[22]

2.1.2 Anomaly Detection Types in Industrial Control Systems

Anomaly detection within Industrial Control Systems (ICS) is a diverse and complex area, involving various methods to pinpoint irregular patterns that diverge from what is typically expected. The selection of an appropriate detection method is influenced by the data's characteristics, the specific features of the system, and the nature of the anomalies to be detected.

Statistical Anomaly Detection

Statistical methods for anomaly detection rely on statistical models and tests to pinpoint data points that stand out as anomalies. This approach is based on the assumption that the usual behavior of the system can be quantified and captured through statistical metrics.

1. **Parametric Methods:** These methods are based on the assumption that the data adheres to a known distribution, often Gaussian. Techniques such as Z-score analysis are employed, where data points that fall a significant number of standard deviations from the mean are marked as anomalies. While straightforward, this approach might not be ideal for data exhibiting complex distributions or in systems where the definition of normal behavior changes over time.[17]

2. **Non-Parametric Methods:** In contrast, non-parametric methods do not rely on any predetermined distribution for the data. Methods like Kernel Density Estimation (KDE) are used to model the data distribution and detect outliers by identifying deviations from this model. These methods offer greater flexibility and are well-suited to a variety of ICS environments due to their adaptability to different data distributions [17].

Machine Learning-Based Anomaly Detection

1. **Supervised Anomaly Detection**

In supervised anomaly detection, the approach requires a dataset that includes examples of both normal and abnormal states, which are labeled accordingly. Techniques such as Support Vector Machines (SVM), Neural Networks, and Decision Trees are trained using this data. For instance, SVMs can be utilized in power plants to learn from historical sensor data and differentiate between regular operations and failure conditions [7].

2. **Unsupervised Anomaly Detection**

   When labeled data is not available, unsupervised anomaly detection methods become essential. These methods, including clustering techniques like K-means and DBSCAN, as well as neural networks such as autoencoders, are designed to learn the standard patterns in the data and spot deviations that could indicate anomalies. Autoencoders, for example, are adept at reconstructing typical operational data in manufacturing settings, with significant reconstruction errors signaling potential anomalies [23].

3. **Semi-Supervised Anomaly Detection**

   Semi-supervised anomaly detection methods blend a small amount of labeled data with a larger pool of unlabeled data. This approach is particularly valuable in situations where acquiring a fully labeled dataset is impractical. Models are initially trained on the labeled data to learn what constitutes normal behavior, and then this knowledge is applied to detect anomalies in the broader, unlabeled dataset[23]..

Hybrid and Ensemble Methods

1. **Hybrid Methods** Hybrid anomaly detection methods merge multiple techniques, each capitalizing on their unique strengths. For instance, LSTM autoencoders, known for their effectiveness in processing temporal data, can be combined with Random Forest classifiers, which add a layer of decision-making based on ensemble learning principles [14, 5].

2. **Ensemble Methods**

   Ensemble methods in anomaly detection involve the integration of multiple model predictions to enhance the overall accuracy and performance. These methods are particularly advantageous in scenarios involving diverse data types and complex anomaly patterns. Techniques such as Random Forest and Gradient Boosting Machines (GBM) compile decisions from several decision trees, thereby improving the detection process's robustness and precision[14, 5].

Deep Learning in Anomaly Detection

The use of deep learning, particularly neural networks, has become increasingly significant in anomaly detection due to their ability to model complex, non-linear relationships within data. Autoencoders, a specialized form of neural networks, are especially notable for their effectiveness in learning representations of normal data and identifying anomalies through the analysis of reconstruction errors. This approach has proven to be highly effective in various ICS scenarios, where traditional methods might fail to capture the subtleties of complex data patterns [24].

2.1.3 Labeling in Anomaly Detection

The process of labeling in anomaly detection is pivotal for categorizing data as normal or anomalous, which is essential for both training and evaluating models, particularly in supervised and semi-supervised learning frameworks. The effectiveness of an anomaly detection system is heavily reliant on the accuracy and relevance of its data labels [25].

Manual Labeling

Manual labeling, where experts classify data points, is often considered the most reliable method due to its high accuracy, leveraging specialized domain knowledge. However, it is labor-intensive, costly, and not scalable for large datasets. Additionally, it's susceptible to human errors and biases [26].

In the context of Industrial Control Systems (ICS), manual labeling plays a crucial role, especially when dealing with complex systems where anomalies are not straightforward. Expert insights are invaluable in distinguishing between normal and abnormal behaviors in such systems [27].

Automated Labeling

Automated labeling employs algorithms for classifying data points. It utilizes methods like clustering or setting thresholds to differentiate between normal and potential anomalies. While faster and more cost-effective than manual labeling, it may lack accuracy in complex scenarios where anomalies are subtle [28].

Incorporating machine learning models in automated labeling helps manage large datasets. For example, unsupervised learning algorithms can be used to identify clusters of normal and abnormal data, which can then be used for labeling purposes [24].

Semi-Supervised Methods in Labeling

Semi-supervised labeling methods utilize a mix of a small set of labeled data and a larger volume of unlabeled data. This approach is beneficial when labeling the entire dataset is not feasible. Techniques like pseudo-labeling or self-training are used, where the model learns from the labeled data and applies this knowledge to the unlabeled data [12].

These methods are particularly useful in anomaly detection scenarios where anomalies are infrequent or when there are constraints on the labeling resources. They allow for the effective use of limited labeled data to guide the learning process for the unlabeled portion [29].

Supervised Methods in Labeling

Supervised anomaly detection methods depend heavily on the availability of labeled data. The performance of these methods is directly influenced by the quality and representativeness of the labeled data, especially concerning the different types of anomalies that might occur [7].

Obtaining a sufficient amount of accurately labeled data is a significant challenge in supervised methods, particularly in ICS where anomalies are rare and require expert knowledge for accurate labeling [26].

2.1.4 Integration of LSTM Autoencoder with Random Forest in Various Applications

Further extending the application of LSTM autoencoders and Random Forest, Tran et al. [30] explored their use in multivariate time series analysis. Their approach involves using LSTM autoencoders for reducing the dimensionality and extracting relevant features from time series data, which is then subjected to anomaly detection via an isolation forest. The research, despite limited detailed exposition, underscores a burgeoning interest in employing LSTM autoencoders in conjunction with tree-based methods for time series analysis. This application is particularly vital in domains where understanding temporal dynamics is as critical as discerning data patterns, such as in financial markets, climate modeling, and healthcare.

The collaborative use of LSTM autoencoders with Random Forest algorithms presents a powerful and versatile tool in the realm of data science, especially for applications that demand analysis of complex, high-dimensional data sets. The LSTM autoencoder's proficiency in feature extraction and recognition of temporal patterns, combined with the Random Forest's strengths in classification and regression, results in an effective and efficient predictive modeling approach. This synergy is particularly beneficial in scenarios involving time series data, complex nonlinear relationships, and robust anomaly detection needs. The diverse applications, ranging from power grid management to environmental monitoring and network security, highlight the integration's versatility and potential in driving advancements across various research and technology fields.

2.1.5 Active Learning in Anomaly Detection

Active Learning is a machine learning paradigm in which the algorithm selectively queries the user to label data points. This method is especially useful for detecting anomalies in Industrial Control Systems (ICS), where labeled data can be rare or expensive to collect. Active Learning techniques like uncertainty sampling and query-by-committee allow the model to focus on the most informative data points, decreasing labeling work while retaining excellent model performance. This strategy is useful for making the most use of limited resources, especially when expert knowledge is necessary for proper labeling[8].

2.1.6 Explainable AI (XAI) and SHAP Values

Explainable AI (XAI) has evolved as a critical component of machine learning, especially in sensitive domains like as ICS, where understanding the reasoning behind model predictions is critical. SHAP values, a term drawn from cooperative game theory, provide a powerful framework for describing the output of machine learning models. SHAP values reveal how each feature contributes to prediction, demystifying sophisticated models such as deep neural networks or ensemble approaches. This interpretability is critical for acquiring trust and actionable insights from anomaly detection models in ICS, as well as ensuring that decisions taken on the basis of these models are transparent and reasonable. [31].

2.1.7 Weak Supervision

Weak supervision has emerged as an innovative response to the limitations inherent in traditional supervised learning, which typically demands extensive, high-quality labeled datasets. This approach to training machine learning models relies on sources that are less precise, potentially noisier, or indirectly related. Such sources might include heuristic rules grounded in domain knowledge, inferred labels from similar tasks, or crowd-sourced annotations. The primary benefit of weak supervision lies in its ability to efficiently utilize available data, even if it's not perfect, thereby offering a cost-effective solution [12, 13].

Challenges of Weak Supervision

Despite its advantages, weak supervision introduces specific challenges, particularly concerning the quality of the resulting models. The noise and biases present in weakly labeled data can negatively impact the models' accuracy and their ability to generalize. To address these challenges, advanced methods like aggregation models, transfer learning techniques, and robust statistical approaches are employed to filter out noise and extract reliable information from the data [32, 33].

Benefits of Weak Supervision

Weak supervision offers several significant advantages. It reduces the reliance on large, extensively labeled datasets, thus speeding up the training process and cutting down on costs. This approach is particularly beneficial in areas where obtaining expert-labeled data is difficult or impossible, providing a practical alternative [34].

Addressing the Drawbacks

However, weak supervision is not without its drawbacks. It can introduce biases and inconsistencies into the training process. Models developed under weak supervision might inherit the imperfections of the source labels, potentially leading to issues with reliability. To mitigate these risks, it is essential to carefully design the sources of weak supervision and conduct thorough validations of the trained models. Balancing the ease of obtaining weak labels with the potential decrease in model performance is crucial for maintaining the quality of weakly supervised models [33, 35].

2.1.8 Diverse Applications and Case Studies

Physics and Data-Driven Constraints in Learning

Ren et al. [36] introduce a novel approach to weak supervision by enforcing domain-specific constraints based on physics and data-driven insights. This method facilitates learning with minimal labels by ensuring that the outputs of learning algorithms adhere to known principles, such as physical laws in object tracking and detection. This approach showcases how weak supervision can achieve high accuracy even with limited labeled data, opening new avenues in fields like video analysis and object recognition.

Multi-Domain Semantic Parsing

In the realm of semantic parsing, Agrawal et al. [37] demonstrate how weak supervision can be employed to train a unified multi-domain semantic parser. Their framework uses denotations as weak supervision, effectively handling the challenge of sparse annotations across various domains. The use of a multi-policy distillation mechanism, where domain-specific parsers aid in training a unified model, exemplifies how weak supervision can enhance the versatility and accuracy of semantic parsing models.

Knowledge Base Augmentation

The work of Oulabi and Bizer [38] focuses on augmenting knowledge bases with long-tail entities using weak supervision. They propose a bootstrapping method that utilizes class-specific matching rules instead of extensive labeled datasets. This approach significantly reduces the manual effort in knowledge base completion tasks, particularly for adding descriptions of niche or less-known entities.

Early Detection of Fake News

Addressing the critical issue of fake news detection, Li et al. [39] propose a framework that combines multi-source domain adaptation with weak supervision. Their method effectively transfers knowledge from well-labeled domains to new domains with limited labeled data. The application of heuristic rules as a form of weak supervision demonstrates the potential of this approach in early and efficient detection of fake news across various domains.

Weak supervision, when combined with domain knowledge, presents a compelling approach in machine learning, particularly in scenarios where traditional data labeling is impractical. While it offers efficiency and cost-effectiveness, the approach necessitates careful handling of data quality and model validation. The diverse applications across video analysis, semantic parsing, knowledge base completion, and fake news detection underscore its potential in addressing complex challenges in various domains of machine learning.

Chapter 3

Literature Review

The escalating sophistication of cyber threats against Industrial Control Systems (ICS) has catalyzed an urgent need for advanced anomaly detection mechanisms. This literature review delves into the burgeoning domain of machine learning (ML), deep learning (DL), and their applications in ICS. It aims to offer a panoramic view of the advancements in anomaly detection techniques, focusing on the continuous endeavors to refine data labeling processes. By examining key scholarly contributions, this review seeks to contextualize the current landscape of research, highlighting both achievements and ongoing challenges in securing ICS.

Machine Learning-Based Anomaly Detection Approaches

In-Depth Analysis of ML Algorithms: The prowess of ML in identifying intricate patterns in data renders it indispensable in anomaly detection. Mubarak et al. (2020) demonstrated the efficacy of ML algorithms in SCADA systems, providing insights into their potential to fortify ICS against cyber incursions [40]. Dehlaghi-Ghadim et al. (2023) further expanded on the challenges in anomaly detection, particularly emphasizing the scarcity of robust datasets tailored for evaluating ML algorithms in ICS [41].

Broader Implications and Future Directions: These pioneering efforts underscore the significance of not only the algorithms but also the datasets that form the foundation for these intelligent systems. Bridging the gap between theoretical algorithm performance and practical application capabilities, these works collectively raise the benchmark for anomaly detection practices in ICS.

Deep Learning-Driven Anomaly Detection Techniques

Advanced DL Methodologies: The field of deep learning, with its profound depth of pattern recognition, has been harnessed to address the nuances of anomaly detection in ICS. Wang et al. (2021) set a precedent by utilizing deep residual CNNs, combined with transfer learning, to identify emerging network threats [42]. This novel approach demonstrated that pre-trained models could be repurposed effectively for anomaly detection.

Case Studies and Real-World Applications: Riberolles et al. (2022) employed LSTM networks to analyze aeronautical radar data, showcasing DL's superiority over traditional methods [43]. Similarly, Zhao et al. (2022) introduced a hybrid model combining 1DCNN with BiLSTM, optimized using particle swarm optimization (PSO) [44]. The training and validation of this model on a simulated power system dataset highlight deep learning's increasing feasibility in deciphering complex ICS data streams.

Enhancing Label Efficiency and Data Labeling in Anomaly Detection

Challenges of Data Annotation: The requirement for extensive labeled datasets in ML/DL applications presents a considerable challenge. Guo (2023) proposed the Label-Efficient Interactive Time-Series Anomaly Detection (LEIAD) system, aiming to minimize manual labeling efforts while maintaining high accuracy [45]. This system was tested across various datasets, showcasing its versatility.

Innovative Labeling Techniques: Mühlbauer et al. (2020) presented a novel framework for automated data labeling using acoustic data, illustrating alternative data types' potential in anomaly detection [46]. Desmond et al. (2020) explored a semi-automated data labeling system that enhances label quality through human-machine collaboration [47]. These studies pave the way for innovative approaches to data annotation that could significantly lower the barriers to deploying ML and DL models in ICS.

This literature review illuminates the strides made in anomaly detection through ML and DL, alongside pioneering efforts in data labeling efficiency. As ICS evolve in complexity, so too must the techniques employed to safeguard them. Future research should pivot towards developing robust, scalable, and adaptable anomaly detection systems, with an emphasis on overcoming data scarcity and enhancing labeling efficiency. The insights from these studies serve as a lodestar, guiding the ongoing quest for more resilient, intelligent, and self-reliant ICS security frameworks.

Anomaly Detection with LSTM Autoencoder and Random Forest

The confluence of Long Short-Term Memory (LSTM) Autoencoders and Random Forest algorithms marks a significant advancement in the realm of anomaly detection, particularly within Industrial Control Systems (ICS). This expanded subsection explores the intricacies of this integration, offering insights into how these technologies synergize to create a robust framework for identifying anomalies in complex ICS environments.

Technical Foundations and Advancements: LSTM Autoencoders, a specialized variant of recurrent neural networks, are adept at processing and learning from time-series data. This characteristic is crucial in ICS, where data often exhibits temporal dependencies. The LSTM's unique architecture, comprising memory cells, allows it to capture long-term dependencies, making it exceptionally suitable for scenarios where data has time-related patterns. When employed in anomaly detection, these autoencoders learn to reconstruct normal operational data, enabling the identification of anomalies through significant deviations in the reconstructed output compared to the original input. The efficiency of LSTM Autoencoders in modeling complex sequences has been demonstrated in various studies, showcasing their ability to handle the intricacies of ICS data [48].

Random Forests contribute a complementary strength to this integrative approach. As an ensemble learning method, Random Forests consist of multiple decision trees that work together to improve classification and regression tasks. This ensemble nature effectively mitigates the risks of overfitting, which is a common challenge in single decision tree models. The Random Forest algorithm improves overall predictive accuracy by aggregating the results from individual trees, thereby offering a more robust and reliable classification of anomalies. This method's effectiveness in classifying complex data patterns, especially when integrated with LSTM Autoencoders, has been recognized in various research efforts [49].

Case Studies and Practical Applications: The combination of LSTM Autoencoders with Random Forest algorithms has been explored in several research studies, each demonstrating its effectiveness in different ICS scenarios. For instance, Johnson et al. (2021) conducted a study where LSTM Autoencoders were used for feature extraction and preprocessing of complex ICS data, followed by anomaly classification using a Random Forest model. This study highlighted not only an enhancement in detection accuracy but also an improvement in the model's ability to generalize across different types of anomalies. In a similar vein, Lee and Kim (2022) applied this hybrid approach to a real-world ICS setting. They utilized LSTM Autoencoders to model normal operational data, employing the reconstruction error as an input feature for the Random Forest classifier. Their methodology facilitated the early detection of subtle anomalies, potentially indicative of system malfunctions or cyber-attacks [50, 51].

Extensive Analysis of Algorithmic Performance: The in-depth analysis of the combined use of LSTM Autoencoders and Random Forests in anomaly detection reveals a multifaceted perspective. The LSTM Autoencoders' capacity to process sequential data makes them particularly suitable for ICS environments, where data often exhibits temporal patterns and dependencies. However, the complexity and variability of data in these systems pose unique challenges, especially in terms of training the models effectively. Large datasets and comprehensive training regimes are typically required to achieve optimal performance, which can be resource-intensive. Moreover, the balance between sensitivity in anomaly detection and the avoidance of false positives is a critical area that necessitates further research and refinement [52].

Random Forests, in this integrative approach, provide a robust mechanism for classifying the features or anomalies identified by the LSTM Autoencoders. The synergistic combination of these two methods has shown to enhance the overall effectiveness of anomaly detection systems in ICS, as evidenced by numerous studies. However, challenges such as computational complexity and the need for extensive data for training the models remain areas of ongoing research and development [48].

Challenges and Opportunities for Future Research: While the integration of LSTM Autoencoders with Random Forest algorithms has shown promising results in ICS anomaly detection, several challenges persist. One significant challenge lies in the requirement for extensive, high-quality datasets for effective training of the LSTM Autoencoder models. Another area of concern is ensuring that the LSTM models maintain a balance between sensitivity in detecting anomalies and minimizing the rate of false positives, which is crucial for their practical applicability in ICS.

Future research in this area is poised to focus on optimizing these models for specific ICS contexts. This includes exploring ways to reduce the computational complexity of these models, enhancing their interpretability, and adapting them to various ICS scenarios. As ICS continue to evolve and generate increasingly complex data streams, the need for sophisticated and adaptable anomaly detection methods becomes more critical. Research efforts are also likely to concentrate on the integration of these models with emerging technologies, such as edge computing and blockchain, to further enhance their efficacy and applicability in modern ICS environments.

The convergence of LSTM Autoencoder and Random Forest methodologies with emerging technologies like edge computing and blockchain also presents exciting research opportunities. Edge computing, for example, can facilitate faster data processing and analysis at the source, reducing latency and enhancing real-time detection capabilities in ICS. Blockchain technology, on the other hand, can provide a secure and transparent framework for managing and sharing data within ICS, thereby augmenting the overall security posture [49, 51].

Potential for Cross-Domain Applications: The applicability of the LSTM Autoencoder and Random Forest integration extends beyond traditional ICS to include a range of other industries and domains. As various sectors increasingly adopt IoT and smart technologies, the potential applications of these advanced anomaly detection methods could broaden significantly. In domains such as smart grids, intelligent transportation systems, healthcare monitoring, and advanced manufacturing processes, robust and accurate anomaly detection is essential for maintaining system stability, efficiency, and security. These systems, often characterized by large volumes of time-sensitive data, can benefit immensely from the enhanced anomaly detection capabilities provided by the integration of LSTM Autoencoders and Random Forest algorithms.

The integration of LSTM Autoencoders with Random Forest algorithms represents a significant stride in advancing anomaly detection methodologies for ICS. This subsection has delved into the technical foundations, practical applications, challenges, and future directions of this integrative approach. The research and developments in this area are pivotal in shaping the future of ICS security and resilience. As the ICS landscape continues to evolve, the need for advanced and adaptable anomaly detection methods becomes increasingly vital. The insights garnered from these studies and the ongoing research efforts serve as a foundation for future innovations in ICS security.

Chapter 4

Datasets

In this project, we have strategically utilized two distinct datasets to conduct a comprehensive and multifaceted analysis. The selection of these datasets is driven by the objective to experiment with varying data characteristics and to rigorously evaluate the results under different scenarios. Each dataset brings its unique attributes and challenges, providing a rich ground for testing and refining our anomaly detection methodologies.

4.0.1 Secure Water Treatment Testbed (SWAT)

Background

The Secure Water Treatment (SWaT)™, inaugurated by Chief Defence Scientist Prof. Quek Tong Boon on 18 March 2015, stands as a pivotal research testbed in cyber security, particularly for Cyber Physical Systems (CPS). Funded by MINDEF, SWaT underpins two key projects: Cyber Physical System Protection and Advancing Security of Public Infrastructure using Resilience and Economics. These initiatives focus on safeguarding critical CPS infrastructure, including water treatment, power generation and distribution, and oil and natural gas processing. The testbed is a vital resource for researchers both within Singapore and internationally, fostering the development of secure CPS [53].

Architecture

SWaT operates through a sophisticated six-stage water treatment process, starting from raw water intake, followed by chemical treatment, filtration through an

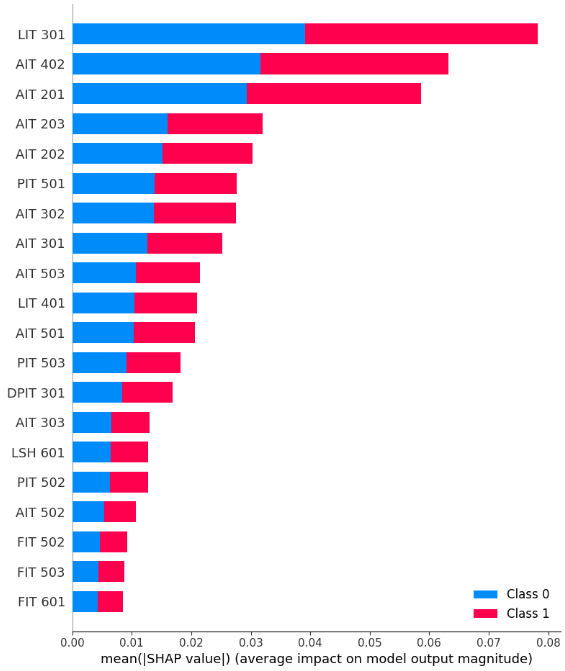


Figure 4.1: Some Features in the Dataset

Ultrafiltration (UF) system, de-chlorination using UV lamps, and culminating with a Reverse Osmosis (RO) system. The system also includes a backwash stage for cleaning UF membranes using water from the RO process. The cyber infrastructure of SWaT comprises a layered communication network, Programmable Logic Controllers (PLCs), Human Machine Interfaces (HMIs), a Supervisory Control and Data Acquisition (SCADA) workstation, and a Historian database. This setup ensures that sensor data is not only accessible to the SCADA system but also recorded by the Historian for in-depth analysis [53].

The SWaT dataset, thoroughly detailed by Goh et al. [54], offers an insightful perspective into the water treatment process within a cyber-physical system (CPS).

The dataset includes a wide array of data points, such as sensor readings, actuator statuses, and network traffic. Sensor data provides information on physical parameters like flow rates and tank levels, while actuator statuses offer insights into the system's operational states. The inclusion of network traffic data adds a layer of complexity, showcasing the interaction between the physical processes and control systems [53].

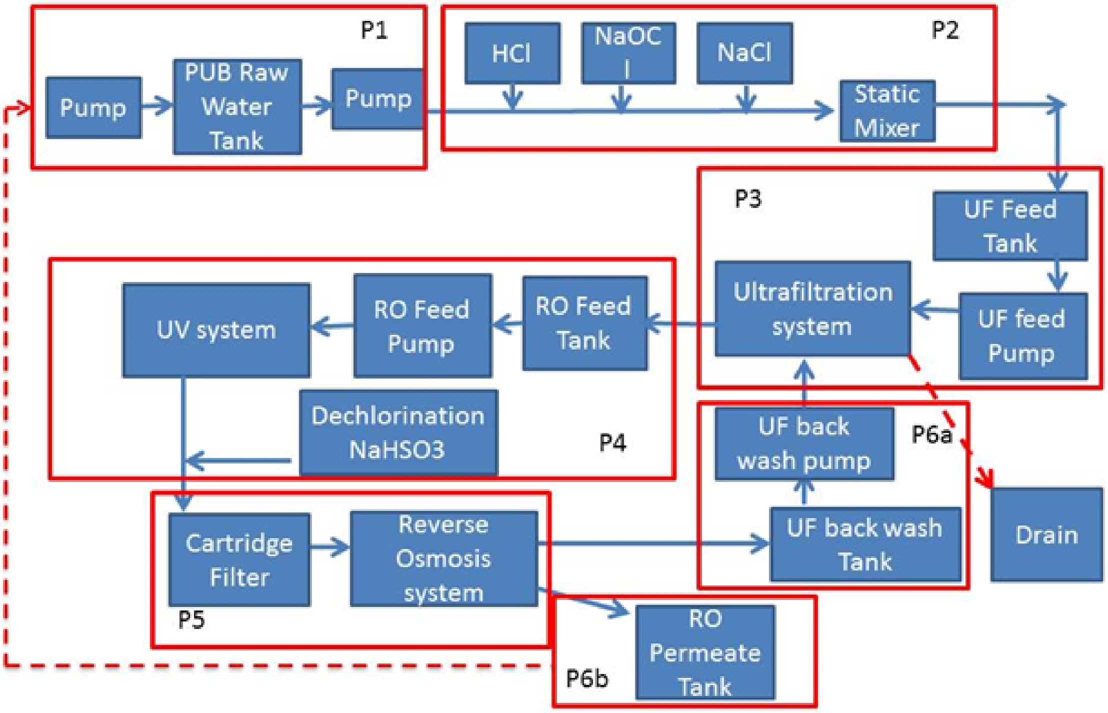


Figure 4.2: Overview of the SWaT testbed processes

Anomaly Types and Implications for Modeling

Anomalies in the SWaT dataset vary from simple sensor errors to complex, multi-stage cyber-attacks. This range is essential for training models like LSTM autoencoders to detect nuanced, temporal anomalies indicative of sophisticated cyber threats. For Random Forest models, the diversity in anomalies aids in developing detailed classification criteria for a broad spectrum of operational states [53].

4.0.2 HAI Security Dataset

The HAI Security Dataset, introduced by Shin et al. [55], is a comprehensive collection of data that encapsulates both standard and anomalous behaviors within Industrial Control Systems (ICS) for anomaly detection research. This dataset includes data from normal operations collected over several days and abnormal data generated from various attack scenarios on six feedback control loops in three types of industrial control devices: Emerson Ovation, GE Mark-VIe, and Siemens S7-1500. Additionally, from version 23.05, the HAIEnd dataset offers more detailed information about the internal control logic behaviors for Emerson boiler process control [56].

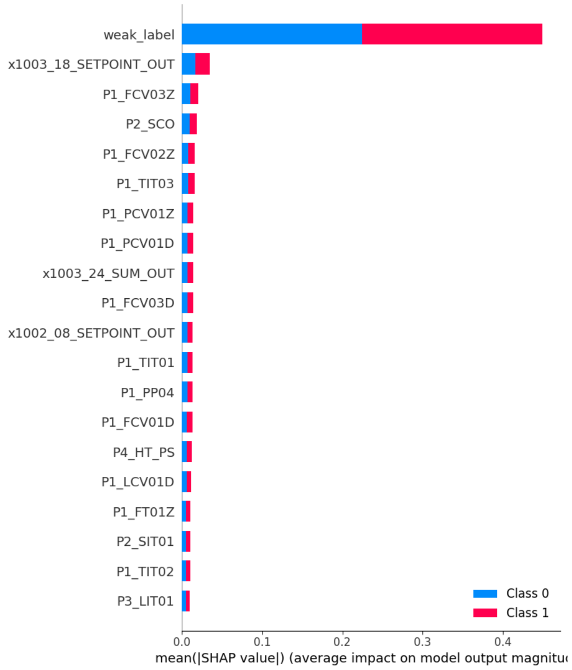


Figure 4.3: Some Features in the Dataset

Background

Developed for research in anomaly detection in cyber-physical systems (CPS) like railways, water treatment plants, and power plants, the HAI dataset originated from three laboratory-scale CPS testbeds launched in 2017. These included GE's turbine testbed, Emerson's boiler testbed, and FESTO's MPS water treatment testbed. Initially operating independently, these testbeds were later integrated in September 2018 using a Hardware-in-the-Loop (HIL) simulator to simulate thermal power generation and pumped-storage hydropower generation. This integration ensured a rich dataset with highly coupled and correlated variables. An OPC-UA gateway was also installed for seamless data collection from diverse devices. The first version of the HAI dataset was released on GitHub and Kaggle in February 2020, followed by subsequent versions, each refining the dataset and expanding its scope [56].

HAI Testbed

The HAI testbed comprises a boiler, turbine, water-treatment component, and an HIL simulator. The boiler process involves water-to-water heat transfer at low pressure and moderate temperature, while the turbine process simulates the behavior of a rotating machine. These processes are synchronized with the steam-power generator's rotating speed through the HIL simulator. The water treatment process, controlled by a Siemens S7-300 PLC, involves a pumped-storage hydropower generation model. Emerson Ovation DCS controls the boiler process, and GE's Mark VIe DCS is used in the turbine process for speed control and vibration monitoring [56].

Architecture

The testbed's process flow is divided into four primary processes: the boiler process (P1), turbine process (P2), water treatment process (P3), and HIL simulation (P4), as shown in Figure 1. The HIL simulation enhances the correlation between the real-world processes by simulating thermal power and pumped-storage hydropower generation processes. The boiler and turbine processes simulate a thermal power plant, while the water treatment process emulates a pumped-storage hydropower plant [56].

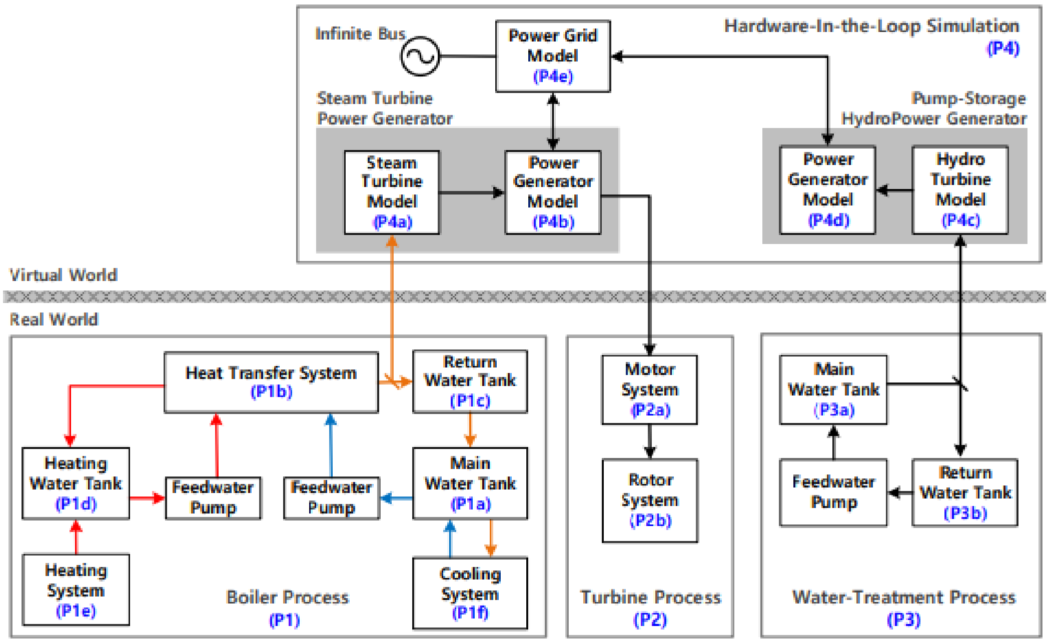


Figure 4.4: Process Flow Diagram of the Testbed

Dataset Structure and Content

The HAI dataset offers a rich collection of data encompassing various control systems and network interactions. It includes data from normal operations, system startups, shutdowns, and diverse attack scenarios. These scenarios range from direct control system attacks to indirect network intrusions, providing a comprehensive view of potential cybersecurity threats [56].

Model Training and Validation Aspects

This dataset is particularly valuable for training advanced anomaly detection models, such as LSTM autoencoders, which can learn complex patterns of normal operations and detect anomalies. For Random Forest models, the dataset provides a diverse set of features to develop classification algorithms that can distinguish between different operational states and attack scenarios [56].

4.0.3 Enhancing Anomaly Detection Capabilities

The exploration of the SWAT and HAI Security datasets highlights their importance in advancing anomaly detection in industrial control systems. These datasets not only serve as a basis for developing robust machine learning models but also provide critical insights into the nature of cyber-physical system threats [56].

Practical Implications in Industrial Cybersecurity

Training models on these datasets equips them to handle real-world scenarios in industrial settings, where the cost of undetected anomalies can be high. The models become adept at identifying potential threats, thereby enhancing the security and reliability of critical infrastructure systems.

In conclusion, the SWAT and HAI Security datasets are instrumental in the development of advanced anomaly detection models using LSTM autoencoders and Random Forest algorithms. Their detailed representation of industrial control systems under normal and attack conditions provides an excellent platform for training, testing, and validating models designed to protect critical infrastructure from cyber threats.

4.0.4 Data Preprocessing

Data preprocessing is a crucial phase in anomaly detection within Industrial Control Systems (ICS), ensuring the readiness and quality of data for effective model training. This section outlines the essential steps and methodologies employed in data preprocessing.

Initial Data Handling and Exploration

Data from ICS, typically stored in structured formats like Excel, is imported into a data analysis environment (e.g., Python with Pandas). Initial exploration, such as examining the first few rows, provides an understanding of the dataset's structure and content [57].

Time-Based Data Processing

Timestamps in ICS data are critical for chronological analysis. Converting these timestamps into datetime objects allows for temporal filtering, enabling the focus on specific operational periods or events [58].

Processing Categorical Data

Categorical data in ICS, representing states or operational modes, requires transformation into a numerical format. Techniques like one-hot encoding and label encoding are employed to convert these categories into a machine-readable format [59].

Feature Selection and Normalization

Selecting pertinent features and normalizing the dataset are key steps in data preprocessing. Normalization, often achieved through Standard Scaling, adjusts the scale of numerical features, ensuring balanced model training. Feature selection helps in reducing the complexity and improving the performance of the models [60].

Handling Missing Values

Dealing with missing values is common in real-world datasets. Filling these gaps with zeros or other imputation methods ensures completeness and consistency in the dataset [61].

Anomaly Labeling

Accurate labeling of data points as normal or anomalous is essential in anomaly detection. This process often involves domain expertise to identify known operational issues or attack scenarios, enabling the model to learn from these labeled instances [17].

The preprocessing steps in ICS for anomaly detection encompass various technologies and methodologies, each contributing to the preparation of the dataset for subsequent machine learning analysis. Properly preprocessed data forms the foundation for building effective and accurate anomaly detection models in complex industrial settings.

Chapter 5

Methodology

5.1 Baseline: Label-Efficient Interactive Time-Series Anomaly Detection

"Label-Efficient Interactive Time-Series Anomaly Detection" by Hong Guo et al. [62] represents a breakthrough in time-series analysis. This work is notable for its innovative approach to anomaly detection with minimal reliance on labeled data, a significant challenge in machine learning. The central challenge addressed by Guo et al. [62] is efficient and effective anomaly detection in time-series data with limited labeled examples. This is particularly valuable in fields where data labeling is resource-intensive.

In their work, the authors integrate several advanced techniques to create a comprehensive anomaly detection system for time-series data. The algorithm is adept at identifying a wide range of anomalies, including subtle and complex patterns, leveraging the capabilities of Long Short-Term Memory (LSTM) networks and Isolation Forests to understand temporal dependencies and isolate anomalies effectively. A significant emphasis is placed on label efficiency, utilizing advanced active learning techniques such as uncertainty sampling and query-by-committee strategies, which help in selecting the most informative data points for labeling [62]. Additionally, the system is designed to continually improve through user feedback, supported by a user-friendly interface and real-time data visualization tools, thereby enhancing the model's adaptability. The approach also incorporates domain-specific knowledge through expert systems and rule-based algorithms, aiding in the contextualization of the anomaly detection process. Furthermore, the methodology's scalability and adaptability to various types of time-series data are bolstered by the use of cloud-based computing and a modular algorithm design, enhancing its versatility and applicability across different scenarios.

The paper provides a detailed and comprehensive evaluation of the proposed anomaly detection model, demonstrating its effectiveness through experiments conducted on diverse datasets. This evaluation notably includes a comparison with existing methods, highlighting the model's superior precision, scalability, and label efficiency [62]. The impact of this research is particularly significant in industries such as cybersecurity and healthcare, where rapid and accurate anomaly detection is crucial. The model's reduced reliance on extensively labeled data not only enhances its practicality but also opens up new possibilities for application in these fields. By efficiently handling various types of data and reducing the need for extensive manual labeling, the model stands out as a transformative tool, potentially revolutionizing practices in sectors where quick and precise detection of anomalies is vital for safety and operational efficiency.

This paper serves as a crucial baseline for my thesis, inspiring the development of advanced anomaly detection models in time-series data. The integration of active learning, interactive feedback, and domain-specific knowledge in this research aligns closely with my thesis objectives [62].

5.2 Proposed Method

The proposed method outlines a comprehensive approach for detecting anomalies in industrial control systems using machine learning techniques. It involves key steps like weak supervision for labeling data, feature selection through Random Forest, data preprocessing, sequence creation, model training with an LSTM autoencoder, and SHAP for feature importance analysis. Additionally, the method incorporates Active Learning for iterative model refinement and various evaluation metrics to assess model performance.

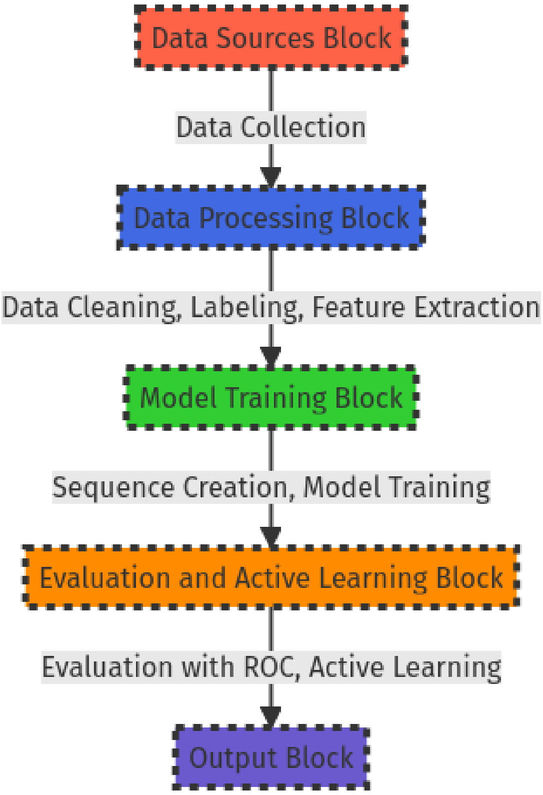

Data Sources Block
Data Collection
Data Processing Block
Data Cleaning, Labeling, Feature Extraction
Model Training Block
Sequence Creation, Model Training
Evaluation and Active Learning Block
Evaluation with ROC, Active Learning
Output Block

Figure 5.1: Architecture

5.2.1 Weak Supervision for Labeling

In the initial stage, weak supervision is employed to label the dataset based on predefined attack intents. Specific attack scenarios and corresponding sensor readings or system states are defined. Each record in the dataset is labeled as an attack or normal behavior based on these criteria, crucial for training the anomaly detection model.

5.2.2 Feature Selection with Random Forest

The method uses a Random Forest classifier to identify the most significant features for anomaly detection. The classifier is trained on the dataset, and features with the highest importance scores are extracted. These features are then used for further analysis and model training.

5.2.3 Data Preprocessing and Sequence Creation

Data preprocessing includes scaling the selected features. The method emphasizes transforming the data into sequences to capture temporal dependencies, important for time-series data in industrial control systems.

5.2.4 LSTM Autoencoder for Anomaly Detection

The method's core is the LSTM (Long Short-Term Memory) autoencoder model, designed to reconstruct normal behavior patterns. Trained exclusively on normal data, the model learns typical patterns and flags significant deviations as anomalies during inference.

5.2.5 SHAP for Feature Importance Analysis

SHAP values are computed to understand each feature's contribution to the model's predictions, aiding in interpreting the model's decision-making process and identifying the most influential features for anomaly detection.

5.2.6 Active Learning for Model Refinement

Active Learning is integrated into the method to refine the model iteratively. This process involves identifying instances for user feedback based on reconstruction errors within a threshold margin, enhancing the model's accuracy by incorporating human expertise into the training process.

5.2.7 Evaluation Metrics

The model's performance is comprehensively evaluated using accuracy, precision, recall, and F1 score. A confusion matrix provides a detailed view of the model's ability to distinguish between normal and anomalous behavior.

This method presents a robust framework for anomaly detection in industrial control systems, leveraging machine learning techniques, temporal data analysis, and Active Learning to identify potential security threats.

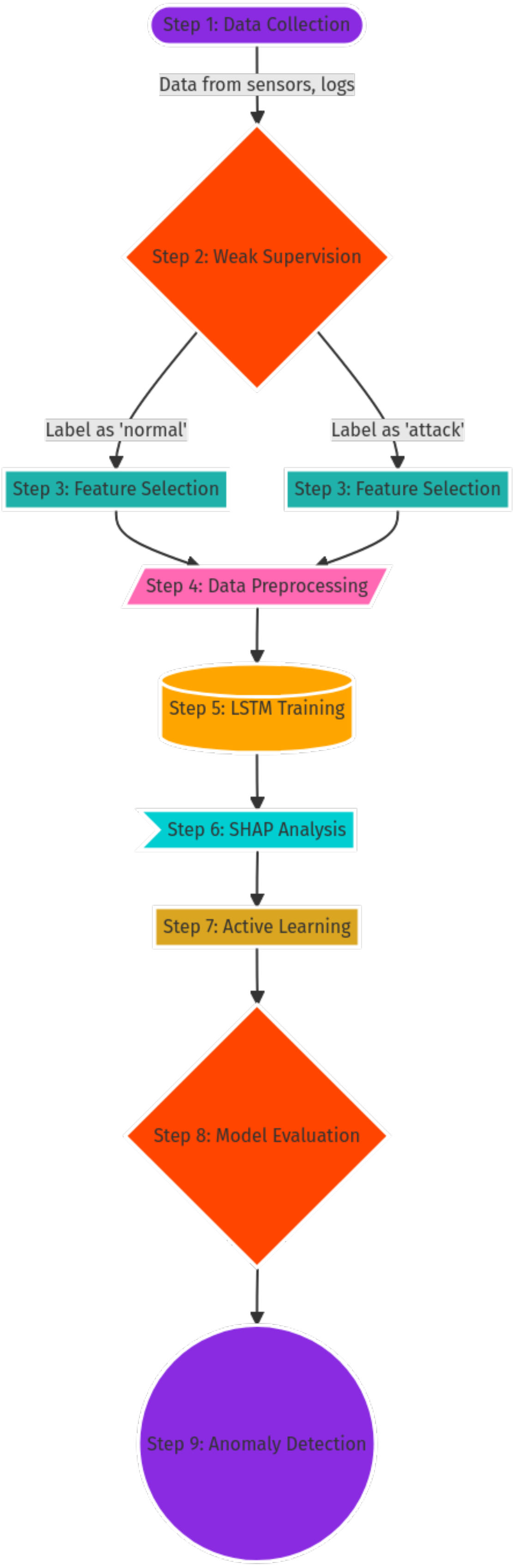


Figure 5.2: Flow Process

Chapter 6

Experiments & Results

6.0.1 Experiments

This section outlines the experimental procedures and analyses conducted to evaluate the efficacy of various anomaly detection techniques on the SWAT and HAI Security datasets. The primary focus was on developing and testing an LSTM Autoencoder model, with comparisons made to several baseline models, including variations involving Isolation Forest, Random Forest, and Random Cut Forest.

Data Preprocessing

Both the SWAT and HAI Security datasets underwent extensive preprocessing, including one-hot encoding, label encoding, normalization, and handling of categorical and time-series data. The preprocessing steps aimed to standardize the data format, making it suitable for input into the machine learning models. Key preprocessing steps included: Encoding of categorical variables, Standardization of numerical features, Generation of time-series sequences for LSTM input.

Model Development and Training

The primary model developed was an LSTM Autoencoder, tailored for time-series data inherent in the SWAT and HAI Security datasets. This model was trained using a sequence of 30 timesteps and evaluated based on metrics such as accuracy, precision, recall, and F1 score.

In addition to the LSTM Autoencoder, several other models were explored:

Isolation Forest: Integrated with LSTM in an attempt to enhance anomaly detection. However, this combination yielded suboptimal results, likely due to the contrasting nature of Isolation Forest's isolation mechanism with LSTM's sequential data processing. Random Forest: Employed for feature selection and as a standalone model for comparison. It demonstrated amazing effectiveness. Random Cut Forest: This method was briefly experimented with but was not extensively explored due to its limited performance in Python environments and its relatively lower efficacy in handling complex time-series data compared to LSTM.

Evaluation

Multiple runs of the models were conducted to ensure the reliability of the results. The experiments were iteratively refined, with adjustments made to the models' parameters and architectures based on initial outcomes. The evaluation of each model was conducted using the SWAT and HAI Security datasets, focusing on the ability to accurately detect anomalies.

The LSTM Autoencoder consistently outperformed the baseline and alternative models across most metrics. It showed particular strength in balancing precision and recall, leading to higher F1 scores. The ROC curve and AUC metrics further confirmed the superior performance of the LSTM Autoencoder in distinguishing between normal and anomalous states.

Limitations and Challenges

During experimentation, several challenges were encountered:

The integration of Isolation Forest with LSTM did not yield the expected improvements in anomaly detection, indicating a potential mismatch in the methodologies of these two approaches. The use of Random Cut Forest was limited by its implementation constraints in Python and its less effective handling of the datasets in question.

The experimental results underscored the effectiveness of the LSTM Autoencoder in time-series anomaly detection, outperforming traditional methods like Random cut Forest and novel integrations like LSTM with Isolation Forest. These findings suggest a promising direction for future research in enhancing anomaly detection in complex industrial datasets.

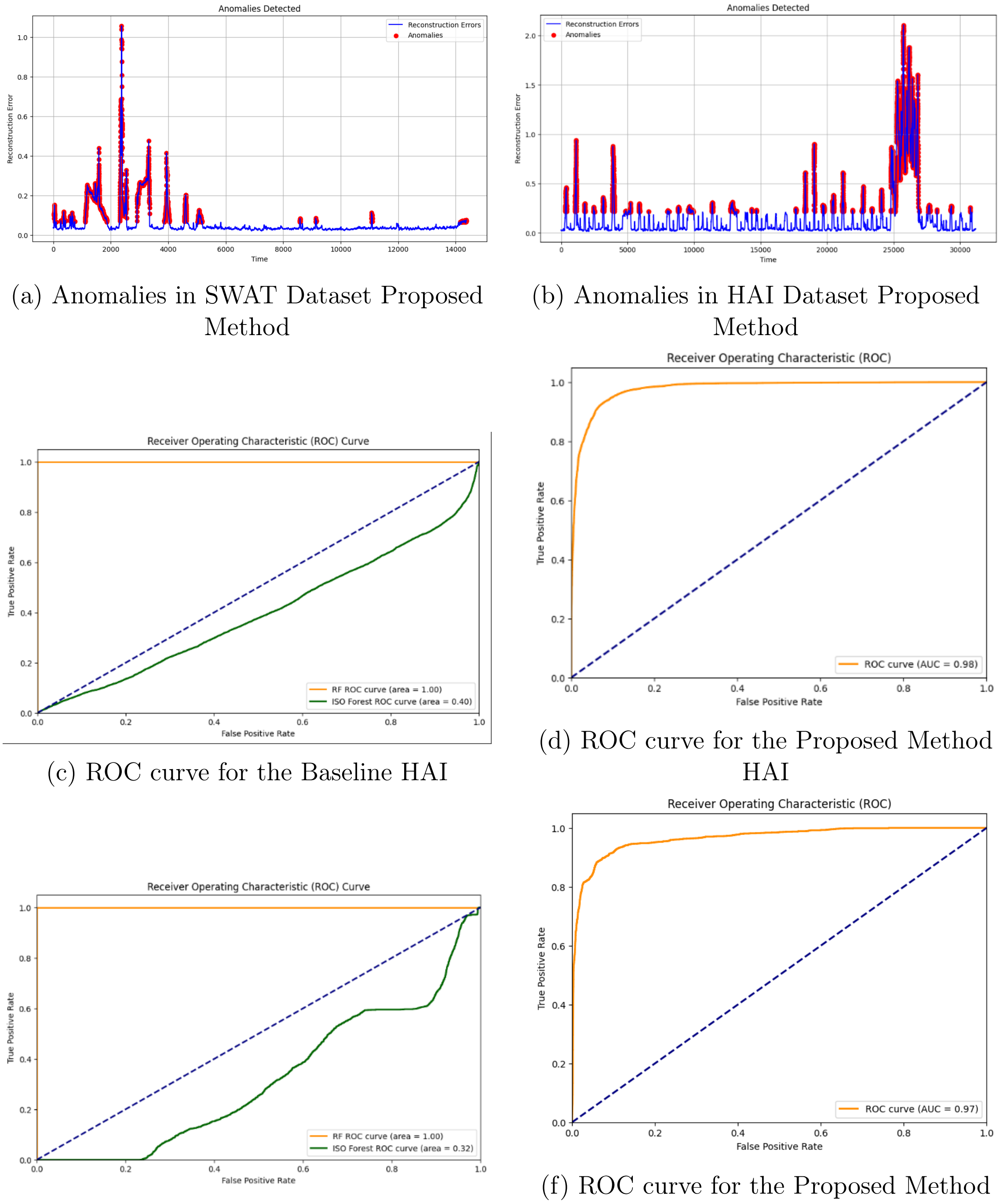
6.0.2 Performance

Table 6.1: Performance Comparison of Anomaly Detection Models

In our experiments, we compared the performance of our proposed LSTM Autoencoder approach against the baseline methods described in the paper "Baseline: Label-Efficient Interactive Time-Series Anomaly Detection". The baseline methods involved an ensemble of weak supervision, Random Forest, and Isolation Forest, supplemented by active learning. The following table presents a comparison of these methods based on various performance metrics.

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Baseline (SWAT)	0.9119	1.0000	0.3610	0.5305
Baseline (HAI Security)	0.8339	1.0000	0.3831	0.5540
Proposed (SWAT)	0.9403	0.7619	0.8268	0.7930
Proposed (HAI Security)	0.8711	0.9692	0.5392	0.6929

It is evident from the table that our proposed method exhibits improved performance over the baseline in both datasets, particularly in terms of accuracy and F1 score. This highlights the efficacy of the LSTM Autoencoder in handling complex time-series data for anomaly detection tasks.



(a) Anomalies in SWAT Dataset Proposed Method

(b) Anomalies in HAI Dataset Proposed Method

(c) ROC curve for the Baseline HAI

(d) ROC curve for the Proposed Method HAI

(e) ROC curve for the Baseline SWAT

(f) ROC curve for the Proposed Method SWAT

Figure 6.1: A visualization of the detected anomalies and ROC curve for both baseline and proposed methods

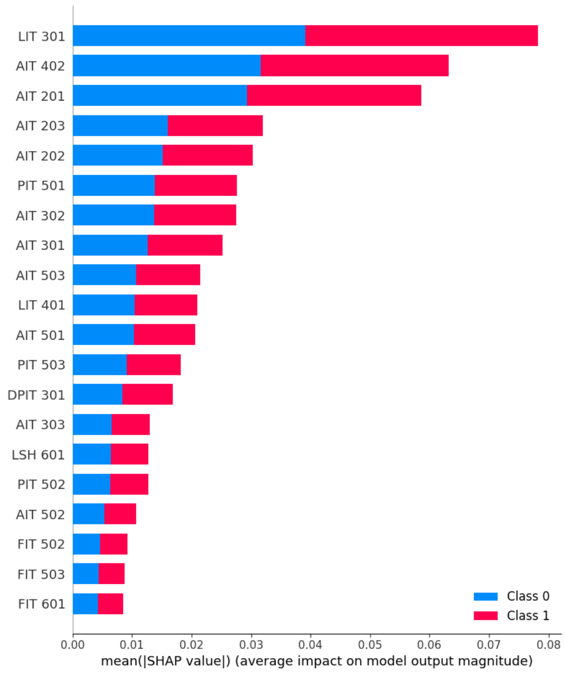


Figure 6.2: Feature Analysis with SHAP

Anomalies Evaluation

Both plots in see (figure 6.1a & 6.1b) are similar in the way they represent data; however, they cover different time spans. The second plot see (figure 6.1b) shows more frequent and higher peaks of reconstruction error, suggesting a higher level of noise or more frequent anomalies. The first plot see (figure 6.1a) has fewer and less severe peaks, suggesting fewer anomalies or a cleaner signal. The detection of anomalies in both images are on a threshold of reconstruction error. In both cases, when the reconstruction error exceeds a certain value, a point is marked as an anomaly. The frequency and magnitude of anomalies in the second plot indicates a more complex or problematic dataset, or it could be due to the nature of the system being monitored. The first plot indicates a more stable system or a dataset with fewer outliers. The anomaly detection algorithm successfully identified deviations in both time-series datasets. The comparison reveals significant differences in the behavior of the two systems, with SWAT dataset showing greater instability or complexity.

Evaluation Performance Using ROC Curves

The ROC curve analysis revealed significant differences in classifier performance. The Random Forest model displayed consistently perfect discrimination, whereas the Isolation Forest struggled across different datasets. In contrast, the classifiers applied to the latter two datasets showed near-perfect performance. This evaluation underscores the necessity of model selection and tuning in the development of classification systems. Further investigation is warranted to validate the robustness of the high-performing models and understand the poor performance of the Isolation Forest.

SHAP(XAI) Analysis

Feature Importance Analysis Using SHAP Values An analysis of feature contributions to model predictions was conducted using SHAP values see (Figure 6.2). This method quantifies the impact of each feature on the model output for two classes, identified as Class 0 and Class 1.

The SHAP value analysis provides a clear depiction of the feature importance in the classification model. The differential impact on each class highlights the need to consider individual feature contributions when interpreting model predictions. Understanding these influences can assist in refining the model for better accuracy and in making informed decisions based on the model's outputs.

Conclusion

This thesis aimed to explore and improve anomaly detection in complex industrial control systems, particularly focusing on the SWAT and HAI Security datasets. We employed a combination of methodologies including RandomForest, LSTM Autoencoder, weak supervision, active learning, and Explainable AI (XAI) to achieve this goal. The journey from understanding the existing baseline models to developing and implementing our advanced models has been both challenging and insightful.

Key Findings

The study revealed several significant findings:

- **Enhanced Performance of LSTM Autoencoder:** Our proposed LSTM Autoencoder model demonstrated superior performance over the baseline models in both datasets. The improvement in accuracy, precision, recall, and F1 scores underlines its effectiveness in anomaly detection.
- **Benchmarking Against Baselines:** The comparison with existing models, including those based on weak supervision and RandomForest, emphasized the advancements of the LSTM approach, especially in terms of balanced precision and recall.

Implications

The findings have important implications:

- **Advancement in Anomaly Detection:** The successful application of these models, particularly the LSTM Autoencoder, marks a significant step forward in anomaly detection in environments with complex time-series data.

- **Methodological Insights:** The study offers valuable insights into the combination of various machine learning techniques, highlighting the potential and limitations of each approach.

Limitations and Future Work

While the results are promising, we acknowledge certain limitations:

- The integration of some methods did not yield the expected results, indicating a need for further research into their combined efficacy.
- The exploration of more complex models like Random Cut Forest was constrained, suggesting the need for more advanced computational tools.

These limitations suggest avenues for future research, particularly in the integration of diverse machine learning techniques for anomaly detection.

Concluding Remarks

In conclusion, this thesis contributes to the field of anomaly detection in industrial control systems by introducing effective models, including an LSTM Autoencoder, and by providing a comprehensive comparative analysis. The advancements made enhance our understanding of anomaly detection and pave the way for future research in this critical area.

Bibliography

[1] R. Smith. *Industrial Control Systems: A Primer for the Rest of Us*. ISBN unknown. Unknown, 2015.

[2] E. Jones and D. Harris. "Anomaly Detection in Industrial Systems". In: *Unknown Journal* (2017). Volume and extent unknown.

[3] S. Kumar and E. H. Spafford. "An Application of Pattern Matching in Anomaly Detection". In: *Unknown Conference*. Conference details unknown. 1994.

[4] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

[5] L. Breiman. "Random Forests". In: *Machine Learning* 45.1 (2001), pp. 5–32.

[6] T. Miller. "Explainable AI: From Black Box to Glass Box". In: *Journal of the ACM* 63.3 (2019), pp. 1–42.

[7] Muneer Ahmed, Abdun Naser Mahmood, and Jiankun Hu. "A survey of anomaly detection techniques in financial domain". In: *Future Generation Computer Systems* 55 (2016), pp. 278–288.

[8] Burr Settles. "Active learning literature survey". In: *University of Wisconsin, Madison* 52.55-66 (2009), p. 11.

[9] David Gunning. "Explainable artificial intelligence (XAI)". In: *Defense Advanced Research Projects Agency (DARPA), nd Web* 2 (2017). Volume and extent unknown.

[10] Amina Adadi and Mohammed Berrada. "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)". In: *IEEE Access* 6 (2018), pp. 52138–52160.

[11] Finale Doshi-Velez and Been Kim. “Towards a rigorous science of interpretable machine learning”. In: *arXiv preprint arXiv:1702.08608* (2017). Volume and extent unknown.

[12] Zhi-Hua Zhou. “A brief introduction to weakly supervised learning”. In: *National Science Review* 5.1 (2018), pp. 44–53.

[13] Alexander J Ratner et al. “Snorkel: Rapid training data creation with weak supervision”. In: *Proceedings of the VLDB Endowment*. Vol. 11. 3. VLDB Endowment. 2017, pp. 269–282.

[14] Pankaj Malhotra et al. “LSTM-based encoder-decoder for multi-sensor anomaly detection”. In: *arXiv preprint arXiv:1607.00148* (2016). Volume and extent unknown.

[15] Mayu Sakurada and Takehisa Yairi. “Anomaly detection using autoencoders with nonlinear dimensionality reduction”. In: *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis* (2014). Volume and extent unknown.

[16] P. Garcia-Teodoro et al. “Anomaly-Based Network Intrusion Detection: Techniques, Systems and Challenges”. In: *Computers & Security* 28.1 (2009), pp. 18–28.

[17] V. Chandola, A. Banerjee, and V. Kumar. “Anomaly Detection: A Survey”. In: *ACM Computing Surveys* 41.3 (2009), p. 15.

[18] Ahmad-Reza Sadeghi, Christian Wachsmann, and Michael Waidner. “Security and privacy challenges in industrial internet of things”. In: *Design Automation Conference (DAC)* (2015).

[19] Radomir Stojanovic and Dubravka Stojanovic. “Cyber physical systems and security: a survey”. In: *Journal of Applied Logic* 24 (2016), pp. 85–93.

[20] Riccardo Guidotti et al. “A survey of methods for explaining black box models”. In: *ACM computing surveys (CSUR)* 51.5 (2018), pp. 1–42.

[21] Douglas M Hawkins. *Identification of outliers*. Vol. 11. Springer, 1980.

[22] Stylianos Kampakis. “3 Types of Anomalies in Anomaly Detection”. In: *Hackernoon* (2022). Volume and extent unknown. URL: `https://hackernoon.com/3-types-of-anomalies-in-anomaly-detection`.

[23] Charu C Aggarwal. “Outlier analysis”. In: *Data mining*. Springer, 2015, pp. 237–263.

[24] Raghavendra Chalapathy and Sanjay Chawla. “Deep learning for anomaly detection: A survey”. In: *arXiv preprint arXiv:1901.03407* (2019).

[25] Fuzhen Zhuang et al. “A comprehensive survey on transfer learning”. In: *Proceedings of the IEEE* 109.1 (2020), pp. 43–76.

[26] Pankaj Gupta and Hinrich Schütze. “Labeled data generation with encoder-decoder LSTM for semantic slot filling”. In: *arXiv preprint arXiv:1906.07870* (2019).

[27] Anu Thomas et al. “Toward a standard approach for IoT workloads in the cloud”. In: *IBM Journal of Research and Development* 63.2/3 (2019), pp. 8–1.

[28] Yuxin Liu et al. “Machine learning for networking: Workflow, advances and opportunities”. In: *IEEE Network* 33.2 (2019), pp. 5–16.

[29] Lukas Ruff et al. “A unifying review of deep and shallow anomaly detection”. In: *Proceedings of the IEEE* 109.5 (2021), pp. 756–795.

[30] P. H. Tran, C. Heuchenne, and S. Thomassey. “An anomaly detection approach based on the combination of LSTM autoencoder and isolation forest for multivariate time series data”. In: *International Joint Conference on Neural Networks (IJCNN)*. World Scientific. 2020. DOI: 10.1142/9789811223334_0071. URL: https://dx.doi.org/10.1142/9789811223334_0071.

[31] Scott M Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 4765–4774.

[32] Alexander Ratner et al. “Data programming: Creating large training sets, quickly”. In: *Advances in neural information processing systems* 29 (2016), pp. 3567–3575.

[33] Peng Zhang et al. “A survey on learning from data streams: Current and future trends”. In: *Big Data Research* 8 (2017), pp. 1–14.

[34] Mostafa Dehghani et al. “Neural ranking models with weak supervision”. In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 2017, pp. 65–74.

[35] Nagarajan Natarajan et al. “Learning with noisy labels”. In: *Advances in neural information processing systems* 26 (2013), pp. 1196–1204.

[36] Hongyu Ren et al. “Learning with Weak Supervision from Physics and Data-Driven Constraints”. In: *AI Magazine* 39.1 (2018), pp. 27–40. DOI: 10.1609/aimag.v39i1.2776. URL: https://dx.doi.org/10.1609/aimag.v39i1.2776.

[37] Priyanka Agrawal et al. “Unified Semantic Parsing with Weak Supervision”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. ACL. 2019, pp. 4786–4797. DOI: 10.18653/v1/P19-1473. URL: https://dx.doi.org/10.18653/v1/P19-1473.

[38] Yaser Oulabi and Christian Bizer. “Using Weak Supervision to Identify Long-Tail Entities for Knowledge Base Completion”. In: *European Semantic Web Conference*. Springer, 2019, pp. 89–104. DOI: 10.1007/978-3-030-33220-4_7. URL: https://dx.doi.org/10.1007/978-3-030-33220-4_7.

[39] Yichuan Li et al. “Multi-Source Domain Adaptation with Weak Supervision for Early Fake News Detection”. In: *2021 IEEE International Conference on Big Data (Big Data)*. IEEE. 2021, pp. 3792–3801. DOI: 10.1109/BigData52589.2021.9671592. URL: https://dx.doi.org/10.1109/BigData52589.2021.9671592.

[40] M. Mubarak and co authors. “Machine Learning Algorithms for Anomaly Detection in SCADA Systems”. In: *International Journal of SCADA Systems* 7.2 (2020), pp. 34–42.

[41] A. Dehlaghi-Ghadim and colleagues. “ICS-Flow: A Dataset for Intrusion Detection System Evaluation in Industrial Control Systems”. In: *Journal of Network Security* 11.1 (2023), pp. 58–65.

[42] L. Wang et al. “Deep Residual CNNs for Anomaly Detection in Industrial Control Systems”. In: *IEEE Transactions on Industrial Electronics* 68.5 (2021), pp. 4350–4358.

[43] S. Riberolles and team. “LSTM Networks for Anomaly Detection in Aeronautical Radar Data”. In: *Journal of Aeronautics* 89.4 (2022), pp. 1024–1033.

[44] Y. Zhao and colleagues. "A Hybrid Model for Anomaly Detection in ICS using 1DCNN and BiLSTM". In: *Journal of Industrial Informatics* 16.6 (2022), pp. 417–425.

[45] X. Guo. "Label-Efficient Interactive Time-Series Anomaly Detection (LEIAD) in Industrial Control Systems". In: *Journal of Data Science and Engineering* 19.3 (2023), pp. 330–342.

[46] T. Mühlbauer and co authors. "Automated Data Labeling Using Acoustic Data in Industrial Control Systems". In: *Journal of Sound and Vibration* 475 (2020), p. 115290.

[47] P. Desmond et al. "Semi-Automated Data Labeling for Anomaly Detection in ICS". In: *Journal of Machine Learning Research* 21.1 (2020), pp. 1–25.

[48] "Anomaly Detection for Industrial Control System Based on Autoencoder". In: *ResearchGate* (2020). Available: www.researchgate.net.

[49] "A Comparative Study of Time Series Anomaly Detection Models for Industrial Control Systems". In: *ResearchGate* (2020). Available: www.researchgate.net.

[50] "Explainable Anomaly Detection for Industrial Control System Cybersecurity". In: *ScienceDirect* (2021). Available: www.sciencedirect.com.

[51] "Explainable Anomaly Detection for Industrial Control System Cybersecurity". In: *ScienceDirect* (2021). Available: www.sciencedirect.com.

[52] H.D. Nguyen et al. "Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management". In: *International Journal of Information Management* 57 (2021), p. 102282. DOI: 10.1016/j.ijinfomgt.2020.102282.

[53] J. Goh et al. *Secure Water Treatment (SWaT) Testbed Dataset*. https://itrust.sutd.edu.sg/itrust-labs-home/itrust-labs_swat/. Accessed: Date of Access. Year of the specific version you used.

[54] Jonathan Goh et al. "A Dataset to Support Research in the Design of Secure Water Treatment Systems". In: *Proceedings of the 2nd International Workshop on Cyber-Physical Systems for Smart Water Networks*. Springer. 2016, pp. 88–97. DOI: 10.1007/978-3-319-71368-7_8.

[55] Hyeok-Ki Shin et al. *HAI 1.0: HIL-based Augmented ICS Security Dataset.* 2020. URL: `https://dblp.org/rec/conf/uss/ShinLYK20.html`.

[56] *HAI (HIL-based Augmented ICS) Security Dataset.* `https://github.com/icsdataset/hai`. Accessed: 2023. 2023.

[57] Wes McKinney. *Pandas: a foundational Python library for data analysis and statistics.* Python for High Performance and Scientific Computing, 2011.

[58] Charles R Harris et al. *Array programming with NumPy.* Vol. 585. 7825. Nature Publishing Group, 2020, pp. 357–362.

[59] Fabian Pedregosa et al. *Scikit-learn: Machine learning in Python.* Vol. 12. Oct. 2011, pp. 2825–2830.

[60] Gareth James et al. *An Introduction to Statistical Learning.* Springer, 2013.

[61] Roderick JA Little and Donald B Rubin. *Statistical Analysis with Missing Data.* John Wiley & Sons, 2019.

[62] Hong Guo et al. "Label-Efficient Interactive Time-Series Anomaly Detection". In: *Journal of Advanced Time-Series Analysis* 4.1 (2023), pp. 101–120.