Doctoral theses at NTNU, 2023:245

Doctoral thesis

Rialda Spahić

Risk-Informed Artificial Intelligence for Autonomous Inspection on Subsea Pipelines

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

Rialda Spahić

Risk-Informed Artificial Intelligence for Autonomous Inspection on Subsea Pipelines

Thesis for the Degree of Philosophiae Doctor

Trondheim, August 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

ISBN 978-82-326-7190-8 (printed ver.)
ISBN 978-82-326-7189-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:245

Printed by NTNU Grafisk senter

Dedicated to my family

”Attention is the beginning of devotion.”

- Mary OLIVER

Abstract

The research in this thesis centers on the image data interpretation capabilities of autonomous underwater systems in the offshore oil and gas industry responsible for visual inspection and monitoring of underwater pipelines and detection of hazards on pipeline surfaces. The main contribution of research is a framework that provides the solutions to overcome the vulnerabilities of artificial intelligence methods during the underwater pipeline inspection by autonomous underwater systems through applied safety engineering.

Increased autonomy in autonomous underwater systems require a greater reliance on artificial intelligence technology for executing pipeline inspection tasks. However, the artificial intelligence for pipeline inspection is limited by several data interpretation challenges. The shortcomings of artificial intelligence for autonomous systems during offshore pipeline hazard inspection can result in catastrophic environmental damage and substantial financial losses for the oil and gas industry. Imbalanced and underrepresented data can cause the artificial intelligence methods, such as machine learning, anomaly detection, and computer vision, to form biases in favor of more represented data with a tendency to reproduce biases learned from data. Underrepresented data can be disregarded as noise during anomaly detection due to their inclination toward efficiency and sacrificing anomalies as tolerable collateral damage. Current methods focus primarily on data content with no regard for the context behind data, yielding conclusions primarily based on correlation and not causation, further causing the occurrence of false alarms during anomaly detection that are a significant drawback during real-time operations. Furthermore, the acute lack of annotated training image data of offshore pipelines and lack of hazard evidence in the data results in the reliance on inexplainable, unsupervised methods. Therefore, one of the main contributions of this research is using the methods for risk and hazard analysis to semi-supervise anomaly detection methods and generate synthetic images of pipeline hazards for extrapolating and annotating the training data. Risk analysis aids anomaly detection in identify the types of anomalies that are recognized as risks, by analyzing low-probability, high-consequence event detection. Furthermore, due to the acute disorganization of categorization and definition of anomalies in current research, this research proposes a redefined anomaly categorization for autonomous underwater systems operations based on hazard behavior and traditional anomaly classification. Finally, this research examines the complex and connected properties of offshore pipeline inspections and offers future directions in rethinking the artificial intelligence methods for pipeline inspection with autonomous underwater systems. The general theme of this thesis lies in risk-informed approaches to address the fundamental challenge of finding early true anomalies and avoiding false alarms when no label exists to inform us of the anomaly or its properties by giving context to anomaly detection methods to comprehend data points not by their labels but by how they relate to one another.

Preface

This thesis is submitted to the Norwegian University of Science and Technology (NTNU) to partially fulfill the requirements for the degree of Doctor of Philosophy. The main work of the Ph.D. thesis was carried out at the Department of Engineering Cybernetics of the Faculty of Information Technology and Electrical Engineering in Trondheim, Norway. The work was accomplished under the supervision of Professor Mary Ann Lundteigen, Professor Vidar Hepsø (Department of Geoscience and Petroleum), and Professor Eric Monteiro (Department of Computer Science).

This thesis is a part of Better Resource Utilization in 21st Century (BRU21) at NTNU, Research and Innovation Program on Digital and Automation Solutions for the Oil and Gas Industry (www.ntnu.edubru21) and supported by Equinor.

This work's target readers include researchers and practitioners interested in the following fields: anomaly detection, machine learning, computer vision, risk analysis, hazard analysis, reliability engineering, image-based subsea pipeline inspection, unmanned autonomous systems, and oil and gas industry engineering. It is assumed that the readers have basic knowledge of machine learning and reliability, preferably related to unmanned autonomous systems.

Trondheim, Norway
May 2023,
Rialda Spahić

Acknowledgements

Starting a PhD, I have heard many times that it is a lonely process. For me, it was everything but lonely, and I have so many people to thank for this incredible and unforgettable adventure.

I would like to express my deepest gratitude to my supervisor Mary Ann Lundteigen for supporting me at every step of this process, and teaching me how to be observant, and understanding the importance of kind discipline. I would like to thank her for all the very late, and very early meetings we had that made it possible for me to walk this journey to its final destination. I could not have been taken this journey without the support of my co-supervisor Vidar Hepsø who thought me to find inspiration in literature and tirelessly provided me the resources I needed from Equinor. I would also like to thank all of the colleagues from Equinor who supported me along the way, and especially to Thor-Andre Aresvik who believed in me and helped me make my dream of doing a PhD come true years before I had the opportunity to join NTNU.

This journey has brought me many travels, from Ireland, Germany, to Brazil and California. I would, therefore, like to express my gratitude to Kameshwar Poolla, my host professor from my research stay at the University of California, Berkeley, for welcoming me and hosting me in the beautiful Bay Area. I would like to express my sincere thanks to BRU21 colleagues, Alexey Pavlov and Eric Monteiro, for encouragement through workshops, events, and gatherings. I have gained close friends while learning about academic writing, therefore, I am thankful to BRU21 for also connecting me to Nataliia Korotkova, and Ewa Laskowska, whose love and kindness will always be a part of this adventure. I believe friends and coffee breaks are an essential part of every PhD, therefore, I am thankful to Ludvig Björklund for many hours of deep conversations and selfless help with Latex.

My PhD journey would not have been possible without my family. I would like to express my deepest gratitude to my father Mensur Spahić, mother Enisa Spahić, and my sister Kasandra Karasalihović, for planting the love for knowledge since my childhood and providing me with endless inspiration and teaching me the freedom of being curious. Thank you for supporting me in dozens of my interests and activities, and for driving me from place to place, just so I can

make all of the hobbies on time. Thank you for your warm patience, support, and for keeping my dreams alive.

Finally, my journey would not have been complete without the support from my husband Tor Arne Nordberg Bergset who has inspired and supported me and cheered for me from the moments of application to PhD, all the way to being my idea machine and my support in every moment, day or night. My husband has made the challenging days of this journey filled with love, kindness and light, and provided me with confidence and strength I needed. Thank you for keeping my spirits up and teaching me that taking time to relax and enjoy, are as equally important as the hours of work. Thank you for continuous ideas, questions, and help that has been invaluable for this journey. I would also like to thank my mother-in-law, Mary-Ann Bergset, for her caring support during tight deadlines in holiday seasons, warmest cups of coffee, and for surprising me by holding onto all of my little notes from brainstorming between holiday meals.

My PhD journey started with a wish to specialise within a specific field, and to learn more. However, looking back at it, it was not only an academic journey, it was a journey of self-discovery. If I were to go back in time, I would do it all over again. I have read once that learning to program teaches you how to think, but I think doing a PhD has thought me to think more reflectively, more openly, and more critically.

I would like to dedicate my PhD thesis to my family and my late grandmother Džemka Musaefendić, who are the reasons for the person I have grown up to become.

With love and gratitude,

Rialda Spahić

Contents

Nomenclature

Acronyms

AD	anomaly detection
AI	artificial intelligence
AUV	autonomous underwater vehicle
CNN	convolutional neural network
CV	computer vision
ML	machine learning
RA	risk analysis
ROV	remotely operated vehicle
SDS	subsea docking station
UAS	underwater autonomous system
UID	underwater intervention drone

Part I

BACKGROUND

Chapter 1

Introduction

A technological explosion and an ever-increasing demand for energy have marked the 21st century. As one of the primary sources of global energy supply, the oil and gas industry transports oil and gas products through vast networks of subsea pipelines. In Norway, the first large pipelines were constructed in the early 1970s, and since then, the Norwegian gas transport system has been expanded to meet the nation's growing energy demands resulting in approximately 8,800 kilometers of pipeline networks [1]. The total length of the pipeline networks is compared to the distance between Oslo in Norway and Bangkok in Thailand [1]. Figure 1.1 shows the intricate network of pipelines that transport gas and oil condensates on the Norwegian continental shelf. Unfortunately, these pipelines are susceptible to environmental factors that can compromise their integrity and further result in the environmental damage through oil and gas product releases and loss of energy supplies to Europe. As a consequence, maintaining and inspecting offshore structures has been one of the industry's greatest challenges. Initially, and for decades after that, operators such as maintenance and inspection engineers, divers, and operators of remotely operated vehicles (ROVs) would travel to remote locations aboard vessels to perform inspection and maintenance and ensure that offshore structures' integrity remains acceptable. However, such a difficult task can result in accidents that harm personnel, the environment, or offshore installations. A continuous and financially sustainable inspection system is a desirable solution for pipeline inspection and monitoring to detect potential sources of harm, also known as hazards [2].

The oil and gas industry has actively developed and tested emerging technologies, from remotely operated to autonomous systems for inspecting remote structures. Remotely operated systems, such as ROVs, eliminate the need to send human operators and divers for inspection tasks. However, they are financially unsustainable, rely on humans to analyze a large amount of incoming data, and do not use the capabilities of artificial intelligence (AI) that the industry strives towards. Autonomous systems, which rely on AI, operate and make decisions independently of human operators and are intended almost entirely to replace human involvement in certain tasks [3], have already demonstrated great potential due to their capacity to collect and analyze vast amounts of data rapidly and in real-time. AI is currently adept at learning patterns and making quick predictions based on learned data but needs help grasping a task's context and understanding it reliably. This challenge is particularly evident in offshore subsea pipeline inspection. The complexity of the subsea environment, the number of properties influencing pipeline degradation, and the limitations of sensors attached to underwater autonomous systems (UAS) are complex

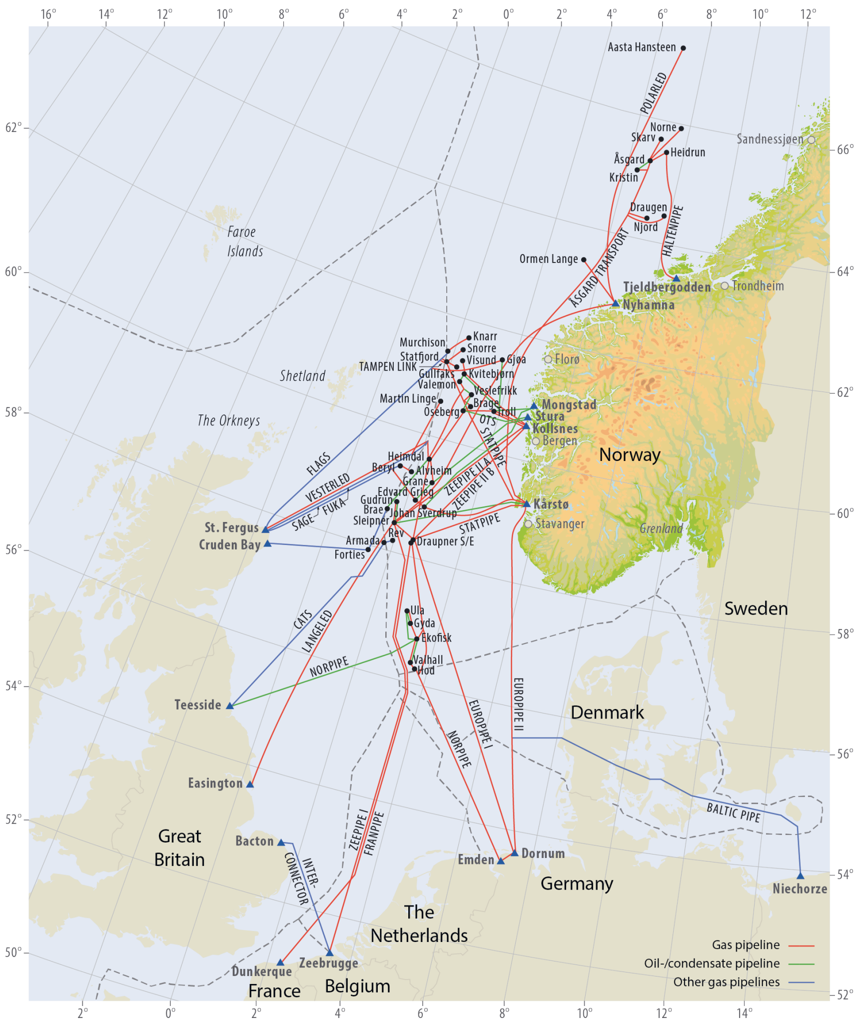

FIGURE 1.1. Pipelines on the Norwegian continental shelf. Image property of Norwegian Petroleum Directorate[1]

networks of challenges that puzzle and intrigue research communities and industries alike.

The Norwegian University of Science and Technology has developed the *Better Resource Utilization in the 21st Century (BRU21)* innovation program on digital and automation solutions for oil and gas in response to the growing need for collaboration between research communities in academia and industry. BRU21 is comprised of multiple research programs, one of which focuses on new business and operational models that investigate the industry's major trends, such as the development and implementation of new technologies across the oil and gas value chain. As a part of *BRU21 New business and operational models* program, this thesis explores and proposes solutions to the challenges of current AI methods employed in UAS for detecting and recognizing hazards, as well as the means by which the context of risk and hazards can be brought to the attention of autonomous systems and increase the potential for safer remote operations.

In this thesis, the terms underwater autonomous system, unmanned autonomous system, autonomous underwater drones or vehicles, are used interchangeably and as research progressed. These terms refer to any type of underwater vehicle (swimming, gliding, or crawling vehicles or drones) that operate autonomously, update and exchange operation data by connecting to a permanent system, i.e, subsea docking station or a designated vessel (described in more detail in Chapter 2).

1.1 Remote, Integrated Operations and Emerging Technologies

Major oil and gas companies that operate globally are pursuing safer and more sustainable future operations enabled by information and communication technology, especially emerging technologies such as highly autonomous systems and sensor networks for monitoring and maintenance tasks [4]. Long-established annual and triennial ship-based monitoring programs for offshore operations include environmental monitoring, sampling of seabed and water sediments, and monitoring of offshore structures [4]. However, technological advancements have enabled flexible, sensor-based, and real-time monitoring of day-to-day operations, such as inspections with ROVs attached to ships (see an ROV launch from ship to the ocean in Figure 1.2).

Enabling advanced technologies necessitates a multidisciplinary approach, requiring contributions from domains as diverse as sensor instrumentation, field development, risk management, and organization development [4]. The incorporation of remote monitoring into day-to-day operations affords the opportunity to act proactively - to take preventative measures and to respond promptly - and, as a result, to achieve safer operations by ensuring early detection of potential incident-causing hazards. In the oil and gas industry, the role of technology and its interaction with people, process, and governance issues has been described through a *capability platform* (illustrated in Figure 1.3)[6].

A capability platform focuses on meeting the capabilities of multiple parties to create economic value through an efficient, adaptable design and network of organizations and individuals who offer complementary goods and services[6]. Figure 1.3 illustrates layers and focus areas of a capability platform:

1. *Business operations:* Addressing the development and execution of work processes and decision support.

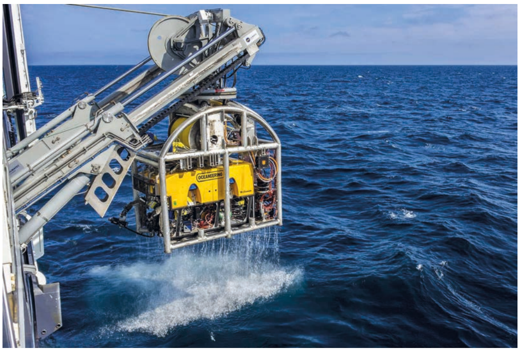

FIGURE 1.2. ROV launch to the ocean. Image property of Oceaneering[5]

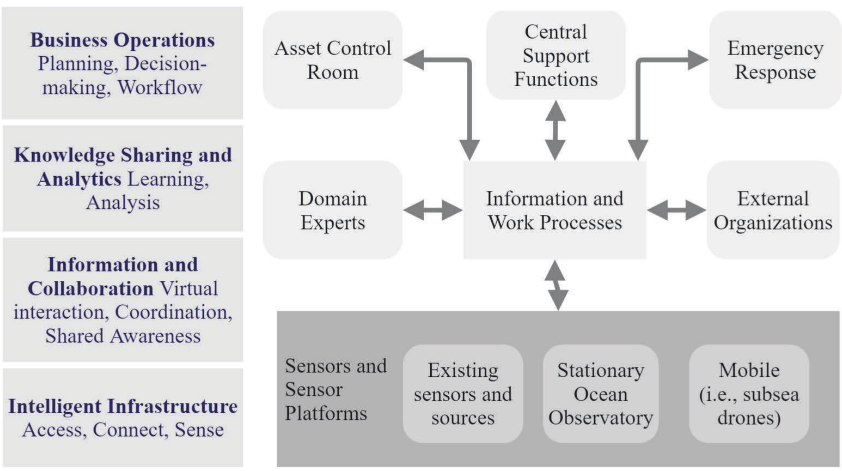

FIGURE 1.3. Capability Platform, adapted from[6]

2. *Knowledge sharing and analysis:* Enabling real-time processing, analysis and information updates for effective operations, including risk management approaches.

3. *Information and collaboration:* Existence of safe and reliable communications infrastructure and safe transfer if sensor data is collected automatically from facilities.

4. *Intelligent infrastructure:* Increase of automatic monitoring through sensing capabilities, including emerging technology such as higher degree of automation and sensing.

The inclusion of real-time monitoring programs for daily operations with a flexible and robust design that is necessary for future systems is provided by sensor technologies and platforms. Sensor technology includes any sensor platforms from existing sources, stationary observatories, to emerging technologies with mobile systems (i.e, ROVs, underwater autonomous drones). A direct implementation of a capability platform can be observed through the implementation of UAS for continuous monitoring and inspection of offshore structures (i.e., pipelines). UAS embraces the focus areas of a capability platform by addressing the business needs and supplementing decision-making through real-time data collection, analysis and information updates, along with risk management capabilities for safe and reliable remote operations.

To support the need for higher automation and sensing technologies, the oil and gas industry is continuously developing autonomous systems for remote monitoring and inspections of offshore structures (see an example of an autonomous drone inspecting a pipeline Figure 1.4). Recent years have marked the immense development for autonomous systems along with the development of permanent or semi-permanent seabed structures that allow the autonomous systems to reside on the seabed and continuously monitor the structures and the environment[7].

As the role of UAS becomes more permanent in industry, it becomes more important to observe and define the relationships and interdependencies these systems form with humans to ensure safe and reliable cooperation. This relationship, known as human-machine teaming, establishes requirements for operating autonomous systems, such as[8]:

- A single task for an autonomous system should provide multiple options for recovery and always allow humans to define an abstraction level for the given task.

- An autonomous system's autonomy should be modified on demand.

- Autonomous systems must be adaptable and tolerant of graceful failure to maintain their own and their surroundings' safety.

Since the autonomous intervention component heavily relies on AI methods, it is essential that these methods comply with safety expectations and regulations and are capable of enabling trustworthy, safe, and reliable operations with autonomous systems such as the UID system. Therefore, the leading motivation of research in this thesis is exploring and addressing the challenges that autonomous systems face due to safety concerns of AI approaches.

1.2 Safety and Regulations for Autonomous Systems and AI

The vast opportunities to enhance safety and environmental sustainability that autonomous systems and AI technologies provide make them highly desirable in the industry and, therefore, develop at a rapid pace. However, as new technologies develop rapidly, they become susceptible to unintended shortcomings and unknown obstacles [9]. These shortcomings require responsible lifecycle planning and design to ensure the reliability and their safe implementation in the industry, and thus ensure the safety of our environment [9]. The European Union Artificial Intelligence Act or EU AI Act [10] is expected to be followed by industries that employ AI in their operations. The EU AI Act broadly divides high-risk AI applications into two categories: physical and software-based applications. Before the AI-based system is put to use, it is expected to be subjected to strict conformity assessments to determine if the system meets requirements of the EU AI Act. Some of the requirements include [10]:

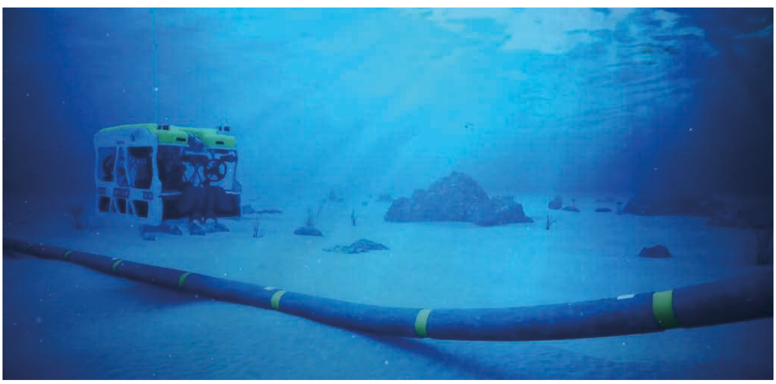

FIGURE 1.4. Autonomous drone inspecting a pipeline. Image is a property of ©Offshore Energy Magazine, 2020.

- Establishment of a risk management system.
- Training, validation, and test data are subjected to adequate data governance and management practices.
- Ensuring the measures that guarantee human supervision of high-risk AI systems.

The International Standard on Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems IEC 61508-1 [11] provides requirements to the design and use of these systems to perform safety functions for conventional technologies, prohibiting the use of AI. However, in the automotive industry, guidelines for AI-based autonomous systems are introduced in SOTIF (ISO/PAS 21448) [12] to ensure safety of autonomous driving. Future functional safety assurance may rely on integrating condition monitoring systems with AI components, such autonomous drones, despite the fact that this is not currently accepted.

Functional safety, as described by ISO/IEC TR5469 on Functional safety for AI [13], is a part of overall safety relating to the Equipment Under Control and the control system that depends on the correct function of the electrical/electronic/programmable electronic safety-related systems and other risk reduction measures.

Safety is defined as freedom from *risk* (ISO/IEC Guide 51:1999, Definition 3.1) [14].*Risk* is further defined as a combination of the probability of occurrence of harm and the severity of that harm (ISO/IEC Guide 51:1999, Definition 3.2), where *hazard* is defined as a potential source of harm (ISO/IEC Guide 51:1999, Definition 3.5) [14].

ISO/IEC TR5469 [13] suggests an architectural pattern for systems using AI technology components, illustrated in Figure 1.5. Since the implementation of AI in safety-related systems can introduce challenges, the architectural pattern proposes three components: AI or machine learning (ML), supervisory component, and a backup decision system. Machine learning, as a type of AI, uses substantial amounts of data to learn patterns and perform tasks, learning more with each increase of data. The AI or ML component is expected to produce diverse results that need to be voted and limited with the help of limiting logic and supervisory component. Supervisory component, that represents the knowledge generated outside of AI component, supervises or limits the AI and is invoked at a failure. Similarly to the backup decision component that is introduced during a failure detection. The architectural pattern for functional safety of AI, Figure 1.5 introduces redundancy in a fault-tolerant system as well as elements that minimally restrict or bound AI operations. Additionally, ISO/IEC TR5469 [13] suggest that the expected behavior of AI systems should be

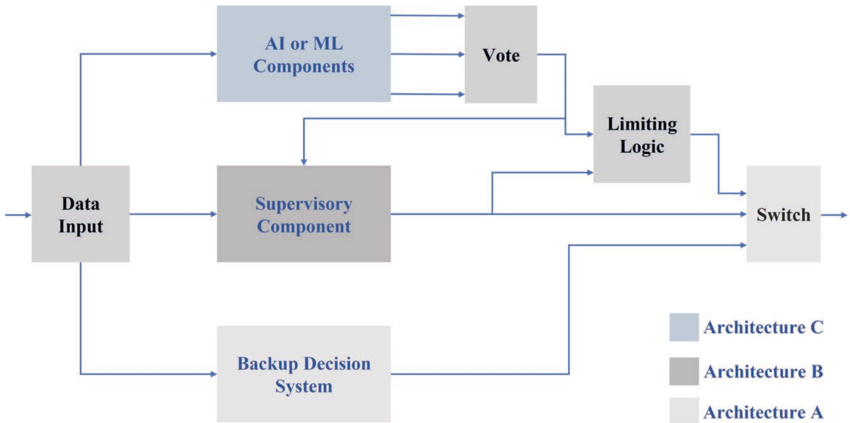

FIGURE 1.5. ISO/IEC TR5469 Architectural patterns for systems using AI, adapted from [13]

possible to evaluate within the distribution of training data. The training data may be also used to potentially predict undesirable behavior of an AI system.

To ensure safety of AI and therefore trust in AI, European AI Strategy [15] suggests AI systems to be:

- *Science-guided* or respectful of scientific knowledge.
- *Uncertainty-aware* or aware of what is unknown.
- *Causal* or understanding of cause and effect of observed situations rather than relying on correlations.

In this way, the European approach to AI is to build trust into AI algorithms and making AI ethical, transparent and unbiased to ensure safety for humanity [15].

1.3 RESEARCH GAPS AND NEEDS

This section briefly outlines and provides an idea of the existing research on the state of subsea pipeline inspection by autonomous systems and the application of AI. Articles relevant to the topics are described in later chapters of this thesis.

The existing research on remote operations and models for subsea pipeline inspection by UAS can be observed from two perspectives: the expectations and challenges of autonomous systems and AI in the industry; and the current role and implications of highly autonomous systems for remote operations subsea pipeline inspection. High autonomy requires extensive use of AI, the study and development of computer systems capable of intelligent behavior and learning from experience, through substantial amounts of data [3].

The application of AI for subsea pipeline inspection presents numerous challenges, including the complexity of algorithms, making it difficult to interpret and validate the results and incorporate risk analysis and specific risk measures. Despite being the integral components for traditional operations and maintenance of offshore structures, risk analysis, and risk factors are still not optimally integrated in data-driven operations [16]. The most common effort to combine risk analysis and data-driven approaches include using probabilistic, Bayesian, and fuzzy logic methods to derive more insight about risk factors from the existing data [17,18]. However, the existing data typically lacks the evidence of hazards, creating imbalanced or biased data and further intensifying the challenge of autonomous systems reliably diagnosing the hazards. Due to the biases in data, autonomous systems run a risk of not learning to detect hazards and failing to report them. Anomaly detection, as a type of AI, is a method for recognizing irregular patterns or occurrences in data, known as anomalies [19]. Anomalies can present undesirable information, such as noise in the environment (i.e., unusual objects at sea floor), or they can present more significant rare occurrence, such as hazards (i.e., damaged pipeline, corrosion). Early detected anomalies can signal warning signs of possible hazards. During hazard detection, it is preferable to detect and identify the hazard warnings as early as possible, while avoiding the undesirable noise that results in reporting false alarms. Image-based hazard detection for subsea pipelines is particularly challenged by the data that is subject to a significant amount of noise, contributing to additional biases and errors during analysis [20,21]. Recent efforts to improve the lack of quality data include collecting data from simulations [22], manual collection of image data with remotely operated vehicles, and manual annotation of data in attempt to create more training data for AI systems [23]. However, the manual approaches are exhausting and expensive. Other efforts at creating better datasets developing methods to color-correct and recover shape from hazy, monochromatic underwater images [24] and methods for automatic annotation of images [25]. In recent years, the field of image manipulation to recover color and shape has prospered and is more commonly used in bio-marine applications. The tools for automatic underwater image annotation are still in their early stages and have not reached maturity for offshore oil and gas structures, still posing challenges in the industry and research society. Consequentially, due to the complex subsea environment and the lack of suitable data, autonomous systems can result in a considerable measure of compromise with regard to the reliability and safety of operations [26,27]. Despite the extensive research on subsea pipeline hazard detection with UAS and the implications of safe AI, the following research gaps need to be addressed for safer remote operations:

RG 1: *Insights from risk and hazard analysis are not sufficiently integrated in AI applications for subsea pipeline hazard detection with UAS.*
The opportunities to integrate risk and hazard analysis as supervisory components for AI and provide supplementary context of risk and hazard to AI in order to detect early warnings of hazards are not sufficiently addressed in existing research.

RG 2: *The existing data for training AI applications for subsea pipeline hazard detection is not adequate enough to reliably represent the complex environment of subsea pipelines for inspection with UAS.*
The lack of adequate training data is a critical challenge in AI applications for subsea pipeline hazard detection. The lack of evidence of hazards in training data creates highly imbalanced training datasets that can contribute to biased results and lead to misdiagnosed or even omitted warning signs of hazards. Furthermore, substantial amount of data is collected during the well-established operations with ROVs that are manually operated with trained personnel. However, it is not sufficiently explored if the autonomously collected data with UAS has the same properties and quality.

RG 3: *Fusion of varied data sources and integration of adaptive sensor technolgies that lead to new opportunities of image-based UAS pipeline hazard detection is not sufficiently discussed.*
The current design of AI applications, such as existing anomaly classification and organization of data sources, do not address the challenges of image-based subsea pipeline hazard detection with UAS and utilize the potential of adaptive sensor technologies.

The identified research gaps serve as a basis for research questions and objectives of this thesis.

1.4 Research Questions

The following research questions are based on the identified research gaps and serve as directions for the objectives of research contributions in this thesis:

RQ 1: How can insights from risk and hazard analysis supervise the results of AI method, such as anomaly detection and classification, and increase the reliability of UAS in detecting early warnings of subsea pipeline hazards?

RQ 2: How can the adequacy of collected data be ensured during autonomous data collection with UAS and how can the training data be enhanced to introduce evidence of hazards necessary for UAS training, while minimizing the manual labor and costs of data collection?

RQ 3: How can the image-based hazard detection with UAS be supplemented by utilizing varied data sources from sensor technologies for adaptive sensing, and how can anomaly classification be reimagined for the future of UAS subsea pipeline hazard detection?

1.5 Objectives and Contributions

This thesis seeks to provide novel and practicable insights into detecting subsea pipeline hazards using UAS and is grounded in a theoretical or conceptual framework and experimental methodology. The main motivation of this thesis is understanding the challenges and blindspots associated with using autonomous systems powered by AI for subsea pipeline inspection and proposes novel approaches for combining traditional risk-based engineering with operation-specific anomaly detection frameworks for detecting and classifying hazards. In addition, this thesis suggests how remote operations may evolve in the future as data, computational capabilities, and autonomous systems advance.

The main objective of this work is to combine traditional risk-based approaches to subsea pipeline detection with anomaly detection and computer vision techniques. The main objective is addressed through contributions C1 - C7, listed below, each addressing research questions as presented in Table 1.1:

C1 Spahic R., Hepsø V., Lundteigen M.A., **Reliable Unmanned Autonomous Systems: Conceptual Framework for Warning Identification during Remote Operations**, *IEEE International Symposium on Systems Engineering (ISSE), September 2021*, DOI: 10.1109/ISSE51541.2021.9582534

C2 Spahic R., Hepsø V., Lundteigen M.A., **Using Risk Analysis for Anomaly Detection for Enhanced Reliability of Unmanned Autonomous Systems**, *Proceedings of the 32nd European Safety and Reliability Conference (ESREL) - Dublin, 2022*, ISBN: 978-981-18-5183-4

C3 Spahic R., Lundteigen M.A., **Manually or Autonomously Operated Drones: Impact on Sensor Data towards Machine Learning**, *IEEE*

International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), 2022, ISBN: 978-1-6654-3445-4

C4 Spahic R., Hepsø V., Lundteigen M.A., **A Novel Warning Identification Framework for Risk-Informed Anomaly Detection**, *Springer Nature Journal of Intelligent and Robotic Systems, 108, 17*. June, 2023. DOI: 10.1007/s10846-023-01887-2

C5 Spahic R., Poolla K., Hepsø V., Lundteigen M.A., **Image-based and risk-informed detection of Subsea Pipeline damage**, *Springer Nature, Discover Artificial Intelligence*. June, 2023. DOI: 10.1007/s44163-023-00069-1

C6 Spahic R., Hepsø V., Lundteigen M.A., **Enhancing Autonomous Systems' Awareness: Conceptual Categorization of Anomalies by Temporal Change During Real-Time Operations**, *The Eighteenth International Conference on Autonomic and Autonomous Systems, 2022*, ISBN: 978-1-61208-966-9

C7 Spahic, R., Lundteigen, M.A., Hepsø, V. **Context-based and image-based subsea pipeline degradation monitoring.** *Springer Nature, Discover Artificial Intelligence 3, 17*. May, 2023. DOI: 10.1007/s44163-023-00063-7

Overall, the scope of this PhD research has not been merely restricted to these main aspects but rather embraces the variety and complexity of encountered challenges and expands to investigate potentials in the future of subsea pipeline inspection through exploration of adaptive sensor systems, marine biology and geographical properties of subsea pipeline environment.

Experiments were conducted to the extent that data and computational power were available. Due to the complexity of the topic and the novelty of challenges, some insights were brought to light through conceptual framework proposals and incorporated into future operating models. The main objective of this thesis is to find feasible ways to address challenges in hazard detection with approaches based on AI and to comply with expectations of explainable, reliable, and safe autonomous operations.

Table 1.1. Research questions addressed by the contributions C1 - C7

Research Question	**Contribution description**	**Contribution**
RQ 1	Mapping the overlapping tasks between risk and hazard analysis to AI methods for anomaly and hazard detection into a conceptual Warning Identification Framework.	C1
RQ 1	Expanding the Warning Identification Framework by introducing hazard analysis as a supervisory component to anomaly detection applications to a traditional data analysis lifecycle.	C2
RQ 2	Analyzing data collected with manually operated drone and comparing it to the analysis of data collected by an autonomous drone under equal circumstances to look for discrepancies between the two modes of drone operation.	C3
RQ 2	Analysis of sensor-collected seismic and seismic tremor data: Supervising anomalies detected with conventional anomaly detection with hazards identified through hazard analysis by domain experts to detect warning signs and eliminate false alarm anomalies. This contribution is and expansion and implementation of the Warning Identification Framework introduced in C2.	C4
RQ 2	Analysis of industry-provided subsea pipeline images: Expanding heavily imbalanced training data with new, synthetic image data of mechanical damage on pipeline that is generated through image manipulation methods, and applying localised anomaly detection to increase explainability of AI methods during mechanical damage detection on subsea pipelines.	C5
RQ 3	Proposal of novel categorization for anomaly classification for the future of UAS subsea pipeline hazard detection.	C6
RQ 3	Proposal of data source fusion and adaptive scheduling to supplement image-based inspections and introducing risk-informed architecture for the future of AI-based subsea pipeline hazard detection with UAS.	C7

1.6 Outline

This thesis consists of eight chapters, Appendix A and Dissemination of research. The contents of the thesis are briefly described as follows:

Chapter 2 describes the function of subsea pipelines, the methods and systems used to inspect them, and the external anomalies that can be visually inspected using autonomous systems.

Chapter 3 examines the experimental methods that are the focus of this thesis, including anomaly detection, classification, computer vision methods, and their blindspots and challenges.

Chapter 4 is based on the contributions C1 and C2 and focuses on theoretical concepts for risk-informed and data-driven UAS operations.

Chapter 5 is based on the contributions C3 and C4 and focuses on reliability of sensor data for machine learning through experimental work

Chapter 6 is based on the contribution C5 and focuses on the visual inspection of subsea pipeline anomalies through analysis of pipeline images provided by the industry:

Chapter 7 is based on the contributions C6 and C7 and explores the future models in subsea pipeline inspection with UAS by exploring opportunities and focus areas for the future of remote operations:

Chapter 8 concludes the research, discusses challenges and learned lessons, and examines future work and possibilities.

Appendix A is supplementary material to Chapter 5, describing attributes used in experimental work for detecting true hazards by analyzing seismic sensor data with anomaly detection.

1.7 References

[1] Ministry of Petroleum and Energy and the Norwegian Petroleum Directorate. The oil and gas pipeline system (1 2023). URL https://www.norskpetroleum.no/en/production-and-exports/the-oil-and-gas-pipeline-system/. Cited on page/s 3, 4.

[2] ISO 31000. Risk management — Guidelines, International Organization for Standardization. Technical report International Organization for Standardization (2018). URL https://www.iso.org/obp/ui/iso:std:iso:31000:ed-2:v1:en. Cited on page/s 3.

[3] Oxford University Press. Oxford Learner's Dictionaries (2021). URL https://www.oxfordlearnersdictionaries.com/. Cited on page/s 3, 11.

[4] Vidar Hepsø, Mona Låte, Geir Gramvik, Ståle Johnsen, Ingunn Nilssen, and Harald Wesenberg. Integrated Environmental Monitoring in Daily Operations. In *SPE Intelligent Energy International* Utrecht, The Netherlands (3 2012). Cited on page/s 5.

[5] Oceaneering. ROV Systems and Services (2023). URL https://www.oceaneering.com/rov-services/rov-systems/. Cited on page/s 6.

[6] John Henderson, Vidar Hepsø, and Øyvind Mydland. What is a Capability Platform Approach to Integrated Operations? In *Integrated Operations in the Oil and Gas Industry: Sustainability and Capability Development* pages 1–19. (2013). doi: 10.4018/978-1-4666-2002-5.ch001. URL http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-2002-5.ch001. Cited on page/s 6.

[7] Daniel Abicht, Jan Christian Torvestad, Pål Atle Solheimsnes, and Karl Atle Stenevik. Underwater intervention drone subsea control system. *Proceedings of the Annual Offshore Technology Conference* **2020-May** (May), 4–7 (2020). ISSN 01603663. doi: 10.4043/30701-ms. Cited on page/s 7.

[8] The MITRE Corporation. A Framework for Discussing Trust in Increasingly Autonomous Systems. *Company White Paper* (June), 2013–2014 (2017). Cited on page/s 7.

[9] IEEE. IEEE 1872.2-2021 Standard for Autonomous Robotics (AuR) Ontology. Technical report Institute of Electrical and Electronics Engineers (IEEE) (2022). Cited on page/s 8.

[10] European Commision European Parliament and The Council. Regulation of the European Parliament and of the Council - Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts 2021/0106/COD. Technical report European Commision Brussels (4 2021). URL https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN. Cited on page/s 8.

[11] IEC61508-1. IEC 61508-1 Functional safety of electrical/electronic/programmable electronic safety-related systems. Technical report INTERNATIONAL ELECTROTECHNICAL COMMISSION (IEC) (4 2010). Cited on page/s 9.

[12] ISO. ISO 21448:2022 Road vehicles — Safety of the intended functionality. Technical report (6 2022). URL https://www.iso.org/obp/ui/#iso:std:iso:21448:ed-1:v1:en. Cited on page/s 9.

[13] ISO/IEC. ISO/IEC TR5469:202x(E) Artificial Intelligence - Functional safety and AI systems. Technical report International Electrotechnical Comission (2022). URL https://www.iso.org/standard/81283.html. Cited on page/s 9, 10.

[14] International Electrotechnical Commission. IEC 903-01: Safety and risk reduction (). URL https://www.electropedia.org/iev/iev.nsf/index?openform&part=903. Cited on page/s 9.

[15] European Commission. A European approach to artificial intelligence (). URL https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence. Cited on page/s 10.

[16] Jung Kwan Seo, Yushi Cui, Mohd Hairil Mohd, Yeon Chul Ha, Bong Ju Kim, and Jeom Kee Paik. A risk-based inspection planning method for corroded subsea pipelines. *Ocean Engineering* **109**, 539–552 (11 2015). ISSN 00298018. doi: 10.1016/j.oceaneng.2015.07.066. Cited on page/s 11.

[17] Xinhong Li, Guoming Chen, Yuanjiang Chang, and Changhang Xu. Risk-based operation safety analysis during maintenance activities of subsea pipelines. *Process Safety and Environmental Protection* **122**, 247–262 (2 2019). ISSN 09575820. doi: 10.1016/j.psep.2018.12.006. Cited on page/s 11.

[18] Mohamed El Amine Ben Seghier, Zahiraniza Mustaffa, and Tarek Zayed. Reliability assessment of subsea pipelines under the effect of spanning load and corrosion degradation. *Journal of Natural Gas Science and Engineering* **102** (April), 104569 (2022). ISSN 18755100. doi: 10.1016/j.jngse.2022.104569. URL https://doi.org/10.1016/j.jngse.2022.104569.

Cited on page/s 11.

[19] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A Survey. *ACM Computing Surveys (CSUR)* **14** (1), 1–22 (7 2009). ISSN 15462218. doi: 10.1145/1541880.1541882. Cited on page/s 11.

[20] Michael Ho, Sami El-Borgi, Devendra Patil, and Gangbing Song. Inspection and monitoring systems subsea pipelines: A review paper. *Structural Health Monitoring* **19** (2), 606–645 (2020). ISSN 17413168. doi: 10.1177/1475921719837718. Cited on page/s 11.

[21] Hongwei Zhu, Weikang Xie, Junjie Li, Jihao Shi, Mingfu Fu, Xiaoyuan Qian, He Zhang, Kaikai Wang, and Guoming Chen. Advanced Computer Vision-Based Subsea Gas Leaks Monitoring: A Comparison of Two Approaches. *Sensors* **23** (5), 2566 (2 2023). ISSN 1424-8220. doi: 10.3390/s23052566. URL https://www.mdpi.com/1424-8220/23/5/2566. Cited on page/s 11.

[22] Simen Eldevik and Frank Borre Pedersen. AI + safety - DNV. Technical report DNV (2018). URL https://www.dnv.com/oilgas/download/artificial-intelligence-ai-and-safety.html. Cited on page/s 11.

[23] Anastasios Stamoulakatos, *et al.* Automatic annotation of subsea pipelines using deep learning. *Sensors (Switzerland)* **20** (3) (2 2020). ISSN 14248220. doi: 10.3390/s20030674. Cited on page/s 11.

[24] Amjad Khan, Syed Saad Azhar Ali, Atif Anwer, Syed Hasan Adil, and Fabrice Meriaudeau. Subsea pipeline corrosion estimation by restoring and enhancing degraded underwater images. *IEEE Access* **6** (July), 40585–40601 (2018). ISSN 21693536. doi: 10.1109/ACCESS.2018.2855725. Cited on page/s 11.

[25] W. T. Nash, C. J. Powell, T. Drummond, and N. Birbilis. Automated corrosion detection using crowdsourced training for deep learning. *Corrosion* **76** (2), 135–141 (2020). ISSN 00109312. doi: 10.5006/3397. Cited on page/s 11.

[26] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1** (5), 206–215 (2019). ISSN 25225839. doi: 10.1038/s42256-019-0048-x. URL http://dx.doi.org/10.1038/s42256-019-0048-x. Cited on page/s 11.

[27] Zhengquan Wang, Yantao Li, Jie Ren, Weichen Xu, and Lihui Yang. Investigating the effects of environment, corrosion degree, and distribution of corrosive microbial communities on service-life of refined oil pipelines. *Environmental Science and Pollution Research* **29** (34), 52204–52219 (2022). ISSN 16147499. doi: 10.1007/s11356-022-19556-6. URL https://doi.org/10.1007/s11356-022-19556-6. Cited on page/s 11.

Chapter 2

Subsea Pipelines and Inspection

The oil and gas industry produces billions of barrels of hydrocarbon resources, such as oil and gas, to meet more than half of the increasing global demand for energy [1]. Extracting hydrocarbons from offshore reservoirs and transporting them necessitates various interdependent systems. One of the most important structures for the long-distance transport of oil and gas are the vast and intricate networks of subsea pipelines. Due to their widespread installation and harsh subsea environment, subsea pipelines are susceptible to natural (i.e., weather and environment conditions, material ageing) and artificial damages (i.e., human error and equipment failure). Therefore it is crucial that the subsea pipelines are environmentally and economically sustainable.

2.1 Subsea Pipeline Degradation and Failures

Monitoring the pipeline to detect potential degradation and fault states is one of the most critical tasks of UAS in pipeline inspection, illustrated in Figure Figure 2.1. Degradation is defined as an undesired deviation in the operational performance of any device, equipment or system from its intended performance that may be caused by internal processes or effects of the environment (IEV 161-01-19) [2]. The UAS are expected to detect failure, such as pipeline damage. Failure is defined as the loss of the capacity to function as required, or an event that results in a fault state (IEV 192-04-01) [2]. A fault state is further defined to be the inability to perform as required, due to an internal state (IEV 192-04-01) [2]. Establishing the degradation and failure leads to repair and restoration functions, towards normal state, or partial repair and restoration from fault to degraded state. An instance of restoration is when the state is restored after a failure (IEC 192-06-23), and a repair instance is an action taken to effect restoration, including localization, diagnosis, and correction of the fault (IEC 192-06-14) [2].

A report from Det Norske Veritas (DNV), who specialize in assurance and risk management for environmental safety shows the failure frequency rates of subsea pipelines, presented in Table 2.1. In their report [3], DNV describe

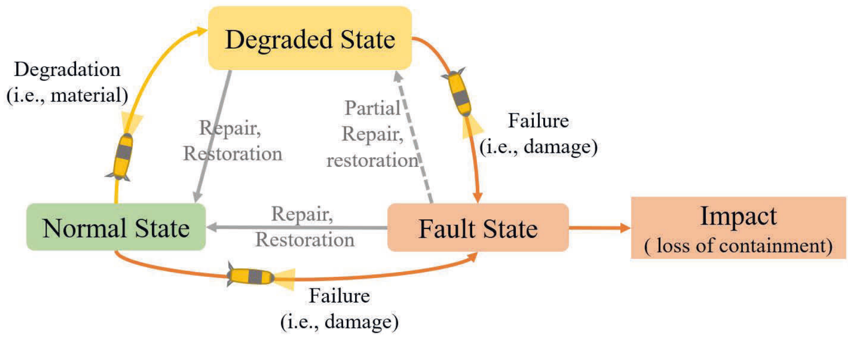

FIGURE 2.1. Role of UAS for Pipeline Inspection

failure as a subset of an incident resulting in loss of containment and leakage. Failure and leak frequency (and loss of containment frequency) are considered equivalent. Pipeline failure frequency is expressed as the number of failures per km pipeline and year.

Bureau of Safety and Environmental Enforcement (BSEE) classifies offshore pipeline failures into five categories: equipment failure, external forces, such as human error, corrosion, weather or natural causes, and vessel, anchor, or trawl damage[4]. Even the smaller surface damages such as local metal loss in forms of abrasions or pipe wall abrasions resulting in dent defects, may accompany the pipe surface damage, further creating weak points on the pipe surface for ruptures and potential leakage[5]. Figure 2.2 illustrates the damage

TABLE 2.1. Offshore Pipeline Failures[3]

Description	Failure frequency	Unit
Well stream pipelines and other pipelines containing unprocessed fluid	Between 2.3 x $^{-3}$ and 4.8 x 10^{-4}	Per km year
Flexible pipelines	2.1 x 10^{-3}	Per km year
Failure frequency from inadvertent dragging of anchors by ships under way	Pipe specific	
Processed oil, gas with pipeline diameter > 24"	Between 1.4 x 10^{-4} and 5.4 x 10^{-6}	Per km year
Processed oil, gas with pipeline diameter ≤ 24"	Between 7.1 x 10^{-5} and 1.7 x 10^{-5}	Per km year

distribution by causes to different levels of oil spills from subsea pipelines. According to the Figure 2.2, smaller scale pipeline failure and ruptures that result in leaks are often caused by corrosion. Pipeline failure analysis and damage monitoring are challenging tasks due to the complex nature in which the damages and failures can occur. The pipeline is built during the manufacturing phase, where defects can already occur at raw material transformation and production, creating weak points in the pipeline structure, particularly where welding occurs [1]. Over time, the pipeline faces new potential threats on the ocean floor. Debris from the environment, such as large marine animals and their carcasses or invasive microbial species, can impact the pipeline over extended periods. The pipeline can also be impacted by other offshore equipment that falls and drags at the seabed, such as ship anchors, drilling installations, or fishing equipment. Ocean currents can wash away the soil beneath the pipeline, causing a segment of the pipeline to become unsupported and unstable, known as free spanning [1]. Chemical reaction and abrasion from the internal fluids, which are frequently acidic and contain sand particles moving at high speeds, can cause corrosion and erosion and develop into external damage [1]. All of these factors necessitate routine pipeline inspections and the implementation of a permanent monitoring system. Continuously monitoring pipelines for potential damage is one method for minimizing pipeline failure.

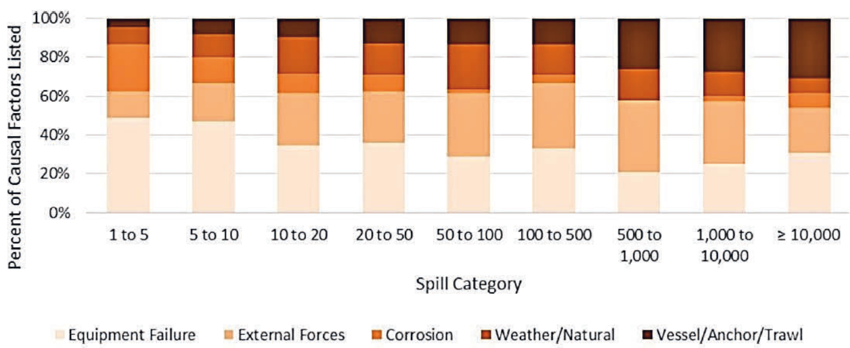

FIGURE 2.2. Distribution of damage causes to different levels of oil spills from subsea pipelines (Spill size in Barrels (bbl)). Adapted from [4,6]

2.2 UNDERWATER AUTONOMOUS SYSTEMS AND INSPECTION METHODS

Pipeline damage and material degradation can occur internally and externally; therefore, numerous inspection and sensing systems have been developed to inspect the internal and external integrity of the pipeline. Unless the sensors have been implemented during pipeline installation, divers or, when personnel safety is a concern, remotely operated vehicles (ROVs) or autonomous underwater vehicles (AUVs) can deliver sensors to subsea pipelines [1]. ROVs are manually operated by a trained operator, and often connected to a nearby surface vessel. AUVs are not manually operated by an operator, rely on pre-programmed instructions and AI-based methods. More recently, subsea docking stations allow AUVs to reside at offshore locations, charge batteries and transmit data. Other systems include intelligent pigs and crawling robots that also introduce pipeline sensors for internal inspection.

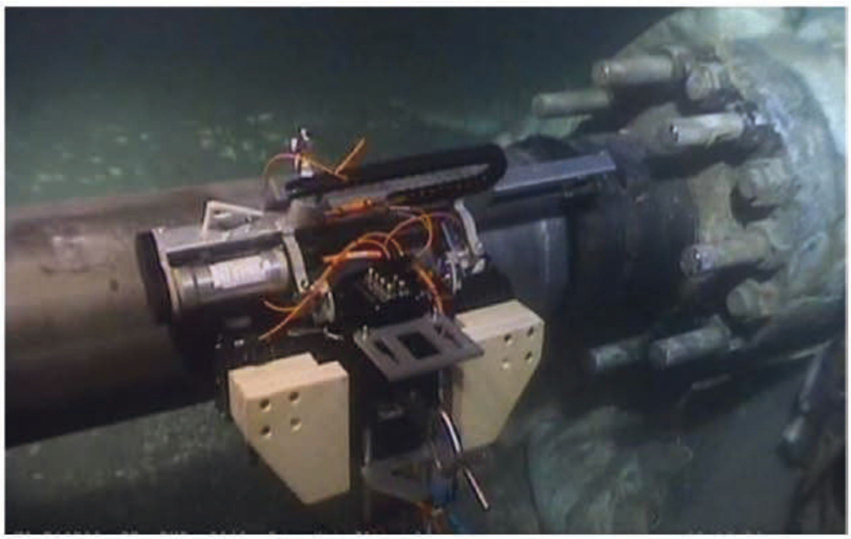

FIGURE 2.3. ROV installing a weld inspection sensor [1] © Sonomatic

Robotic vehicles, such as ROVs and AUVs, have demonstrated outstanding potential in the offshore industry for various applications. Due to their tested and proven technology, ROVs already have a permanent place in remote offshore infrastructure inspection, sensor delivery, exploration of the subsea environment, and even rescue missions. Most offshore pipeline inspections with ROVs involve a submersible component powered and controlled by a tether cable [7]. Despite some disadvantages that ROVs have, expensive oper-

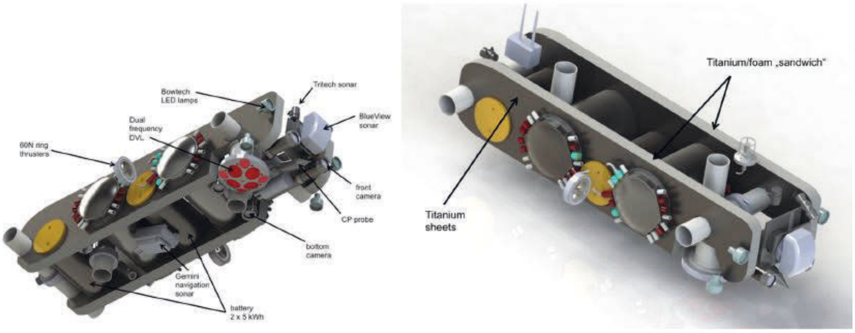

FIGURE 2.4. Flatfish AUV for long-term underwater operations, adapted from Albiez *et al.*[8]

ations that involve sending vessels offshore for tether cable connection and extensive personnel training to operate them, the ROVs have become one of the most applied subsea inspection methods. Figure 2.3 shows an example of an ROV installing a sensor for weld inspection of subsea pipeline.

The offshore oil and gas industry recognizes the immense potential of semi-autonomous and, particularly, fully underwater autonomous systems (UAS) capable of tetherless communication, such as AUVs or autonomous underwater drones, swarms of drones, gliders, and autonomous benthic landers (observational platforms that sit on the seabed). UAS relies on AI methods for inspection tasks, whereas a ROV relies on the training of the operator who

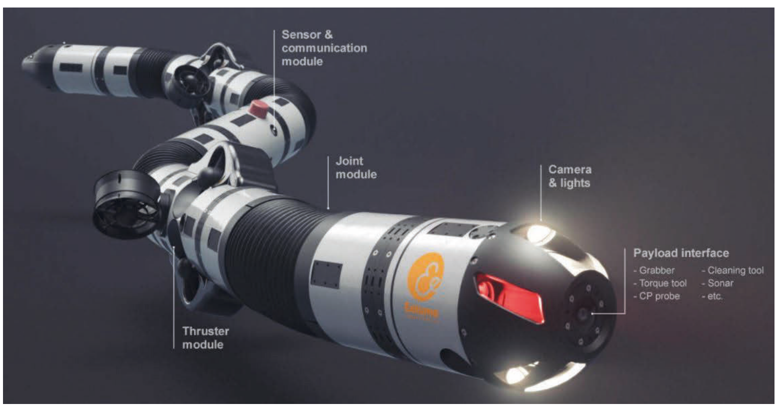

FIGURE 2.5. Eelume Subsea Intervention Drone. Image property of ©Eelume Subsea Intervention Eelume AS[9]

manually operates the ROV and inspects the pipelines. Figure 2.4 shows a subsea resident AUV, Flatfish AUV, for long-term underwater operations, equipped with sonars, cameras, and probes. Figure 2.5 shows a subsea intervention vehicle Eelume designed to carry out subsea inspection, maintenance and repair with autonomous robotic arms and flexible body [9].

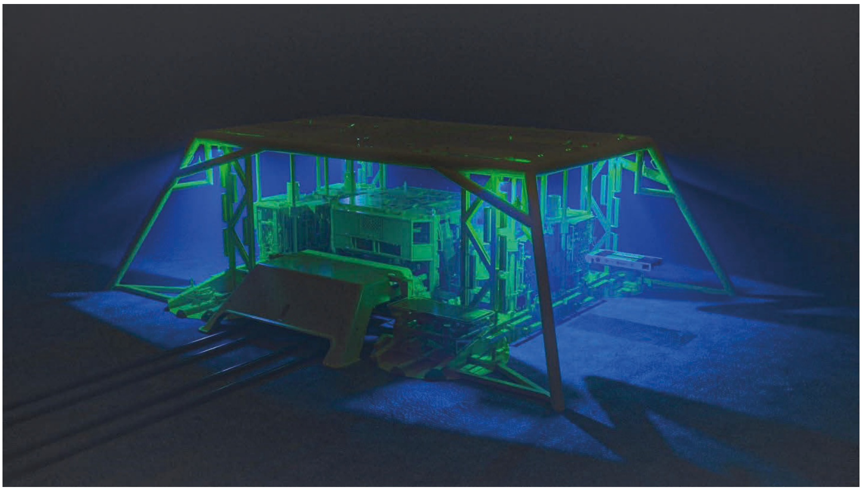

FIGURE 2.6. Equinor's UID System. Image is a property of © EnergyVoice

The development of subsea docking stations (SDS) has made it possible for UAS to remain submerged for months, recharge their batteries, and transmit and receive data. Over the recent years, for autonomous offshore operations, Equinor has been designing and developing an Underwater Intervention Drone (UID) subsea control system to combine the advantages of SDS and underwater vehicles [10]. Figure 2.6 shows an example of the UID system developed by Equinor. Figure 2.7 shows the technology building blocks of the UID system that consists of the SDS and the drone. The SDS consists of the components such as inductive connectors, electronic modules, and interface component that consist of facility-specific and drone-specific interfaces. The drone component needs to be capable of docking and undocking to the SDS, autonomous maneuvering (i.e., tracking the pipeline), and autonomous intervention. Additionally, drone is expected to provide pilot supervised intervention capabilities through tethered communication, free-space optical, or acoustic communication [10]. The highlighted component, autonomous intervention, provides the capacity for greater autonomy, typically through artificial intelligence. It is anticipated that the drone will be able to autonomously detect sources of harm or hazards on facilities that require attention, such as pipeline damage, leaks, or corrosion [10].

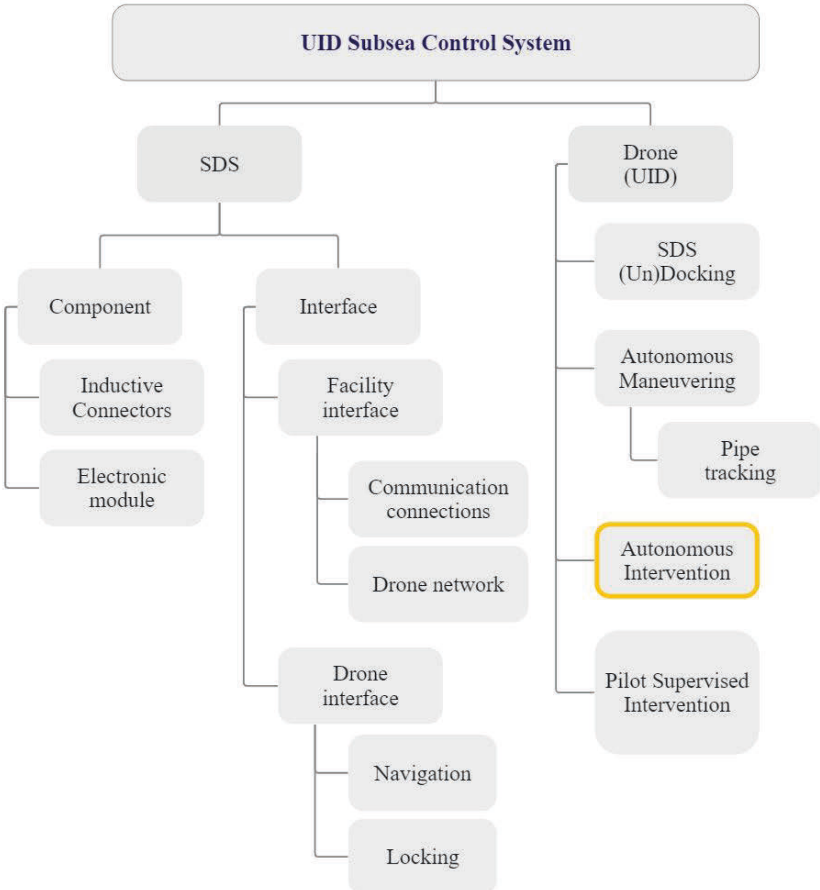

Figure 2.7. UID System Building Blocks, adapted from[10]

There are multiple levels of autonomy for UAS that perform inspection tasks, which are categorized according to industry expectations[11] and include data deliberation and risk management capabilities that allow the system to perform decision-making during inspection:

Level 1 - During inspection tasks, UAS warns human operator if it approaches it's operation design limits, and systematizes incoming sensor data, providing the operator with task-relevant information.

Level 2 - UAS is capable of decision-making during inspection tasks by assessing data from all connected devices, continuously assesses risk and adjusts warnings.

Level 3 - UAS is capable of complex decision-making during inspection tasks by assessing historic data, comparing it to ongoing observations, assessing risks and providing mitigative options.

Level 4 - UAS reasons on distributed data, conducts high-level planning, and provides the best risk management policy.

UAS can continuously gather the data about the environment and ongoing operation, and perform near-real-time and real-time analysis with the help of sophisticated data analysis methods, particularly artificial intelligence (AI). The intention is to replicate human-level capabilities of assessing the situation, providing insight, and making decisions. AI consists of multiple branches, such as, among others, machine learning (ML) and deep learning that utilize large amounts of sensor-collected data to learn and predict patterns; computer vision that interprets image and video inputs; fuzzy logic that imitates human-interpretive logical decisions. Although AI can potentially replace humans in complex tasks, perform high-level functions, and increase the safety of remote operations, AI continues to suffer from immature technology and unproven reliability for safety-critical operations. Most of the challenges with AI stem from inadequate data used to train the AI systems in learning correct, unbiased patterns and conducting reliable, trustworthy decisions that are on par with human operators. More on AI methods and challenges in remote offshore operations is described in Chapter 3 Experimental methods and Artificial Intelligence.

Multiple phenomena, including electromagnetic, acoustic, and radio-graphic, can be observed in a non-destructive manner (without hurting or damaging the pipeline during testing) to determine the pipeline's condition. An electromagnetic method, Magnetic Flux Leakage (MFL), which detects flaws in the material's surface and potential leaks, is one of the most commonly employed tools for pipeline inspection [12]. Other commonly used electromagnetic sensors include variants of MFL, such as the Hall Effect, Electrical/field signature mapping, Eddy current inspection, and their variants, whose employment focuses on testing pipeline networks for local corrosion, cracking, and erosion [1]. Although efficient, most of these methods lack speed of detection and precision, and area coverage depends on the sensor type. Acoustic tools, such as ultrasound inspection, Guided Wave Testing, Acoustic Emission, and Sonar Mapping, depend on the use of elastic waves to detect small cracking. Although sonar inspection can also efficiently detect foreign objects dragging on the seabed, it relies on flat surfaces and may incorrectly assume the seabed is flat, potentially leading to misguided results [13]. In general, acoustic sensors are often limited in range and speed in detecting damages [14]. Another promising way of sensing the pipeline integrity and detecting damages is through Fiber Optic Sensors (FOS). FOS are filaments of silica glass or plastic that transmit light via total internal reflection [15]. The benefit of FOS is that they are lightweight and can be permanently installed along long pipeline distances for distributed real-time sensing. They provide accurate sensing, but the FOS technology is currently immature and the implementation of it is exceptionally costly.

The subsea environment is extraordinarily complex and ever changing, and subsea pipeline networks can suffer from various interconnected challenges. Maintenance requirements of subsea pipelines, including pipeline inspection and monitoring, can be a challenging task and requires the use of multiple tools or synergy of sensors to achieve the most reliable insights into subsea pipeline integrity.

2.3 External Anomaly and Risk-Based Analysis

Reliability engineering and risk assessment are essential components of design, maintenance, and inspection for subsea infrastructure. *Reliability engineering* is described as an engineering application that ensures that a component, product, or process performs as intended without failure for a specified period in a given environment[16–18]. To ensure the reliability of UAS pipeline inspection, it is necessary to address issues with AI-based methods, such as the absence of data-based hazard evidence. In order to identify potential hazards and risks that may be encountered during pipeline inspection, risk assessment is an essential component of reliable UAS. *Risk assessment* includes a structured analysis and identification of potential hazards, their causes, and consequences, as well as a description of the risk and representation of uncertainties[19–21]. It is essential to deconstruct the definitions of risk and hazard in this context, to understand what properties are described during their analysis. *Risk* is the effect of uncertainty on objectives, which can be positive, negative, or both, and results in either threats or opportunities[20]. Risk is often expressed in terms of sources, sequences of events, consequences, and likelihood or probability of occurrence, where a potential source of harm is termed *hazard*[20]. The majority of offshore platforms are currently designed using risk assessment to reduce and mitigate potential risks in a timely manner[22]. Extended inspection times can reduce any necessary reaction time and cause damage to subsea equipment. Identifying risks is one of the first steps that form the basis of risk analysis. Before conducting risk identification, the necessary operational data, such as pipeline type, failure modes, weather and environmental conditions, and maintenance properties, should be gathered[23]. An example of steps for the risk-based subsea pipeline corrosion inspection are illustrated in Figure 2.8. Data gathering is the first step towards risk-based inspection, including information on the design, inspection methods, operating environment of subsea pipeline. With gathered data and due to the various types and degrees of corrosion that can occur on subsea pipelines, the following tasks would include the characterization of defects through thickness measurements and an initial screening phase to determine whether a detailed analysis is necessary[22]. It is necessary to establish risk acceptance criteria. The risk acceptance criteria are risk reduction objectives and aid in maintaining confidence in the subsea pipeline's structural integrity. The general acceptance criteria comply with the operation's safety objectives and specify acceptable limits for risks to personnel and environmental safety as well as economic viability. After the acceptance criteria is defined, the probability of failure needs to be established, including defining the corrosion damage, its physical properties, the distribution of corrosion damage, and classification. After the corrosion properties have been evaluated and described, the conse-

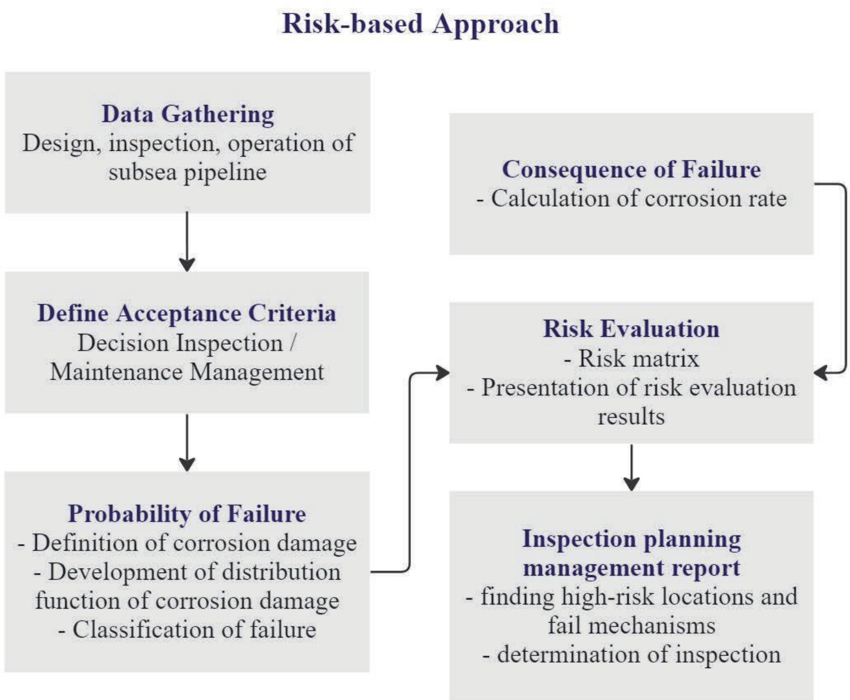

Figure 2.8. Subsea pipeline corrosion: Flowchart of the risk-based inspection planning, adapted from[22]

quence of failure can be determined through corrosion rate and burst pressure calculation. Finally, risk evaluation is derived and presented in a risk matrix. Figure 2.9 illustrates an example of a risk matrix describing the risk level and consequence severity by risk occurrence likelihood that is color-coded. Color green in the risk matrix represents insignificant consequences, yellow and orange represent minor and moderate consequences, and darker red colors represent major and severe consequences. Color coding reflects As Low As Reasonably Possible (ALARP) principle for broadly acceptable, conditionally acceptable, and intolerable regions of risk[24]. With these results, as illustrated in Figure 2.8, the inspection planning and management report is developed to find fail mechanisms and high-risk locations and determine the inspection and maintenance plans.

With visual inspection being the most common technique for pipeline inspection by ROVs and UAS[25], the success of the inspection depends on detecting unusual or unexpected situations that can be rare and negligible or situations that are rare yet hazardous. These rare occurrences are known as

Risk Level		Consequence Severity				
		Insignificant 1	Minor 2	Moderate 3	Major 4	Severe 5
Likelihood	Very High 5	5	10	15	20	25
	High 4	4	8	12	16	20
	Medium 3	3	6	9	12	15
	Low 2	2	4	6	8	10
	Very low 1	1	2	3	4	5

FIGURE 2.9. An example of risk matrix, adapted from[23]

anomalies. Therefore, anomaly criteria for visual inspection depend not only on the descriptions of anticipated risks and hazards but also on the degree to which these can be observed via image and video data. In contrast to an ROV, where anomaly detection relies on the training of the operator who manually controls the ROV, an UAS relies on AI methods for anomaly detection and classification. In both situations, proper classification and criteria for expected anomalies are necessary. Figures 2.7 - 2.11 show examples of anomalies on subsea pipelines, provided by operators from oil and gas industry, where an anomaly represents a deviation from normal state during data-driven or image-driven inspection and is recognized as a potentially degraded or fault state.

Table 2.2 shows a list of anomalies by their category and criteria description, followed by a reference to figures that illustrate examples of the described anomalies. Tadjiev[25] derived the anomaly criteria for visual inspection by collecting industry experiences and requirements, as well as lessons learned and best practice guidance. Further pipeline inspection necessitates using specialized tools and sensors when anomalies are not evident from a visual inspection. However, there is no widely accepted standard for anomaly criteria for visual inspection by ROV and UAS, and the task remains operation- or even operator-specific[25]. In addition, what is considered to be an anomaly depends on the context of the situation, meaning the anomalies depend on a substantial number of factors, such as the seabed environment, soil type, pipe design, material type, age, depth, and marine life, as well as the interrelationships between these factors. Standardization can be achieved through consistent operator reporting, enabling consistent quality inspection data and a more precise understanding of anomaly occurrences.

Table 2.2. Anomaly Criteria for Visual Inspection of Subsea Pipeline[25]

Anomaly category	Criteria	Figure reference
Inadequate Cathodic Protection	Disconnected or heavily depleted anodes	Not available
External corrosion	External corrosion on exposed surface or staining on outer sheath when no obvious exposed surface is visible	Figure 2.10
External damage or deformation	Evidence of deformed pipeline, abrasion, cut, tear, burst	Figure 2.11
Debris	Any debris or objects in contact with pipe or blocking visibility of pipe or pipe components such as fishing equipment, anchors, boulders	Figure 2.12
Inadequate support	Evidence of missing support, such as eroded seabed, riser self-trenching	Not available
Loss of primary containment	Evidence of fluid leakage	Not available
Marine growth	Presence of marine growth or coral colonies covering > 50% of surface area or prevents meaningful inspection	Figure 2.13
Layout disarrangemen	Interference of pope with other subsea equiment (risers, mooring lines, anchors), overbending, sliding	Not available
Seabed movement	Burial of pipes, gas vent valves, scour around seabed structures	Figure 2.14
Other	Unexpected anomalies, undescribed, unidentified and suspicious events	Not available

FIGURE 2.10. Example of external corrosion on subsea pipeline and pipe equipment[25] © Anonymous Operator

FIGURE 2.11. Example of external damage (ruptures)[25] © Anonymous Operator

FIGURE 2.12. Example of external debris - misplaced objects[25] © Anonymous Operator

FIGURE 2.13. Example of marine growth disrupting inspection[25] © Anonymous Operator

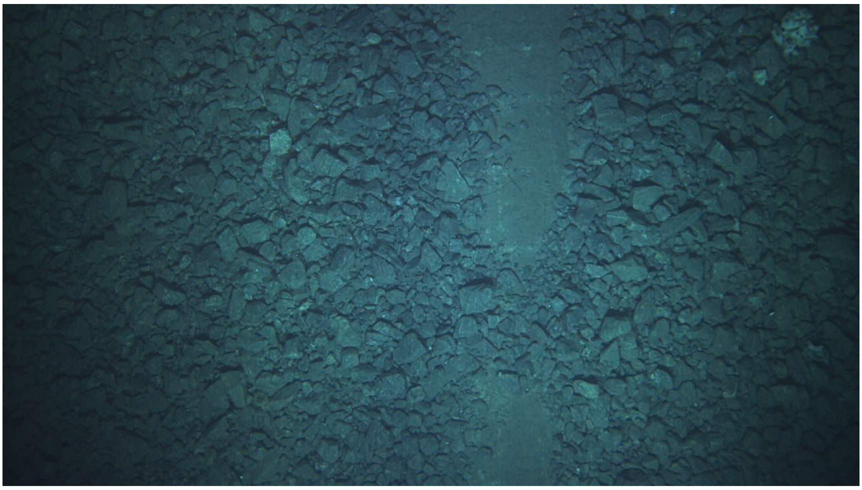

Figure 2.14. Example of buried pipeline © Equinor

2.4 References

[1] Michael Ho, Sami El-Borgi, Devendra Patil, and Gangbing Song. Inspection and monitoring systems subsea pipelines: A review paper. *Structural Health Monitoring* **19** (2), 606–645 (3 2020). ISSN 17413168. doi: 10.1177/1475921719837718. Cited on page/s 19, 21, 22, 26.

[2] International Electrotechnical Commission. IEC 60050 (2 2015). URL https://www.electropedia.org/iev/iev.nsf/. Cited on page/s 19.

[3] Erling Håland, Andreas Falck, Marianne Hauso, and Espen Funnemark. Recommended Failure Rates for Pipelines 2017-0547, Rev. 2. Technical report DNV GL AS Oil & Gas Safety Risk Management Høvik (8 2017). Cited on page/s 19, 20.

[4] Cheryl McMahon Anderson, Melinda Mayes, and Robert LaBelle. Update of occurrence rates for offshore oil spills. Technical report epartment of Interior Bureau of Ocean Energy Management and Department of Interior Bureau of Safety and Environmental Enforcement Herndon, VA (2012). Cited on page/s 20, 21.

[5] S. Vishnuvardhan, A. Ramachandra Murthy, and Abhishek Choudhary. A review on pipeline failures, defects in pipelines and their assessment and fatigue life prediction methods (2 2023). ISSN 03080161. Cited on page/s 20.

[6] Michael Ho, Sami El-Borgi, Devendra Patil, and Gangbing Song. Inspection and monitoring systems subsea pipelines: A review paper (3 2020). ISSN 17413168. Cited on page/s 21.

[7] Christian Mai, Simon Pedersen, Leif Hansen, Kasper L. Jepsen, and Zhenyu Yang. Subsea infrastructure inspection: A review study. In *USYS 2016 - 2016 IEEE 6th International Conference on Underwater System Technology: Theory and Applications* pages 71–76. Institute of Electrical and Electronics Engineers Inc. (4 2017). ISBN 9781509057986. doi: 10.1109/USYS.2016.7893928. Cited on page/s 22.

[8] Jan Albiez, *et al.* FlatFish - a compact subsea-resident inspection AUV. In *OCEANS 2015 - MTS/IEEE Washington*. Institute of Electrical and Electronics Engineers Inc. (2 2016). ISBN 9780933957435. doi: 10.23919/oceans.2015.7404442. Cited on page/s 23.

[9] Eelume AS. Eelume Subsea Intervention (). URL https://eelume.com/#the-eelume-concept. Cited on page/s 23, 24.

[10] Daniel Abicht, Jan Christian Torvestad, Pål Atle Solheimsnes, and Karl Atle Stenevik. Underwater intervention drone subsea control system. *Proceedings of the Annual Offshore Technology Conference* **2020-May** (May), 4–7 (2020). ISSN 01603663. doi: 10.4043/30701-ms. Cited on page/s 24, 25.

[11] Francesco Scibilia, Knut Sebastian Tungland, Anders Røyrøy, and Marianne Bryhni Asla. Energy industry perspective on the definition of autonomy for mobile robots. In *Communications in Computer and Information Science* volume 1056 CCIS pages 90–101. Springer International Publishing (2019). ISBN 9783030356637. doi: 10.1007/978-3-030-35664-4{_}9. URL http://dx.doi.org/10.1007/978-3-030-35664-4_9. Cited on page/s 25.

[12] Yan Shi, Chao Zhang, Rui Li, Maolin Cai, and Guanwei Jia. Theory and application of magnetic flux leakage pipeline detection. *Sensors (Switzerland)* **15** (12), 31036–31055 (12 2015). ISSN 14248220. doi: 10.3390/s151229845. Cited on page/s 26.

[13] L-3 Communications SeaBeam Instruments. Multibeam Sonar Theory of Operation. Technical report L-3 Communications SeaBeam Instruments East Walpole, MA (2000). URL https://www3.mbari.org/data/mbsystem/sonarfunction/SeaBeamMultibeamTheoryOperation.pdf. Cited on page/s 26.

[14] S Ravi, S Karthikraj, D Sabareesan, and R Kishore. Pipeline Monitoring Using Vibroacoustic Sensing-A Review. *International Research Journal of Engineering and Technology* (2016). ISSN 2395 -0056. URL www.irjet.net. Cited on page/s 26.

[15] Hang Zhou Yang, Xue Guang Qiao, Dong Luo, Kok Sing Lim, Wuyi Chong, and Su-

laiman Wadi Harun. A review of recent developed and applications of plastic fiber optic displacement sensors. *Measurement: Journal of the International Measurement Confederation* **48** (1), 333–345 (2014). ISSN 02632241. doi: 10.1016/j.measurement.2013.11.007. Cited on page/s 26.

[16] ISO/DIS20815. ISO/DIS 20815 2018 Petroleum, petrochemical and natural gas industries - Production assurance and reliability management. Technical report ISO (10 2018). Cited on page/s 28.

[17] IEC60300-3-4. International Standard International Electrotechnical Commission IEC 60300-3-4:2022 Dependability management - Part 3-4: Application guide - Specification of dependability requirements. Technical report IEC International Electrotechnical Commission (3 2022). Cited on page/s 28.

[18] D.R. Kiran. Reliability Engineering. In *Total Quality Management* pages 391–404. Elsevier (1 2017). doi: 10.1016/B978-0-12-811035-5.00027-1. URL https://linkinghub.elsevier.com/retrieve/pii/B9780128110355000271. Cited on page/s 28.

[19] E. Zio. The future of risk assessment. *Reliability Engineering and System Safety* **177** (March), 176–190 (2018). ISSN 09518320. doi: 10.1016/j.ress.2018.04.020. Cited on page/s 28.

[20] ISO 31000. Risk management — Guidelines, International Organization for Standardization. Technical report International Organization for Standardization (2018). URL https://www.iso.org/obp/ui/iso:std:iso:31000:ed-2:v1:en. Cited on page/s 28.

[21] ISO/IEC31010. IEC 31010:2019: Risk management — Risk assessment techniques. Technical report ISO/IEC 2009 (11). Cited on page/s 28.

[22] Jung Kwan Seo, Yushi Cui, Mohd Hairil Mohd, Yeon Chul Ha, Bong Ju Kim, and Jeom Kee Paik. A risk-based inspection planning method for corroded subsea pipelines. *Ocean Engineering* **109**, 539–552 (11 2015). ISSN 00298018. doi: 10.1016/j.oceaneng.2015.07.066. Cited on page/s 28, 29.

[23] Xinhong Li, Guoming Chen, Yuanjiang Chang, and Changhang Xu. Risk-based operation safety analysis during maintenance activities of subsea pipelines. *Process Safety and Environmental Protection* **122**, 247–262 (2 2019). ISSN 09575820. doi: 10.1016/j.psep.2018.12.006. Cited on page/s 28, 30.

[24] Marvin Rausand. Risk Assessment Theory, Methods, and Applications. John Wiley and Sons Inc Hoboken, New Jersey (2011). ISBN 9780470637647. doi: 10.10029781118281116. Cited on page/s 29.

[25] Damir Tadjiev. Anomaly criteria for general visual inspection of subsea flexible pipes. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE* **4**, 1–9 (2020). doi: 10.1115/OMAE2020-19044. Cited on page/s 29, 30, 31, 32.

Chapter 3

Experimental methods

Inspection of subsea oil and gas pipelines relies on collecting and analyzing sensor data, typically using data-driven and artificial intelligence (AI) approaches, such as machine learning (ML) and computer vision. These methods can extract features from sensor-collected data, classify images captured by camera-equipped UAS, and identify anomalous patterns in vast data collections. Machine learning enables computer systems to become more intelligent as they encounter more data [1]. Various data-driven approaches, such as anomaly detection, image classification, and image segmentation, can detect and analyze pipeline surface hazards when observing UAS-captured images. The following sections describe the general theory of anomaly detection and classification, as well as the different computer vision approaches, their developments, and their function in subsea pipeline hazard inspection that have been discussed and applied during this research.

3.1 Anomaly Detection and Classification

Anomaly detection is a method for identifying any deviations in data (i.e., sensor data, image data) from expected behavior. These deviations are referred to in different terms, such as anomalies, outliers, exceptions, aberrations, depending on the application domain [2]. Depending on the application, the detected anomalies may indicate noise, or anomalies as hazards. The distinction lies in the fact that *noise* is an occurrence that impedes data analysis, misleads findings, or misleads the machine learning methods learning process from the data [3], and therefore it is removed after it has been detected. *Anomaly*, on the other hand, is the reverse of noise; it can reflect something unique and frequently critical, such as pipeline hazards in subsea inspection applications, network intrusions in cybersecurity applications, or diseases in medical applications. In the context of subsea pipeline inspection, an anomaly represents the deviation in observed data that indicates a hazard (i.e., pipeline damage as a deviation from the pipeline's normal state), whereas noise represents any deviations in observed data that do not require inspection or do not represent hazards (i.e., soft debris on the pipeline surface). Apart from noise detection and noise removal, anomaly detection is related to novelty detection. *Novelties* are previously unobserved data patterns. The main distinction between novelty and an anomaly is that after a novelty has been identified, it is often incorporated into the standard data model. An anomaly typically never conforms to the standard or normal data[2]. Anomaly detection methods are typically tied to domain-specific applications and challenges because what constitutes an anomaly relies on the domain problem[2]. Therefore the key components of an anomaly detection, as illustrated on Figure 3.1, are often described within a specific domain (i.e., cybersecurity, medical, offshore operations). This means that the anomalies have problem-specific characteristics that describe types of data, expected anomaly types, available labels, and expected outputs[2].

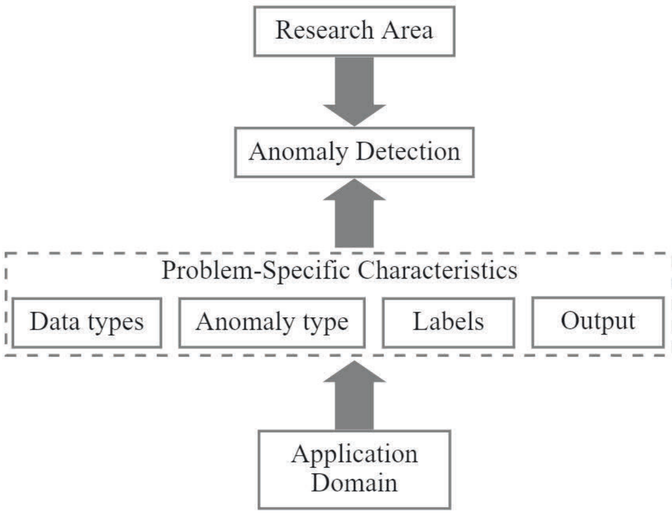

FIGURE 3.1. Key components for anomaly detection, adapted from[2]

Point, contextual, and collective anomalies are the most general types of anomalies[3]. *Point* anomalies, also known as global anomalies, are single data points that generally differ from the rest of the data points in a dataset. *Contextual* anomalies only occur under specific conditions, such as season or location (i.e., snow during summer would be considered an anomalous event, while winter snow would be a normal event). *Collective* anomalies are anomalous data points only when they appear in groups or collections, not as individual data points. Detecting anomalies typically happens during a data preprocessing stage in machine learning applications so that the anomalies are identified and addressed before the training stage where learning from data patterns occurs.

Depending on the availability of labels in the data or the existence of training data, anomaly detection methods can there are three main modes by which anomalies are detected: supervised, semi-supervised, and unsupervised.

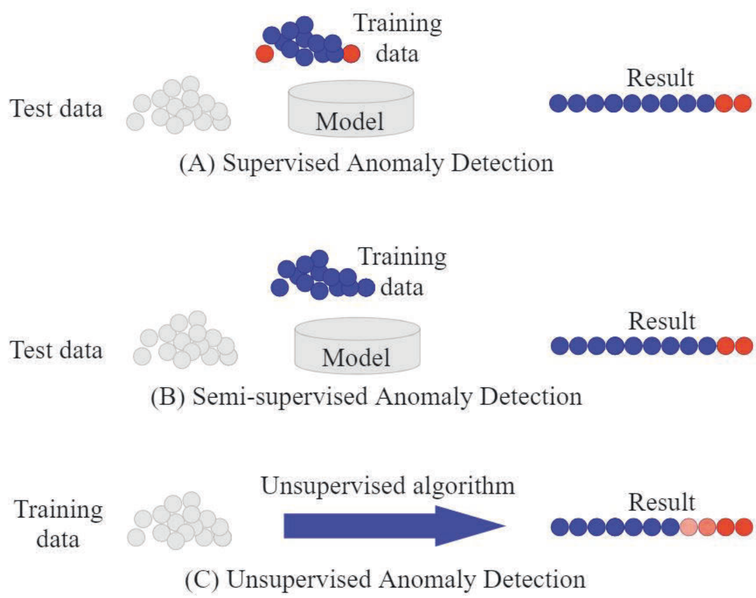

FIGURE 3.2. Modes of anomaly detection based on label data availability, adapted from[4]

Supervised anomaly detection assumes the availability of a training dataset comprising labels for normal and anomalous data points (see Figure 3.2 (A)). Any unknown data point is compared to the model to determine the class or label to which it belongs, normal or anomaly. A significant issue is that there are considerably fewer anomalous than normal data points in the training data, thus creating imbalanced data and making it more challenging for the model to detect less represented data points - anomalies[2,4].

Semi-supervised anomaly detection assumes that only normal-class data points are labeled in the training data, as illustrated in Figure 3.2 (B). Since they do not require labels for the anomaly class, their applicability is greater than supervised approaches[2,4].

Unsupervised anomaly detection, illustrated in Figure 3.2 (C), work without any labeled data and implicitly assume that normal occurrences in the test data are significantly more frequent than anomalies. If this assumption is untrue, such methods have a high rate of false alarms (i.e., reporting points that are not truly anomalous). Due to the cost and complexity of obtaining labeled data, unsupervised methods are the most prominent [2,4].

3.1.1 Blindspots and Challenges

Although anomaly detection appears to be a simple process, its application is nevertheless hindered by numerous obstacles, despite the substantial research conducted to address them:

- The noise in data is difficult to distinguish from significant anomalies that may indicate something important. This challenge is particularly evident in image data with complex backgrounds or poor visibility.
- The lack of training or labeled data requires the use of unsupervised methods that are difficult to explain, and suffer from high false alarm rates [2]. The false alarm rates disrupt the trust in operators observing the anomaly detection methods, and may disrupt the final conclusions of application's functioning.
- Identifying a region in data to be considered normal, in the form of boundaries, can be difficult, especially if anomalies are located near or on the normal boundary.
- The low representation of anomalies in datasets often leads to biases. Due to the method's inclination towards efficiency, these anomalies may get ignored [5].
- Anomalies can be fluid and dynamic from domain to domain and under different circumstances within the same domain. So, the anomaly detection methods may need to actively change to adhere to the situation's dynamic definition of anomaly.

Many other algorithm-specific issues include requirements for high computational power, and inability to reliably detect anomalies in higher dimension and large datasets [6].
Due to these obstacles, solving the anomaly detection problem in its most general form requires extensive and detailed work. Many elements influence the application of anomaly detection, such as the nature of the data, the availability of labeled data, and the type and behavior of expected anomalies.

3.2 COMPUTER VISION

As a type of artificial intelligence, computer vision consists of methods that specialize in analyzing visual data, such as images and videos[7]. In recent years, computer vision has gained popularity with applications like street view for autonomous vehicle driving, satellite and map images, smart surveillance security cameras for object and movement detection, and autonomous drone navigation and object recognition. For offshore applications, underwater computer vision is increasingly popular in the research community and industry with autonomous underwater systems that explore the seabed, inspect offshore structures, or monitor marine life[7]. The most common computer vision tasks are image classification (see Figure 3.4), object detection (see Figure 3.5), image segmentation (see Figure 3.6). Underwater images frequently suffer from image

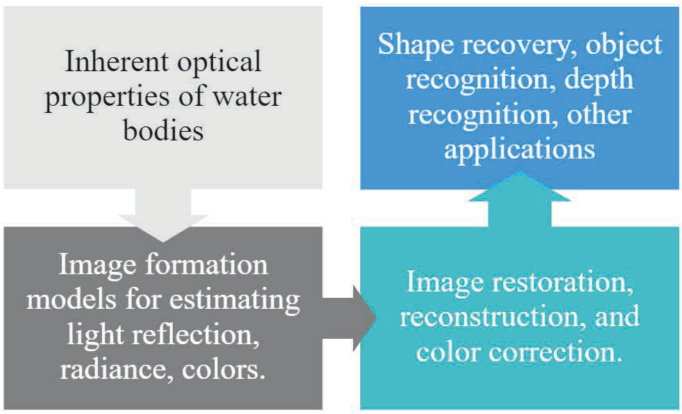

FIGURE 3.3. Computer vision preprocessing tasks, adapted from[7]

degradation through loss of color, shape, and visibility; as a result, underwater computer vision typically involves some form of image preprocessing to reconstruct shape in blurred images, color from many blue and green hues from water, and overall visibility. The process preceding classification, segmentation, and recognition is illustrated in Figure 3.3. The optical properties of water bodies, such as the radiance, can be estimated by correlating the water's optical properties with other variables, such as the optical characteristics of the camera lens or the detector specifications[7]. Despite being influenced by environmental changes, these properties are predictable enough to distinguish between bodies of water. This makes it possible to obtain a color-corrected, photometrically invariant image for computer vision applications such as recognition, segmentation, and shape recovery. Image classification is one of the earliest ap-

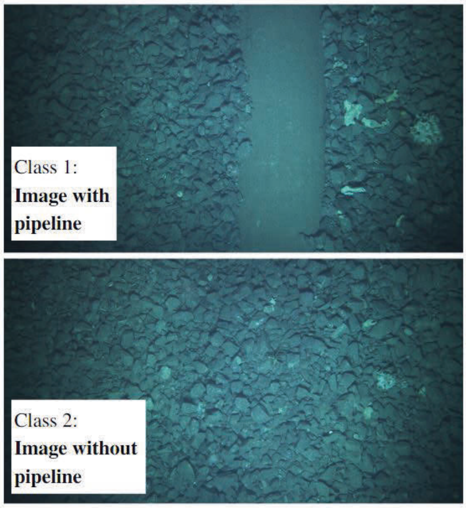

FIGURE 3.4. Subsea pipeline object detection and classification. Image property of © Equinor.

plications of computer vision for differentiating images, typically learned from a set of training images labeled into classes. Figure 3.4 illustrates an example of binary image classification where the entire image is observed and classified into an image with or without a pipe. However, the image classification does not separate individual objects on the image.

Another common task of computer vision is object detection and classification, in which individual objects, rather than the entire image, are observed. Object detection typically draws a box around a detected object but does not follow the precise object shape, as illustrated in Figure 3.5.

Image segmentation is the process of separating or segmenting individual image objects that can be classified individually [8]. Human experts label or mark the majority of image segments. Figure 3.6 illustrates an example of image segmentation, in which individual objects on an image are precisely contoured to detect the exact shape of an object. In recent years an increasing number of AI-based automated labeling approaches have emerged that can segment objects following the pixels of object corners [8].

The typical classification process, either of entire images, objects or segments on the image, can be broken down as follows [9]:

1. Information collection via camera.

2. Data preparation that may consist of color correction, anomaly detection, and noise elimination (dependent on the application needs).

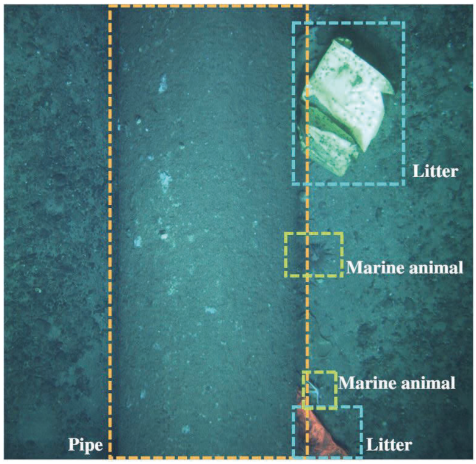

FIGURE 3.5. Subsea pipeline image classification. Pipeline images property of © Equinor.

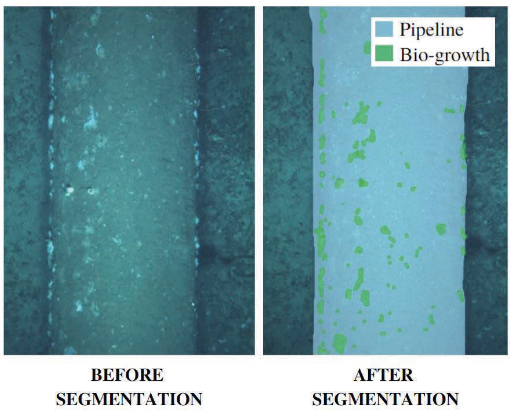

FIGURE 3.6. Subsea pipeline before and after image segmentation. Pipeline image property of © Equinor.

3. Separation of the data into a training set for creating the models and allowing the classification algorithm to learn the image patterns, and a test set for evaluating the model's accuracy.

4. Evaluation of the model's accuracy by testing it on data from the test set that was not used for training.

Figure 3.7 presents the evaluation methods for evaluating the classification

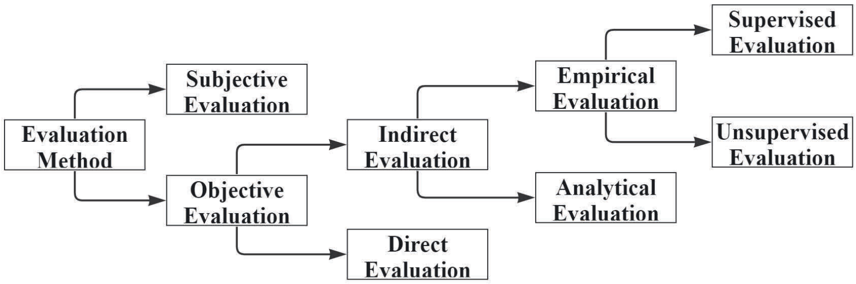

FIGURE 3.7. Evaluation methods, adapted from[8]

results of computer vision methods. The most practical evaluation method is a subjective evaluation, in which human evaluators evaluate segmentation results[8]. A disadvantage of this approach is that the visual or qualitative evaluations are subjective and depend on each evaluator, as each evaluator uses unique criteria to evaluate the quality of the segmented image. The direct evaluation assesses the computer vision algorithm's computational power, speed, and execution time. Analytical and empirical evaluations are a form of indirect evaluation and the most common approaches. The accuracy of image classification is observed via various algorithm metrics, such as accuracy and precision[8]. In case of a supervised evaluation, a reference image is used to obtain the accuracy and precision metrics of the results. Unsupervised evaluation methods do not require reference images and evaluate segmentation images by calculating human-recognized criteria representing preferred results. Obtaining reference images is generally difficult, time-consuming, and expensive (i.e., labeling or segmenting objects on thousands of images)[8]. Without a reference image, the unsupervised method does not provide a more reliable evaluation accuracy than the supervised method.

3.2.1 *Blindspots and Challenges*

Visibility is one of the main challenges in underwater computer vision where light absorption, scattering, refraction, currents and water turbulence often make it challenging to capture a good quality underwater image[7]. While image classification may be the least challenging to create training data for, the underwater images still make the detection of patterns, colors and shapes, difficult due to visibility and monochromatic nature. Image classification, object detection, and image segmentation tasks suffer largely from lack of labelled data. Therefore, computer vision applications typically resort to unsupervised approaches, neural networks and deep learning. Many of these technologies are black boxes whose results are difficult to explain, which is a significant disadvantage [10]. Advanced underwater computer vision is in the early stages of research for reliable application. Improving AI methods and focusing on their reliability ensures computer vision's constant development.

3.3 References

[1] Oxford University Press. Oxford Learner's Dictionaries (2021). URL https://www.oxfordlearnersdictionaries.com/. Cited on page/s 37.

[2] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A Survey. *ACM Computing Surveys (CSUR)* **14** (1), 1–22 (7 2009). ISSN 15462218. doi: 10.1145/1541880.1541882. Cited on page/s 37, 38, 39, 40.

[3] Abir Smiti. A critical overview of outlier detection methods (11 2020). ISSN 15740137. Cited on page/s 37, 38.

[4] Markus Goldstein and Seiichi Uchida. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLOS ONE* **11** (4) (2016). doi: 10.1371/JOURNAL.PONE.0152173. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152173. Cited on page/s 39, 40.

[5] Karima Makhlouf, Sami Zhioua, and Catuscia Palamidessi. On the applicability of ML fairness notions. *arXiv* pages 1–32 (2020). ISSN 23318422. Cited on page/s 40.

[6] Jinjiang Wang, Peilun Fu, and Robert X. Gao. Machine vision intelligence for product defect inspection based on deep learning and Hough transform. *Journal of Manufacturing Systems* **51**, 52–60 (4 2019). ISSN 02786125. doi: 10.1016/j.jmsy.2019.03.002. Cited on page/s 40.

[7] Salma P. González-Sabbagh and Antonio Robles-Kelly. A Survey on Underwater Computer Vision. *ACM Computing Surveys* (1 2023). ISSN 0360-0300. doi: 10.1145/3578516. Cited on page/s 41, 44.

[8] Zhaobin Wang, E. Wang, and Ying Zhu. Image segmentation evaluation: a survey of methods. *Artificial Intelligence Review* **53** (8), 5637–5674 (12 2020). ISSN 15737462. doi: 10.1007/s10462-020-09830-9. Cited on page/s 42, 44.

[9] Ankita Singh and Pawan Singh. Image Classification: A Survey. *Journal of Informatics Electrical and Electronics Engineering (JIEEE)* **1** (2), 1–9 (11 2020). ISSN 25827006. doi: 10.54060/JIEEE/001.02.002. URL https://jieee.a2zjournals.com/index.php/ieee/article/view/2. Cited on page/s 42.

[10] Vanessa Buhrmester, David Münch, and Michael Arens. Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *Machine Learning and Knowledge Extraction* **3** (4), 966–989 (12 2021). ISSN 25044990. doi: 10.3390/make3040048. Cited on page/s 45.

Part II

Results

Chapter 4

Risk-informed and Data-Driven UAS Operations

This chapter is based on the following articles:

- Spahic, Rialda; Hepsø, Vidar; Lundteigen, Mary Ann. (2021) Reliable Unmanned Autonomous Systems: Conceptual Framework for Warning Identification during Remote Operations. 2021 IEEE International Symposium on Systems Engineering (ISSE).
  DOI: 10.1109/ISSE51541.2021.9582534

- Spahic, Rialda; Hepsø, Vidar; Lundteigen, Mary Ann. (2022) Using Risk Analysis for Anomaly Detection for Enhanced Reliability of Unmanned Autonomous Systems. Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022).
  DOI: 10.3850/978-981-18-5183-4-R08-03-390-cd

All authors contributed to the research conception of the two articles. Rialda Spahic performed material preparation, literature analysis, and manuscript writing. Mary Ann Lundteigen performed writing reviews and supervision of all prior drafts of the manuscript. Vidar Hepsø contributed to the literature gathering and concept visualisation of the research.

The two articles focus on identifying concepts from multiple safety-concerned disciplines (risk assessment, reliability engineering, resilience engineering, and human-machine teaming) to supervise anomaly detection and classification methods for subsea pipeline hazard detection. The discussed challenges are trust calibration, explainability of algorithms, data biases, and the inadequacies of anomaly detection methods to identify data biases and efficiently distinguish noise from meaningful anomalies. The following sections describe the motivation and main contribution of the two articles, a conceptual framework for identifying warnings. The framework is a novel, risk-informed approach to validating anomaly detection results and extracting meaningful information from detected anomalies, which serve as early warning signs of potential hazards during AUS operations. The application of the proposed framework is presented in Chapters 5, and 6.

4.1 Reliable Unmanned Autonomous Systems: Conceptual Framework for Warning Identification during Remote Operations

Abstract

In the offshore industry, unmanned autonomous systems are expected to have a permanent role in future operations. During offshore operations, the unmanned autonomous system needs definite instructions on evaluating the gathered data to make decisions and react in real-time when the situation requires it. We rely on video surveillance and sensor measurements to recognize early warning signals of a failing asset during the autonomous operation. Missing out on the warning signals can lead to a catastrophic impact on the environment and a significant financial loss. This research is helping to solve the issue of trustworthiness of the algorithms that enable autonomy by capturing the rising risks when machine learning unintentionally fails. Previous studies demonstrate that understanding machine learning algorithms, finding patterns in anomalies, and calibrating trust can promote the system's reliability. Existing approaches focus on improving the machine learning algorithms and understanding the shortcomings in the data collection. However, recollecting the data is often an expensive and extensive task. By transferring knowledge from multiple disciplines, diverse approaches will be observed to capture the risk and calibrate the trust in autonomous systems. This research proposes a conceptual framework, Warning Identification Framework, that captures the known risks and creates a safety net around the autonomy-enabling algorithms to improve the reliability of the autonomous operations.

4.1.1 Introduction

The advancements in technology are changing the way the industry handles risk. What used to be a tedious or dangerous job for a human can be replaced by an unmanned autonomous system (UAS). This replacement can enhance safety, work efficiency, and knowledge of the operating environment. As a type of artificial intelligence [1], machine learning (ML) is at the forefront of research in the context of reliable UAS. Autonomy is "an unmanned system's own ability of integrated sensing, perceiving, analyzing, communicating, planning, decision-making, and acting/executing to achieve its goals" [2]. Recent research on autonomous systems identifies common challenges within risk and trust of ML algorithms that enable autonomy [3], [4]. Formal definitions of technical tests and evaluation of UAS [5] highlight challenges of lacking the quantitative definitions of emergent behavior, human trust, reliability and resilience [3]. However, to ensure the UAS's ability to act and make decisions to achieve the mission's goal, it is critical to explore the concept of calibrated trust [6]. Calibrated trust is the process of adjusting the trust level of human operators with the actual reliability of a system [6], - or trusting the machine will do as intended within a specific environment [7].

ML allows computing systems to learn how to do tasks from significant amounts of data, rather than being programmed (human instructed) [1]. Therefore, there is a rising need to understand how ML capabilities can be integrated into existing systems engineering, and design processes [3]. The performance of ML algorithms can measure the majority of UAS's capabilities, inevitably measuring the system's reliability. However, the software and system reliability engineering for UAS incorporating ML is not a trivial task. ML integration is experiencing significant limitations, including black box algorithms or algorithm explainability [8], scalability, and limited structural approach to problem-solving. The data, often impacted by biases, is another limitation related to ML. Bad quality data can lead the ML algorithms to result in poor predictions or decisions, and eventually, unintended harm [9].

During offshore operations, the UAS relies on integrated sensors and video input for surveillance, intervention, and inspection of the assets and the environment. The role of the UAS is to recognize warning signals from the environment or the inspected asset, trigger warning signals, and report them to the offshore control center or operator control rooms in real-time. Unintended ML outcomes can significantly impact the environment, the asset, and the UAS itself. The environmental disruptions can stay unnoticed and develop to critical states, such as disruptive water states or chemical leaks. Similarly, unobserved corrosion, chemical leaks, material degradation, cracks, misplaced objects, and biological growth on assets are just a few examples of the potential issues. This problem can lead to a catastrophic impact on the environment and significant financial loss for the industry. Knowing how to respond and prepare the data for anticipated insights is a challenge in dynamic operations. The industry needs more knowledge on reliable, and time-efficient UAS operations [10].

The contribution of this section is a Warning Identification Framework (WIF) for UAS incorporating ML. The WIF incorporates managing resilience, ensuring the system's ability to plan, prepare and react to the potential occurrence of unwanted and disruptive events. While designing this framework we consolidate knowledge on reliability and resilience engineering, risk assessment, and human-machine teaming approach to UAS. In this section, we:

1. Provide a multidisciplinary approach to the safety concerns of current systems incorporating ML through the lenses of risk assessment's future.

2. Propose a global framework based on a shared understanding of gaps in ML of a particular application instead of solutions based on specific ML algorithm enhancements or global change of data gathering processes.

4.1.2 Motivation and Related Applications

Trust Calibration

Recent research shows potential in alleviating risks, enhancing reliability, and influencing trust in autonomous systems. Reliability is an ability to perform as required, without failure, for a given time interval, under given conditions (IEC 192-01-24)[11]. The reliability of an autonomous system directly impacts trust. However, over-trust and under-trust often occur in highly dynamic environments and can pose serious safety and efficiency concerns[6]. Over-trust in the system implies that the human operator overestimated the reliability of a system. Under-trust in the system implies that the human operator estimates that the system should not be trusted with a given task. Okamura et al.[6] describe the trust calibration in autonomous systems in a dynamic environment as an essential process for successful collaboration between humans and systems. Trust calibration incorporates system reliability and continuous system transparency. Okamura et al.[6] argue that trust is a latent construct and therefore challenging to measure. The authors[6] observe human behavior to determine the trust calibration status. They experimented with a drone simulator and observed seventy participants who performed inspection tasks manually or relied on the inspection by an autonomous drone. In the experiment, the participants observed the changing weather conditions in the drone simulator. The participants were required to actively make decisions whether they trust or rely on the autonomous drone to perform inspection tasks within the environmental conditions presented on the simulator. The experiment's goal was to capture the under-trust and over-trust of the participants in the autonomous drone operations. The experiment demonstrated successful detection of miscalibration of trust and adjustment of participants' behaviors, showing trust gaps in collaboration between humans and autonomous systems. The results showed that understanding how the system functions and makes decisions is crucial when trusting the autonomous systems.

Explainable Machine Learning

ML is taking over many high-stakes decision-making throughout society[8]. The author[8] defines *black box ML algorithms* as either functions that are too complicated for any human to comprehend or as proprietary functions. Past research highlights that developing explainable algorithms will mitigate some of the problems caused by the black box algorithms [8]. Rudin [8] argues that trying to explain black box algorithms rather than developing explainable ones can support a bad practice and therefore cause harm to society.

The author [8] singles out some of the most prominent challenges of black box and explainable algorithms that are summarised as follows:

1. *Complexity:* There is a belief that black box algorithms result in top predictive performance when compared to the explainable algorithms that are easier to understand. The author claims that when the data is structured and contains meaningful features, complex classifiers (such as neural networks, random forest, boosted decision trees) and more straightforward classifiers (such as logistic regression and decision lists) perform similarly. Complexity does not imply accuracy, which is also valid for computer vision or image processing algorithms that are often particularly complex.

2. *Faithfulness of explainable algorithms:* Explainable ML algorithms provide interpretations that are not faithful to what the black box algorithm computes. The explainable algorithm does not mimic the black box algorithm but instead tries to interpret it as accurately as possible and provide an explainable alternative to the black box algorithm. The difficulty in creating this interpretation can lead to misalignment with the black box algorithm and endanger the trust in the black box algorithm. Rudin [8] proposes calling these interpreted algorithms 'summary statistics', 'summary predictions' or 'trends of the algorithm' to avoid confusion with the belief that the interpretation should mimic the black box algorithm.

3. *Challenge to incorporate risk estimation within black box algorithms:* The database is a definite collection of data or information that the algorithms learn from and train on to make predictions. Black box algorithms are often incompatible with the situation where information outside the database needs to be combined with a risk assessment. Rudin [8] argues that the black box algorithms are challenging to calibrate with additional information on estimated risk manually. Another downside of these algorithms is that it is not transparent as to what the risk estimation is.

4. *Explainability leading to human error:* Additional explanations to the black box algorithm can lead to complicated decision-making and leave space for other human error.

5. *Hidden patterns in data:* There is a myth that only black box algorithms can uncover hidden patterns in data. This myth can lead to less trust

in the performance of explainable algorithms. The author [8] claims that if the pattern were significant enough, it would be possible to obtain it with an explainable algorithm.

6. *Explainability is difficult to design and develop:* Creating explainable algorithms for specific domains often involves constraints on data dimensions, meaning that explainability requires low-dimensional space. It is challenging to troubleshoot the algorithm or agree on the explainable algorithm's reasoning process for a specific domain. The main challenge lies in the difficulty of developing and designing explainable algorithms.

Explainable ML algorithms lead to increased transparency that is crucial in measuring the fairness of the advanced system's decision-making processes. The *fairness* notion tells if the output of a predicting system is fair or discriminating [12]. Fairness is a rising problem due to the predictive system's tendency towards efficiency and sacrificing anomalies as tolerable collateral damage [12].

Errors and Biases in Machine Learning

There is a growing worry about the errors of ML in sensitive domains [13]. Pleiss et al. [13] describe cases of ML errors due to biases in data that have directly impacted human lives. The authors examine the cases of ML classification algorithms and frameworks that constrain these algorithms such that no false-positive or false-negative predictions affect any classified group or that there exists fairness in the classified groups. Their study demonstrated unsettling results that any algorithm with one error constraint (i.e., equal false-negatives across groups) is almost equal to randomizing the percentage of predictions for an existing classifier.

Knowing when to react is critical during remote operations. A timely reaction can prevent accidents saving the environmental impact and significant amounts of money. Galaz et al. [4] provide recent research of machine intelligence risks that include algorithmic bias and harms, unequal access and benefits, cascading failures and disruptions, mis- and disinformation, and trade-offs between efficiency and resilience. The authors imply that many foreseeable risks can be acted upon proactively. However, they do not propose actions or algorithms to intervene with ML outcomes' foreseeable risks. The authors [4] highlight that these risks are related to algorithmic biases and their allocative harms. The authors group these biases into training data bias, transfer context bias, and interpretation bias:

1. Training data bias is the erroneous data from which machines learn.

2. Transfer context bias occurs when using ML algorithms and dataset created in/for one environment in another.

3. Interpretation bias is a conflict between ML interpreted results and expected or needed results for further functioning of a system.

Suresh et al.[9] discuss important choices generated over extensive data and build a framework for understanding unintended consequences of ML. The authors identify 'biases' as the most common reason from which unwanted ML consequences arise. The bias represents an unintended or even malicious property of the data[9]. The authors[9] curate through recent work of known ML issues and identify six sources of harm that represent a framework for understanding the unintended ML consequences:

1. Historical bias occurs when the machine learns on historical or available data samples that do not reflect an accurate picture of the world.

2. Representation bias occurs when there is an imbalanced representation of all the data samples in the data set.

3. Measurement bias occurs when what we choose to measure does not relate well to the data samples the machine learns on or when the ML task is oversimplified.

4. Aggregation bias occurs when using a one-size-fits-all algorithm for cases with different conditional distributions.

5. Evaluation bias occurs when the evaluation or benchmark data for the ML algorithm does not represent the target measurement.

6. Deployment bias occurs when there is a mismatch between the problem an algorithm is intended to solve and how the algorithm is used.

The authors[9] advise tweaking ML algorithms to mitigate aggregation and evaluation biases in data. They indicate that the framework can communicate knowledge on ML outcomes and possibly facilitate productive solutions on dealing with the harmful consequences.

Applications

A significant number of applications are developed for autonomous systems incorporating ML for mitigating risks during operations. As a major task in offshore operations, the UAS are increasingly popular to gather information for risk assessment of the assets or the environment. Condition-monitoring data is often used as additional information for evaluating risk[14]. In offshore operations, monitoring of assets can give real-time information on degradation of the asset material, and the condition-monitoring data provides information on individual degradation process[14]. Some of the applications regarding data assessment on degradation processes such as oxidation, corrosion, fatigue, crack growth are[14–16]. Improved design and tweaking of ML algorithms and reconsideration of data gathering and pre-processing methods are the most notable research topics for enhancing the reliability of autonomous systems, and understanding error measurements[5,17–21].

Anomaly detection is an essential process for recognizing unexpected events in the data during operations. Liu et al.[22] explore background biases for anomaly detection in surveillance videos. Their study shows that the algorithms are biased to capture a considerable amount of background information as the basis of predictions. The authors[22] argue that background bias is a problem that exists in the majority of the action recognition algorithms, particularly in deep neural networks. They propose a trainable, area-guided framework for the anomaly detection algorithms to recognize anomalous regions and learn the essence of the anomaly instead of simply remembering the background[22]. Related concerns around anomaly detection algorithms are prominent in research, such as trade-offs and analysis of the algorithms[23], bottleneck identification[24], and large-scale anomaly detection in surveillance videos[25].

4.1.3 *Multidisciplinary Approaches to Risk and Reliability of Autonomous Systems*

Risk Assessment

Risk assessment is a discipline that incorporates structured analysis and identification of possible hazards/threats, their causes and consequences, risk description, quantification, and representation of uncertainties[14]. The terms *risk* and *warning* are often used together or interchangeably. According to[1], *risk* is the possibility of something bad happening at some time in the future, a situation that could be dangerous or have a bad result. Moreover, a *warning* is a statement or an event telling somebody that something bad or unpleasant may happen in the future so that they can try to avoid it[1]. Additionally, a warning is a sign that indicates approaching or threatening risk and may require immediate intervention. Therefore, it is crucial to understand a specific environment or assets' potential risks of failure to understand warning signs and act upon them. The risk assessment should provide a coherent increase of the awareness on risk and attention to safety. The fourth industrial revolution, particularly the internet of things, big data, and artificial intelligence that enables autonomy, changes how we design and develop systems and monitor our environment. This complex network of cooperative systems provides opportunities to improve the systems that monitor, intervene, and inspect the environment or the industrial assets to become more efficient, faster, more flexible, and resilient. However, these systems also generate new weaknesses, hazards and create new risk, somewhat due to new and unknown functional dependencies in and among the systems [14]. [26] describe the industry perspective on the definition of autonomy and divide autonomy into six levels, from no automation to a fully autonomous system that does not require human interaction. The authors [26] highlight that the fully autonomous system is multidimensional and incorporates autonomy/automation, data deliberation, and risk assessment. Data deliberation signifies the system's capability to continuously gather data from the environment, analyze it, and compare historical data to make predictions. Risk assessment signifies the system's capability to continuously assess the risk and adjust the criticality of the warnings accordingly, deciding the best risk mitigation policy [26].

Naturally, the digital future is shaping the future of risk assessment. According to [14], six underlying factors impact the advancement of risk assessment:

1. Knowledge, information, and data available for analyzing and computing the risk are continuously growing.

2. Modeling capabilities and computational power are continuously advancing, making more accessible simulations and large-scale data analysis.

3. The increasing complexity of the advancing systems made of heterogeneous elements (hardware, software, human) leads to system behaviors challenging to predict or explain.

4. The risk assessment extends to cover managing risk in a systematic way that includes the occurrence of the risk, prevention, mitigation, emergency crisis management, and restoration [14].

5. Recognition that risk varies over time and accordingly, the effectiveness of the mitigation measures changes.

6. Cyber-physical systems require solid frameworks for safety and security assessment.

Zio [14] highlights that description of the risk and future risk assessment is conditioned on available knowledge. However, it is equally important to address the incomplete knowledge or the unknowns within the risk assessment. According to the available knowledge, Flage et al. [27] classify the events in risk assessment to:

1. Unknown-unknown events that are new and unknown to everyone.

2. Unknown-known events that are new to risk analysts but have been recognized by someone else.

3. Known-unknown events with weak background knowledge and justified indications that a new, unknown type of event or scenario could occur in the future.

4. Known-known events that are known to the analysts performing the risk assessment and for which there is existing evidence.

In autonomous systems that incorporate ML, unknown events require novelty detection and anomaly detection approaches. *Novelty detection* is the task of classifying test data that differ in some respect from the data that are available during training[28]. Anomaly detection detects the anomalies unrelated to the training data[29]. Both anomalies and novelties occur rarely and are dealing with unexpected events in the data. We can argue that the most dangerous events are the unknown ones because otherwise, we can take action to prevent them. Accordingly, Flage et al.[27] argue that known-unknown events are representative of *known risks* that become apparent in *new conditions*. However, the unknown-unknown, unknown-known, and known-known events can be associated with negligible probabilities of occurrence.

Reliability Engineering for UAS

The system reliability engineering and reliability assessment are practical ways to manage risk and support decision-making for safe, reliable, and efficient operation of complex engineering systems[30]. According to[31], *reliability engineering* is an engineering discipline for applying scientific know-how to a component, product, plant, or process in order to ensure that it performs its intended function, without failure, for the required time duration in a specified environment. Reliability engineering involves an iterative process of reliability assessment and improvement, and the relationship between the two processes[32]. Autonomous systems can change their behavior in response to unanticipated events during operation[33]. However, assessing the reliability of an autonomous system varies depending on the autonomy levels of the system. Previous research on autonomy levels includes the work of Huang et al.[34] who developed Autonomy Levels for Unmanned Systems that specifies metrics to assess autonomous systems capabilities. As the enablers of autonomy, the reliability engineering approach to ML algorithms is similar to traditional software reliability assessment. Abstractly, ML performs perception tasks and informed decision-making; thus, most systems that incorporate ML will naturally include standard software components[3]. Reliability growth modeling that characterizes how the reliability of a system increases during testing[3] is one of the standard approaches to software reliability assessment. In ML, the reliability growth measures the accuracy as a fraction of correct predictions divided by a total number of predictions [3]. Consequently, reliability and accuracy in ML are commonly synonymous terms [3].

According to [35], there are four technical components of reliable software:

1. Fault prevention - avoiding faults during design and development of systems through enforcement of good design methods.
2. Fault removal - the process of enforcing formal inspection and testing systems until eliminating all visible faults while not creating any new faults.
3. Fault tolerance - the survival attribute of a system.
4. Fault/failure forecasting - the process of establishing reliability models, failure data, fault/failure relationships, analysis, and interpretation of system behavior.

A reliable system has a capability to function until the system desists under *expected* circumstances. Moreover, a reliable system is a representation of the resilience engineering results.

Resilience Engineering for UAS

Resilience engineering brings together the system safety concepts, reliability of a system, analysis and handling uncertainties, risks, and survivability of a system. According to [1], resilience is the ability to recover quickly after something unpleasant, such as shock or an injury, the ability to return to its original shape. Hollnagel [36], who was at the forefront of resilience engineering, has developed three premises of resilience engineering that showcase limitations and issues in resilience engineering:

1. The conditions of performance are underspecified.
2. Unfavorable events can be attributed to a combination of normal performance uncertainties.
3. Safety management cannot be based on error probabilities and calculations.

These premises demonstrate the limitations within current safety engineering and pose guidelines for the continuous evolution of resilience engineering.

Vachtsevanos et al. [37] illustrate the expected basic functioning of a resilient system through anticipation of undesirable events, the monitoring of performance, and the response to warnings or threats (see Figure 4.1). This kind of system implies proactive measures and readiness to adapt to the variability of

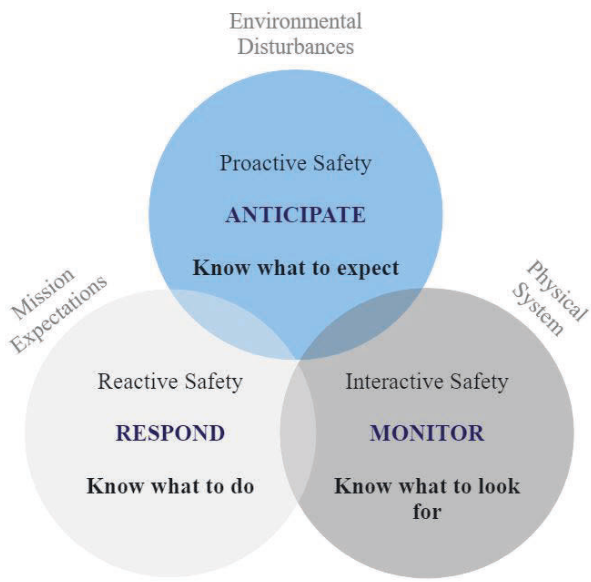

FIGURE 4.1. Basic functions of a resilient system, adapted from[37]

circumstances making it less susceptible to a hazardous environment. As a result, a resilient UAS is flexible and capable of returning to a normal functioning state after experiencing disturbances.

Human-Machine Teaming Perspectives

Human-Machine Teaming (HMT) is a relationship between humans, the machine, and their interdependencies. The goal of HMT is to build trustworthy, transparent, predictable, adaptable, and reliable systems that incorporate artificial intelligence, to create effective human-machine teams[7]. HMT requirements[7] for an adaptable autonomous system include:

1. Multiple options or paths for recovery from a single problem (among which allowing humans to specify problem at different levels of abstraction);
2. On-demand adjustment of autonomy;
3. System degradation and failure resistance (the system shall be tolerant and fail gracefully maintaining its safety[7]).

For highly effective HMT, the most relevant requirements during the development and design stage of the autonomous systems are to ensure safe and effective systems during operations in complex, contested, unanticipated, and dynamic environments[7]. Calibrated trust (i.e., trusting the autonomous system will do what it is supposed to do within a particular environment) and shared understanding (i.e., shared perception between human-to-machine and machine-to-human) are fundamental HMT concerns. A long-term strategy is to achieve an intuitive, shared, and bidirectional information flow between humans and machines[7].

4.1.4 New Warning Identification Framework

This section proposes an early concept of a Warning Identification Framework (WIF) to guide the planning of UAS incorporating ML in addressing the known risks and recognizing the warning signals accordingly (see Figure 4.2). The UAS incorporating ML can understand the operating environment and decide their reactions to the changes in the environment. During asset surveillance, the UAS can detect anomalies in sensor measurements that can suggest possible risks or early warning signals. A risk indicates the possibility of asset disturbance, and a warning signifies the early sign of a disturbance that can require immediate reaction. The anomalous events during remote operations, such as a measured crack on the pipeline during surveillance for the offshore oil and gas industry, can be extreme and unlikely. The rarity of such measurements leads to very little evidence in data. The rare measurement can be dismissed or even unnoticed by the anomaly detection ML algorithm (as discussed by[22]). The autonomous systems' ability to detect warnings or risks is not merely about building a tool; it is about creating a long-term strategy. The UAS needs to have the possibility to react to these warnings when the situation requires it.

The process of development and integration of ML into a system is referred to as the *ML lifecycle*[38]. The ML lifecycle consists of four stages: Data Management, Model Learning, Model Verification (as the activities during which machine-learned models are produced), and Model Deployment (the deployment of ML component along with the other software components in the system)[38]. The Data Management stage is responsible for the acquisition of the data that can be used to predict future data or to perform other kinds of decision making under uncertainty[39]. The planning of the Data Management stage is often underestimated. However, with the trust, reliability, and explainability issues that ML encounters, it is critical to have a clear understanding of the purpose of ML incorporated into a more extensive system. The WIF intends to address trust calibration, errors and biases, and explainability for the UAS that depend on ML algorithms (as shown in Figure 4.2). This framework bases on concepts and mitigation methods from Risk Assessment, Reliability Engi-

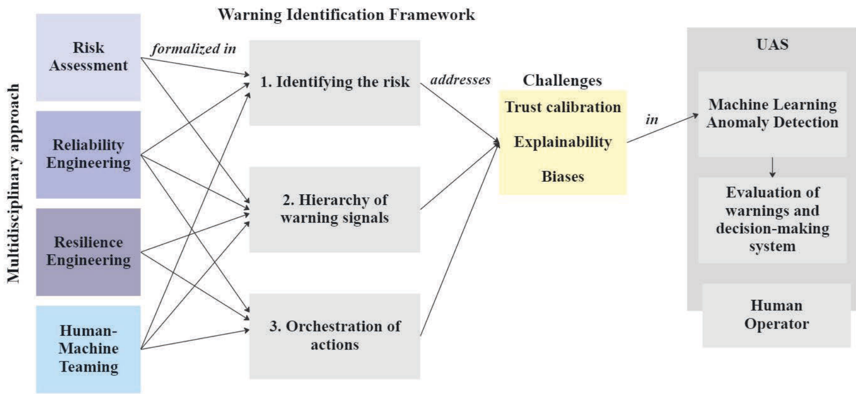

FIGURE 4.2. Multidisciplinary approach formalized in WIF to address challenges in UAS ML, adapted from [40]

neering, Resilience Engineering, and Human-Machine Teaming (as shown in Figure 4.2). Two of the factors of future risk assessment, according to [14], are the recognition of knowledge and data growth and the need for solid frameworks for the safety assessment of cyber-physical systems.

Therefore, the WIF consists of three segments:

1. Identifying the Risk The first step in WIF is identifying the risk of experiencing disturbances in the form of rare anomalies (i.e., concerning pipeline surveillance) from available knowledge, historical insights, and domain expert inputs. In this step, Risk Assessment provides insights into risk definition based on available knowledge [14] and focusing on known events that become apparent in new conditions [27]. Known risks or vulnerabilities provide knowledge on the sequence of events that can lead to the asset or environmental disruption, frequency of occurrence of these events, and consequences of the disruption. These factors are a part of formal characterizations and representations of risk described in [41]. An extended definition, by [42], describes the knowledge of risk through defining the set of disturbance scenarios, set of consequences and, quantified uncertainties. Furthermore, a reliable system is capable of normal functioning under expected or ordinary circumstances. These circumstances are a part of the risk scenario definition. This step requires developing models based on existing knowledge to *identify* risks.

2. Hierarchy of Warning Signals Hirerachization or ranking of the warning signals is a description of the sequence of the events that may evolve into a disturbance that requires immediate intervention. This hierarchy provides the early-to-late-warning evolution of a disturbance by defining the criticality of a warning signal. Adjusting the criticality of warning signals is a part of Risk Assessment. This adjustment allows for fault forecasting, as a characteristic of a reliable system that incorporates the analysis of warning signal relationships. Finally, as a resilient system, analysis and adjustment of the criticality of warning signals allow the system to incorporate a shared understanding of anticipation and monitoring of the disruptions. This step requires domain experts to develop models based on existing knowledge for *describing* risks.

3. Orchestration of Actions Knowing how to respond to the emerging disturbance is one of the critical elements of reliable UAS [26,37]. The orchestration of UAS actions is an essential task in remote operations. This step incorporates the reliable system capabilities to prevent, remove or tolerate the disruptions and a resilient system capacity to respond to the emerging situation. This step satisfies the requirement and expectation of HMT for an autonomous system to adjust the autonomy on demand. The ability for the UAS to systematically and intelligently recognize and act upon warning signals gives the system the capability of being proactive and reactive. A proactive UAS expects and captures weak signals before anomalies occur. A reactive UAS communicates and responds to the emerging situation.

Finally, the three steps of WIF satisfy the HMT requirements for an explainable functioning of a system with a shared understanding of intentions and multiple approaches to a single problem.

Warning Identification Process Inspired by the conceptual model of Process Performance Indicator (PPI) [14], Figure 4.3 illustrates the *Warning Identification Process (WIP)*. PPI reflects on the degree of system objective satisfaction and describes the disruptive events leading to unwanted disruptions of the operation. The WIP demonstrates the development and identification of the warning signals by the UAS, guided by the WIF Hierarchy of Warning Signals. Displayed are stages of warning from *one* to *four*, where *one* represents the earliest stage of the warning sensed by UAS, and *four* represents the latest and recognized warning that requires action. *Anomaly Response State* is the period from when the UAS detects an anomaly until it recognizes it as a warning. During the *Warning Response Trigger*, the UAS takes action and responds to the disturbance following the WIF Orchestration of the actions. Finally, the *Response State* is when the UAS returns to the natural flow of the operation, between the recognized disturbance and a new expected disturbance.

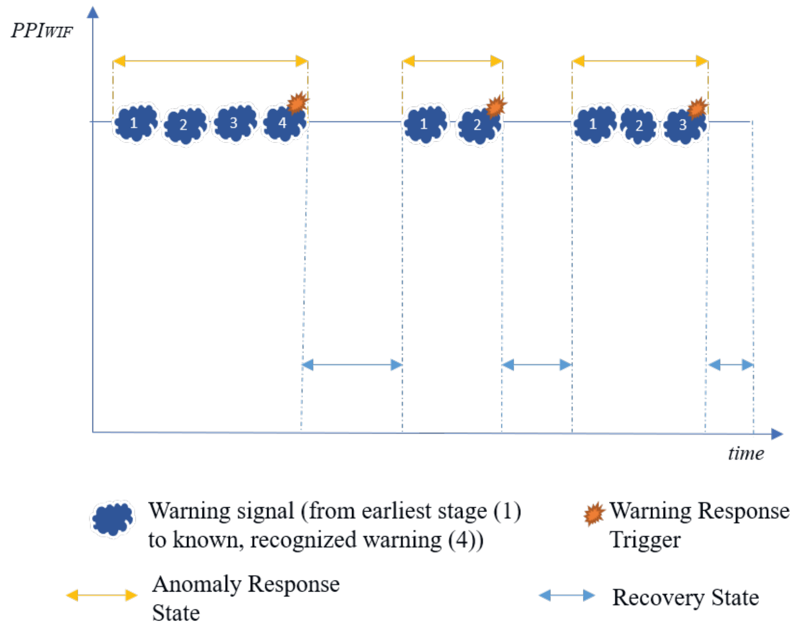

FIGURE 4.3. Warning Identification Process

4.1.5 Application and Contribution Summary

The complexity of algorithms that enable autonomy makes it challenging to control, identify and characterize potential disruptions and react to the consequences. As a future factor in risk assessment,[14] highlights the need for extending the frameworks of risk assessment for complex, interconnected systems that support critical infrastructures. Disruptive events and safety barrier failures typically occur due to degradation processes[14]. Introduction of the condition-monitoring or surveillance data in WIF can give insight into the disruptive process, such as degradation, and prioritize the monitored variables. The WIF can complement the ML processes incorporated in UAS towards condition monitoring, surveillance, and intervention-based risk assessment. The proposed WIF provides a scalable, explainable and structural approach to dynamic risk assessment alongside ML. Examples of case studies for the proposed Warning Identification Framework are presented in Chapters 5 and 6.

4.2 Using Risk Analysis for Anomaly Detection for Enhanced Reliability of Unmanned Autonomous Systems

Abstract

Unmanned Autonomous Systems (UAS) are intended to improve the safety of offshore operations by residing on the seabed and monitoring and inspecting assets and the environment. The UAS can collect and analyze data in real-time through sensor measurements and video analysis, warning onshore operators of data anomalies that indicate potentially hazardous environmental events. Due to the rarity of hazardous events, data on them are scarce, resulting in a data imbalance between normal and anomalous occurrences. Consequently, it has become increasingly challenging for UAS to recognize potentially hazardous circumstances. Thus, the UAS may overlook early warning signs of hazardous events or may overwhelm operators with trivial, resource-intensive information in the form of false alarms. Recent research has addressed data imbalances by simulating underrepresented data, extrapolating it using causal knowledge, or adding parameters to data and methods as a form of semi-supervision. However, in this research, we examine risk analysis as a tool for providing a semi-supervised approach to anomaly detection. We emphasize the overlapping properties of risk analysis and anomaly detection within the objectives of a highly autonomous system. Finally, we apply the derived insights to anomaly detection in sensor data to lower the likelihood of false alarms or missed signals.

4.2.1 *Introduction*

Offshore operations at remote oil and gas platforms rely on support functions such as monitoring, inspection, and maintenance to support the safe and optimal functioning of highly engineered assets and their surrounding environment. These support functions frequently require underwater labor in restricted, low visibility or chemically contaminated environments, requiring continuous resource-intensive response capability [26]. Unmanned autonomous systems (UAS) are intended to take a permanent part in offshore operations, replacing the need for personnel and vessels in remote and potentially dangerous locations and thereby improving remote operations' safety. Equipped with sensors and cameras, the UAS, such as underwater intervention drone, monitor and inspect the assets and environment by residing on the seabed, collecting and analyzing the data in real time. Sensor data from UAS reshapes our perception of the environment and how the offshore industry manages risk by enabling UAS to detect unusual occurrences through irregularities in data, known as anomalies. Anomalies can represent early warning signs of potentially hazardous events, such as material degradation on the pipeline surface, a developing fracture, misplaced objects, or biological growth. The anomalous observations can be used as inputs to risk analysis to identify hazards as potential sources of harm. Adequately detecting the anomalies with anomaly detection and classification methods enables the UAS to react to the emerging situation and activate the autonomous alarm management to alarm the operators if and when the situation requires it. The alarms are used to warn operators of a malfunctioning piece of equipment, a process deviation, or an unanticipated state requiring operator involvement [43]. It is vital for UAS to detect early warning indications of a failing asset or substantial environmental changes, as failure to do so can have a harmful effect on the environment and result in significant financial loss.

Since the hazardous occurrences are relatively rare, there is insufficient evidence or balance in data to adequately identify hazardous occurrences, making it challenging for anomaly detection methods to capture the criticality of potentially rising hazardous events. These methods may overlook the anomalous data by sacrificing the anomalies for efficiency and disregarding them as tolerable collateral damage [12]. On the contrary, because of the inefficient ordering of anomaly criticalities, anomaly detection methods may overwhelm operators with low-significance, resource-wasteful information in form of false alarms. Ideally, the anomaly detection methods would have sufficient understanding or context of the environment to warn the operators of anomalous occurrences as soon as they are detected while minimizing the false alarms [44].

Recent research explores numerous approaches to expanding underrepresented data instances in imbalanced datasets, i.e., through extrapolation with causal knowledge, expanding the underrepresented data with the one reproduced by simulations of rare events, or setting additional parameters within data and methods as a form of semi-supervision. Numerous applications of integrated risk analysis and anomaly detection have been investigated, with anomaly detection results being used as inputs to risk analysis. The frequent application of anomaly detection is the identification of anomalous trends in data to help prevent or reduce the risk of undesired events [45], predict the risks [46], or enhance risk identification [47]. However, the properties of anomaly detection and risk analysis are comparable, and the objectives of these methods often overlap. This overlap is overlooked in recent research.

Therefore, this section introduces a novel approach to anomaly detection by examining risk analysis as a tool for providing a semi-supervised approach to anomaly detection, thereby providing an opportunity to lower the probability of false alarms or missed signals caused by data imbalance. We first compare the weaknesses of existing methods for hazard detection in imbalanced data, such as simulation, extrapolation with causal knowledge, and decision boundaries, and identify the patterns in their advantages and disadvantages. We then search for gaps in which anomaly detection can benefit from existing knowledge resulting from risk analysis. By linking risk analysis to anomaly detection, we concurrently address the absence of risk context in complex anomaly detection methods while minimizing the impact of imbalanced data and thus adhere to the essential EU guidelines for trustworthy intelligent systems [48].

4.2.2 *Motivation and Related Work*

In the context of an unmanned autonomous system, autonomy refers to a system's or component's ability to operate without external control [37]. In contrast to an automated system that performs strictly programmed tasks, an autonomous system decides and executes actions driven by the intended objective, enabled through artificial intelligence (AI) approaches (i.e., ML, computer vision) [26]. Thereby, by creating systems that accomplish tasks that require intelligence when performed by people [49], AI is essential for autonomy. Full autonomy of a system relies on the advancement of enabling technologies and associated data processing. The integral development of autonomy is adapting the control of UAS in critical situations to reliably respond to dynamic environments and provide appropriate detection, identification, and prediction of hazards that may impact the safety of the environment and assets.

Challenges: data and enabling technologies

The goals of achieving a reliable, highly autonomous UAS are impeded by several factors, according to [37]:

1. Significant operator engagement is necessary without a reliable level of autonomy, posing critical new issues in human-machine interfaces and mixed-initiative control.

2. Acquiring high levels of autonomy in an uncertain, unstructured, hazardous, and dynamic environment necessitates using ML techniques, facing numerous systems engineering difficulties.

3. Today's widely used ML techniques are inherently unpredictable. They lack the necessary frameworks to provide proof of safety and predictability, whereas industrial applications require predictable behavior and a strong guarantee of safety.

4. UAS operate in uncertain and noisy environments subjected to hazards, jeopardizing their operational integrity, necessitating the development of methods that ensure resilient and reliable UAS operations.

Additionaly, the rarity of hazardous occurrences leads to their underrepresentation in data, creating imbalanced data. Combined with noisy environments, the methods become biased towards recognizing the broadly represented data instances while discarding the underrepresented ones as noise. However, data recollection is typically an expensive and time-consuming task [40]. As a result, a particular emphasis should be placed on bringing context to data and finding means for supervising or semi-supervising the methods used to handle imbalanced data. The challenge of imbalanced data in anomaly detection systems frequently leads to a distorted perspective of data distribution, resulting in unreliable conclusions. The most common approaches to addressing this challenge are to extrapolate data with artificially generated circumstances through simulations or computations from existing causal and physics knowledge, or to determine decision boundaries by which normal and anomalous data would be distinguished.

Simulating Underrepresented Data Fortunately, hazardous events with high consequences are uncommon. Therefore, simulating the physical world is a widely used technique for collecting missing data describing the high consequence hazardous events to train ML models and address the data imbalance. While the extensive study has been conducted on data collection via simulation, this technique for balancing the data is not always suitable. [50] highlight the complexity of imbalanced data and argue that it is challenging to construct an adequate dataset for training ML models in engineering contexts becuase a system is generally comprised of several sensors that can continuously gather data, most of which is healthy normal data with little to no evidence of hazardous occurrences. Conducting simulations to acquire data on hazardous occurrences can be costly. For instance, researchers need to acquire adequate equipment and computing power to simulate the occurrences, establish a test benchmark, and collect the data. The data on hazardous occurrences obtained by simulation in a virtual environment is often insufficiently practical, as the simulations may not represent the complex physical world in its entirety.

Extrapolating underrepresented data with causal knowledge [51] discuss high-consequence and low-probability scenarios not being well captured by data-driven models and suggest using causal and physics-based knowledge to extrapolate the scarce data. The authors [51] propose separating the components of data-generating processes that are stochastic from those that are deterministic or guided by well-defined principles. Accordingly, stochastic elements can be leveraged to augment the robustness against empirically observed variations for relevant scenarios. Thus, by understanding the deterministic processes that generate data, some physical constraints can be applied and employed to extrapolate beyond the bounds of existing observed data with higher confidence. They suggest that integrating data-driven and causal models to forecast potentially catastrophic events in advance of their occurrence can enhance the precision of real-time risk-based decisions.

Decision boundaries Sensor systems are critical components of modern networked digital infrastructures, such as autonomous systems' environmental monitoring [52]. As a result, a considerable amount of the world's data is in the form of streaming, time-series data, where anomalies provide essential information in critical situations [44]. Since not all potential anomalies are known in advance, most data-driven anomaly detection techniques depend on developing a model of the system's normal behavior. This dependence can potentially ease the occurrence of noise or false alarms during anomaly detection [53].Decision boundaries are frequently seen in classification and supervised algorithms that utilize labeled data [54].

The discussed approaches vary in their disadvantages. Simulation and extrapolation of imbalanced data are computationally and resource-intensive and frequently lack real representation of the physical world. As an alternative, setting decision boundaries lacks the context of underlying causes of anomalous and hazardous occurrences, necessitating extensive iterations to determine the correct boundaries and exposing the model to the risk of overfitting, which can create difficulties during real-time analysis. However, the lack of risk context is the overreaching drawback to the three discussed approaches. This finding leaves an opportunity to combine the strengths of existing approaches and fill in the gap of missing risk context. We can exploit the advantages of utilizing accessible causal knowledge gathered through risk analysis insights and employed in establishing decision boundaries.

4.2.3 Risk Context within Anomaly Detection

[55] defines risk as the effect of uncertainty on objectives. The effect is a deviation from the expected that can be positive, negative or both, and can create or result in opportunities and threats. Risk is usually expressed in terms of risk sources, potential events, their consequences and their likelihood of occurrence [55]. *Risk analysis*, as an element of risk assessment, systematically uses available information to identify hazards and estimate risk. [56] defines hazard as a potential source of harm. Thus, risk analysis can be viewed as a tool for informing future

welfare decision-making, as risk is concerned with what might happen in the future [57]. The majority of systems incorporate barriers that can assist prevent or limit hazardous events and their consequences, encircle and contain the hazard, serve to safeguard the asset and separate the hazard from the asset [58]. Similarly, the identification of a hazard is a critical goal of *anomaly detection* since it plays a vital role in ensuring the reliability and safety of industrial operations by alerting the system of potentially hazardous occurrences in the environment, monitored asset, or system itself.

Overlapping roles in UAS

Because the oil and gas industry prioritizes safety during remote operations, we can consider safe operations an overarching goal of a UAS operation. Risk analysis, anomaly detection, and autonomous operations generally follow the same process: analyzing data, identifying and assessing risk/anomaly, determining the criticality of risk/anomaly, and determining appropriate mitigation steps. The common objective and overlapping processes of risk analysis, anomaly detection, and high-autonomy UAS operation, such as during pipeline inspection and monitoring, involve reporting the risk analysis, and deciding the best risk management policies. While risk analysis would provide information on potential hazards derived from causal knowledge (i.e., from material engineers), the anomaly detection method would look for irregular patterns analyzed during video inspection of the material surface. Risk analysis incorporates the analysis of barriers that encircle and divide the hazard from the safe operation and can be used to define decision boundaries within anomaly detection methods, encircling and distinguishing normal data from anomalous data. Additionally, risk analysis provides insight into the sequence of events leading up to a hazardous occurrence, as well as the frequency and likelihood of that occurrence, which can assist in differentiating between significant and non-significant anomalies. Similarly, consequence analysis that stems from risk analysis can aid in determining the criticality of observed anomalies and the appropriate mitigation actions. Therefore, the potential of risk analysis for reliable anomaly detection is substantial.

4.2.4 Contribution Summary

Figure 4.4 illustrates the integration of high autonomy, data deliberation, and risk analysis that models an objective-oriented (perceptive), risk-oriented (reliable), and independent (predictable) UAS. Concentrating on high autonomy enables independence from operators, which frees up resources and boosts the potential to manage complex missions involving vast volumes of data. Data

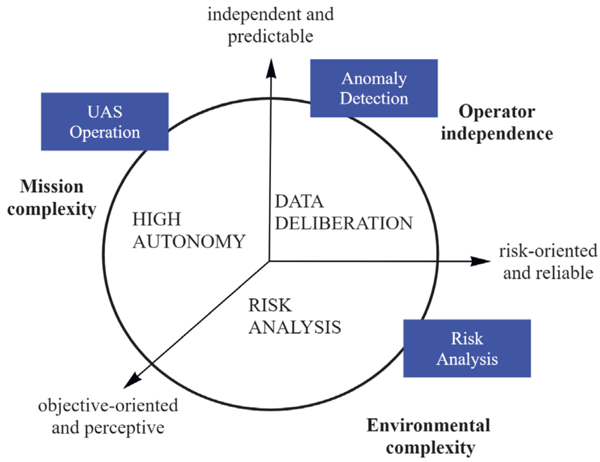

FIGURE 4.4. Reliable UAS

deliberation entails analyzing and comprehending the vast volumes of data using AI approaches, mainly ML and anomaly detection. Risk analysis brings context and reliability to autonomous operations, bridging the objective and risk-oriented operations, thus balancing safety and efficiency. To achieve higher predictability and thus reliability, the essential task is to use the derived knowledge from risk analysis to identify the criticality of detected anomalies while avoiding shortcomings of anomaly detection methods. Major shortcoming involves overwhelming the system with noise, and minor to insignificant data in the form of false alarms, resulting in wasted resources (i.e., insignificant biological growth on pipeline surfaces). Contrastingly, the method may favor the efficiency while classifying anomalies as noise and averting potentially relevant data that might indicate early hazardous occurrence (i.e., early surface fracture). Anomalies that overwhelm the data become a part of normal data; hence anomaly detection methods may camouflage such occurrences [44,59] (i.e., substantial material degradation on pipeline surface).

To utilize existing knowledge and semi-supervise the anomaly detection methods, we propose using the existing knowledge on identified hazards through RA. Figure 4.5 illustrates the benefits of RA task insights in AD processes for efficient criticality detection and thus decision making by the UAS. The information derived from hazard labeling aids anomaly detection in distinguishing noise or false alarms from more meaningful information, such as hazards. Furthermore, criticality labeling derived from risk analysis aids in classifying hazards, enabling the efficient prioritization of mitigative actions by UAS. This process initiates the potential of heavily data-oriented methods, such as AD, to result in more educated and less presumptuous predictions.

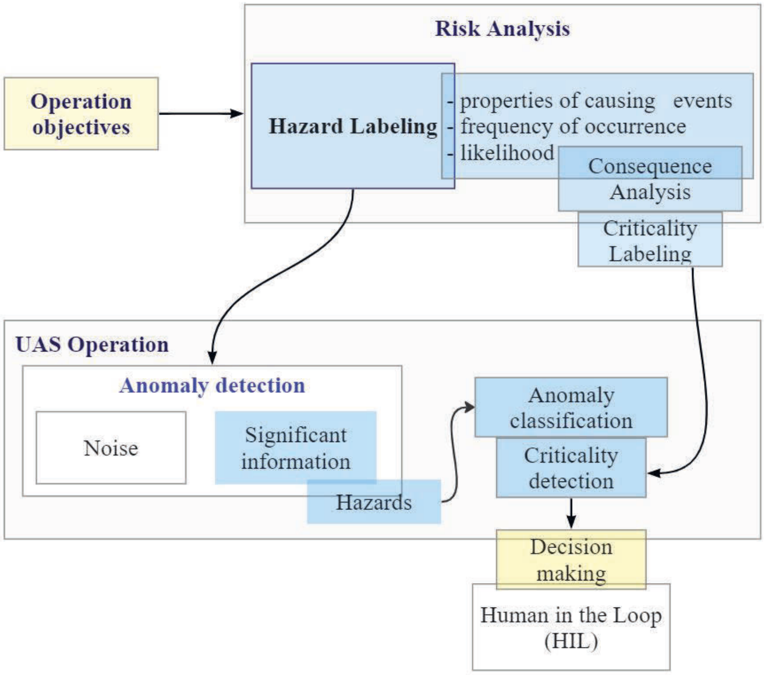

FIGURE 4.5. Methodology scheme, adapted from [60]

4.3 Conclusions and Key Contributions

This section highlights the key contributions and concludes the chapter and the presented articles.

Implementing ML techniques in a standardized practice that incorporates reliability is still a matter of early development. During remote UAS operations, any unintended misbehaviors of the UAS can have severe environmental and financial consequences. With an increase in UAS employment in remote offshore operations, we observe a noticeable need to validate and improve the ML processes that enable autonomy, further supporting critical decisions during the UAS operations. The proposed framework, the Warning Identification Framework, attempts to improve the warning signal detection of UAS during remote operations, address the shared understanding of UAS ML intentions, and prevent unintentional consequences of ML.

Anomalous observations can be abundant in datasets due to the extensive use of sensors and, as a result, vast data collection. Therefore, they may be a source of noise in data analysis. Due to the abundance of unlabeled and biased datasets that lack context, training autonomous systems to detect anomalies with meaningful insight while minimizing false alarms is becoming exceedingly challenging. Despite ample research on anomaly detection, there is a gap in knowledge on their relationship to overlapping risk analysis processes. The existing approaches to addressing this challenge, through simulation, data extrapolation, and decision boundaries, show great potential but lack the necessary context to distinguish significant from insignificant anomalies. Simultaneously, recent research singles out the need for more trustworthy autonomous systems while highlighting the lack of explainable and contextual ML methods.

The key contributions of the presented articles are the identification of components from Risk Assessment, Reliability Engineering, Resilience Engineering, and Human-Machine Teaming that aid in comprehending the directions for addressing the common data-driven challenges associated with hazard identification. The contributions of these articles tackle the critical shortcomings of data-driven methods: the complexity of algorithms, difficulty to explain or reason results, and to integrate risk assessment. This creates the opportunities to design risk-informed data-driven models and applications. These findings have shaped the knowledge of some of the critical obstacles for anomaly detection employed in UAS and led to shaping a conceptual risk-informed warning identification framework as a novel approach to observing well-known shortcomings in using AI-based and data-driven systems for remote operations.

4.4 References

[1] Oxford University Press. Oxford Learner's Dictionaries (2021). URL https://www.oxfordlearnersdictionaries.com/. Cited on page/s 50, 51, 56, 59.

[2] Hui-Min Huang. Autonomy levels for unmanned systems (ALFUS) framework. Technical Report September National Institute of Standards and Technology Gaithersburg (2004). URL https://doi.org/10.6028/NIST.sp.1011-I-2.0. Cited on page/s 50.

[3] Aiden Gula, Christian Ellis, Saikath Bhattacharya, and Lance Fiondella. Software and system reliability engineering for autonomous systems incorporating machine learning. In *Proceedings - Annual Reliability and Maintainability Symposium* volume 2020-Janua (2020). ISBN 9781728136899. doi: 10.1109/RAMS48030.2020.9153595. Cited on page/s 50, 51, 58, 59.

[4] Victor Galaz, *et al.* Machine intelligence, systemic risks, and sustainability. Technical Report 274 Stockholm, Sweden (2021). Cited on page/s 50, 54.

[5] Fil Macias. The Test and Evaluation of Unmanned and Autonomous Systems. *ITEA Journal* **29**, 388–395 (2008). Cited on page/s 50, 56.

[6] Kazuo Okamura and Seiji Yamada. Calibrating Trust in Autonomous Systems in a Dynamic Environment. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society* pages 1–6 (2020). Cited on page/s 51, 52.

[7] Patricia Mcdermott, Cindy Dominguez, Nicholas Kasdaglis, Matthew Ryan, Isabel Trahan Mitre, and Alexander Nelson. Human-Machine Teaming Systems Engineering Guide. Technical report The MITRE Corporation (2018). URL https://www.mitre.org/publications/technical-papers/human-machine-teaming-systems-engineering-guide. Cited on page/s 51, 60, 61.

[8] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1** (5), 206–215 (2019). ISSN 25225839. doi: 10.1038/s42256-019-0048-x. URL http://dx.doi.org/10.1038/s42256-019-0048-x. Cited on page/s 51, 52, 53, 54.

[9] Harini Suresh and John V. Guttag. A framework for understanding unintended consequences of machine learning. In *Equity and Access in Algorithms, Mechanisms, and Optimization* (2019). doi: https://doi.org/10.48550/arXiv.1901.10002. Cited on page/s 51, 55.

[10] S O Johnsen, T Bakken, A A Transeth, S Holmstrøm, M Merz, E I Grøtli, S R Jacobsen, and R Storvold. Safety and security of drones in the oil and gas industry. Technical report Equinor (2019). Cited on page/s 51.

[11] International Electrotechnical Commission. IEC 60050 (2 2015). URL https://www.electropedia.org/iev/iev.nsf/. Cited on page/s 52.

[12] Karima Makhlouf, Sami Zhioua, and Catuscia Palamidessi. On the applicability of ML fairness notions. *arXiv* pages 1–32 (2020). ISSN 23318422. Cited on page/s 54, 66.

[13] Geoff Pleiss, Manish Raghavan, Felix Wu, Jon Kleinberg, and Kilian Q. Weinberger. On fairness and calibration. *Advances in Neural Information Processing Systems* **2017-Decem** (Nips), 5681–5690 (2017). ISSN 10495258. Cited on page/s 54.

[14] E. Zio. The future of risk assessment. *Reliability Engineering and System Safety* **177** (March), 176–190 (2018). ISSN 09518320. doi: 10.1016/j.ress.2018.04.020. Cited on page/s 55, 56, 57, 62, 63, 64.

[15] Michele Compare, Fabio Martini, Sara Mattafirri, Fausto Carlevaro, and Enrico Zio. Semi-Markov Model for the Oxidation Degradation Mechanism in Gas Turbine Nozzles. *IEEE Transactions on Reliability* **65** (2), 574–581 (2016). ISSN 00189529. doi: 10.1109/TR.2015.2506610. Cited on page/s 56.

[16] Zhiguo Zeng, Rui Kang, and Yunxia Chen. Using PoF models to predict system reliability considering failure collaboration. *Chinese Journal of Aeronautics* **29** (5), 1294–1301 (2016).

ISSN 10009361. doi: 10.1016/j.cja.2016.08.014. URL http://dx.doi.org/10.1016/j.cja.2016.08.014. Cited on page/s 56.

[17] Dario De Dominicis and Domenico Accardo. Software and sensor issues for autonomous systems based on machine learning solutions. In *2020 IEEE International Workshop on Metrology for AeroSpace, MetroAeroSpace 2020 - Proceedings* pages 545–549 (2020). ISBN 9781728166360. doi: 10.1109/MetroAeroSpace48742.2020.9160292. Cited on page/s 56.

[18] Christoph A. Thieme and Ingrid B. Utne. Safety performance monitoring of autonomous marine systems. *Reliability Engineering and System Safety* **159** (March 2016), 264–275 (2017). ISSN 09518320. doi: 10.1016/j.ress.2016.11.024. URL http://dx.doi.org/10.1016/j.ress.2016.11.024. Cited on page/s 56.

[19] Brad Cline, Radu Stefan Niculescu, Duane Huffman, and Bob Deckel. Predictive maintenance applications for machine learning. *Proceedings - Annual Reliability and Maintainability Symposium* (2017). ISSN 0149144X. doi: 10.1109/RAM.2017.7889679. Cited on page/s 56.

[20] Yi Ping Fang and Enrico Zio. An adaptive robust framework for the optimization of the resilience of interdependent infrastructures under natural hazards. *European Journal of Operational Research* **276** (3), 1119–1136 (2019). ISSN 03772217. doi: 10.1016/j.ejor.2019.01.052. Cited on page/s 56.

[21] Jinhao Zhang, Mi Xiao, and Liang Gao. An active learning reliability method combining Kriging constructed with exploration and exploitation of failure region and subset simulation. *Reliability Engineering and System Safety* **188** (January), 90–102 (2019). ISSN 09518320. doi: 10.1016/j.ress.2019.03.002. URL https://doi.org/10.1016/j.ress.2019.03.002. Cited on page/s 56.

[22] Kun Liu and Huadong Ma. Exploring background-bias for anomaly detection in surveillance videos. *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia* pages 1490–1499 (2019). doi: 10.1145/3343031.3350998. Cited on page/s 56, 61.

[23] S. Elbaum, S. Kanduri, and A. A. Amschler. Anomalies as precursors of field failures. *Proceedings - International Symposium on Software Reliability Engineering, ISSRE* **2003-Janua**, 108–118 (2003). ISSN 10719458. doi: 10.1109/ISSRE.2003.1251035. Cited on page/s 56.

[24] Olumuyiwa Ibidunmoye, Francisco Hernández-Rodriguez, and Erik Elmroth. Performance anomaly detection and bottleneck identification. *ACM Computing Surveys* **48** (1), 1–35 (2015). ISSN 15577341. doi: 10.1145/2791120. Cited on page/s 56.

[25] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. *arXiv* pages 6479–6488 (2018). ISSN 23318422. Cited on page/s 56.

[26] Francesco Scibilia, Knut Sebastian Tungland, Anders Røyrøy, and Marianne Bryhni Asla. Energy industry perspective on the definition of autonomy for mobile robots. In *Communications in Computer and Information Science* volume 1056 CCIS pages 90–101. Springer International Publishing (2019). ISBN 9783030356637. doi: 10.1007/978-3-030-35664-4{_}9. URL http://dx.doi.org/10.1007/978-3-030-35664-4_9. Cited on page/s 57, 63, 65, 67.

[27] R. Flage and T. Aven. Emerging risk - Conceptual definition and a relation to black swan type of events. *Reliability Engineering and System Safety* **144**, 61–67 (8 2015). ISSN 09518320. doi: 10.1016/j.ress.2015.07.008. Cited on page/s 57, 58, 62.

[28] Marco A.F. Pimentel, David A. Clifton, Lei Clifton, and Lionel Tarassenko. A review of novelty detection. *Signal Processing* **99**, 215–249 (6 2014). ISSN 01651684. doi: 10.1016/j.sigpro.2013.12.026. Cited on page/s 58.

[29] Marc Masana, Idoia Ruiz, Joan Serrat, Joost van de Weijer, and Antonio M. Lopez. Metric Learning for Novelty and Anomaly Detection. *arXiv* (2018). URL http://arxiv.org/abs/1808.05492. Cited on page/s 58.

[30] Ajit Kumar Verma, Srividya Ajit, and Durga Rao Karanki. Reliability and Safety Engineering. Springer-Verlag London 2 edition (2016). ISBN 978-1-4471-6269-8. doi: 10.1007/978-1-4471-6269-8. URL https://www.springer.com/gp/book/9781447162681.

Cited on page/s 58.

[31] D.R. Kiran. Reliability Engineering. In *Total Quality Management* pages 391–404. Elsevier (1 2017). doi: 10.1016/B978-0-12-811035-5.00027-1. URL https://linkinghub.elsevier.com/retrieve/pii/B9780128110355000271. Cited on page/s 58.

[32] Butterworth-Heinemann. Chapter 7 - Reliability Engineering. In Sam Mannan, editor, *Lees' Loss Prevention in the Process Industries (Fourth Edition)* pages 131–203. Butterworth-Heinemann Oxford fourth edi edition (2012). ISBN 978-0-12-397189-0. doi: https://doi.org/10.1016/B978-0-12-397189-0.00007-0. URL https://www.sciencedirect.com/science/article/pii/B9780123971890000070. Cited on page/s 58.

[33] David Watson and David Scheidt. Autonomous systems. *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)* **26**, 368–376 (2005). Cited on page/s 58.

[34] Hui Min Huang, Kerry Pavek, Brian Novak, James Albus, and Elena Messina. A framework for Autonomy Levels for Unmanned Systems (ALFUS). *AUVSI's Unmanned Systems North America 2005 - Proceedings* pages 849–863 (6 2005). Cited on page/s 58.

[35] Michael R. Lyu. Handbook of Software Reliability Engineering volume 37. The McGraw-Hill Companies, Inc. United States of America (1996). ISBN 0070394008. URL http://linkinghub.elsevier.com/retrieve/pii/S0167923603000204. Cited on page/s 59.

[36] Erik Hollnagel. Resilience Engineering: A New Understanding of Safety. *Journal of the Ergonomics Society of Korea* **35** (3), 185–191 (6 2016). ISSN 1229-1684. doi: 10.5143/jesk.2016.35.3.185. URL http://dx.doi.org/10.5143/JESK.2016.35.3.185http://jesk.or.krelSSN:2093-8462. Cited on page/s 59.

[37] George Vachtsevanos, Benjamin Lee, Sehwan Oh, and Michael Balchanos. Resilient Design and Operation of Cyber Physical Systems with Emphasis on Unmanned Autonomous Systems. *Journal of Intelligent and Robotic Systems: Theory and Applications* **91** (1), 59–83 (7 2018). ISSN 15730409. doi: 10.1007/s10846-018-0881-x. URL https://doi.org/10.1007/s10846-018-0881-x. Cited on page/s 59, 60, 63, 67.

[38] Rob Ashmore and Colin Paterson. Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges. Technical report ACM Comput. Surv. (5 2019). Cited on page/s 61.

[39] Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. Technical report The MIT Press (2012). Cited on page/s 61.

[40] Rialda Spahic, Vidar Hepso, and Mary Ann Lundteigen. Reliable Unmanned Autonomous Systems: Conceptual Framework for Warning Identification during Remote Operations. In *ISSE 2021 - 7th IEEE International Symposium on Systems Engineering, Proceedings* Vienna (9 2021). Institute of Electrical and Electronics Engineers Inc. ISBN 9781665431682. doi: 10.1109/ISSE51541.2021.9582534. Cited on page/s 62, 68.

[41] Stanley Kaplan and B. John Garrick. On The Quantitative Definition of Risk. *Risk Analysis* **1** (1), 11–27 (1981). doi: 10.1111/j.1539-6924.1981.tb01350.x. URL https://doi.org/10.1111/j.1539-6924.1981.tb01350.x. Cited on page/s 62.

[42] Terje Aven and Ortwin Renn. Risk Management. In *Risk Management and Governance: Concepts, Guidelines and Applications* pages 121–158. Springer Berlin Heidelberg Berlin, Heidelberg (2010). ISBN 978-3-642-13926-0. doi: 10.1007/978-3-642-13926-0{_}8. URL https://doi.org/10.1007/978-3-642-13926-0_8. Cited on page/s 62.

[43] Stig Ole Johnsen, Trond Bakken, Aksel Andreas Transeth, Sture Holmstrøm, Mariann Merz, and Esten Ingar Grøtli. Safety and security of drones in the oil and gas industry. In Francesco Di Maio Piero Baraldi and Enrico Zio, editors, *30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference (ESREL2020 PSAM15)* pages 811 – 818 Singapore (2020). Research Publishing, Singapore. doi: 10.3850/978-981-14-8593-0. Cited on page/s 66.

[44] Alexander Lavin and Subutai Ahmad. Evaluating Real-time Anomaly Detection Algo-

rithms - the Numenta Anomaly Benchmark. In *IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015* pages 38–44 Miami, Florida, USA (2015). Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICMLA.2015.141. Cited on page/s 66, 69, 71.

[45] Muhammad Fahim and Alberto Sillitti. Anomaly Detection, Analysis and Prediction Techniques in IoT Environment: A Systematic Literature Review. *IEEE Access* 7, 81664–81681 (2019). ISSN 21693536. doi: 10.1109/ACCESS.2019.2921912. Cited on page/s 66.

[46] Mahsa Salehi and Lida Rashidi. A Survey on Anomaly detection in Evolving Data. *ACM SIGKDD Explorations Newsletter* **20** (1), 13–23 (2018). ISSN 1931-0145. doi: 10.1145/3229329.3229332. URL https://dl.acm.org/doi/abs/10.1145/3229329.3229332. Cited on page/s 66.

[47] Kevin Sheridan, Tejas G. Puranik, Eugene Mangortey, Olivia J. Pinon, Michelle Kirby, and Dimitri N. Mavris. An application of dbscan clustering for flight anomaly detection during the approach phase. *AIAA Scitech 2020 Forum* **1, F** (2020). doi: 10.2514/6.2020-1851. Cited on page/s 66.

[48] Tambiama Madiega. EU guidelines on ethics in artificial intelligence: Context and implementation PE640.163 (9 2019). URL https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2019)640163. Cited on page/s 67.

[49] Ray. Kurzweil and Diane Jaroch. The age of intelligent machines. MIT Press Cambridge, MA, US (1990). ISBN 978-0-262-11121-8. Cited on page/s 67.

[50] Tianci Zhang, Jinglong Chen, Fudong Li, Kaiyu Zhang, Haixin Lv, Shuilong He, and Enyong Xu. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Transactions* **119**, 152–171 (1 2022). ISSN 0019-0578. doi: 10.1016/J.ISATRA.2021.02.042. Cited on page/s 68.

[51] Simen Eldevik and Frank Borre Pedersen. AI + safety - DNV. Technical report DNV (2018). URL https://www.dnv.com/oilgas/download/artificial-intelligence-ai-and-safety.html. Cited on page/s 68.

[52] L. Erhan et al. Smart anomaly detection in sensor systems: A multi-perspective review. *Information Fusion* **67** (September 2020), 64–79 (2021). ISSN 15662535. doi: 10.1016/j.inffus.2020.10.001. Cited on page/s 69.

[53] Peng Li, Oliver Niggemann, and Barbara Hammer. On the identification of decision boundaries for anomaly detection in CPPS. In *Proceedings of the IEEE International Conference on Industrial Technology* volume 2019-Febru pages 1311–1316. Institute of Electrical and Electronics Engineers Inc. (2 2019). ISBN 9781538663769. doi: 10.1109/ICIT.2019.8754997. Cited on page/s 69.

[54] Salima Omar, Asri Ngadi, and Hamid H. Jebur. Machine Learning Techniques for Anomaly Detection: An Overview. *International Journal of Computer Applications* **79** (2), 33–41 (10 2013). doi: 10.5120/13715-1478. Cited on page/s 69.

[55] ISO 31000. Risk management — Guidelines, International Organization for Standardization. Technical report International Organization for Standardization (2018). URL https://www.iso.org/obp/ui/iso:std:iso:31000:ed-2:v1:en. Cited on page/s 69.

[56] ISO:51. Safety aspects - Guidelines for their inclusion in standards ISO/IEC Guide 51:2014(E). Technical report International Organization for Standardization and the International Electrotechnical Commissio (2014). Cited on page/s 69.

[57] Marvin Rausand. Risk Assessment Theory, Methods, and Applications. John Wiley & Sons Inc Hoboken, New Jersey (2011). ISBN 9780470637647. doi: 10.1002/9781118281116. Cited on page/s 70.

[58] Snorre Sklet. Safety barriers: Definition, classification, and performance. *Journal of Loss Prevention in the Process Industries* **19** (5), 494–506 (9 2006). ISSN 0950-4230. doi: 10.1016/J.JLP.2005.12.004. Cited on page/s 70.

[59] Len Feremans et al. Pattern-Based Anomaly Detection in Mixed-Type Time Series. *Machine Learning and Knowledge Discovery in Databases* **11906**, 240–256 (2020). ISSN 16113349. doi: 10.1007/978-3-030-46150-8{_}15. Cited on page/s 71.

[60] Rialda Spahic, Hepsø, Vidar, and Mary Ann Lundteigen. Using Risk Analysis for Anomaly Detection for Enhanced Reliability of Unmanned Autonomous Systems. In Maria Chiara Leva, Edoardp Patelli, Luca Podofillini, and Simon Wilson, editors, *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)* pages 273–280 Singapore (2022). Research Publishing, Singapore. doi: 10.3850/978-981-18-5183-4{_}R08-03-390-cd. URL https://rpsonline.com.sg/rps2prod/esrel22-epro/html/toc.html. Cited on page/s 72.

Chapter 5

Reliability of Sensor Data for Machine Learning

This chapter is based on the following two articles:

- Spahic, Rialda; Lundteigen, Mary Ann, *Manually or Autonomously Operated Drones: Impact on Sensor Data towards Machine Learning* IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (2022)
  DOI: 10.1109/CIVEMSA53371.2022.9853685

- Spahic, Rialda; Hepsø, Vidar; Lundteigen, Mary Ann. *A Novel Warning Identification Framework for Risk-Informed Anomaly Detection*, *Springer Nature* Journal of Intelligent and Robotic Systems (June, 2023)
  DOI: 10.1007/s10846-023-01887-2

All authors contributed to the research conception of the two articles. Rialda Spahic performed material preparation, literature analysis, data analysis and manuscript writing. Mary Ann Lundteigen performed writing reviews and supervision of all prior drafts of the manuscript. Vidar Hepsø contributed to the literature gathering and concept visualisation of the research for the second listed article.

Analyzing sensor-collected data during remote operations and identifying the types of obstacles that methods and data may face is one of the preliminary tasks required to investigate the reliability of machine learning. The first article presented in this chapter compares the data collected by underwater drones operated manually and autonomously to determine the differences in data quality. The second article examines a set of seismic data and a chosen anomaly detection method to determine whether reported anomalies are hazards or noise by comparing the anomalies to hazard assessment methods provided by domain exports for the analyzed dataset.

5.1 Manually or Autonomously Operated Drones: Impact on Sensor Data towards Machine Learning

Abstract

The growing need for autonomous systems in offshore industries has contributed to the increased use of machine learning methods. These systems promise to improve safety in operations. However, the methods as enablers of autonomy are susceptible to various failures while interpreting data and making decisions. Several studies have highlighted the lack of research on the reliability and resilience of autonomous systems powered by these standard methods. Recent research provides sets of data interpretation methods. Despite the popularity of machine learning, there is a significant drop in knowledge when these methods result in failures. These failures further support autonomous systems in making wrong decisions. For autonomous systems, resilience and safety management should be an integrated functionality for recovery from risky situations and reporting of incidents. This research proposes an overview of machine learning methods for interpreting sensor data captured by drones operated manually and autonomously. We apply Isolation Forest for anomaly detection analysis and evaluate the Decision tree, Random forest, kNN, Logistic Regression, SVM, and, Naive Bayes for classification analysis. The methods are chosen based on their adequacy and comparative research prevalence. Comparison between the two drone operation modes contributes to understanding the reliability level for autonomously collected data. This research's results provide an evaluation of machine learning methods' performance across sensor data.

5.1.1 *Introduction*

Autonomous systems (AS) have shown potential in enhancing safety in the industry by replacing human activities during dangerous operations in remote environments. These systems can perform tasks that require little to no human intervention by actively interpreting the real-time collected data. Timely decisions based on previously learned knowledge come from historical insights and domain expert inputs. There is a considerable number of machine learning (ML) methods that enable autonomy. However, these methods need to be reliable and trusted within safety-critical circumstances. The potential of ML depends on the data that the AS collects. In particular, sensor data can be overwhelming for the methods to answer with desired results. This data is most often varying from sound, video, image, pressure, temperature, and gas sensor measurements. The environment can potentially overwhelm the collected data with extensive noise that impacts the final decisions and results that AS provide.

During operations and asset surveillance, researchers are often interested in occurrences in the environment that are distinct from expected operating times, such as the presence of material degradation, misplaced objects, or biological growth. It is possible that during ML analysis, considerable amounts of data would lack distinct samples that are interesting to research. Therefore, methods can discard the data outside of the ordinary as noise. It is important to curate the data to avoid disregarding vital information hidden in the noise. Over the last decade, the industry interest in employing AS to perform tasks and reduce human efforts has continuously increased.

This section compares ML performance on the sensor data collected by a manual and an autonomously operated underwater drone. Data collected by the same drone under different operating modes can widen our comprehension of autonomy dependability. We analyze the data through anomaly detection and classification methods. Anomaly detection identifies abnormalities in the data, contributing to fault prevention and predictive maintenance [1]. Classification methods, known as classifiers, are learning tasks that predict the data category of given data points. By applying these methods to the manually and autonomously collected sensor data, we build machine learning models that provide us with an insight into the reliability of the methods that enable autonomy.

5.1.2 Motivation and Related Work

There is a considerable amount of research on finding the best methods to evaluating sensor data. Related work of this section presents semi-organized comparative research. We single out the studies within the context of autonomous systems and highlight the motivation of using ML methods. In the following paragraphs, we discuss the sensor data analysis challenges through applied anomaly detection and predictive capabilities of classification methods in AS.

Anomaly Detection in Sensor Data

As earlier mentioned, environments in which drones operate can be noisy and disruptive. Anomaly or outlier detection is an important research area that contributes to fault prevention and predictive maintenance [1]. Erhan et al.[1] review anomaly detection methods employed in sensor systems. Authors highlight the data volumes, network efficiencies, information fusion, and biases as some of the anomaly detection challenges. Due to the ample employment of sensors in smart devices such as the Internet of Things, the authors [1] argue that the sensor systems have become dominant generators of data. The authors identify different anomaly detection methods. Their research contributes to the study of sensor systems constraints and their impact on machine learning and anomaly detection. Authors also classify anomalies based on their source. These are typically sensor recordings that are distinct from expected behavior. Erhan et al.[1] point out that real-world data is necessary to validate the effectiveness of anomaly detection methods. However, anomalies occur unexpectedly and can be scarce in real-world data. Therefore, it can be challenging to generate them artificially [1].

Anwar *et al.*[2] propose a novel ML framework using feature extraction and SVM with varying kernels. The motivation behind their research is to eliminate disruptive sounds such as birds, airplanes, or thunderstorms as anomalies. This elimination would provide the detection of nearby amateur drones more accurate. Authors approach the problem by gathering real-time acoustic data and classifying the noise with Mel frequency cepstral coefficients, Linear predictive cepstral coefficients, and SVM. SVM has proved to be an efficient method for classifying noisy environments using small batches of data. In this research, ML promised a cost-effective and accurate tool with minimized chances of misclassification between classes [2].

Classification Methods in Sensor Data

Increased interest in autonomous systems has led to an increase in the use of machine learning methods. Choi and Cha[3] explore the application of traditional ML methods employed in Unmanned Aerial Vehicles (UAVs) for autonomous operations. The authors explain that the collected data can show the method's performance more realistically when the testing environment is heterogeneous, consisting of various operational circumstances. They also advise testing the models in smaller batches of non-ideal settings to track AUVs' performance under disturbances.

Moustafa and Jolfaei[4] propose an autonomous Intrusion Detection Scheme (IDS) for real-time complex attack scenarios from drone networks. They use the predictive capabilities of ML for autonomous detection of malicious events in drone networks. Their research compares the following methods to classify cyber-attacks in drone networks: Decision tree, k-Nearest Neighbor (kNN), Naive Bayes, Support Vector Machine (SVM), and Deep learning Multi-layer Perceptron. The authors have synthetically created three different attack scenarios for testing vulnerabilities and recognizing attacks on time. The authors [4] depicted a concept of targeted awareness towards different settings and in-

volved detecting false alarm samples. In this research, the Decision tree has proven to be the best classifier, followed by multi-layer perceptron and kNN. The least performing classifier was Naive Bayes. The authors mention the opportunity to extend their research for more complex networks, simulating more sophisticated attacks. However, there is a usual lack of justification of the method failure in the context of their model.
The sensor data collected by the AS can be challenging to evaluate. Consequently, the employment of machine learning has a remarkable impact on the performance and efficiency of the AS. De Dominicis and Accardo[5] highlight that among the benefits of machine learning, there are three main issues: security, certification, and cost. To establish strong evidence behind these methods, the authors[5] suggest having a quantitative assessment about the system performance after introducing ML. The advice is to carry out a benchmarking analysis comparing 'novel methods' with the traditional solutions. The comparison should encompass prediction capabilities, robustness, integrity, and reliability.

5.1.3 Data and Methods

Research Purpose and Expectations

This research analyzes the differences between the data collected by a manual and autonomous drone operation by applying anomaly detection and classification methods. The expected result is that the differences between manual and autonomous operations are minimal due to the same sensors and pre-programmed mission plans. Therefore, the autonomously operated drone should provide the same level of reliability as the manually operated drone. The second expectation is that the non-linear classifiers that perform well on high dimensional and correlated data, such as Random Forest, will prevail over the linear methods (SVM, Logistic Regression, and Naive Bayes) due to the dataset's dimensionality, non-variability, and imbalance.

Research Data

The data used for this research, collected by Castellini *et al.*[6], is multivariate data containing sensor measurements of six data acquisition campaigns performed by underwater drones for water monitoring. The authors explored lakes and rivers of different locations in Spain and Italy for data collection. There are 11 monitored features in the dataset that results in 20,187 total samples. We have separated these features into general information of the area, water measurements, and drone measurements (see Table I). The available information of the site contains area location and time during drone exploration. Water mea-

surements are specific sensor data that monitor water temperature, dissolved oxygen in the water, and electrical conductivity. Finally, drone-specific measurements monitor the drone's internal state, such as battery voltage, signals to propellers, and direction of the drone's bow.
Additionally, the data of each campaign is labeled by [6] based on the drone operation status, drone curving, location in the water, and the status of the water flow. These four labels consist of set values. In the dataset, each value is represented by a number:

- Drive values: autonomous (2) , manual (1), unlabeled (0);
- Flow values: upstream (3), downstream (2), no water flow (1), unlabeled (0);
- Curves values: turning (2), no turning (1) ;
- Water values: out of water (2), in the water (1), unlabeled (0);

For this research, we merged the six data acquisition campaigns into one complete dataset. We then divided the complete dataset based on the 'Drive' label into manual and autonomous datasets. Division by drive allows exploring the data collected by the drone when it is manually operated and compare it to the data collected during autonomous operation.
We select the Flow label as the *ground truth* (GT). GT is a measurement that classification methods predict. The GT Flow contains values for water monitoring that yield essential contextual information for the domain experts [7].
The sensor data represented in this section is non-varied data with limited sensor inputs collected by simple drone missions. Additionally, the ground truth consists of a more significant number of samples within upstream and unknown water flow values than downstream and no water flow values. The imbalanced representation of the ground truth values in this data can restrict the performance of ML models.

Research Methods

Different methods allow the evaluation of model performance to justify the best methods across this dataset. We select the ML methods following their adequacy and comparative research prevalence. We identify the abnormalities in the data through the anomaly detection method, Isolation Forest. Following the anomaly detection, we analyze the data features by applying feature selection and ranking to understand the relationships among the dataset's features and their relationship with the ground truth. We apply data validation with

TABLE 5.1. Feature Description of the Dataset by Castellini *et al.*[6]

Feature	Category	Description
Latitude	General	Latitude of the area
Longitude	General	Longitude of the area
Altitude	General	Height above sea level
Date and time	General	5,6 h of runtime
EC	Water	Water electrical conductivity
Temperature	Water	Water temperature
DO	Water	Water dissolved oxygen
m0 current	Drone	Signal to propeller 0
m1 current	Drone	Signal to propeller 1
Heading	Drone	Compass direction in which the drone's bow is pointed
Voltage	Drone	Drone's battery voltage

the hold-out method to divide the data into training and testing datasets to prepare for classification. Finally, we compare different classification methods to analyze the predictive performance of the model. In this section, we justify the selected methods for analyzing the data.

Anomaly Detection with Isolation Forest This model-based approach to anomaly detection isolates anomalies with low computational requirements. Isolation Forest works well in high-dimensional problems, and deals well with many irrelevant attributes[8]. Liu *et al.*[8] highlight the problem of anomalies being few, making them prone to isolation. Isolation Forest partitions instances repeatedly and recursively until they are isolated, producing shorter paths for anomalies[8]. The method does not use distance or density measures to detect outliers that eliminate computational costs making it a good fit for large and non-linear datasets.

Feature Selection and Ranking Choosing a reduced feature set makes the model easier to interpret, removes inessential information, reduces the dataset's size, and lowers the possibility of overfitting[9]. Overfitting the model is an error that occurs when the training on the data results with high accuracy. However, the testing results with poor accuracy and is typically caused by high variance in the data.

Lasso Regularization The Least Absolute Shrinkage and Selection Operator (Lasso) is a powerful regularization and feature selection method. This method applies the regularization or shrinking process by penalization of the coefficients of regression features [9]. The features regularized to zero are pruned from the model. The model, therefore, has the potential of reduced variance without a considerable increase of bias.

Filter Method with Pearson Correlation Filter methods choose features through statistical tests and correlation by ranking them on their usefulness to the model [9].

Data Validation with Hold-Out Method This validation method divides the data into two non-overlapping sets, training and testing. The hold-out is the testing set and can contain any percentage of the original dataset. The time for learning in the hold-out method is lesser than in comparable cross-validation methods [10]. The hold-out can eliminate the problem of overfitting, avoid uneven distribution, and introduce a clear division of data with stratification [10].

Decision Tree classifier performs well on nonparametric, complex datasets. This method classifies samples into branch-like elements and constructs an inverted tree to make decisions [11]. However, the Decision tree can result in overfitting when working on small datasets or datasets with strongly correlated features.

Random Forest A popular classifier, Random forest, constructs multiple decision trees with randomly selected subsets of features and training data. It is less sensitive to overfitting because of the considerable number of decision trees produced randomly. Random forest performs well on datasets with high dimensionality and highly correlated data, making this method a promising approach in heterogeneous research [12].

k-Nearest Neighbor (kNN) Classifier kNN forms around finding similarities in data. Therefore data quality is crucial to this method. KNN calculates the nearest points in data and nominates the sign of majority. Choosing the k number is often considered arbitrary; however, a larger value of k number can reduce the effects of anomalous points [13]. Due to its sensitiveness to data quality, the method performs the best with smaller data batches with eliminated anomalies.

Logistic Regression This classifier produces quick outputs that can be interpreted as probability and therefore used for ranking. Logistic Regression is not sensitive to overfitting; however, it underperforms on non-linear data.

Support Vector Machine (SVM) The SVM classifier has the advantage of performing well within high-dimensional space [14]. It fits a hyperplane that separates classes in data and positions every new data point within this hyperplane. However, this method is computationally expensive, slow on extensive data, and challenging to interpret.

Naive Bayes Naive Bayes represents decision-making under uncertainty, or probabilistic approach to deduction [13]. This simple method is computationally fast, easy to interpret, and performs well with high-dimensional data. However, Naive Bayes will underperform if the data features are highly correlated or calculate the probability of zero if an unknown class in test data appears [15].

Implementation, Results and Discussion

For this research, we implemented the models using the sklearn module[1] for Python with default hyperparameters. For every experiment, we analyzed the complete dataset containing 20,187 samples, manual and autonomous datasets. After removing the non-labeled or unknown Drive value, the manual set results in 7,586, and the autonomous set in 7,530 samples.

Anomaly Detection Results Isolation Forest for anomaly detection outputs compelling results for the three datasets. According to the results (see Table II), the distribution of anomalies is comparable. A similar number of samples, approx.10%, of each dataset are identified as anomalies. Uniform distribution of detected anomalies promises a comparable data reliability level by manually or autonomously operated drones. Almost all of the identified anomalies have a GT value of 0. In the autonomous operation dataset, all anomaly samples belong to GT value 0. The anomaly detection method removed 100% of the samples with GT value 0, leaving only GT value 3 in the data.

[1] Scikit-learn Machine Learning in Python: https://scikit-learn.org/stable/

Table 5.2. Anomalies Analyzed with Isolation Forest

Analyzed data	Number of anomalies	% of anomalies
Complete dataset	2019	10.0014
Manual operation	759	10.0052
Autonomous operation	753	10.0000

Similarly, in the manual operation dataset, only 0.0052% of the anomalies are not samples with GT value 0. This small number of non-zero anomalies are scattered around the manual dataset without showing a significant pattern. Autonomous and manual operation datasets display uniform distribution of anomalies, making them nearly equivalent in performance. However, GT value 0 represents unlabeled or unknown water flow which can be essential contextual information for the domain experts [7], particularly when analyzing sensor data performance. Hence, we retain the samples with GT value 0 in the dataset.

Feature Selection and Correlation Ranking Results Pearson correlation results in uneven distribution of highly correlated features regarding the GT. For the complete dataset, there are seven highly correlated features: electrical current (ec), drive, water, altitude, longitude, latitude, and water temperature. For the manual dataset, only four features highly correlate to the GT: altitude, longitude, latitude, and water temperature. Lastly, for autonomous data, highly correlated features are voltage, altitude, longitude, latitude, m0 current and, m1 current. A high correlation between features can impact the classification, such as biased predictions due to a strong relationship of two or more features. The impact of these results is visible in the classification analysis.
Contrastingly, the Lasso Regularization method resulted in a more uniform set of selected features regarding their importance in the dataset: ec, dissolved oxygen, temperature, altitude, and heading. Furthermore, selected features are penalized to a significantly low coefficient, nearly pruned from the dataset. This method typically penalizes correlated features, potentially removing important information and creating unstable models [16]. With this information, we retain the entire set of features for the classification analysis.

Data Validation We use the hold-out method for the data validation, where 60% of the data is split for training and 40% for testing the model. GT values' distribution is undeviating for train and test sets (Table III and Table IV). The autonomous operation dataset does not contain the GT values 2 and 1, and there is a heavy imbalance of the existing values, 3 and 0. The complete and manual operation datasets also contain differences between the GT values.

However, a more uniform distribution of the GT values can contribute to the model's lessened sensitivity for the data's bias.

Table 5.3. Distribution of the ground truth values on train set

Ground truth value	Complete	Manual	Autonomous
3	6880	2357	4316
2	514	532	0
1	383	381	0
0	4335	1281	202

Table 5.4. Distribution of the ground truth values on test set

Ground truth value	Complete	Manual	Autonomous
3	4499	1569	2862
2	341	323	0
1	259	261	0
0	2976	882	150

Supervised Classification Results Choosing a metric is likely the most critical phase in the project. Figure 5.1 illustrates the steps necessary for choosing an appropriate performance metric for imbalanced datasets. The metric is used to evaluate and compare all models. Choosing the incorrect metric can result in selecting the incorrect algorithm. The measure must reflect the most critical facts about a model or its forecasts for the project or its stakeholders [17]. Furthermore, essential indicators of models' performance are the trade-offs in the data: bias and variance. Bias in the data indicates the inaccuracy of the model's prediction compared to the data's actual values. The biases can occur during the training phase, where the model is 'simplified' to make the GT easier to predict. Alternatively, high variance indicates that the method learned the noise instead of the output. The high variance can cause overfitting. High variance and low bias relate to the high model complexity. The optimal model performance is the crossing point of bias error with variance error. Results of feature selection, feature correlation, and imbalance of the GT values can

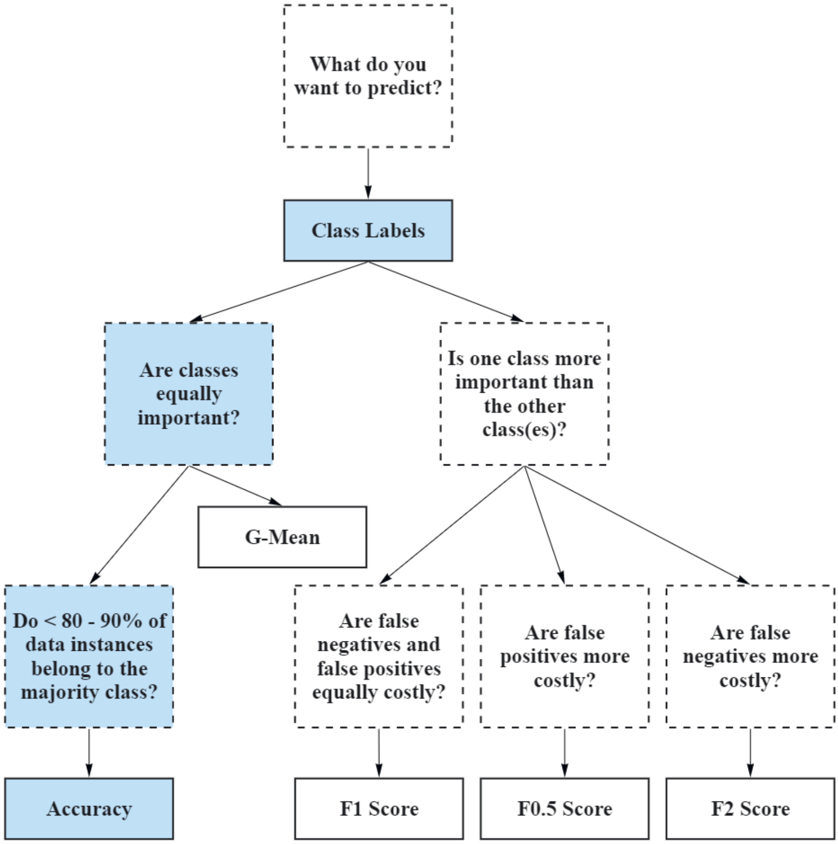

FIGURE 5.1. Imbalanced data: Choosing Performance Metrics, adapted from [17]

set expectations for the prediction capabilities. The test data results (see Figure 5.2) of the three datasets show high accuracy for all classifiers. Accuracy, the selected performance metric, describes the measure of correctly classified records. The results are presented in a box plot, Figure 5.2, showing the spread of the accuracy scores across data validation for each algorithm.

Manually-Collected Sensor Data Classification There is a considerable difference between linear and non-linear methods for the manually-collected dataset. As expected, non-linear methods, Decision Tree and Random Forest, performed with higher accuracy than all three linear methods. The prevalence of GT value '0' contributes to good prediction results. However, through observation of the confusion matrices resulted by linear methods, it is evident that the less represented GT values' prediction is erroneous. Generally, the

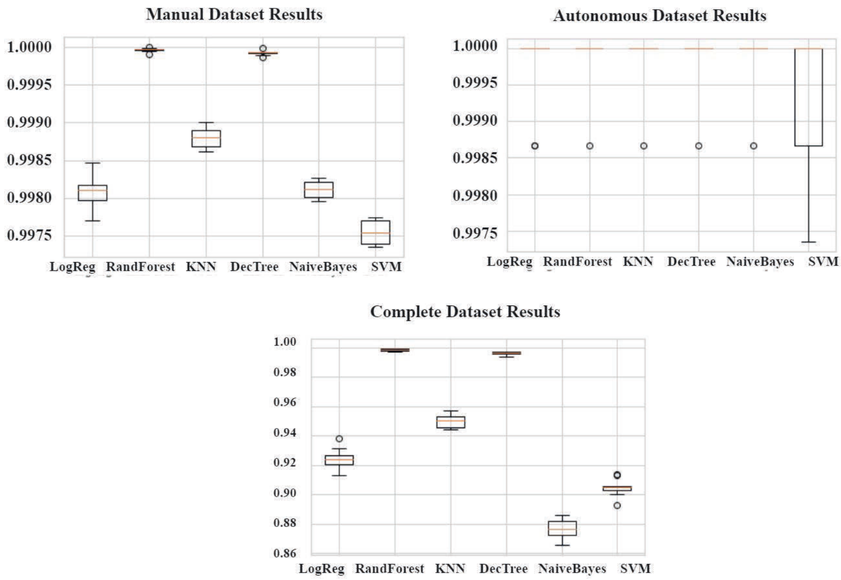

FIGURE 5.2. Algorithm Comparison: Autonomous, Manual and Complete Datasets

performance of non-linear methods on this model is adequate, making Random Forest the most reliable classifier for this dataset.

Autonomously-Collected Sensor Data Classification The autonomously-collected data illustrates significantly different results when compared to the manual data. The GT in the autonomous operation set contains only two values, 0 and 3. The 100% accuracy on the testing set can happen if the test set overlaps with the training set. However, in this case, the test and training sets are separate and not overlapping. In earlier anomaly detection results, the Isolation Forest has eliminated the samples with GT value 0, which can indicate a clear difference between these two values in the dataset. The poor distribution of the GT values results in an uncomplicated model that predicts with 100% accuracy. Arguably, the autonomous drone operated when the operation is unobserved (value 0) or even exclusively within the selected upstream environment (value 3). This model requires data complex enough to avoid bias and with a significantly less data imbalance.

Complete Sensor Data Classification A complete dataset exhibits both previous analysis' results as a combination of manually and autonomously collected data. The high correlation of features in the data causes the model's high performance with Decision trees and Naive Bayes classifiers. Other classifiers that are less sensitive to high correlation, such as Logistic Regression or kNN, are sensitive to data quality, such as data imbalance in this dataset.

The results of the autonomous operation model do not meet reliability expectations. Therefore, repeated data collection methods in more complicated scenarios can improve the balance and complexity of the dataset [7]. Alternatively, manually-collected data proved to be inherently different from the autonomously-collected data. As a novel contribution, we suggest that future data is collected from planned manual and autonomous drone missions in more complex environments, recording the same sensor measures. A human operator of the manual drone should follow the same path as the preprogrammed autonomous drone. Following these requirements, we can obtain consistent data from both operations and avoid significant data differences.

5.1.4 Contribution Summary

This section analyzed the difference between manual operation and an autonomous operation of an underwater drone. We explicitly identified the similarities and the differences between the two operation modes through anomaly detection and classification methods during the analysis. Our research recognized the vital role of sensor data variations of different operation modes in the context of prevalent machine learning methods' performance and identified the gaps in which these methods underperformed. Unfortunately, unbalanced data is pervasive in research and industry, resulting in skewed classification results and reduced reliability for machine learning methods.

5.2 A Novel Warning Identification Framework for Risk-Informed Anomaly Detection

Abstract

Cyber-physical systems are taking on a permanent role in the industry, such as in oil and gas or mining. These systems are expected to perform increasingly autonomous tasks in complex settings removing human operators from remote and potentially hazardous environments. High autonomy necessitates a more extensive use of artificial intelligence methods, such as anomaly detection, to identify unusual occurrences in the monitored environment. The absence of data characterizing potentially hazardous events leads to disruptive noise displayed as false alarms, a common anomaly detection issue for hazard identification applications. Contrastingly, disregarding the false alarms can result in the opposite effect, causing loss of early indications of hazardous occurrences. Existing research introduces simulating and extrapolating less represented data to expand the information on hazards and semi-supervise the methods or by introducing thresholds and rule-based methods to balance noise and meaningful information, necessitating intensive computing resources. This research proposes a novel Warning Identification Framework that evaluates risk analysis objectives and applies them to discern between true and false warnings identified by anomaly detection. We demonstrate the results by analyzing three seismic hazard assessment methods for identifying seismic tremors and comparing the outcomes to anomalies found using the unsupervised anomaly detection method. The demonstrated approach shows great potential in enhancing the reliability and transparency of anomaly detection outcomes and, thus, supporting the operational decision-making process of a cyber-physical system.

5.2.1 Introduction

The environment's safety is ever more reliant on cyber-physical systems that have applications in, among others, intelligent drones, remote sensing, and smart sensor systems. These systems are taking on permanent roles in various industries such as oil and gas, energy, and mining. They are replacing various human operations and carrying out critical responsibilities, including inspecting and monitoring remote, possibly hazardous environments. The increasing growth of sensor-collected data grows a need for artificial intelligence (AI) and data-oriented technologies along with the requirements for more autonomous systems that are safer, more perceptive, and more financially viable. Autonomy is described as the capability of a system to operate independently from external factors [18]. Increased autonomy necessitates a more significant usage of AI [19] methods that copy intelligent human behavior [20]. With various sensors, the cyber-physical systems can efficiently gather data during ongoing operations and use AI methods to analyze the data in real-time and gain situational awareness. Consequentially, increased autonomy has the potential to replace constant human supervision. As a form of AI, machine learning (ML) uses high volumes of data to learn how to execute tasks rather than being programmed to do them, allowing computing systems to become more intelligent as they encounter additional data [20]. Similarily, anomaly detection, as a data-oriented method, detects unusual trends in data that can give insight into potentially hazardous occurrences. Detecting critical trends in good time allows for the opportunity to take necessary corrective actions in advance to ensure safe operations. Considering the variety of hazards that can affect these systems, many techniques might increase their ability to operate safely under all conditions. Therefore, AI technology must be reliable in order to responsibly integrate it into existing systems and operations.

The challenges inherent in unsupervised anomaly detection emphasize the necessity for further research into semi-supervised or alternative methodologies [21]. Although sensor data and data-driven methods are becoming essential in many safety-critical or high-risk engineering systems, data-driven methods may not be sufficient to ensure safety because they lack the underlying causal knowledge [22]. Additionally, benchmarking and comparing anomaly detection methods is eminently challenging. Due to these challenges, early warning indicator of potential hazardous events may be missed, possibly placing assets or the environment in jeopardy during operations [23].

In cyber-physical systems, particularly autonomous systems, that form decisions based on data-oriented methods, the safety and responsibility of the methods and the data that trains the methods cannot be overemphasized. Therefore, this section summarizes the development and evaluation of a Warning Identification Framework (WIF) through a case study. The described framework in this section is an extension of Chapter 4 that summarize articles [24] and [25] and focus on theoretical concepts of risk-informed and data-driven operations. The purpose of the WIF is to facilitate the decision-making of a cyber-physical system that uses anomaly detection methods to identify warning signs of an ongoing operation. Such applications include autonomous underwater drones for inspecting pipelines and observing potential surface corrosion or cracking or intelligent sensor systems for monitoring drilling operations in mines and listening for potential seismic tremors, shaking of the ground under the stress of mining or drilling. To facilitate the decision-making of a cyber-physical system, another objective of WIF is to address the interrelated challenges of unlabeled, contextless, biased data, unsupervised methods, and consequentially unreliable anomaly detection results. WIF is anchored in risk analysis and comprises three main steps: characterization, analysis, and ranking of warning impacts detected through anomaly detection. To compare the standard hazard assessment and anomaly detection methods, we examine unlabeled seismic data with varied sensor values for tracking seismic tremors and three distinct hazard assessment methods for identifying low, medium, and high-impact hazardous occurrences.

The following is a summary of the primary contribution of this section, Warning Identification Framework:

1. Novel risk assessment perspective on seismic hazard identification's training and assessment role in unsupervised anomaly detection approach.

2. Identification of overlapping methods and roles in risk assessment and anomaly detection.

3. Preliminary results of three seismic hazard identification methods and their assessment role for unsupervised anomaly detection results.

5.2.2 Background

Anomalies (in literature often interchangably referred to as outliers, novelties, abnormalities, discordants or deviants) are occurrences in a dataset that are odd in some sense and do not fit the dataset's general or expected trend [26]. They apply to a wide range of desired and undesired phenomena, appearing as static occurrences, time-related events, single and grouped occurrences. Despite being interchangebly used, the terms anomaly and outlier are distinguished in some studies [27–29] . For an example, Hawkins [28] provides a definition of an outlier: "An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism." In the recent study on anomaly classification, Foorthius [30] describes that the definition of anomaly is vague and dependent on the application domain due to the wide variety of ways anomalies manifest themselves. In order to understand how unsupervised anomaly detection methods in relationship with knowledge from risk analysis can be utilized to improve true anomaly discovery and potentially avoiding missed out early warning signals, it is important to understand how anomalies manifest themselves.

5.2.3 Risk and Risk Analysis

Risk is defined as the effect of uncertainty on objectives, where the effect can be positive, negative, or both, resulting in opportunities and threats [31]. Typically, the risk is expressed in terms of risk sources, future occurrences, their effects, and the probability they will occur. Earlier guidelines for the inclusion of safety aspects in standards [32] define risk as a combination of the probability of occurrence of harm and the severity of that harm, where harm is an injury or damage to people's health, property, or environment [32].

In his book, Risk Assessment Theory, Methods, and Applications, Rausand [33] describes risk analysis as one of the three main elements of risk management (see Figure Figure 5.3), the continuous process to reveal, analyze, and assess potential hazardous events in a system, and identify and introduce efficient risk control measures to eliminate or reduce possible harm [33]. The risk analysis is responsible for:

- the identification of hazards and threats related to the system of interest;
- the identification of potential hazardous events that may occur;
- the identification of causes of hazardous events;
- the identification of barriers and safeguards to prevent or reduce the hazardous events and assessment of their reliability;
- the identification of accident scenarios related to each hazardous event and their consequences.

The other two main elements of risk management are [33]:

1. Risk evaluation for assessing risk picture, comparison of the risk with established risk acceptance criteria, considerations of alternative systems.
2. Risk control and risk reduction for making decisions regarding introducing new risk-reducing measures, implementing the measures, monitoring, and communicating the risk.

Risk analysis systematically uses available information to identify hazards and estimate risk where the hazard is a potential source of harm [32]. Therefore, risk analysis can be observed as a tool to inform decision-making concerning future welfare since the risk is always related to what can happen in the future [33]. As illustrated in Figure 5.3, the analysis of risk is carried out to answer the following questions:

- Hazard identification: *What can go wrong?*

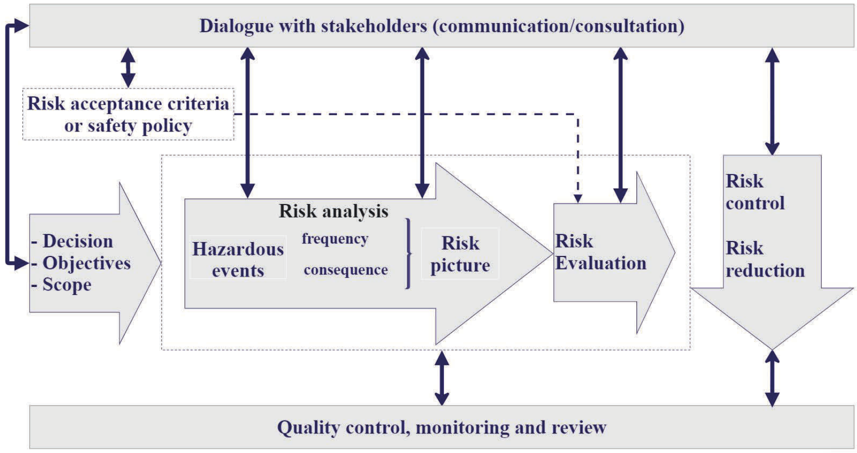

FIGURE 5.3. Elements of Risk Management, adapted from [33]

- Frequency analysis: *What is the likelihood of that happening*?
- Consequence analysis: *What are the consequences*?

The risk analysis results have great potentials for assessing or improving the data for anomaly detection models. These potentials may be observed using the risk analysis for rule-based labeling in classification problems or transferring its knowledge to train the new models. As described by [34], transfer learning comprises different techniques that aim to gather the knowledge gained at the source problem to develop a new model using the gathered knowledge, thus minimizing the efforts of developing that new model. To the best of our knowledge, these potentials have not been leveraged effectively in prior studies (see Section 5.2.4) that would tackle the existing challenges in the anomaly detection methods.

Hazard Identification

Identifying the hazard is a critical first step toward preventing or mitigating it. Certain hazards require a triggering event to grow into a hazardous event, whilst others may develop into a hazardous event gradually [33]. A triggering event is an event or situation that must occur in order for a hazard to cause an accident [33]. Hazard identification techniques determine [35]:

- Possible cause of the harm;
- How the harm will manifest itself;

- What measures are in place to avoid or mitigate harm;
- The extent to which the harm is tolerable;
- What further actions or resources are required to avoid or mitigate harm.

The What-If Checklist, the Hazard and Operability Study, and the Failure Modes and Effects Analysis are three of the most often used techniques for hazard detection [35]. Knowing what can go wrong and identifying the properties of hazards is a crucial step in labeling the training sets for supervised classification or anomaly detection.

Consequence Analysis

A consequence is an adverse event that may occur due to a hazard [35]. As a result, consequence analysis examines the predicted impacts of incident outcome situations regardless of their frequency or likelihood. There is a specific amount of energy or material released in the event of containment failure. This is referred to as the source term [35]. Assume the effects are instantaneous, as with an explosion. In that situation, the analysis uses inputs such as the material type, the release pressure, and other factors to determine the impact effects. If the effects are delayed, the source term characteristics are used as inputs in a dispersion analysis followed by an analysis of the impact effects. Anomaly detection can detect anomalies representing significant information about the ongoing operation or anomalies that do not require any insight or resource allocation. Consequence analysis provides critical information on the impact of hazards or anomalies that can aid operators in allocating necessary resources.

Likelihood Analysis

Risk cannot be accurately assessed without first analyzing the likelihood of an event occurring, which can be challenging. Analyzing the likelihood becomes progressively more challenging for complex systems, and hazard scenarios [35]. The likelihood of often occurring events may be evaluated and validated using statistical analysis that requires large amounts of data. The common methods for likelihood analysis are fault propagation modeling methods - event tree analysis and fault tree analysis.The situations, conditions, and protective mechanisms, together referred to as intermediate events that should have prevented the accident, are listed, along with their associated probability of occurrence. In anomaly detection, the likelihood and frequency analysis bring invaluable information on the underlying knowledge of detected anomaly or hazard. Although not all detected anomalies require reaction response or allocation of resources, knowing the likelihood or frequency of certain undesired events is smaller or larger under a particular operational context may eliminate the need for conjecture when classifying or labelling observed anomalies.

Warning Management

While warning management is not explicitly included in risk analysis, it is necessary to employ risk analysis insights as a layer of protection. A warning is used to notify the operator of a malfunctioning piece of equipment, a process deviation, or an unexpected state that demands operator intervention [35]. Alarms assist the process in remaining within normal operating parameters and ensuring its safety, differentiating between negligible, tolerable, and unacceptable risks . A risk level that is considered acceptable suggests that the risk level is usually recognized as insignificant [33]. Typically, additional risk-reduction measures are not necessary. Tolerating a risk, or tolerable risk, implies that

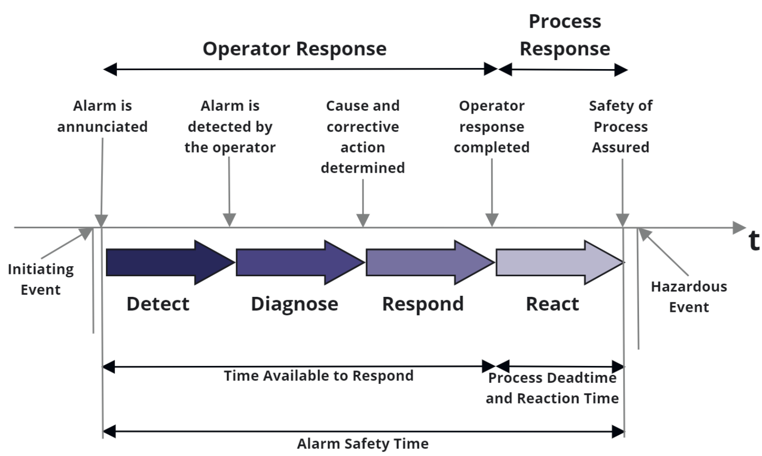

Figure 5.4. Operator and Process Reaction Time, adapted from [35]

we do not perceive it as negligible or something to be overlooked, but rather as something to be monitored and mitigated further as and when possible [33]. Except in exceptional circumstances, activities with an unacceptable level of risk are considered unsuitable, regardless of their advantages. Activities that create such risk would be prohibited, or the risk would have to be mitigated at all costs [33]. To assist in determining which alarms should be addressed first, each warning is assigned a priority, often based on the severity of the potential consequences.

Figure 5.4 depicts the operator response to warning. The operator must be capable of promptly detecting, diagnosing, and appropriately responding to the warning to avoid a hazardous event. A warning management system is a crucial component of cyber-physical systems involved in safety-critical activities. In increasingly complex systems, an autonomous system, such as a UAS, is anticipated to conduct detection, diagnostics, response, and reaction depending on the scenario. UAS, such as underwater drone or smart sensor systems, offer warnings to the operators if the ongoing activity requires further attention. In this instance, the UAS can autonomously determine if a given monitored occurrence is an early warning indicator and whether to sound a warning using data-driven approaches, particularly AI.

5.2.4 Challenges

Missing Context and Data Imbalance

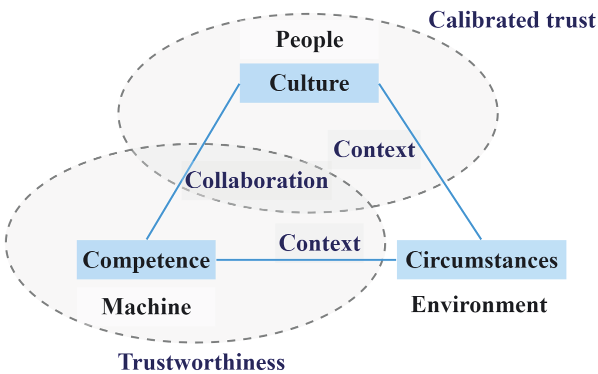

FIGURE 5.5. A Framework for Discussing Trust in Increasingly Autonomous Systems, adapted from[36]

We observe a growing interest in research within the context of ML strategies for knowledge sharing and organizing, such as[37],[38],[39]. Righteously so, we witness a more permanent role of autonomous systems and ML in the industry. However, integrating ML into existing systems involves heterogeneous teams of, amongst others, software engineers, data analysts, and domain specialists. Every domain specialist and analyst in a heterogeneous team developing a software system that integrates ML should comprehend the context underlying ML methods and data. This context specifies the relationship between code and data, as well as the relationship between data and intent of the operation. Lacher et al.[36] point out that the context is critical to a system's capacity to respond satisfactorily as it becomes increasingly autonomous. In a *framework* *for discussing trust in increasingly autonomous systems* by[36], context is represented as a binding point between people, environment and the machine (i.e., the autonomous drones) (as seen in Figure 5.5). *People* have varying perspectives of the machine which is based on their roles and greatly influenced by their culture (such as age and professional affiliation). The operation's context is established by the *environmental factors* because the machine will identify the situations based on the data received from sensor inputs. *The machine* is designed to perform required tasks at a high level of performance, which can be observed, measured or assessed. The results of these assessments will have an impact on people's confidence in a machine's competence. If the machine produces expected results, the human confidence in the machine will increase, answering the question of the machine's reliability. Lacher et al.[36] conclude that most of the machines will have a degree of human collaboration and the degree of trustworthiness between people and machines is a cultural, organizational and sociological challenge. According to[36], calibrated trust is founded in our perception and expectation of system performance, which has become an engineering, social, cultural, and organizational challenge. Yet, as machines become increasingly complex, trustworthiness becomes more challenging to maintain due to the difficulty to understand the functioning and set the expectations on the machine performance.

Hayes et al.[40] offers an example of an anomaly detection algorithm missing context in the circumstance of a sensor reading detecting that a particular electrical box consumes an abnormally high quantity of energy. However, when examined in the context of the sensor's location, present weather conditions, and time of year, it is well within normal boundaries. There are various explanations for these shortcomings, which we loosely divide into two categories: technical and people/process-related. The technical reasons as the often unpredictable malfunctions of the system. However, the people/process-driven reasons for ML shortcomings are due to the more complex methodologies that the individuals or teams use to organize and transfer knowledge, including designing, developing, and maintaining the systems that employ ML. Lee et al.[41] argue that the shortcomings, particularly due to biases caused by imbalances in data, can be removed not by niche methods but rather by informing the appropriate mitigation strategy, whether technical or people/process-driven. Nevertheless, the previous studies inform that practitioners struggle to integrate newly proposed tools and methods into existing processes[41]. Authors[41] suggest that identification and categorization of different types of shortcomings, such as biases, can help to understand the roots of the unintended ML outcomes.

Due to the wide range of anomalies that can disrupt operations and the large amount of data produced by environmental sensors, real-time anomaly detection is becoming more challenging. Imbalanced or underrepresented data, such as high consequence and low probability hazardous event data, is particularly problematic because the data processing methods form biases in favor of more represented data. Classification methods, entrusted with effectively predicting outcomes from the sensor data, tend to reproduce these biases[42]. Furthermore, underrepresented data can be disregarded as noise due to the anomaly methods' inclination toward efficiency and sacrificing anomalies as tolerable collateral damage[43]. False alarms, or noise, are another known drawback of anomaly detection[44]. False alarms fall into two categories: false positives and false negatives[45]. When a normal or non-hazardous event is recorded as a hazardous event, this is called a false positive. A consequence of false positives is that a potentially hazardous event may go undiscovered due to prior false positives. A false negative is defined as the inability to notice a hazardous event. Due to the high proportion of false alarms created by anomaly detection, it is difficult to correlate specific alarms with the events that triggered them[45]. Additionally, current methods for anomaly detection focus primarily on data content, with no regard for the context behind the data[40]. These methods yield conclusions that are based on correlation without causation. *Causation* is the situation in which one event, *a cause*, causes another event to happen *an effect*. *A correlation* is the situation in which two or more events appear to be related. Therefore, basing conclusions solely on correlations is one of the critical problems in data analysis[22], as it might result in misleading predictions. However, many datasets lack labels or supervision that provides additional information and context about the data[46] making the training and testing of anomaly detection methods even more challenging.

Trust Imbalance

Many judgments made by cyber-physical systems in various scenarios are based on its analysis of the environment[47]. The biased and unjust consequences of data-driven methods are frequently the result of opaque or black-box methods that lack transparency. As a result, anomaly detection methods have recently piqued the interest of industry and academics in the hopes of gaining greater transparency and offering more context to the data and the anomaly detection methods[30]. The three of the biggest challenges of evaluating these systems are user acceptance and trust, adequate evaluation, and defining autonomy comprehensively and quantitatively[48]. Autonomous drones, for example, must operate safely and be resilient in changing environments and complex scenarios. The ability to successfully manage disturbances and emergent needs during the system's mission - resilience - determines the efficacy and reliability of autonomous systems[49]. A resilient and reliable system can alter its functioning in advance of or in response to changes and disturbances, allowing it to continue working even after a severe incident or in the face of persistent stress, primarily by being proactive on safety[50]. Hollnagel has outlined the three fundamental functions of a resilient system[18]:

1. *Anticipate disturbances*, prospective threats (Hollnagel uses the terms *threat* and *hazard* synonymously), and any other destabilizing conditions. This function enables the system to forecast the future and adjust risk tolerance.

2. *Monitor performance*, risks and threats while constantly improving its own risk identification model. This function enables the detection of non-permanent transient impacts that, despite not being permanent, can still cause failures and accidents.

3. *Respond to threats*, whether they are regular, irregular, unexpected or unexampled. This function denotes a resilient system's preparedness, flexibility, and adaptability.

O'Neil[51] argues that data-driven methods should be prejudice-free, produce objective results, judge according to universal norms, and eliminate biases. However, since the methods are based on historical data, they not only incorporate biases, they reinforce them[51]. Since highly autonomous systems rely heavily on data-driven methods, these systems must include human-centered features to ensure that they society, industry, and the economy while adhering to ethical norms.

Existing Approaches to the Challenges

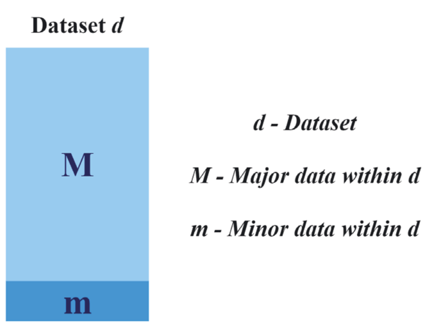

FIGURE 5.6. Major and minor data in a dataset

With supervised anomaly detection and labeled datasets, discriminating between anomalous and non-anomalous data is supposed to be straightforward. The dataset contains labels for anomalous and non-anomalous data points, enabling anomaly detection methods to classify the data more precisely. Distinctively, unlabeled data are analyzed using distances, density, and trends between data points. However, the difficulty of underrepresented data, or minor data (see Figure Figure 5.6), has recently eroded faith in supervised methods as well. The classifier or anomaly detection method is unable to distinguish between more represented or major and minor classes in the data, favoring the major data and thus overlooking the minor data, potentially omitting critical information.

Numerous approaches to the problem of underrepresented data have been developed, and the following paragraphs allude to recent studies while summarizing the approaches as follows:

- Extrapolating minor data through simulations, causal and physics data
- Setting decision boundaries and thresholds for normal data
- Semi-Supervised and Rule-Based Anomaly Detection and Classification

Extrapolating Minor Data and Simulations

Due to the rarity of hazardous occurrences with high consequences, there is a sparse indication of their properties in the sensor-collected data during environment or asset monitoring. This lack of data necessitates using simulations to replicate the natural world and generate artificial hazards and hazardous events, further extrapolating imbalanced datasets with the artificially generated data from simulations so that machine learning models can train on less imbalanced data. Eldevik et al.[22] highlight in their work on AI and safety that data-driven models alone are insufficient. Although sensor data and data-driven models are becoming an integral part of a growing number of safety-critical and high-risk engineering systems, the high-consequence and low-probability scenarios are not well reflected by data-driven models. The authors[22] propose the use of causal and physics-based knowledge for extrapolation robustness. The data-generating processes consist of stochastic and deterministic elements, providing an opportunity to utilize the deterministic processes, or those governed by known principles, and extrapolate the naturally underrepresented data with the existing causal and physics-based knowledge. The authors[22] argue that the combination of data-driven models and the causal knowledge of industry experts is essential for decision-making processes within AI systems.

The method of simulating, or extrapolating with causal and physics-based knowledge, has significant drawbacks, including runtime and curse of dimensionality[22]. For a high-consequence system, a model used to inform risk-based decisions must predict potentially catastrophic scenarios prior to the occurrence of the scenario. However, the runtime of these complex models is often significant, commonly taking up to several days, making it nearly impossible to initiate necessary analysis in real or near real-time. Alternately, the models can be run in advance, but this again necessitates sophisticated processes with many inputs, restricting the possibility of simulating every possible condition that an actual system can encounter before its operation. Eldevik et al.[22] emphasize that data-driven models should incorporate risk assessment into the decision-making processes as we rapidly progress toward more autonomous systems that employ AI for making safety-critical decisions.

In addition to computationally expensive simulations, Zhang et al.[52] argue that simulation experiments can be expensive to conduct in laboratories and frequently meet physical limitations for simulating the real world (i.e., simulating scenarios in the ocean vs. in a laboratory water pool). The primary limitation of simulations, whether virtual or in laboratories, is their inability to reliably mimic the complex interactions between the environment, the asset, and the ongoing hazard in the case of a hazardous occurrence.

Decision Boundaries and Thresholds

Anomalies are characterized in terms of previous behavior. This suggests that a novel behavior may first appear anomalous but ceases to be anomalous if it persists, establishing a new normal pattern[53]. Lavin et al.[53] define anomaly windows to aid in early detection. Each window is a collection of data points centered on a ground truth label for an anomaly. The earlier a detector can reliably identify anomalies, the better, which implies that these windows should be as large as possible. The trade-off with exceedingly large windows is that unreliable or random detections would be reinforced regularly[53]. This technique allows for a large window of opportunity for early detection and allows for partial credit for detections made shortly after the ground anomaly. The authors[53] emphasize that various applications may place a greater emphasis on true positives than on false negatives and false positives. For instance, in a manufacturing plant, a false negative may result in machine failure and costly production disruptions. Similarly, a false positive may necessitate an in-depth examination of the data by a technician.

Li et al.[54] developed a novel data-driven approach to anomaly detection in cyber-physical systems by establishing a decision boundary to classify new observations using a geometric structure non-convex hull. Convex hull-based methods define a closed boundary around the normal data points. These methods make no assumptions about the underlying distribution. The convex hull-based methods do not require extensive parameter tuning, making them useful for boundary-based anomaly detection[54]. Since not all potential anomalies are known in advance, most data-driven anomaly detection techniques depend on developing a model of the system's normal behavior. This dependency may reduce the likelihood of noise or false alarms occurring during anomaly detection. The points within the convex hull are normal, whereas the points on its periphery are anomalous. However, convex hull-based algorithms produce many erroneous classifications when the input normal data is not convex[54]. The authors[54] demonstrated that incorporating a non-convex hull as a decision boundary for anomaly detection in data with non-convex forms achieved significant improvements over typical convex hull-based approaches.

Shin et al.[55] studied data bias caused by underrepresented classes in datasets. They advised using decision boundaries to increase the accuracy of anomaly detection generative adversarial network (AnoGAN) results produced from low-quality data. The primary challenge encountered by the authors[55] is *the subjective nature of establishing the decision boundary*. They evaluated the proposed method's success using the Area Under the Curve (AUC) and the F-measure through testing multiple arbitrary values for the decision boundary. AUC evaluates a classifier's ability to discriminate between classes. In contrast, F-measure evaluates the performance of a binary classification model based on predictions for the positive class. The proposed model presented in the[55] research has a slightly greater AUC and F-measure value (0.023 and 0.0231, respectively) than the initially tested AnoGAN result. While decision boundaries are frequently seen in classification and supervised algorithms that utilize labeled data, such as SVM[56], a similar approach can be applied to unlabeled data using semi-supervision.

The disadvantage of decision boundaries or thresholds is their construction. The boundaries are constructed either by an algorithm that learns from data patterns or by assuming a geometrical shape (i.e., convex hull-based methods[54]). Forming context- or application-specific boundaries, as opposed to dataset-specific ones, is one approach to mitigate the disadvantages and establish more reliable decision boundaries.

Semi-Supervised and Rule-Based Anomaly Detection and Classification

Rule-based classification is a method for classifying or labeling data points using conditions such as 'if-then.' The benefit of rule-based classification resides in its interpretability and approach to generalization, rather than labeling each data entry individually. Nonetheless, this strategy requires manual inputs from domain experts and can soon become a complex task when applied to extensive data and unstructured sets.

Deng et al.[57] explored a rule-based semi-supervised approach to anomaly detection due to a lack of labels in data and, consequentially, an emphasis on unsupervised methods that produce incomprehensible results. The authors[57] observed the challenge in selecting appropriate labels when training models for anomaly detection due to the vague definition of an anomaly being *a data point that does not share a similar pattern with the rest of the data population*. Their approach to applying rule-based classification in anomaly detection consisted of visually presenting identified anomalies and allowing users to select, label, and describe the anomalies. Although this approach yields reliable and interpretable results, it becomes a complex task when data is scaled up. While the manual labeling and conditioning of anomalous points show promising results in preventing false alarms or mistaking frequently occurring anomalous points for normal points, the process makes the system less automated and more reliant on the continual engagement of domain experts.

A more automated yet interpretable method for anomaly detection is to have the model learn from normal data and report unusual deviations, a semi-supervision process. In this instance, the model's reliability depends on the quality of the normal data it is trained on—the likelihood of frequently occurring anomalies being misinterpreted as normal increases significantly.

5.2.5 Warning Identification Framework

The Warning Identification Framework (WIF) aims to support the decision-making of a cyber-physical system that uses anomaly detection methods to detect warning signals during an ongoing operation. WIF targets anomalies with a low likelihood of occurring but can have severe consequences. Typically, such anomalies are underrepresented in data, necessitating that WIF addresses data biases, a lack of labeled data, and a lack of context in data and anomaly detection methods to provide reliable results. The motivation behind WIF lies in key aspects of multiple disciplines towards operations of autonomous and intelligent sensing systems, adapted from[24]:

- Aspects of future *Risk Assessment*:
  - The recognition of knowledge, the growth of data, and the requirement for robust frameworks for the safety assessment of cyber-physical systems[58].
  - Focus on new events that become apparent in new conditions.
- Aspects of future *Reliability Engineering*, an engineering discipline for applying scientific know-how to a component, product, plant, or process

in order to ensure that it performs its intended function, without failure, for the required time duration in specified environment [59]:

  - Fault prevention, removal, and tolerance.
  - Fault forecasting.
  - Reliable functioning under expected circumstances.

- Aspects of future *Resilience Engineering*, a discipline that brings together the system safety concepts, reliability of a system, analysis and handling uncertainties, risks, and survivability of a system (where a resilient system can recover quickly after a shock or an injury) [60]:

  - Anticipation of hazardous events.
  - Monitoring of hazardous events.
  - Responding to hazardous events.

- Aspects of future *Human-Machine Teaming*, a relationship between humans, the machine, and their interdependencies aiming to build trustworthy, transparent, predictable, adaptable, and reliable systems that incorporate AI [61]:

  - On-demand adjustment of autonomy.
  - Explainable functioning of a system.
  - Shared understanding of intentions.
  - Multiple approaches to a single challenge.

Anomaly detection is frequently used in applications to identify unusual data patterns that might harm the system. In comparison, risk analysis identifies hazards as potential sources of harm. Risk analysis and anomaly detection have comparable objectives. As illustrated in Figure 5.7, the two disciplines share a common interest in identifying low probability events that may result in high consequences and require extensive data analysis. Therefore, the combination of risk analysis and anomaly detection provides a risk-informed approach to anomaly detection.

The interest in anomaly detection in combination with risk analysis is dependent on the capacity of anomaly detection to provide anomalous points that may be used to identify potential hazardous events, hazards, and threats. The combination of anomaly detection and risk analysis is particularly interesting for autonomous warning management. However, this section demonstrates that the process's reverse order is equally interesting, particularly in addressing the challenges caused by imbalanced data that contributes to poor anomaly

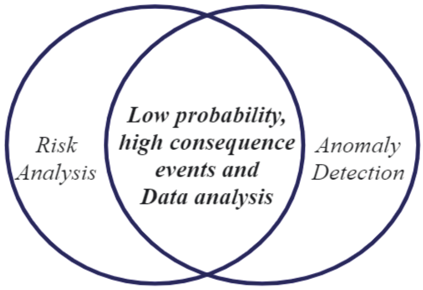

FIGURE 5.7. Risk analysis and anomaly detection overlap

detection outcomes. By using insight from risk analysis, such as the list of possible hazards and their properties, anomaly detection can be guided in detecting the true anomalies that can be of interest for further inspection. The causal analysis, accompanied by an identified sequence of events leading to the potentially hazardous event, can aid in the detection of anomalies. With the analysis of the severity of potential consequences, the detected anomalies can be prioritized, consequentially decreasing the number of false alarms. In light of this, we propose selecting anomaly detection methods that consider the likelihood that true anomalies will occur infrequently. One such method is Isolation Forest, which attempts to eliminate reporting of noise by isolating rare points in the dataset on the assumption that there are fewer true anomalies. We divide the process of using risk analysis as a supervisory component to anomaly detection into three steps, with an assumption that historical data exists for risk analysis as an input to WIF (as illustrated in Figure 5.8):

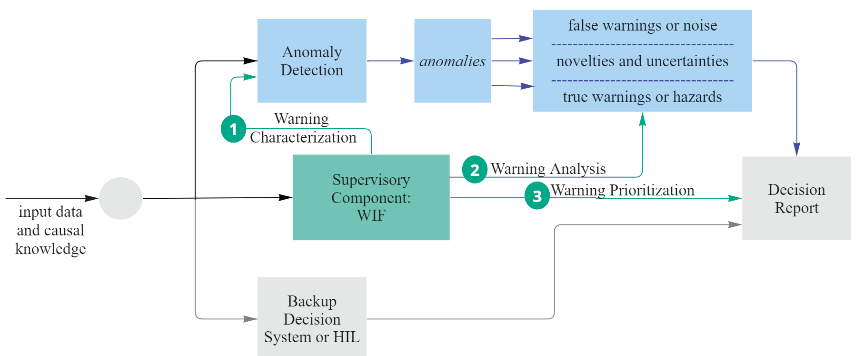

FIGURE 5.8. Architectural pattern for systems using WIF

5.2.6 Step 1: Warning Characterization

Given the context and circumstances of the planned operation, such as the operation goals, the assets, expected environmental compounds, location, time, and season, the first step is to answer the question, *"What can go wrong during the given operation and given the context and circumstances?"*. By answering this question, the warning characterization step, through risk analysis, aids in setting the objectives of anomaly detection, as illustrated in Figure 5.8. Context and circumstances are crucial for minimizing false alarms during anomaly detection. Since not all occurrences are anomalous under all circumstances, distinguishing hazards and their contextual occurrences makes it easier to overlook expected or insignificant disturbances detected by anomaly detection methods. In addition, it is essential to identify the events or circumstances that contribute to a hazardous event, known as triggering events. While some hazards develop gradually, others occur due to another event, a trigger, typically a technical failure or human error [33]. Furthermore, while a single anomalous phenomenon may not suggest a hazardous occurrence, a collection of several phenomena may. All known or expected variables that may constitute a hazardous occurrence, or an early sign of one, should be included in the step of warning characterization.

5.2.7 Step 2: Warning Analysis

After determining what can go wrong and compiling a list of hazardous and potentially hazardous occurrences, the second step is to answer the question *"How does the hazard manifest?"* to gain a more profound knowledge of hazards. It is essential to collect as many attributes as possible that can explain the hazard, such as the sequence of events that may lead to their occurrence, frequency, and the likelihood of appearance. The sequence of events can highlight changes in environmental components that may lead to a hazardous event. Inner corrosion in a gas pipeline is an example of a hazard that builds gradually until a gas leak, a hazardous event, occurs [33]. Accordingly, the sequence of events may consist of multiple sensor measurements with specific values and properties that are informative of hazardous occurrence, as determined by domain experts. Rausand [33] describes the first occurrence in a sequence of events that will lead to undesirable outcomes as an initiating event or the event that disrupts the normal operations of the system and may necessitate a response in order to prevent subsequent undesirable outcomes.

The second part of understanding how hazards manifest is by answering, "What is the likelihood of this hazard occurring, and how frequently does it occur?". While frequency tells us about the number of times an event has happened within a specific timeframe, likelihood answers how probable it is that it will occur[62]. Knowing the frequency and likelihood of a hazard is valuable for dismissing false alarms and recognizing circumstances under which the hazardous events are more likely to occur. As illustrated in Figure 5.8, the warning characterization step, through risk analysis, aids the anomaly detection outcome in distinguishing noise from hazards or true and false warnings detected in a group of anomalies. Nonetheless, unsupervised anomaly detection leads to identifying novelties that may have been overlooked during risk analysis.

5.2.8 Step 3: Warning Prioritization

Knowing whether to respond is the objective of the third step. Figure 5.8 illustrates the critical component of decision-making of a cyber-physical system responsible for autonomously reporting hazards during ongoing operations. Warning priority is derived from consequence analysis and is responsible for determining the impact of an identified hazardous occurrence. The impact of a hazard prioritizes a response by the autonomous system to notify the operators or supervisory system if and when the situation necessitates it, allowing for early warnings with minimal false alarms.

Figure 5.8 formalizes the three phases of WIF into an architectural pattern for systems employing WIF and utilizing anomaly detection (or comparable ML approaches) for safety-related decision-making. This type of architecture permits decisions to be risk-informed instead of based on ML-discovered patterns that depend on often unreliable data. Risk analysis through WIF represents a supervisory component for anomaly detection, provided by domain specialists examining historical data and causal knowledge. Incorporating a supervisory component increases the opportunities to address the challenges associated with anomaly detection, such as a high number of false alarms and the inability to differentiate noise from hazards, and other general challenges associated with machine learning methods too, such as bias, lack of context, and lack of explainability. WIF enables anomaly detection to distinguish false alarms, true alarms (potentially hazardous occurrences), uncertainties, and novelties. *Uncertainties and novelties* represent anything unknown. While some publications use the terms *anomalies* and *outliers* interchangeably, other sources[63–65] use the term *outliers* to denote uncertainties or novelties captured by anomaly detection. As part of the architectural pattern for WIF-based systems, as illustrated in Figure 5.8, it is recommended to include a backup decision plan that requires human intervention, Human In the Loop (HIL), if the system fails to operate autonomously.

The suggested methods for each WIF step depend on the data, case study, and objectives. The methods for our seismic data case study are described in the following paragraphs.

5.2.9 *Case Study*

The application of the WIF is demonstrated using data acquired by the geophysical station supporting system towards estimating the rock burst hazard using seismic and seismoacoustic techniques[66]. Seismic hazard is one of the most challenging natural hazards to detect and anticipate[67] and can result in devastating consequences during underground activities such as mining and drilling. One of the primary responsibilities of geophysical stations is to determine the current level of seismic hazard, especially the probability of high-energy, destructive seismic tremors that might cause rock bursts during underground activities. For example, rock bursts pose a significant risk to humans on-site during mining operations and can destroy longwalls and damage equipment. The complexity of seismic processes and the imbalanced distribution of favorable "hazardous state" and unfavorable "non-hazardous state" data points pose a significant challenge for predicting seismic hazards using machine learning approaches[67]. The original Seismic dataset is a 19-attribute binary classification dataset. It is an unbalanced dataset in which the positive (hazard) class is in the minority and considered an anomaly class. In contrast, the negative (no hazard) class is considered normal[68]. The list of seismic dataset attributes is presented in Appendix A 8.3.

The prediction horizon of the data is eight hours. This eight-hour shift indicates that the prediction methods (anomaly detection and classifiers) make seismic hazard predictions one shift in advance. Continuous data collection necessitates the aggregation of raw data prior to analysis. The aggregation process replaces a series of measurements recorded at eight-hour intervals with a single value. For instance, aggregating measurement data collected over 100 shifts yields a sequence of records or vector of variables $x_1, x_2, ..., x_{100}$, where x_t is a vector of aggregated measurement values characterizing the eight-hour interval or one shift, as denoted in the dataset. After two-month data collection process and aggregation, the seismic dataset consists of 2584 instances.

Seismic Data Hazard Assessment

Three hazard assessment methods are performed for the seismic data: seismic hazard assessment, seismoacoustic hazard assessment, and seismoacoustic hazard assessment based on only the registration of maximum energy from a geophone[67]. The main aim of the three hazard assessments is to predict increased seismic activity, which can cause a rockburst. There are four distinct categories of rockburst hazard: no hazard, low hazard, moderate hazard, and high hazard. The following are the primary assessment factors influencing the hazardous occurrence probability and the condition of rockburst hazard [69,70]:

- Coal seam thickness;
- The distance between a coal seam and a probable seismogenic layer;
- Maximum seismic energy of tremors from a particular coal seam.

The seismic hazard assessment method The essence of seismic hazard assessment is observing changes in seismic activity and identifying an increase or decrease in the degree of hazard relative to a previously determined degree [67]. Seismic hazard assessment utilizes qualitative assessment (for low seismic activity) or quantitative assessment (for high seismic activity) based on the strength of seismic tremors. The level of seismic activity is calculated by the quantity and magnitude of seismic tremors recorded in the vicinity of an observed longwall during a specific period (a shift) [67]. Table 5.5 provides the foundation for quantitative hazard assessment.

Table 5.5. Basis of hazard assessment for quantitative method, adapted from [67]

Rockburst hazard	Caved faces	Roadways
a No hazard	1. No tremors or single tremors with energies E of the order of 10^2J – 10^3J $E_{\max} \leq 10^4$J 2. $\sum E < 10^5$J per 5m of longwall advance	1. No tremors or single tremors with energies E of the order of 10^2J $E_{\max} \leq 10^3$J 2. $\sum E < 10^3$J per 5m of longwall advance
b Low hazard	1. Occurrence of tremors with energies E of the order of 10^2J – 10^5J $1 \cdot 10^4\text{J} < E_{\max} \leq 5 \cdot 10^5$J 2. $1 \cdot 10^5\text{J} \leq \sum E < 10^6$J per 5m of longwall advance	1. Occurrence of single tremors with energies E of the order of 10^2J – 10^3J $E_{\max} \leq 5 \cdot 10^3$J 2. $1 \cdot 10^3\text{J} \leq \sum E < 10^4$J per 5m of longwall advance
c Moderate hazard	1. Occurrence of tremors with energies E of the order of 10^2J – 10^6J $5 \cdot 10^5\text{J} < E_{\max} \leq 5 \cdot 10^6$J 2. $1 \cdot 10^6\text{J} \leq \sum E < 10^7$J per 5m of longwall advance	1. Occurrence of tremors with energies E of the order of 10^2J – 10^4J $5 \cdot 10^3\text{J} < E_{\max} \leq 5 \cdot 10^5$J 2. $1 \cdot 10^4\text{J} \leq \sum E < 10^5$J per 5m of longwall advance
d High hazard	1. Occurrence of tremors with energies E of the order of 10^2J – 10^6J $E_{\max} > 5 \cdot 10^6$J 2. $\sum E \geq 10^7$J per 5m of longwall advance	1. Occurrence of tremors with energies E of the order of 10^2J – 10^5J $E_{\max} > 10^5$J 2. $\sum E \geq 10^5$J per 5m of longwall advance

The seismoacoustic hazard assessment method The seismoacoustic method for assessing seismic hazard is based on the relationships between seismoacoustic emission and seismic hazard. In the seismoacoustic method, the following criteria are essential for assessing earthquake risk:

- recording of the seismoacoustic emission;
- the number of pulses recorded by geophones or denoted by seismic energy.

TABLE 5.6. Seismoacoustic method for hazard assessment, adapted from[67]

Time	$25 \leq$ DEV ≤ 100	$100 <$ DEV ≤ 200	Decrease of activity/energy after an increase of activity/energy such as 100 <DEV$\leq$ 200			DEV> 200	Decrease of activity/energy after an increase of activity/energy such as DEV>200		
			1 shift	2 shifts	>2shifts		1 shift	2 shifts	>2shifts
1 shift	a	b	a	a	current hazard state −1 after every 3 changes of activity/energy drop	c	c	c	current hazard state −1 after every 3 changes of activity/energy drop
2 shifts	a	c	b	b		d	d	d	
3 shifts	b	c	c	c		d	d	d	

Changes in recorded seismoacoustic activity and energy are the primary evaluation criteria. In addition, deviations (denoted as DEV in Table 5.6) of values calculated during subsequent time intervals also influence the classification of one of the four seismic hazard states (a,b, c, and d for no, low, medium, and high-impact hazards). Identifying the hazard level is based on the percentage changes in activity/energy value deviations (see Table 5.6).

Anomaly Detection for Seismic Data

In order to achieve the most credible results, it is essential to select the anomaly detection method that corresponds to the data description from among the vast number available. Our approach is firstly to determine if the data is Gaussian. If the data is Gaussian, anomalies often reside away from the peak of the normal distribution[71]. The normality test of the seismic dataset, performed with Python Library for statistical calculations *Shapiro-Wilk Test for Normality*.

[72], in our case study indicates that the seismic data is not Gaussian with p-value approaching 0. The data distribution contains more information than the covariance matrix, which measures how much two random variables change together and is helpful for normal data but less for non-Gaussian data. Plotting

noise and artificial anomalies is more difficult for this type of data. Correlation between attributes or their relevance to one another is an additional essential characteristic of data. Figure 5.9 demonstrates the heat map illustrating the magnitude of the correlation between attributes.

For this dataset we have selected an anomaly detection method *Isolation Forest* that isolates anomalies with minimal computational needs.[8] provides a comprehensive summary of each step of the Isolation Forest algorithm and the underlying equations. Since Isolation Forest is capable of isolating outstanding data points efficiently, it can also be used to determine if these points share similarities with hazards, i.e., if hazards also appear as outstanding points and if they are apparent to both domain experts and anomaly detection methods. If they are difficult to detect, it indicates that the anomalies may not share contextual properties with hazards, so the autonomous approach may need to be modified accordingly. This knowledge can be used to extract what is not apparent for anomaly detection to detect the hazards and determine what properties to introduce to increase apparency and improve autonomous detection, since not every anomaly is a hazard and vice versa. In our case study, Isolation Forest provides a suitable testing environment for measuring an al-

FIGURE 5.9. Heat map - correlations between dataset attributes

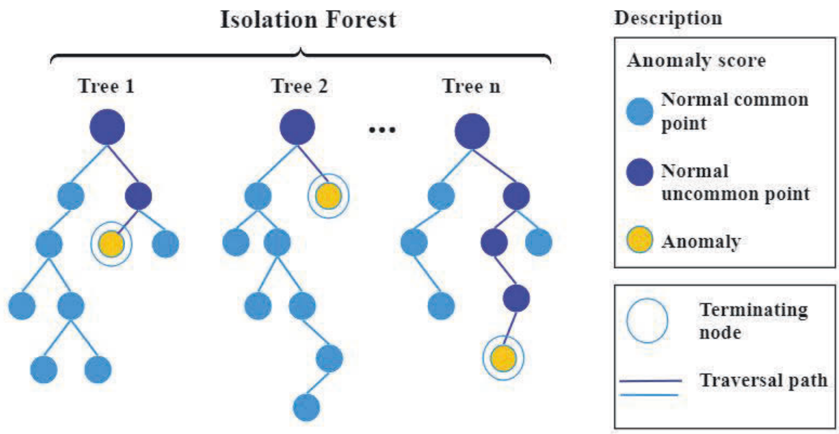

FIGURE 5.10. Isolation Forest, illustrated. Adapted from[73]

gorithm's capacity in isolating anomalies and determining whether they are comparable to the anomalies that a human domain expert would identify as seismic hazards. Isolation Forest effectively solves high-dimensional problems with multiple non-correlated attributes by constantly and recursively splitting instances until they are isolated by their normal common, normal uncommon, and anomalous occurrence[8], as illustrated in Figure 5.10. Isolation Forest does not rely on distance or density metrics to identify anomalies, eliminating processing expenses and making it suitable for nonlinear datasets. It is expected that there are fewer true anomalies in the dataset; therefore, they are more susceptible to isolation, eliminating the overabundance of registered noise or false alarms. It is a widely applied and one of the most developed unsupervised anomaly detection methods. The efficiency of Isolation Forest is in the way it builds a normal data point profile and isolates the points that do not fit that profile, taking advantage of anomalous properties and uncommon values, as illustrated in Figure 5.10. Algorithm Part 1, 2, and 3 show the algorithm details of Isolation Forest split into three parts: initialization of a forest, initialization of a single tree (more of which construct a forest), and calculation of traversal path length, a path between the tree node and the isolated anomaly. A group of isolation trees finds anomalies as points with path lengths, with numerous trees functioning as "domain experts" to target the anomalies[8]. Additionally, the Isolation Forest does not need to separate the majority of the training sample consisting of normal examples.

As described in the Algorithm Part 1 and 2, the trees are produced by iteratively splitting the data until instances are isolated or a predetermined tree

Algorithm Part 1: Creating a forest, adapted from [8]

Input: X - input data, t - number of trees, ψ - sub-sampling size
Output: a set of t *iTrees*

```
Initialize Forest
set height limit l = ceiling(log2 ψ)
for i = 1 to t do
    X' ← sample(X, ψ)
    Forest ← Forest ∪ iTree(X', 0, l)
    return Forest
end for
```

Algorithm Part 2: Creating a tree, adapted from [8]

Input: X - input data, e - current tree height, l - height limit
Output: an *an iTree*

```
if e ≥ l or X ≤ 1 then
    return exNode { Size ← | X | }
else
    let Q be a list of attributes in X,
    randomly select an attribute q ∈ Q,
    randomly select a split point p from max and min values of q in X,
    X_l ← filter(X, q ¡ p),
    X_r ← filter(X, q ≥ p),
    return inNode {
    Left ← iTree(X_{l,e} + 1, l),
    Right ← iTree(X_{r,e} + 1, l),
    SplitAtt ← q,
    SplitValue ← p
end if
```

Algorithm Part 3: Calculating path length, adapted from [8]

Input: x - an instance, T - an iTree, e - current path length; to be initialized to zero when first called
Output: path length of x

```
if T is an external node then
    return e + c (T.size) {where c is average search path}
end if
a ← T.splitAtt
if X_a < T.splitValue then
    return PathLength(x, T.left, e + 1)
else X_a ≥ T.splitValue
    return PathLength(x, T.right, e + 1)
end if
```

height is attained, resulting in a partial model. The algorithm automatically determines the tree height limit based on the sub-sampling size, which is denoted as the height limit variable. Finding the average height limit is necessary because shorter-than-average path lengths are more likely to be anomalies. Sub-sampling size ψ that controls the data size is reliably detected by Isolation Forest, keeping the performance, processing time, and memory size optimal. Algorithm Part 3 depicts the evaluating stage in which an anomaly score s is derived from the expected path length for each test instance, which is obtained by passing instances through each tree in the forest. A single path length is determined by counting the number of edges e from the root node to a terminating node as an instance traverses a tree.

When a single path is obtained for each tree in the forest, an anomaly score Equation s is derived following the 5.1, where $h(x)$ denotes the path length, $E(h(x))$ is the normalized $h(x)$ from a collection of isolation trees, and $c(n)$ is the average of path lengths. Finally, the data are then sorted in descending order to identify the most significant anomalies.

$$s = 2 - \frac{E(h(x))}{c(n)} \tag{5.1}$$

WIF Steps: Application of Risk Definition

Step 1: Warning Characterization Answering the question *"What can go wrong during the given operation and given the context and circumstances?"* necessitates domain expert observations. For the seismic dataset, this is answered through three hazard assessment methods. Table 5.5 and Table 5.6 provide the hazards as the events that *can go wrong*. These hazardous events serve as the ground truth for testing the capability of anomaly detection method to detect the same events as anomalies. The results of the hazard assessment methods are shown in Table 5.7. The three methods do not yield the same amount of hazardous and non-hazardous states. Upon closer inspection, the number of equal instances of the non-hazardous state resulting from seismic and seismoacoustic hazard assessment methods is 1071, and the number of equally denoted hazardous states is 393. As suggested by [67], knowledge of the present hazard state is essential for production process management and industrial safety. However, assessing and predicting seismic hazards is a highly complex procedure with a substantial element of randomness.

Treating each result of hazard assessment methods as different ground truths, the information derived by domain expert observations and known as

Table 5.7. Number of hazardous and non-hazardous instances labeled by hazard assessment methods

	Hazard Assessment Methods		
State	Seismic	Seismoacoustic	Seismoacoustic by max energy
Non-hazardous	1682	1580	2342
Hazardous	902	1004	242

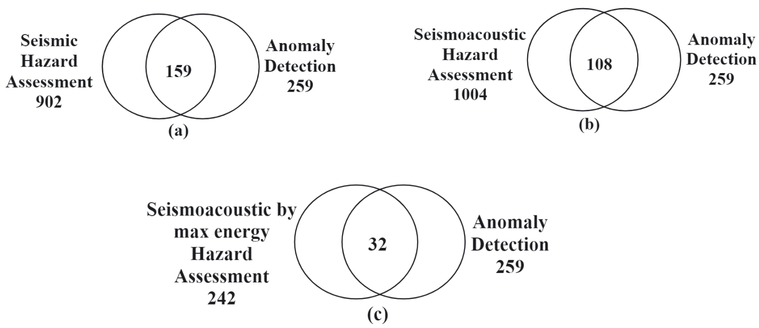

Figure 5.11. Hazards identified by anomaly detection and (a) seismic hazard assessment, (b) seismoacoustic hazard assessment, (c) seismoacoustic by max energy hazard assessment methods

the absolute truth, the anomaly detection results show significantly different numbers (see Figure 5.11). Out of the 259 detected anomalies, 159 are the hazardous state identified by seismic (Figure 5.11 (a)), and 108 are by the seismoacoustic hazard assessment method (Figure 5.11(b)). These results lead to an early conclusion that approximately half of the anomalies detected by the anomaly detection method are considered hazardous, and the other detected anomalies are of no significance. Compared to the results of seismoacoustic by max energy results of 242 hazardous states, anomaly detection has identified only 32 (see Figure 5.11 (c)). The Figure 5.11 illustrates the critical difference and the main shortcoming of the anomaly detection method, the inability to independently detect true hazards and a substantial number of false alarms. The confusion matrices in Table 5.8, Table 5.9, and Table 5.10 provide additional insight into these results. Despite three distinct seismic hazard assessment methods representing hazard occurrences, seismic, seismoacoustic, and seismoacoustic by maximum energy, Isolation Forest demonstrated an insufficient understanding of hazards. This evidence may prompt an early proposition that

TABLE 5.8. Confusion Matrix with Seismic Hazard Assesment as True Values, Isolation Forest Anomalies as Test Values

		Anomalies by Isolation Forest: Positive	Anomalies by Isolation Forest: Negative	Total
Hazards by Seismic	Positive	159	743	902
	Negative	100	1582	1682
	Total	259	2325	

TABLE 5.9. Confusion Matrix with Seismoacoustic Hazard Assesment as True Values, Isolation Forest Anomalies as Test Values

		Anomalies by Isolation Forest: Positive	Anomalies by Isolation Forest: Negative	Total
Hazards by Seismoac.	Positive	108	896	1004
	Negative	151	1429	1580
	Total	259	2325	

unsupervised anomaly detection may not be appropriate for seismic hazard detection despite its widespread use for unusual patterns and threat detection and that seismic hazard assessment is required as an element of supervision.

Step 2: Warning Analysis Warning Analysis intends to detect patterns in which the hazards may occur and the likelihood of their occurrence. Conditional probability P, Equation 5.2, is the likelihood that an event A or outcome will occur given the occurrence of a prior event or outcome B, C, D[62]. Multiplying the likelihood of the preceding event by the updated probability of the subsequent, or conditional, occurrence yields the conditional probability.

TABLE 5.10. Confusion Matrix with Seismoacoustic by Max Energy Hazard Assesment as True Values, Isolation Forest Anomalies as Test Values

		Anomalies by Isolation Forest: Positive	Anomalies by Isolation Forest: Negative	Total
Hazards by Max Energy	Positive	32	210	242
	Negative	227	2115	2342
	Total	259	2325	

$$P(A \mid B, C, D) = \frac{P(A \cap B)}{P(B, C, D)} \tag{5.2}$$

Table 5.11 shows the results of conditional probability of each nonhazardous occurrence and hazard, categorized by their impact. Table 5.11 represents the probability for each state (either no hazard, low, medium, high-impact) given the occurrence of the other states. The impact of the hazard is derived following the causal knowledge described in Table 5.5 and Table 5.6 on hazard detection patterns with the task of hazard prediction based on the association between the energy of recorded seismic tremors and seismoacoustic activity with the probability of seismic tremor occurrence[67].

TABLE 5.11. Conditional probability of hazard occurring, by hazard assessment methods

Conditional Probability by Hazard Impact, expressed in percentages			
State	Seismic	Seismoacoustic	Seismoacoustic by max energy
No hazard	65.09	61.14	90.63
Low-impact hazard	34.90	36.99	8.20
Medium-impact hazard	1.80	1.85	1.16
High-impact hazard	0	0	0

Even though this case study has an exact number of seismic hazards, it is not always expected that seismic tremor monitoring operations will have sensor data for identifying hazards analyzed by domain experts. In the event of not having an exact number of hazards derived from sensor data by domain experts, knowing the frequency of a hazardous occurrence in a given environment may help compensate for situations in which a large number of anomalies are reported in order to determine the likelihood of the anomaly being a true hazard or a false alarm.

Step 3: Warning Prioritization The three approaches to hazard assessment for seismic data provide the impact of the hazard on four levels (hazard impacts derived following the Table 5.5):

1. No hazard

2. Low-impact hazard
3. Medium-impact hazard
4. High-impact hazard

Table 5.12 shows the number of hazards, categorized by their impact, detected by the three hazard assessment methods. In comparison, Table 5.13 represents the number of anomalies detected by the unsupervised anomaly detection method, Isolation Forest, where each hazard assessment method is used to categorize the hazards and their impacts among the detected anomalies.

TABLE 5.12. Hazard impacts by hazard assessment methods

Hazard Assessment Methods: Hazard Impact			
State	Seismic	Seismoacoustic	Seismoacoustic by max energy
No hazard	1682	1580	2342
Low-impact hazard	902	956	212
Medium-impact hazard	0	48	30
High-impact hazard	0	0	0

As presented in Table 5.12, during the hazard assessment, there were no records of high-impact hazardous occurrences. The seismic hazard assessment method has identified only low-impact hazards, and no medium or high impact hazards. According to the impact, the reactions during operations can be prioritized.

Anomaly detection has provided poor results concerning the identification of various levels of hazard impacts, presented in Table 5.13. For low impact-hazards, anomaly detection has, on average, identified only 14,2% of the low-impact hazards, and for medium-impact hazards, only 14,5% of the cases on average. These results indicate that unsupervised anomaly detection cannot reliably identify seismic hazards and distinguish them based on their severity impact. Therefore, a form of supervision, as demonstrated with different hazard assessment approaches, is necessary to introduce.

Case Study Summary and Opportunities for a Generalized Framework

The case study architecture of WIF, applied to identify seismic hazards among detected anomalies by unsupervised anomaly detection, is illustratated in Figure 5.12. Isolation Forest, a method for unsupervised anomaly detection, analyzes unlabeled seismic sensor data and detects a group of anomalies. However, within the detected anomalies, there is yet no knowledge of which ones are false warnings and the ones that are true warnings or hazards. In this case study, domain expert knowledge is leveraged through hazard assessment criteria based on three methods: seismic, seismoacoustic, and seismoacoustic with maximum energy. Hazards, or true warnings, can be extracted from the given dataset and compared to anomalies detected by the unsupervised method to determine if the unsupervised method can capture the properties of hazards and report them as anomalies while ignoring false warnings. These hazard anomalies can be prioritized based on their impact, such as none, low, medium, and high.

As the use of data-driven and machine learning methods increases, the problem of unintended and harmful behavior of machine learning systems resulting from poor design of real-world AI systems becomes increasingly apparent[74]. Unsupervised anomaly detection, classification, and other data-driven machine learning methods face well-known challenges:

- biased data,
- false positives and false negatives (false alarms),
- prioritization of anomaly reporting for anomaly detection applications,
- lack of context that is tied to all of the previous challenges, and
- lack of explainability of the results produced by unsupervised methods

Table 5.13. Hazard impacts identified by anomaly detection methods, with hazard assessment methods as the ground truth

Anomalies detecting hazard impacts			
State	Seismic	Seismoacoustic	Seismoacoustic by max energy
Low-impact hazard	166	104	29
Medium-impact hazard	0	6	5
High-impact hazard	0	0	0

Introducing a supervisory component to data-driven systems is a step toward providing context to the method, reducing biases and false reporting, adding prioritization knowledge, and improving the explainability of the results as they are derived from more traditional risk, and hazard assessment approaches. The approach studied in this section can be generalized by observing risk assessment methods and properties of hazards for a specific operation, where hazard properties may serve as a class label by which the unsupervised data-driven method can be validated. According to a technical report and recommendations on AI and safety by ISO/IEC, 2022[75], providing explainable algorithms and results and validating them in the real world characterizes the future of AI-related systems and safety.

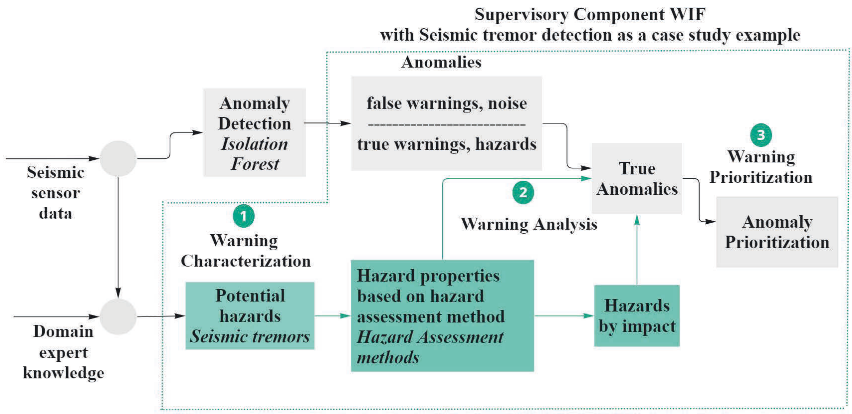

FIGURE 5.12. Case Study Summary Architecture

5.2.10 Discussion

The results of the case study show shortcomings of an unsupervised anomaly detection method through a clear difference between the identified hazards by three seismic hazard assessment methods and their results with unsupervised anomaly detection method Isolation Forest. The findings from Step 1 Warning Characterization provide crucial insights into unsupervised anomaly detection and seismic hazard assessment differences. During safety-critical operations, such as seismic hazard monitoring, it is essential to assess the difference between the discovered hazardous states and adapt our expectations for the implementation of anomaly detection. Since it is not expected that an operation will have labeled training dataset, and identified hazards list by domain experts for each case, the analysis of the data becomes unsupervised. This step showed unexpected differences between the number of seismic hazards identified by domain expert inputs and anomalies identified by unsupervised anomaly detection. An unexpectedly low number of detected anomalies proved to be hazards when categorized following the hazard identification criteria presented in Table 5.5 and Table 5.6. This leads to an assumption that unsupervised anomaly detection, despite being used for detection of threats and unusual patterns, cannot be trusted to detect all seismic hazards. An additional layer of context, namely through hazard assessment methods, is necessary to distinguish the anomalies that are only data discrepancies and offer no significance, from the ones that are hazardous.

The results of conditional probability obtained in Step 2, Warning Analysis, and based on hazard assessment methods, are crucial to setting the expectation of the occurrence of a hazard of varying degrees of impact. The seismic and seismoacoustic hazard assessment methods resulted in the highest probability of a non-hazardous event, followed by a low-impact hazard and a low probability of medium-impact hazard occurrence. The data in this case study provided no evidence of high-impact hazards, resulting in an expected no probability of high-impact hazardous occurrences. In comparison, the seismoacoustic by maximum energy method resulted in the highest probability of 90.53% of non-hazardous occurrence, followed by a low probability of low-impact and medium-impact hazards. This step showed a limitation in the case study data where the lack of high-impact hazard evidence resulted in a 0 % probability of such hazards occurring. This imbalance in hazard impacts can lead to biases during anomaly detection or hazard identification methods.

Further analysis in Step 3, Warning Prioritization, categorized the identified hazards in the varying degrees of impact: no hazard, low-impact, medium-impact, and high-impact. The anomaly detection method resulted in fewer identified hazards than the hazard assessment methods. This step showed the low reliability of the anomaly detection method as an autonomous hazard identification approach.

The case study results have validated the assertion that unsupervised anomaly detection generates a considerable amount of false alarms, that may waste operator response resources if the methods are used as a part of an autonomous drone or smart-sensor system. These results provide valuable insight into the possibilities of addressing the shortcomings of unsupervised anomaly detection methods for seismic hazard identification, where risk assessment approaches, such as hazard identification, can play a crucial role.

5.2.11 Contribution Summary

Recent research provides different approaches to handling discussed challenges through simulations, rule-based classification, and decision boundaries. However, these approaches do not address the explainability of the data-driven methods and introduce new complexities. The results and contributions of this section can be summarized as follows:

1. A novel outlook on utilizing existing domain knowledge in seismic tremors through seismic hazard assessment methods as a supervisory component for unsupervised anomaly detection through the Warning Identification Framework based on risk assessment, resilience and reliability engineering, and future human-machine teaming expectations.

2. Identification of overlapping tasks for risk assessment and anomaly detection objectives that can be utilized in addressing the shortcomings of anomaly detection results.

3. A case study examining the sensor-obtained seismic data for monitoring seismic tremors and analyzing three different hazard identification methods in comparison to unsupervised anomaly detection for hazard identification.

During our analysis, we identified significant anomaly shortcomings in detection methods to detect hazardous occurrences by their levels of impact and to distinguish anomalies of no significance from the anomalies that represent hazardous occurrences. The results of this research show significant opportunities in utilizing risk assessment insights to tackle the shortcomings of unsupervised anomaly detection methods and aid a more reliable and transparent hazard detection.

5.3 Conclusions and Key Contributions

This section highlights the key contributions and concludes the chapter and the presented articles.

It is anticipated that cyber-physical and intelligent sensor systems will play a permanent role in industrial operations, including monitoring, inspecting, and intervening with assets and the environment, necessitating greater autonomy for making significant decisions in near-real and real-time. Current challenges include a lack of context, the underutilization of causal knowledge, and an excess of imbalanced data. We discussed the growing need for employing data-driven methods in a more explainable, transparent, and reliable practice.

The key findings of the two papers presented in this chapter confirmed the challenges inherent in sensor data. They uncovered the most significant algorithmic shortcoming resulting from a lack of context: the abundance of noise compared to significant anomalies as hazards. The analysis of seismic hazard data provided solid justification for using traditional methods to add the context of risk to semi-supervised anomaly detection results.

5.4 References

[1] L. Erhan et al. Smart anomaly detection in sensor systems: A multi-perspective review. *Information Fusion* **67** (September 2020), 64–79 (2021). ISSN 15662535. doi: 10.1016/j.inffus.2020.10.001. Cited on page/s 81, 82.

[2] Muhammad Zohaib Anwar, Zeeshan Kaleem, and Abbas Jamalipour. Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications. *IEEE Transactions on Vehicular Technology* **68** (3), 2526–2534 (2019). ISSN 00189545. doi: 10.1109/TVT.2019.2893615. Cited on page/s 82.

[3] Su Yeon Choi and Dowan Cha. Unmanned aerial vehicles using machine learning for autonomous flight; state-of-the-art. *Advanced Robotics* **33** (6), 265–277 (2019). ISSN 15685535. doi: 10.1080/01691864.2019.1586760. URL https://doi.org/10.1080/01691864.2019.1586760. Cited on page/s 82.

[4] Nour Moustafa and Alireza Jolfaei. Autonomous detection of malicious events using machine learning models in drone networks. *DroneCom 2020 - Proceedings of the 2nd ACM MobiCom Workshop on Drone Assisted Wireless Communications for 5G and Beyond* pages 61–66 (2020). doi: 10.1145/3414045.3415951. Cited on page/s 82.

[5] Dario De Dominicis and Domenico Accardo. Software and sensor issues for autonomous systems based on machine learning solutions. In *2020 IEEE International Workshop on Metrology for AeroSpace, MetroAeroSpace 2020 - Proceedings* pages 545–549 (2020). ISBN 9781728166360. doi: 10.1109/MetroAeroSpace48742.2020.9160292. Cited on page/s 83.

[6] Alberto Castellini, Domenico Bloisi, Jason Blum, Francesco Masillo, and Alessandro Farinelli. Multivariate sensor signals collected by aquatic drones involved in water monitoring: A complete dataset. *Data in Brief* **30**, 105436 (2020). ISSN 23523409. doi: 10.1016/j.dib.2020.105436. URL https://doi.org/10.1016/j.dib.2020.105436. Cited on page/s 83, 84, 85.

[7] Alberto Castellini, Giovanni Beltrame, Manuele Bicego, Jason Blum, Matteo Denitto, and Alessandro Farinelli. Unsupervised activity recognition for autonomous water drones. *Proceedings of the ACM Symposium on Applied Computing* pages 840–842 (2018). doi: 10.1145/3167132.3167396. Cited on page/s 84, 88, 92.

[8] Fei Tony Liu, Kai Ming Ting, and Zhi Hua Zhou. Isolation forest. In *Proceedings - IEEE International Conference on Data Mining, ICDM* pages 413–422. IEEE (2008). ISBN 9780769535029. doi: 10.1109/ICDM.2008.17. Cited on page/s 85, 115, 116, 117.

[9] Valeria Fonti. Feature Selection using LASSO. *VU Amsterdam* pages 1–26 (2017). ISSN 2169-3536. Cited on page/s 85, 86.

[10] Sanjay Yadav and Sanyam Shukla. Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. In *Proceedings - 6th International Advanced Computing Conference, IACC 2016* number Cv pages 78–83. IEEE (2016). ISBN 9781467382861. doi: 10.1109/IACC.2016.25. Cited on page/s 86.

[11] Yan Yan Song and Ying Lu. Decision tree methods: applications for classification and

prediction. *Shanghai Archives of Psychiatry* **27** (2), 130–135 (2015). ISSN 10020829. doi: 10.11919/j.issn.1002-0829.215044. Cited on page/s 86.

[12] Mariana Belgiu and Lucian Drăgu. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* **114**, 24–31 (2016). ISSN 09242716. doi: 10.1016/j.isprsjprs.2016.01.011. Cited on page/s 86.

[13] Mohammed J. Islam, Q. M. Jonathan Wu, Majid Ahmadi, and Maher A. Sid-Ahmed. Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers. In *International Conference on Convergence Information Technology (ICCIT 2007)* pages 1541–1546. IEEE (2008). ISBN 0769530389. doi: 10.1109/iccit.2007.148. Cited on page/s 86, 87.

[14] Anissa Bouzalmat, Jamal Kharroubi, and Arsalane Zarghili. Comparative study of PCA, ICA, LDA using SVM classifier. *Journal of Emerging Technologies in Web Intelligence* **6** (1), 64–68 (2014). ISSN 17998859. doi: 10.4304/jetwi.6.1.64-68. Cited on page/s 87.

[15] Jerome Boyer. Evaluate and select a machine learning algorithm - IBM Garage Practices (2021). URL https://www.ibm.com/garage/method/practices/reason/evaluate-and-select-machine-learning-algorithm/. Cited on page/s 87.

[16] Laura Toloşi and Thomas Lengauer. Classification with correlated features: Unreliability of feature ranking and solutions. *Bioinformatics* **27** (14), 1986–1994 (2011). ISSN 13674803. doi: 10.1093/bioinformatics/btr300. Cited on page/s 88.

[17] Jason Brownlee. Framework for Imbalanced Classification Projects (3 2020). URL https://machinelearningmastery.com/framework-for-imbalanced-classification-projects/. Cited on page/s 89, 90.

[18] George Vachtsevanos, Benjamin Lee, Sehwan Oh, and Michael Balchanos. Resilient Design and Operation of Cyber Physical Systems with Emphasis on Unmanned Autonomous Systems. *Journal of Intelligent and Robotic Systems: Theory and Applications* **91** (1), 59–83 (2018). ISSN 15730409. doi: 10.1007/s10846-018-0881-x. Cited on page/s 94, 103.

[19] Ibrahim Habli John McDermid, Yan Jia. Towards a Framework for Safety Assurance of Autonomous Systems. In *AISafety@IJCAI*. Artificial Intelligence Safety 2019 (2019). doi: orcid.org/0000-0003-4745-4272. URL https://eprints.whiterose.ac.uk/150187/. Cited on page/s 94.

[20] Oxford University Press. Oxford Learner's Dictionaries (2021). URL https://www.oxfordlearnersdictionaries.com/. Cited on page/s 94.

[21] Katherine Fraser, Samuel Homiller, Rashmish K. Mishra, Bryan Ostdiek, and Matthew D. Schwartz. Challenges for Unsupervised Anomaly Detection in Particle Physics. *arXiv* (2021). URL https://arxiv.org/abs/2110.06948v1. Cited on page/s 94.

[22] Simen Eldevik and Frank Borre Pedersen. AI + safety - DNV. Technical report DNV (2018). URL https://www.dnv.com/oilgas/download/artificial-intelligence-ai-and-safety.html. Cited on page/s 94, 102, 104, 105.

[23] Rialda Spahic, Vidar Hepso, and Mary Ann Lundteigen. Enhancing Autonomous Systems' Awareness: Conceptual Categorization of Anomalies by Temporal Change During Real-Time Operations. In *The Eighteenth International Conference on Autonomic and Autonomous Systems* pages 25–30 Venice, Italy (5 2022). International Academy, Research and Industry Association (IARIA). Cited on page/s 94.

[24] Rialda Spahic, Vidar Hepso, and Mary Ann Lundteigen. Reliable Unmanned Autonomous Systems: Conceptual Framework for Warning Identification during Remote Operations. *2021 IEEE International Symposium on Systems Engineering (ISSE)* pages 1–8 (9 2021). doi: 10.1109/ISSE51541.2021.9582534. URL https://ieeexplore.ieee.org/document/9582534/. Cited on page/s 94, 107.

[25] Rialda Spahic, Hepsø, Vidar, and Mary Ann Lundteigen. Using Risk Analysis for Anomaly Detection for Enhanced Reliability of Unmanned Autonomous Systems. In Maria Chiara Leva, Edoardp Patelli, Luca Podofillini, and Simon Wilson, editors, *Proceedings of the 32nd*

European Safety and Reliability Conference (ESREL 2022) pages 273–280 Singapore (2022). Research Publishing, Singapore. doi: 10.3850/978-981-18-5183-4{_}R08-03-390-cd. URL https://rpsonline.com.sg/rps2prod/esrel22-epro/html/toc.html. Cited on page/s 94.

[26] Charu C. Aggarwal. An Introduction to Outlier Analysis. In *Outlier Analysis* chapter 1, pages 1–34. Springer, Cham (2017). ISBN 978-3-319-47577-6. doi: 10.1007/978-3-319-47578-3{_}1. Cited on page/s 95.

[27] Ayman Taha and Ali S. Hadi. Anomaly detection methods for categorical data: A review. *ACM Computing Surveys* **52** (2) (5 2019). ISSN 15577341. doi: 10.1145/3312739. Cited on page/s 95.

[28] D. M. Hawkins. Identification of Outliers. *Identification of Outliers* (1980). doi: 10.1007/978-94-015-3994-4. Cited on page/s 95.

[29] R J Beckman and R D Cook. Outlier … … …. s. *Technometrics* **25** (2), 119–149 (1983). doi: 10.1080/00401706.1983.10487840. URL https://doi.org/10.1080/00401706.1983.10487840. Cited on page/s 95.

[30] Ralph Foorthuis. On the nature and types of anomalies: a review of deviations in data. *International Journal of Data Science and Analytics* **12** (4), 297–331 (2021). ISSN 23644168. doi: 10.1007/s41060-021-00265-1. Cited on page/s 95, 102.

[31] ISO 31000. Risk management — Guidelines, International Organization for Standardization. Technical report International Organization for Standardization (2018). URL https://www.iso.org/obp/ui/iso:std:iso:31000:ed-2:v1:en. Cited on page/s 96.

[32] ISO:51. Safety aspects - Guidelines for their inclusion in standards ISO/IEC Guide 51:2014(E). Technical report International Organization for Standardization and the International Electrotechnical Commissio (2014). Cited on page/s 96.

[33] Marvin Rausand. Risk Assessment Theory, Methods, and Applications. John Wiley & Sons Inc Hoboken, New Jersey (2011). ISBN 9780470637647. doi: 10.1002/9781118281116. Cited on page/s 96, 97, 99, 110.

[34] Gabriel Michau and Olga Fink. Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. *Knowledge-Based Systems* **216**, 106816 (3 2021). ISSN 0950-7051. doi: 10.1016/J.KNOSYS.2021.106816. Cited on page/s 97.

[35] Eric William Scharpf, Harold W. Thomas, and Todd R. Stauffer. Practical SIL Target Selection: Risk Analysis Per the IEC 61511 Safety Lifecycle. exida.com LLC Sellersville, Pennsylvania 2 edition (2012). ISBN 978-1-934977-03-3. Cited on page/s 97, 98, 99.

[36] Lacher and Andrew R. A Framework for Discussing Trust in Increasingly Autonomous Systems. Technical report The MITRE Corporation (2017). Cited on page/s 100, 101.

[37] Rolando Garcia, Vikram Sreekanti, Neeraja Yadwadkar, Daniel Crankshaw, Joseph E Gonzalez, and Joseph M Hellerstein. Context: The Missing Piece in the Machine Learning Lifecycle. In *Common Model Infrastructure* London, UK (2018). Cited on page/s 100.

[38] Sebastian Schelter, Felix Biessmann, Tim Januschowski, David Salinas, Stephan Seufert, and Gyuri Szarvas. On Challenges in Machine Learning Model Management. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* pages 5–13 (2018). URL http://sites.computer.org/debull/A18dec/p5.pdf. Cited on page/s 100.

[39] Behrouz Derakhshan, Alireza Rezaei Mahdiraji, Ziawasch Abedjan, Tilmann Rabl, and Volker Markl. Optimizing Machine Learning Workloads in Collaborative Environments. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* pages 1701–1716. Association for Computing Machinery (6 2020). ISBN 9781450367356. doi: 10.1145/3318464.3389715. Cited on page/s 100.

[40] Michael A. Hayes and Miriam A.M. Capretz. Contextual anomaly detection in big sensor data. In *Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014* pages 64–71. Institute of Electrical and Electronics Engineers Inc. (9 2014). ISBN

9781479950577. doi: 10.1109/BigData.Congress.2014.19. Cited on page/s 101, 102.

[41] Michele Seng Ah Lee and Jatinder Singh. Risk Identification Questionnaire for Detecting Unintended Bias in the Machine Learning Development Lifecycle. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (1 2021). Cited on page/s 101.

[42] David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. Fairness through causal awareness: Learning causal latent-variable models for biased data. *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency* pages 349–358 (1 2019). doi: 10.1145/3287560.3287564. Cited on page/s 102.

[43] Karima Makhlouf, Sami Zhioua, and Catuscia Palamidessi. On the applicability of ML fairness notions. *arXiv* pages 1–32 (2020). ISSN 23318422. Cited on page/s 102.

[44] R. Sekar et al. Specification-based Anomaly Detection: A New Approach for Detecting Network Intrusions. In *Proceedings of the 9th ACM conference on Computer and communications security - CCS '02* page 265–274 New York, NY, USA (2002). Association for Computing Machinery. ISBN 1581136129. doi: 10.1145/586110.586146. Cited on page/s 102.

[45] Animesh Patcha and Jung Min Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks* **51** (12), 3448–3470 (8 2007). ISSN 1389-1286. doi: 10.1016/J.COMNET.2007.02.001. Cited on page/s 102.

[46] Christina Gopfert, Shai Ben-David, Olivier Bousquet, Sylvain Gelly, Ilya O Tolstikhin, and Ruth Urner. When can unlabeled data improve the learning rate? In *Annual Conference Computational Learning Theory*. Association for Computational Learning (2019). Cited on page/s 102.

[47] Maximilian Henne, A. Schwaiger, and Gereon Weiss. Managing Uncertainty of AI-based Perception for Autonomous Systems. *AISafety@IJCAI* (2019). Cited on page/s 102.

[48] Structures J Phillip Durst and Wendell Gray. ERDC/GSL SR-14-1 "Levels of Autonomy and Autonomous System Performance Assessment for Intelligent Unmanned Systems". Technical report The US Army Engineer Research and Development Center (ERDC) (2014). URL www.erdc.usace.army.mil. Cited on page/s 102.

[49] Curtis Marshall, Blake Roberts, and Michael Grenn. Intelligent control & supervision for autonomous system resilience in uncertain worlds. In *3rd International Conference on Control, Automation and Robotics, ICCAR 2017* pages 438–443. Institute of Electrical and Electronics Engineers Inc. (6 2017). ISBN 9781509060870. doi: 10.1109/ICCAR.2017.7942734. Cited on page/s 102.

[50] Erik Hollnagel, David D. Woods, and Nancy Leveson. Resilience Engineering: Concepts and Precepts . ASgate Publishing Ltd (2007). ISBN 9780754649045. Cited on page/s 103.

[51] Cathy O'Neil. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishers New York (2016). ISBN 978-0553418811. Cited on page/s 103.

[52] Tianci Zhang, Jinglong Chen, Fudong Li, Kaiyu Zhang, Haixin Lv, Shuilong He, and Enyong Xu. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Transactions* **119**, 152–171 (2022). ISSN 00190578. doi: 10.1016/j.isatra.2021.02.042. Cited on page/s 105.

[53] Alexander Lavin and Subutai Ahmad. Evaluating Real-time Anomaly Detection Algorithms - the Numenta Anomaly Benchmark. In *IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015* pages 38–44 Miami, Florida, USA (2015). Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICMLA.2015.141. Cited on page/s 105.

[54] Peng Li, Oliver Niggemann, and Barbara Hammer. On the identification of decision boundaries for anomaly detection in CPPS. In *Proceedings of the IEEE International Conference on Industrial Technology* volume 2019-Febru pages 1311–1316. Institute of Electrical and Electronics Engineers Inc. (2 2019). ISBN 9781538663769. doi: 10.1109/ICIT.2019.8754997.

Cited on page/s 105, 106.

[55] Dong Hoon Shin, Roy C. Park, and Kyungyong Chung. Decision Boundary-Based Anomaly Detection Model Using Improved AnoGAN from ECG Data. In *IEEE Access* volume 8 pages 108664–108674. Institute of Electrical and Electronics Engineers Inc. (2020). doi: 10.1109/ACCESS.2020.3000638. Cited on page/s 106.

[56] Salima Omar, Asri Ngadi, and Hamid H. Jebur. Machine Learning Techniques for Anomaly Detection: An Overview. *International Journal of Computer Applications* **79** (2), 33–41 (10 2013). doi: 10.5120/13715-1478. Cited on page/s 106.

[57] J Deng and E T Brown. RISSAD : Rule-based Interactive Semi-Supervised Anomaly Detection. In *EuroVis 2021* Chicago, IL, U.S.A. (2021). The Eurographics Association. doi: 10.2312/evs.20211050. Cited on page/s 107.

[58] E. Zio. The future of risk assessment. *Reliability Engineering and System Safety* **177** (March), 176–190 (2018). ISSN 09518320. doi: 10.1016/j.ress.2018.04.020. Cited on page/s 107.

[59] D.R. Kiran. Reliability Engineering. In *Total Quality Management* pages 391–404. Elsevier (1 2017). doi: 10.1016/B978-0-12-811035-5.00027-1. URL https://linkinghub.elsevier.com/retrieve/pii/B9780128110355000271. Cited on page/s 108.

[60] Erik Hollnagel. Resilience Engineering: A New Understanding of Safety. *Journal of the Ergonomics Society of Korea* **35** (3), 185–191 (6 2016). ISSN 1229-1684. doi: 10.5143/jesk.2016.35.3.185. URL http://dx.doi.org/10.5143/JESK.2016.35.3.185http://jesk.or.kreISSN:2093-8462. Cited on page/s 108.

[61] Patricia Mcdermott, Cindy Dominguez, Nicholas Kasdaglis, Matthew Ryan, Isabel Trahan Mitre, and Alexander Nelson. Human-Machine Teaming Systems Engineering Guide. Technical report The MITRE Corporation (2018). URL https://www.mitre.org/publications/technical-papers/human-machine-teaming-systems-engineering-guide. Cited on page/s 108.

[62] Tore Schweder and Nils Lid Hjort. Confidence, likelihood, probability: An invitation. In *Confidence, Likelihood, Probability* pages 1–22. Cambridge University Press (2 2016). doi: 10.1017/CBO9781139046671.002. URL https://www.cambridge.org/core/product/identifier/CBO9781139046671A008/type/book_part. Cited on page/s 111, 120.

[63] Kishan G Mehrotra, Chilukuri K Mohan, and Huaming Huang. Anomaly Detection. In *Anomaly Detection Principles and Algorithms. Terrorism, Security, and Computation* volume XXII pages 21–32. Springer 1 edition (2017). Cited on page/s 111.

[64] Charu C. Aggarwal. Applications of Outlier Analysis. In *Outlier Analysis* pages 399–422. Springer, Cham (2017). doi: 10.1007/978-3-319-47578-3{_}13. URL https://link.springer.com/chapter/10.1007/978-3-319-47578-3_13. Cited on page/s 111.

[65] Markos Markou and Sameer Singh. Novelty detection: A review - Part 1: Statistical approaches. *Signal Processing* **83** (12), 2481–2497 (2003). ISSN 01651684. doi: 10.1016/j.sigpro.2003.07.018. Cited on page/s 111.

[66] M Sikora and P Mazik. Towards the better assessment of a seismic hazard—the Hestia and Hestia map systems. *Mechanization and Automation of Mining* **3** (457), 5–12 (2009). Cited on page/s 112.

[67] Jozef Kabiesz, Beata Sikora, Marek Sikora, and Lukasz Wrobel. Application of rule-based models for seismic hazard prediction in coal mines. *Acta Montanistica Slovaca* **18** (4), 262–277 (2013). Cited on page/s 112, 113, 114, 118, 121.

[68] Saket Sathe and Charu Aggarwal. LODES: Local density meets spectral outlier detection. *16th SIAM International Conference on Data Mining 2016, SDM 2016* pages 171–179 (2016). doi: 10.1137/1.9781611974348.20. URL https://epubs.siam.org/terms-privacy. Cited on page/s 112.

[69] M. Bukowska. The probability of rockburst occurrence in the Upper Silesian Coal Basin area dependent on natural mining conditions. *Journal of Mining Science 2006 42:6* **42**

(6), 570–577 (11 2006). ISSN 1573-8736. doi: 10.1007/S10913-006-0101-0. URL https://link.springer.com/article/10.1007/s10913-006-0101-0. Cited on page/s 113.

[70] Zhen lei Li, Xue qiu He, Lin ming Dou, and Gui feng Wang. Rockburst occurrences and microseismicity in a longwall panel experiencing frequent rockbursts. *Geosciences Journal 2018 22:4* **22** (4), 623–639 (7 2018). ISSN 1598-7477. doi: 10.1007/S12303-017-0076-7. URL https://link.springer.com/article/10.1007/s12303-017-0076-7. Cited on page/s 113.

[71] Joana Frontera-Pons, Miguel Angel Veganzones, Frederic Pascal, and Jean Philippe Ovarlez. Hyperspectral Anomaly Detectors Using Robust Estimators. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **9** (2), 720–731 (2 2016). ISSN 21511535. doi: 10.1109/JSTARS.2015.2453014. Cited on page/s 114.

[72] The SciPy community. scipy.stats.shapiro — SciPy v1.9.1 Manual (2022). URL https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html. Cited on page/s 114.

[73] Dina Mohamed, Ayman El-Kilany, and Hoda M.O. Mokhtar. A Hybrid Model for Documents Representation. *International Journal of Advanced Computer Science and Applications* **12** (3), 317–324 (2021). ISSN 21565570. doi: 10.14569/IJACSA.2021.0120339. Cited on page/s 116.

[74] Dario Amodei, Chris Olah, Google Brain, Jacob Steinhardt, Paul Christiano, John Schulman, Openai Dan, and Mané Google Brain. Concrete Problems in AI Safety. *Computer Science ArXiv* (6 2016). URL https://arxiv.org/abs/1606.06565v2. Cited on page/s 123.

[75] ISO/IEC. ISO/IEC TR5469:202x(E) Artificial Intelligence - Functional safety and AI systems. Technical report International Electrotechnical Comission (2022). URL https://www.iso.org/standard/81283.html. Cited on page/s 124.

Chapter 6

Subsea Pipeline Visual Inspection of Anomalies

This chapter is based on the article:

- *Image-based and risk-informed detection of Subsea Pipeline damage* by Rialda Spahic, Kameshwar Poola, Vidar Hepso, and Mary Ann Lundteigen. Discover Artificial Intelligence, Springer Nature. June, 2023.
  DOI: 10.1007/s44163-023-00069-1

All authors contributed to the research conception. Rialda Spahic performed material preparation, literature and data analysis, and manuscript writing. Kameshwar Poolla and Mary Ann Lundteigen performed writing reviews and supervision of all prior drafts of the manuscript. Vidar Hepsø contributed to the concept visualisation of the research.

This research explores and addresses the prevalent challenges of image-based subsea pipeline hazard detection, such as the lack of training data, insufficient evidence of hazards in data resulting in heavy data imbalance, complex subsea environment reducing image quality, and the lack of explainability of methods for image analysis. The main contributions of this research are the analysis of external hazards on raw subsea pipeline images taken by subsea drones and provided by the oil and gas industry, generation of synthetic data through seamless blending of existing hazards and expansion of training data with synthetically generated images. Additionally, to increase the explainability of convolutional neural network for image analysis, we apply localized anomaly detection that highlights the most discriminatory regions of images. Finally, we formalize the success and potential of our approach in a methodology that expands on traditional data analysis lifecycle.

6.1 Image-Based and Risk-Informed Subsea Pipeline Hazard Detection

Abstract

As one of the most important assets in the transportation of oil and gas products, subsea pipelines are susceptible to various environmental hazards, such as mechanical damage and corrosion, that can compromise their structural integrity and cause catastrophic environmental and financial damage. Autonomous underwater systems (AUS) are expected to assist offshore operations personnel and contribute to subsea pipeline inspection, maintenance, and damage detection tasks. Despite the promise of increased safety, AUS technology needs to mature, especially for image-based inspections with computer vision methods that analyze incoming images and detect potential pipeline damage through anomaly detection. Recent research addresses some of the most significant computer vision challenges for subsea environments, including visibility, color, and shape reconstruction. However, despite the high quality of subsea images, the lack of training data for reliable image analysis and the difficulty of incorporating risk-based knowledge into existing approaches continue to be significant obstacles. In this section, we analyze industry-provided images of subsea pipelines and propose a methodology to address the challenges faced by popular computer vision methods. We focus on the difficulty posed by a lack of training data and the opportunities of creating synthetic data using risk analysis insights. We gather information on subsea pipeline anomalies, evaluate the general computer vision approaches, and generate synthetic data to compensate for the challenges that result from lacking training data, and evidence of pipeline damage in data, thereby increasing the likelihood of a more reliable AUS subsea pipeline inspection for damage detection.

6.1.1 *Introduction*

Monitoring and inspection are essential for operational subsea oil and gas pipelines. However, subsea oil and gas operations are complex, with a range of structures and systems, in complex and harsh subsea environment. As a critical asset for transporting oil and gas products over vast distances, subsea pipelines are exposed to a variety of environmental hazards. Hazard is defined as the source of harm [1]. Exposure to environmental hazards can damage the pipelines and cause severe personnel, environmental, and financial damage [2]. Therefore, proper inspection and maintenance of subsea pipelines are essential tasks for their safe and reliable functioning and operations. In case of an unexpected event, continuous monitoring (i.e., pressure drop monitoring for leak detection) notifies the pipeline shutdown system with the supervisory role of an operator [3]. Despite the worldwide safety record of subsea pipelines, comprehending and responding appropriately to complex situations as well as anticipating their consequences are crucial for the safety of offshore operations [4]. Since sending human operators offshore can be dangerous and expensive, autonomous underwater systems (AUS) are intended to assist human operators in inspecting offshore structures, especially long and vast subsea pipelines. With the development of subsea docking stations that allow AUS to reside on the seabed for months, trained operators have the flexibility and opportunity to use AUS to inspect pipelines when the situation calls for it [5].

Autonomy, as described by [6], is the capacity to act and make decisions without external assistance. For AUS, autonomy is typically achieved through artificial intelligence (AI) systems, the computer systems designed to mimic intelligent human behavior [6], by analyzing large amounts of incoming data collected in near-real-time or real-time by sensors and cameras attached to the AUS. For damage detection scenarios, the dominant AI approaches include [7], [8]:

- *Computer vision* methods for analyzing image data,
- *Machine learning* methods that learn from large amounts of data to find patterns, and
- *Anomaly detection* methods that identify and report irregularities, or anomalies, in data patterns.

In addition, *risk assessment and analysis* are common and well-established approaches for identifying *what can go wrong* in operations and offering a list of hazards, as potential sources of harm, the likelihood, sequence of events and consequences of hazards [9].

In recent years, due to the success of remotely operated vehicles (ROVs) that are manually controlled, pipeline inspection research has considered the potentials of AI technologies employed by AUS, such as underwater drones. Therefore, there is an increase in interest for the potential of image-based inspection by computer vision techniques through cameras attached to AUS, such as image classification, object detection, and image segmentation [2,10–16]. However, the existing research for image-based inspections with AUS is particularly oriented toward image color and shape reconstruction and unsupervised methods due to the complexity of underwater conditions, poor visibility, and a significant lack of training data.

Despite the abundance of available research, the remaining obstacles to reliable operations with AUS stem from the underrepresentation of evidence of pipeline damage in data, which contributes to data imbalances that can lead to inaccurate data analysis results and misleading data pattern findings. In addition, there is a significant lack of training data for computer vision and data-driven methods to learn the patterns of potential dangers in order to detect them efficiently and reliably. Unfortunately, a significant number of the detected anomalies represent insignificant data, also known as noise, which further mislead the data analysis conclusions and disrupt the AUS operations decision-making system.

In this section, we focus on analyzing industry-provided subsea pipeline images captured by underwater drones for external damage detection, introducing risk-informed training processes for the anomaly detection methods and evaluating the detected anomalies by isolating potential the anomalies that represent pipeline damage. The focus of this section is on utilizing risk analysis knowledge and semi-supervising computer vision methods for subsea pipeline images for early identification of pipeline damage while separating them from insignificant anomalies (noise and false alarms). The objective is to provide the missing training data while limiting the amount of manual labor to annotate the training images, and therefore to limit the frequency of false alarms generated by autonomous systems and to identify pipeline damage as early as feasible while increasing the scope of anomaly detection capabilities during visual monitoring and inspection. Therefore, the contributions of this section can be summarized as:

- Analysis of external damage on subsea pipelines on raw, industry-provided data.
- Generation of synthetic data through a seamless blending of known anomalies, as defined by risk assessment and analysis methods, for a more reliable computer vision and anomaly detection.
- Review of computer vision challenges, such as monochromatic images and large images that necessitate extensive computational power to analyze.
- Proposal of a methodology to address the lack of training data, imbalanced data, and data quality for image-based subsea pipeline damage detection.

6.1.2 Related Work

The efficiency and reliability of damage detection are vastly enhanced by computer vision. During visual inspection, environmental conditions and appropriate image collection are essential for obtaining high-quality images for image analysis [17]. Computer vision is a type of real-time, in-line detection that requires the analysis of vast quantities of data, often including redundant information, and a high-dimensional feature space. The primary obstacles of general computer vision applications are the computation speed required for real-time operations and the detection intelligence required to differentiate between significant and redundant information [17]. Recent efforts in computer vision have centered on general algorithms for the efficacy and precision of visual inspections [17,18], the necessity of integrating multiple detection technologies [19], and the improvement of real-time performance with less computational power [18,20]. The restrictions of computational power are particularly critical in applications with autonomous systems, such as underwater drones and other mobile vehicles [21]. However, underwater computer vision for subsea structures inspection is facing additional challenges, such as poor visibility, and lack of training data [21,22]. Subsea pipelines are exposed to various environmental factors that can compromise their integrity and contribute to various types of damage. Due to this, substantial research has been conducted on inspecting subsea pipelines to look for damage.

Zhou et al. [7] described the challenge of locating anomalies during subsea exploration. Using a context-enhanced autoregressive network that learns semantic dependence based on conditional probability to identify the anomaly in low-visibility underwater images weighted by both image reconstruction loss and feature similarity loss, they proposed a deep-learning-based anomaly detection framework to identify unknowns in a complex underwater environments for autonomous robots. With sufficient training data with images of marine animals, they successfully demonstrated their method for detecting marine animals as anomalies on a large, imbalanced dataset.

Samnejad et al. [23] explored ways to reduce the time-to-value and overall cost of the subsea pipeline inspection by replacing the laborious task of manually searching for anomalies through unorganized data with an efficient workflow through a set of neural network methods and substantial computational power from cloud-based services. The authors [23] presented a digital solution that integrates the value of visual data collected and aggregated over decades of inspection campaigns with computer vision technologies to detect and classify structure and equipment anomalies autonomously. However, the 20,000 images for the training dataset were annotated manually, requiring intense labor.

Bastian et al. [24] visually inspected and characterized external corrosion in pipelines located on land using a convolutional neural network (CNN). They proposed a CNN for detecting and classifying corrosion on four levels: no corrosion, low, medium, and high corrosion. Despite high accuracy and promising results, the authors [24] encountered several issues that made CNN misclassify corrosion, such as leaves, deposits on the pipeline, and the corrosion-like landscape surrounding the pipelines. They highlighted the need for pipeline images to contain background information, or context, for training. Among the classified corroded pipelines, there were images with background clutter that the CNN model could not distinguish. They emphasized the importance of pipeline images containing context or background information for training purposes a recommended a more localized pipeline inspection approach for more reliable results in differentiating corrosion levels. On land pipelines, however, image-based damage detection encounters fewer challenges with hazy, monochromatic images than on subsea pipelines, making the subsea pipeline inspection task more challenging.

Khan et al.[25] investigated methods for estimating subsea pipeline corrosion based on the color of the corroded pipeline. The authors[25] encouraged incorporating the color correction methods into a robotic system for subsea pipeline corrosion inspection, even in real-time to address the visibility challenges for underwater images. They proposed an algorithm for image restoration and enhancement to reduce blur and improve the color and contrast of underwater images that were tested on experimentally collected and publicly available hazy underwater images.

6.1.3 Problem Description

Underwater computer vision for offshore inspections with autonomous systems is receiving greater attention and the methods need to mature for reliable and safe anomaly detection operations. The primary challenges that pique the interest of both the research community and the industry are:

- **Imbalanced data** is a frequent obstacle in data-driven analysis, such as with most machine learning and anomaly detection techniques. The difficulty is most apparent in anomaly detection applications where anomalies may reflect important information, such as potential pipeline damage. Due to the scarcity of damage evidence in everyday operations, the collected data consists of the vast majority of non-anomalous situations, making it difficult for algorithms to learn patterns about anomalies, report them, and not eliminate them as noise, which is the information that misleads data analysis[26].

- **Training data** is generally sparse in AI-based data-driven approaches. There is a saturation of applications tested with accessible training data; nevertheless, unsupervised algorithms that do not require annotated data are becoming increasingly popular as more data becomes available[27]. Yet, due to the complexity and inexplicability of these techniques, there is a growing interest in discovering automated methods to annotate massive amounts of data and save laborious manual effort. Creating training data is being explored from different perspectives, among others, generating data from simulations, using AI tools for automatic annotation, or through transfer learning where data is learned from one application and tested on a different one.

- **Image quality and visibility** are computer vision applications' most persistent and obvious obstacles. Due to the nature of water as a medium, underwater photos frequently need to be corrected to avoid incorrect lighting and color, causing them to appear predominantly blue or green. In addition, seawater may include a high concentration of plankton and other marine organisms that can obscure photographs. For subsea pipelines, layers of material such as sand and biological deposits referred to as fouling and biofouling, limit the view of the pipeline surface, and inhibit inspection. Hence, many underwater computer vision applications concentrate on reconstructing the image's color, shape, and overall item visibility.

- **Computing power** is another challenge for computer vision applications, because images are often very large and need substantial computing power and processing time. A weakness of prominent neural network algorithms is the necessity to resize or downscale images to improve processing speed, which may result in a substantial loss of information from the resized images. Sliding-window approaches are used in applications where the larger regions of image need analysis without substantial resizing or in case of substantial information loss due to resizing[28].

Autonomous systems powered by computer vision have great potential to detect subsea pipeline damage. However, as offshore operations prioritize the reliability and maturity of emerging technologies, it is necessary to investigate options for generating more training data and reducing the need for Black-box algorithms to be closer to permanently employing autonomous underwater systems for remote operations. It is also important to determine if the image resizing, which is often required to reduce needed computational power during image analysis poses a considerable information loss and reduces the chances of reliable anomaly detection.

6.1.4 Anomalies as Risk Factors

General visual inspection of subsea pipelines, traditionally performed by ROVs is one of the most common inspection methods for determining the pipeline's integrity and identifying areas of increased risk [29]. The operators who manually control the ROV during the pipeline inspection are trained and experienced in detecting anomalies on and around pipelines. The following is a set of the common anomaly criteria for general visual inspection of subsea pipelines established by the best practices in industry [29]:

1. Any evidence of **fluid leakage**.
2. Any external **corrosion** on the exposed metal or outer sheath.
3. Any external **damage, deformation, and bending** on the pipe surface, anodes or other components.
4. Any **debris** blocking the visibility of the pipeline, including litter and other seabed debris, and sediments, is known as fouling. The visibility is also impeded by an abundance of **marine growth**, known as marine fouling or biofouling. The anomaly is considered if more than 50% of the surface is covered within 10 meters. Additionally, debris considered an anomaly are **objects in the nearby vicinity**, up to 1 meter, of a pipeline that can cause damage or obstruct visibility, such as large boulders.
5. **Ineffective pipeline support**, including ineffective seabed support.

Accordingly, Table 6.1 shows a summary of anomalies as risk factors that can contribute to pipeline failure. Table 6.1 illustrates each risk factor's potential damage analyzed, from extensive to minor damage, and compared to its expected occurrence probability, from most probable to least probable occurrence of damage [30].

Table 6.1 shows the general representation of anomalies and the expectation of their occurrence probability, However, the exact probability and anomalies that are identified as damage are typically calculated within a specific operation context. It is crucial for the UAS that detects anomalies to have information or knowledge of the major risk contributing factors associated with the subsea pipelines to adjust expectations and reporting in regions where the likelihood of the most extensive damage potential is higher.

6.1.5 Case Study

Data Description

The dataset for this case study consists of an imbalanced set of 164 subsea pipeline images captured with an autonomous underwater drone, provided by domain experts from the oil and gas industry. There are 126 images without

TABLE 6.1. Risk factors contributing to pipeline failure, adapted from[30]

Potential Hazard	Damage Potential			Probability of Occurence		
	Extensive	Moderate	Minor	Most Probable	Expected Occurance	Least Probable
Leakage, explosion						
External Corrosion						
Material Deficiency						
Debris						
Marine Fouling						
Object Dragging (Anchor, boulder)						
Erosion, soil transport and bottom phenomena						

anomalies and 38 images with anomalies or mechanical damage on the surface of the pipeline. We used 35 additional images without anomalies to generate synthetic mechanical damage images. This was done to balance out the anomalous and non-anomalous images and test if the synthetic data is sufficiently realistic to improve the network learning process. The images are in high resolution and do not require shape or color recovery. However, the nature of the mechanical damage makes it difficult to distinguish the damage from marine growth on the pipeline surface, as both share irregular patterns and similar colors, posing a challenge to distinguish between small-scale damage and marine growth. The original size of each image was 4096 x 2304 pixels, however, due to computational resources required during CNN training, the images were reduced to 224 x 224 pixels where mechanical damage is still visible on the pipeline.

Image Classification with Neural Networks

One of the elements of data analysis through machine learning is the discovery of discriminant data features. Discovering discriminant data features in images can be particularly challenging and requires complex methods inspired by visual cortex processing in the brain that are capable of learning a substantial number of features and extracting patterns[31]. We will focus on a deep learning method CNN or convolutional neural networks. The CNN model consists of convolutional layers whose primary function is to learn and extract the features required for efficient image comprehension [31]. The objective of the convolutional layer, modeled over neuronal cells, is to extract features such as edges, colors, texture, and gradient orientation. Convolutional layers, see Figure 6.1, are composed of convolutional filters or kernels. The kernels are convolved across the width and height of the input image. CNN intuitively learns filters that are activated upon encountering edges, colors, textures, and other image properties. The pooling layer performs nonlinear downsampling of convolved features and reduces the computational power necessary to process the data by reducing dimensionality [31]. The output of pooling is the subdivision of its input into a collection of rectangle patches. Depending on the pooling method selected, each patch is replaced with a single value [31]. There are two main types of pooling, maximum and global average pooling. *Global average pooling* is the more interpretable of the two types because it enforces correspondence between feature maps and categories through the creation of micro-networks [32]. Global average pooling is a structural regularizer that prevents overfitting, a phenomenon in which the CNN model provides accurate predictions for training data but not test data. *Maximum pooling*, or Max pooling performs linear separation, and provides a maximum network that is more potent and achieves higher performance with less computational power by assuming that instances of latent concept lie within a convex set [32].

Although CNN is considered a less explainable approach in image analysis applications, numerous efforts have been made to enhance its explainability. Particularly for image classification and object detection tasks, *localized anomaly detection* is one of the most effective methods for explaining which local regions of an image have been selected for classification. Typically, local regions are depicted using attention maps, which highlight feature regions deemed (by the trained model) crucial for satisfying the training criteria [33]. An example of an attention mask is a highlighted class region on the image, such as mechanical

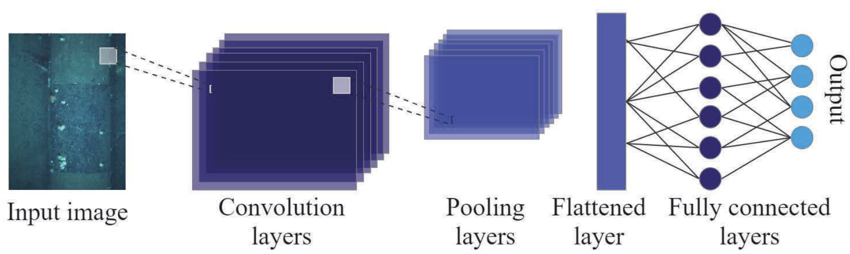

Figure 6.1. Building blocks of CNN, adapted from [31]

damage, which helps to explain why this image has been classified by CNN as mechanical damage or anomaly. Localized anomaly detection is crucial not only for determining if the classification occurred for the correct reason but also for understanding CNN's learning patterns and identifying noise during classification (i.e., analyzing highlighted regions that do not represent the accurate class).

Evaluation Metrics

The evaluation metrics are used to assess the general performance of a trained method, such as a classifier that classifies two or more classes from a given set of data [34]. Various metrics can be evaluated based on the application's requirements. *Accuracy* is one of the most common metrics that counts the total amount of correct classifications on the unseen data. The correct and incorrect classification results can also be illustrated with a confusion matrix, such as in Table 6.2.

TABLE 6.2. Confusion Matrix for Binary Classification, adapted from [34]

Confusion Matrix	**Actual Positive Class**	**Actual Negative Class**
Predicted Positive Class	True Positive (TP)	False Negative (FP)
Predicted Negative Class	False Positive (FP)	True Negative (TN)

The confusion matrix consists of the total numbers of correctly and incorrectly predicted classes, and the numbers of actual classes, to determine true and false positive and negative predictions [34]. *True positive (TP)* and *true negative (NP)* represent the total number of accurately predicted classes, where the predictive method (i.e., classifier) accurately predicted the instances of positive class, and the instances of a negative class. Alternatively, *false positive (FP)* and *false negative (FN)* represent the total number of incorrectly predicted positive, and negative classes. Typical evaluation metrics that are calculated through a confusion matrix are accuracy, error rate, sensitivity, specificity, precision, recall, F-measure, and averaged measures of each of these metrics [34].

$$Accuracy\ (acc) = \frac{tp + tn}{tp + fp + tn + fn} \tag{6.1}$$

$$Error\ Rate\ (err) = \frac{fp + fn}{tp + fp + tn + fn} \tag{6.2}$$

$$Sensitivity\ (sn)\ or\ Recall\ (r) = \frac{tp}{tp + fn} \tag{6.3}$$

$$Specificity\ (sp) = \frac{tn}{tn + fp} \tag{6.4}$$

$$Precision\ (p) = \frac{tp}{tp + fp} \tag{6.5}$$

$$F\text{-}Measure\ (FM) = \frac{2 * p * r}{p + r} \tag{6.6}$$

Accuracy, calculated with Equation 6.1, measures the ratio of correct predictions from the total number of predicted instances[34]. However, accuracy does not represent a reliable evaluation metric when the dataset is imbalanced. Due to low representation of certain classes, many predictive models are unable to learn the patterns of poorly represented data and the inaccurate prediction becomes nearly invisible as compared to the prevalent number of highly represented classes. Accuracy of a predictive model can be high even when all of the underrepresented classes are predicted incorrectly. Depending on the needs of an application, other evaluation metrics are measured to determine the reliability of the model. Error rate measures the ratio of incorrect predictions from a total number of evaluated instances and it is calculated with Equation 6.2. Sensitivity or Recall, calculated with Equation 6.3, measures the proportion of correctly classified positive patterns, whereas Specificity (see Equation 6.4) measures the proportion of correctly classified negative patterns[34]. With Equation 6.5, Precision determines correctly classified positive patterns from the total predicted patterns of a positive class. Finally, F-Measure, calculated with Equation 6.6, measures the harmonic mean between recall and precision[34].

Generating Synthetic Anomalies

Global image editing, such as resizing, shape reconstruction, and color correction, is a typical preprocessing step for image analysis tasks. However, achieving local changes that are restricted to a region of an image, such as object replacement, distortion, blending, cloning, and texture changes, can provide opportunities to manipulate images and create new, seamless, and realistic images. To balance the dataset and provide additional training data for image analysis, we generate synthetic anomalies, mechanical damage on pipeline surface, using the computationally efficient Poisson equation for local seamless blending. With the Poisson equation, we blend an extracted anomaly from anomalous images and seamlessly blend it into another image without anomalies.

$$\mathbf{v} = \nabla g \tag{6.7}$$

$$\triangle f = \triangle g\ over\ \Omega,\ \ with\, f \mid \delta\Omega\ = f^* \mid_{\delta\Omega} \tag{6.8}$$

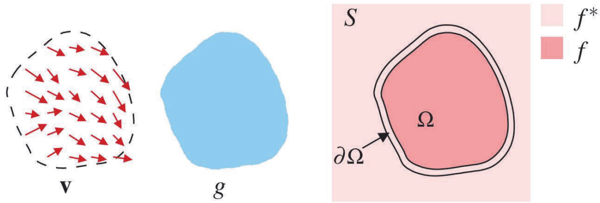

FIGURE 6.2. Guided image interpolation, adapted from [35]

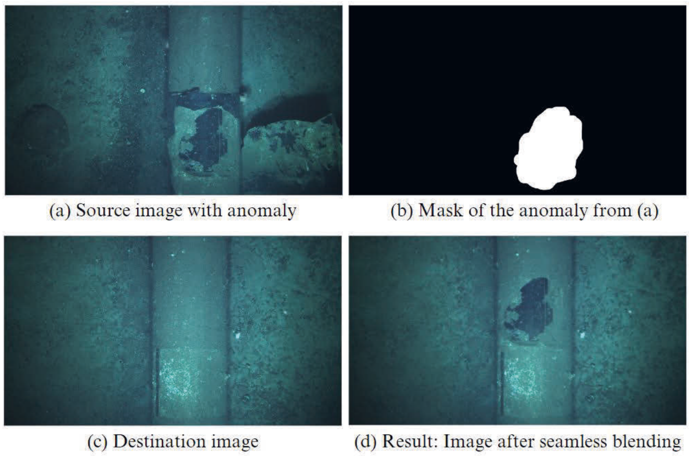

FIGURE 6.3. Seamless blending of mechanical damage on a subsea pipeline: (a) Source image with an anomaly from which a mask (b) and seamlessly interpolated onto destination image (c), resulting in (d)

Perez et al. [35] described and proposed a method for seamless object blending. The seamless blending method is based on a Poisson partial differential equation with Dirichlet boundary conditions that specify the Laplacian of an unknown function over the domain of interest and the unknown function values at the domain's boundary. This allows an object to be seamlessly interpolated onto another object. Figure 6.2, described by Equations 6.7 and 6.8, illustrates a guided interpolation in terms of a function f that interpolates in domain Ω the destination function f^* within a closed subset S with boundary $\delta\Omega$, guided by vector $\mathbf{v}$, as a gradient field of a source function g [35].

A detailed mathematical description of the process is offered in [35]. Seamless cloning and insertion of an object relies on importing the gradients where the most common option for the guidance field $\mathbf{v}$ is a gradient field extracted directly from the image source (i.e, color information from the source image).

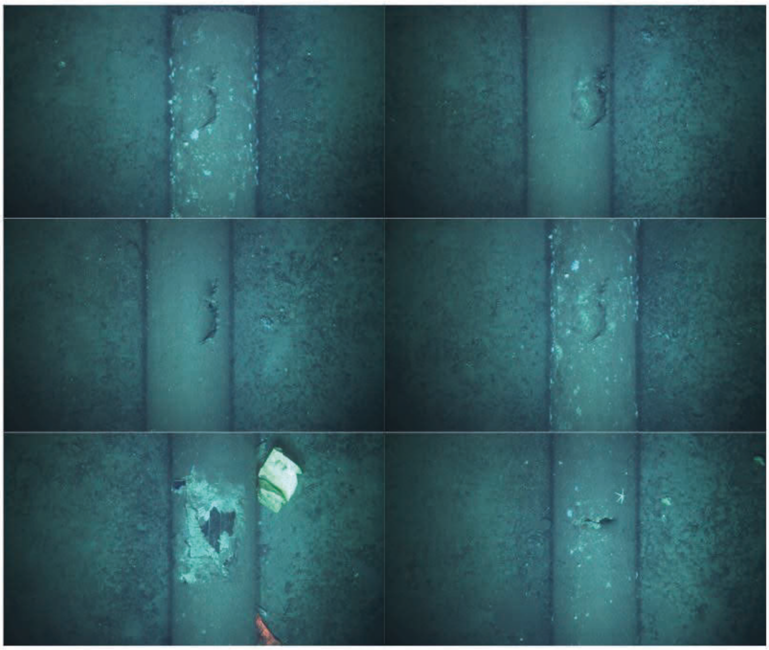

FIGURE 6.4. Other examples of synthetic anomalies

Gradient field performs non-linear mixing or seamless blending, between source and destination images and selects the more dominant features for blending (color, texture, etc.). Equation 6.7 is used to guide the interpolation of this source image, which is denoted by g, after which the final reading for the function f is described by Equation 6.8.

Figure 6.3 shows the process of seamless blending on an image of a subsea pipeline. A source image (Figure 6.3 (a)) has a mechanical damage anomaly on the pipeline that is masked off using an open-source annotation tool for machine learning and image analysis applications. We used Label Studio [36] for this purpose to achieve a precise mask image as shown in Figure 6.3 (b). Annotation or labeling of images with Label Studio [36] was performed by marking a local region on the image. The marked region contains the bounding box and is assigned a label. Exported labels of the labeled regions are then exported as mask images. The source and mask, along with the position of the local region (i.e. position on the pipeline surface) on the destination image (Figure 6.3 (c)) where the blending will occur (other changes such as resizing and reshaping of source/mask object can be made at this point) are provided for seamless blending. Finally, the resulting image is obtained as a synthetic anomaly, as depicted in Figure 6.3 (d). Figure 6.4 shows other images with synthetic anomalies.

Obtaining the mask images, which requires hand-labeling of anomalies with the knowledge of anomalies as risk factors, is the most labor-intensive aspect of creating synthetic anomaly images. However, once the masks have been obtained, the remaining steps are automated to produce batches of synthetic images. The reshaping and placement of the anomalies are randomized so that they do not appear in identical or similar forms. Nonetheless, generated synthetic anomaly images are manually inspected to identify any unrealistic or incorrect results.

Image Classification without Synthethic Training Data

CNN Global Average Pooling and Maximum Pooling on two-dimensional images have been implemented through Keras, a Python-based application programming interface for deep learning that runs on the machine learning platform TensorFlow[37,38]. We analyzed the available data without added

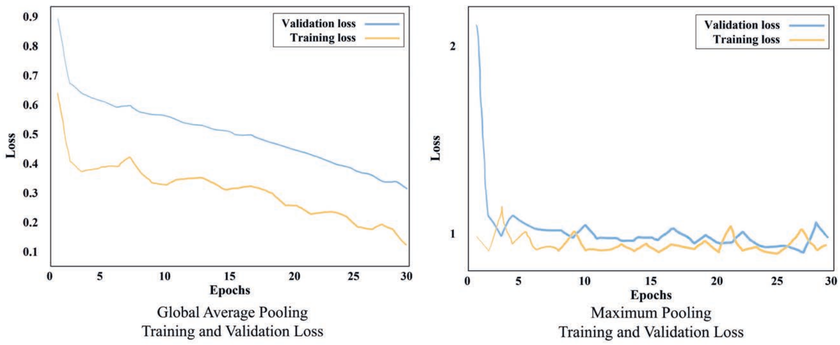

FIGURE 6.5. Training and Validation Loss by Global Average and Max Pooling, without Synthetic Training Data

synthetic anomaly images to test the level at which CNN can classify the normal from anomalous images. The total number of images in the dataset without added synthetic mechanical damage is 164, out of which there are 126 normal images, and 38 of anomalous images with mechanical damage. We split the dataset into 80% for training, and 20% for testing. For the training, we have set the CNN to train over 30 epochs. During each epoch, one cycle of CNN training, all images are processed forward and backward to the CNN. Figure 6.5 shows the training and validation losses. Training loss measures how well the model fits the training data, while validation loss measures how well the model fits new data. The left graph in Figure 6.5 shows the training and validation loss lowering with the epochs, indicating that the model is getting better with learning. However, the graph on the right in Figure 6.5, displaying losses for

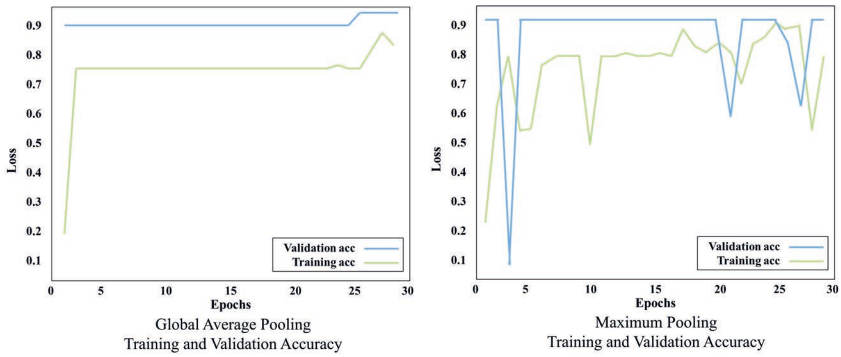

FIGURE 6.6. Training and Validation Accuracy by Global Average and Max Pooling, without Synthetic Training Data

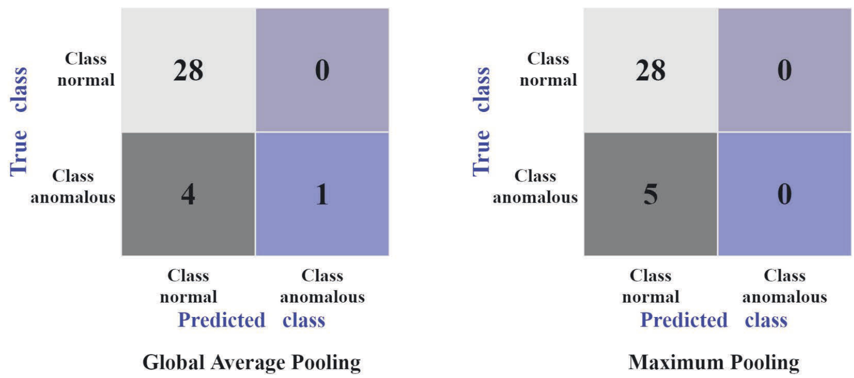

FIGURE 6.7. Confusion Matrix by Global Average and Max Pooling, without Synthetic Training Data

The Maximum Pooling model, shows a mismatched pattern for training and validation, indicating that as the model is struggling to learn the pattern with epochs. These trends are also visible through the accuracy, Figure 6.6, and particularly when observed in the resulting confusion matrix, Figure 6.7. The confusion matrix in Figure 6.7 shows that Global Average Pooling resulted in four incorrectly classified anomalies and only one correctly classified anomaly. Maximum Pooling, however, was not able to learn the trends of anomalous class and did not classify any images as anomalies.

Image Classification with Synthetic Training Data

This section describes the results achieved with added synthetic anomalies through analysis with CNN Global Average Pooling and Maximum Pooling on two-dimensional images[37,38]. Total number of images in the dataset with

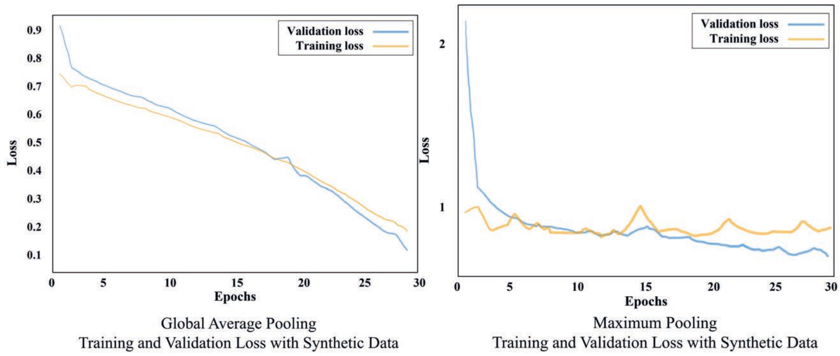

FIGURE 6.8. Training and Validation Loss by Global Average and Max Pooling, with Added Synthetic Training Data

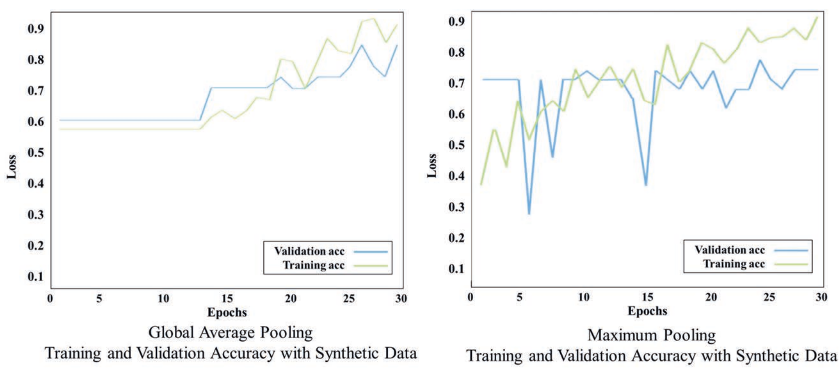

FIGURE 6.9. Training and Validation Accuracy by Global Average and Max Pooling, with Added Synthetic Training Data

added synthetic mechanical damage, is 199, out of which there are 126 normal images and 73 anomalous images with mechanical damage where original and synthetic images are mixed. We split the dataset into 80% for training, and 20% for testing and set the CNN to train over 30 epochs. Figure 6.8 shows the training and validation loss for Global Average Pooling, and Maximum Pooling, with added synthetic data. Unlike Maximum Pooling the loss for Global Average

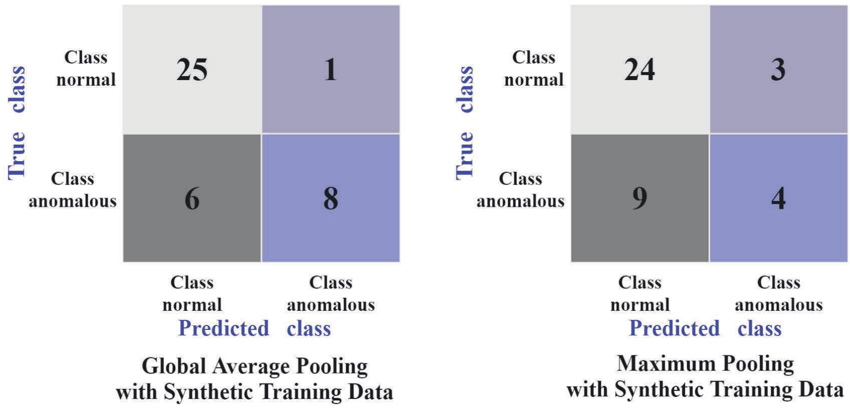

FIGURE 6.10. Confusion Matrix by Global Average and Max Pooling, with Added Synthetic Training Data

Pooling shows a good result, with a promising learning trend with the epochs. This is also observed in Figures Figure 6.9 and Figure 6.10 where the accuracy improves for both, training and validation over the epochs, in both cases Global Average, and Maximum Pooling CNN. In the confusion matrices, Figure 6.10, both approaches show that the network was able to learn patterns of anomalous images. With additional synthetic training data, the CNN model has learned the pattern of anomalies more successfully, which is the most optimistic result. In the case of Global Average Pooling, eight anomalies were classified correctly, and six incorrectly. For Maximum Pooling, four anomalies were classified correctly, and nine incorrectly. Maximum Pooling showed difficulty to classify small-sized anomalies (such as the anomaly in Figure 6.12). In both cases, there is a high accuracy rate for classifying images without anomalies. When synthetic anomalies are added to the training data, the normal and anomaly classes become more balanced, and the CNN model has more anomaly data to learn from.

Localised Anomaly Detection

Localized anomaly detection highlights the anomaly on the evaluated image. The highlighted part of the image illustrates how CNN classified the image into noromal and anomalous regions. Figure 6.11 illustrates the examples of localized mechanical damage on accurately classified subsea pipeline anomalies, we see three different regions highlighted with red boxes:

(a) Localized damage on the pipeline without any noise.

(b) Localized damage on the pipeline surface, and dislocated anode cover on the sides of the pipeline.

(c) Localized damage on the pipeline surface, and noise in the corner of the image.

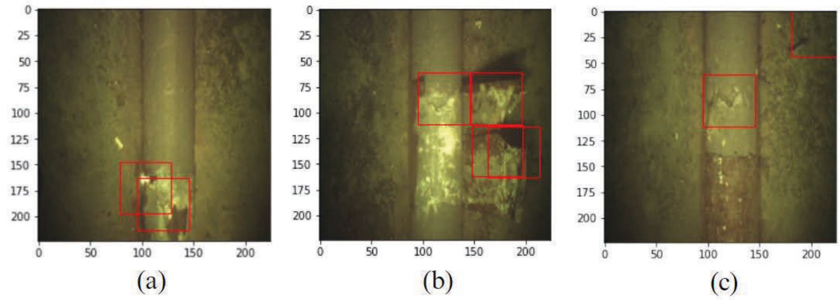

FIGURE 6.11. Localized Mechanical Damage

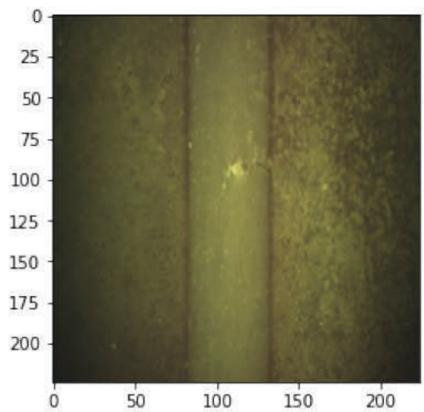

FIGURE 6.12. Inaccurate classification of undersized anomalies. True label: Anomaly; Predicted: Non-anomalous (normal)

Figure 6.11 (a) shows a clean image of highlighted damage as the most desirable outcome. However, two cases Figure 6.11 (b) and (c) have resulted in additional highlighted regions that do not represent mechanical damage. The highlighted regions give insight into possible noise levels that result in inaccurately classified anomalies. Similarly, Figure 6.12 shows one of the inaccurately classified images where an undersized anomaly is not recognized and captured by CNN.

6.1.6 Resulting Methodology

The case study and its objectives are summarized in the proposed resulting methodology presented in Figure 6.13. The resulting methodology proposes

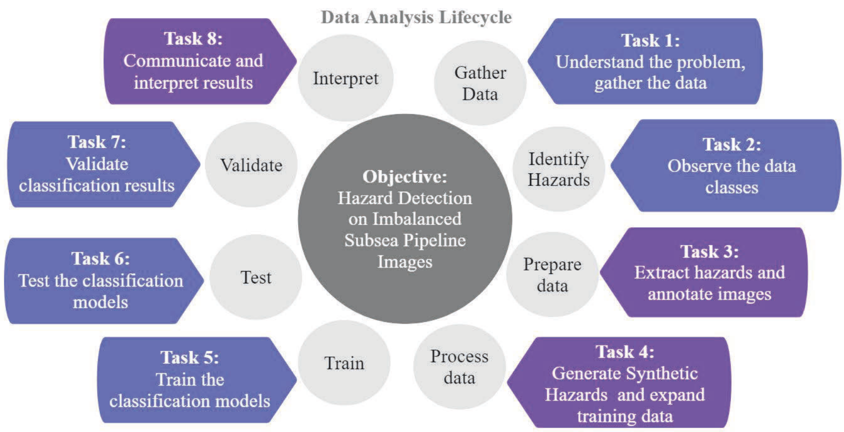

FIGURE 6.13. Resulting Methodology

the eight-task data analysis lifecycle for pipeline damage detection on images of imbalanced subsea pipelines. Tasks 3, 4, and 8 are the most novel contributions to a traditional data analysis lifecycle:

1. The first task is to understand the objective, the problem and gather the data.

2. As the objective is to detect pipeline damage, the second task is to observe the data, identify the anomalies that are pipeline damage within the data, and determine the imbalances between the anomaly and no-anomaly data classes.

3. Once the anomaly and no-anomaly classes have been determined, the third step is to prepare the data by extracting images with pipeline damage from the dataset, masking, and annotating images in preparation for the next step.

4. The fourth task is processing the data which entails generating synthetic damage by seamless blending and image manipulation. This step allows us to expand the training data with additional evidence of pipeline damage.

5. Once the training data is complete, the fifth task consist of training the classification models.

6. After the training is complete, the sixth task is testing the classification models.

7. Utilizing appropriate evaluation metrics, the seventh task is the validation of classification outcomes.

8. Finally, the eighth task is to communicate and interpret the classification results. One of the efforts at interpretation is the application of localized anomaly detection that provides more precise insight into damage detection and possible errors. The last task is particularly important for complex image analysis algorithms that are challenging to explain.

The proposed methodology is based on the case study presented in this section and the primary challenges identified in image analysis and damage detection, such as a lack of training data and the difficulty explaining Black-box algorithms.

6.1.7 Discussion

Despite the small data size, the resulting methodology that includes generating synthetic anomalies to balance the heavily imbalanced data and employing localized anomaly detection has proven to be a promising strategy for addressing the lack of training data, imbalance, and explainability issues that are commonly encountered in image analysis. The subsea images present additional difficulties with visibility, color, and resizing which is especially evident in cases of small and less evident anomalies that are challenging to detect. The resizing of the images has contributed to loss of information resulting in small and less evident anomalies to be less visible. However, resizing of the images is necessary because the computational requirement is a critical challenge. Analysis of large, high quality images requires significant computational resources. Therefore, resizing of images is necessary and during this process, information may be lost. Despite considerable image compression, seamless blending, manipulation, and generation of anomalies allow for the realistic and straightforward expansion of data as required. Moreover, since there is a general absence of high-quality data on subsea pipelines, this method of creating synthetic images may prove useful in industry for generating new data with minimal effort and sharing the data openly and anonymously, while maintaining the realism of the images.

6.1.8 Conclusion and Key Contributions

This section highlights the key contributions and concludes the chapter and the presented article.

As one of the most important assets in the transportation of oil and gas products, subsea pipelines are vulnerable to environmental hazards that can compromise their structural integrity and result in catastrophic environmental damage and financial loss. Autonomous underwater systems (AUS) are expected to assist subsea pipeline inspection and enhance damage detection. However, image-based inspections with computer vision and anomaly detection methods for detecting anomalies, such as pipeline damage, continue to face numerous obstacles that reduce their reliability. These obstacles include visibility, color reconstruction, and shape reconstruction. The lack of training data for image analysis impedes reliable subsea pipeline inspection. In this section, we analyzed images of subsea pipelines provided by the industry and generated a set of synthetic images using seamless blending techniques. We compared the outcomes of convolutional neural networks trained on data with and without synthetic anomalies. In addition, localized anomaly detection during CNN training and validation increases explainability by highlighting regions of classification impact. Finally, we demonstrated the potential of our approach of augmenting the data with synthetic anomalies and presented the tasks in a new methodology that expands the traditional data analysis lifecycle. The proposed methodology shows a potential in training AUS for more reliable damage detection, and assisting pipeline inspection tasks.

6.2 References

[1] ISO/IEC. Safety aspects - Guidelines for their inclusion in standards ISO/IEC Guide 51:2014(E). Technical report International Organization for Standardization and the International Electrotechnical Commissio (2014). Cited on page/s 134.

[2] Michael Ho, Sami El-Borgi, Devendra Patil, and Gangbing Song. Inspection and monitoring systems subsea pipelines: A review paper (3 2020). ISSN 17413168. Cited on page/s 134, 135.

[3] Sirous Yasseri. Selection of Leak Detection Systems by Aggregation of Experts' Judgment. In *International Conference on Ocean, Offshore and Arctic Engineering* San Francisco, California, USA (10 2014). ASME The American Society of Mechanical Engineers. Cited on page/s 135.

[4] Sirous F. Yasseri and Hamid Bahai. Safety in Marine Operations. *International Journal of coastal and offshore engineering* **3** (3), 29–40 (12 2018). doi: 10.29252/IJCOE.2.3.29. Cited on page/s 135.

[5] Daniel Abicht, Jan Christian Torvestad, Pål Atle Solheimsnes, and Karl Atle Stenevik. Underwater intervention drone subsea control system. *Proceedings of the Annual Offshore Technology Conference* **2020-May** (May), 4–7 (2020). ISSN 01603663. doi: 10.4043/30701-ms. Cited on page/s 135.

[6] Oxford University Press. Oxford Learner's Dictionaries (2021). URL https://www.oxfordlearnersdictionaries.com/. Cited on page/s 135.

[7] Yang Zhou, Baihua Li, Jiangtao Wang, Emanuele Rocco, and Qinggang Meng. Discovering unknowns: Context-enhanced anomaly detection for curiosity-driven autonomous underwater exploration. *Pattern Recognition* **131** (11 2022). ISSN 00313203. doi: 10.1016/j.patcog.2022.108860. Cited on page/s 135, 137.

[8] Rialda Spahic, Hepsø, Vidar, and Mary Ann Lundteigen. Using Risk Analysis for Anomaly Detection for Enhanced Reliability of Unmanned Autonomous Systems. In Maria Chiara Leva, Edoardp Patelli, Luca Podofillini, and Simon Wilson, editors, *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)* pages 273–280 Singapore (2022). Research Publishing, Singapore. doi: 10.3850/978-981-18-5183-4{_}R08-03-390-cd. URL https://rpsonline.com.sg/rps2prod/esrel22-epro/html/toc.html. Cited on page/s 135.

[9] Marvin Rausand. Risk Assessment Theory, Methods, and Applications. John Wiley and Sons Inc Hoboken, New Jersey (2011). ISBN 9780470637647. doi: 10.10029781118281116. Cited on page/s 135.

[10] Hongwei Zhu, Weikang Xie, Junjie Li, Jihao Shi, Mingfu Fu, Xiaoyuan Qian, He Zhang, Kaikai Wang, and Guoming Chen. Advanced Computer Vision-Based Subsea Gas Leaks Monitoring: A Comparison of Two Approaches. *Sensors* **23** (5), 2566 (2 2023). ISSN 1424-8220. doi: 10.3390/s23052566. URL https://www.mdpi.com/1424-8220/23/5/2566. Cited on page/s 135.

[11] Salma P. González-Sabbagh and Antonio Robles-Kelly. A Survey on Underwater Computer Vision. *ACM Computing Surveys* (1 2023). ISSN 0360-0300. doi: 10.1145/3578516. Cited on page/s 135.

[12] Alexander G. Rumson. The application of fully unmanned robotic systems for inspection of subsea pipelines. *Ocean Engineering* **235** (9 2021). ISSN 00298018. doi: 10.1016/j.oceaneng.2021.109214. Cited on page/s 135.

[13] Jinjiang Wang, Peilun Fu, and Robert X. Gao. Machine vision intelligence for product defect inspection based on deep learning and Hough transform. *Journal of Manufacturing Systems* **51**, 52–60 (4 2019). ISSN 02786125. doi: 10.1016/j.jmsy.2019.03.002. Cited on page/s 135.

[14] Marco Jacobi and Divas Karimanzira. Underwater pipeline and cable inspection using autonomous underwater vehicles. In *OCEANS 2013 MTS/IEEE Bergen: The Challenges of the Northern Dimension* (2013). ISBN 9781479900015. doi: 10.1109/OCEANS-Bergen.2013.6608089. Cited on page/s 135.

[15] W. T. Nash, C. J. Powell, T. Drummond, and N. Birbilis. Automated corrosion detection using crowdsourced training for deep learning. *Corrosion* **76** (2), 135–141 (2 2020). ISSN 00109312. doi: 10.5006/3397. Cited on page/s 135.

[16] Nicholas Carlevaris-Bianco, Anush Mohan, and Ryan M. Eustice. Initial results in underwater single image dehazing. In *MTS/IEEE Seattle, OCEANS 2010* (2010). ISBN 9781424443321. doi: 10.1109/OCEANS.2010.5664428. Cited on page/s 135.

[17] Zhonghe Ren, Fengzhou Fang, Ning Yan, and You Wu. State of the Art in Defect Detection Based on Machine Vision. *International Journal of Precision Engineering and Manufacturing - Green Technology* **9** (2), 661–691 (3 2022). ISSN 21980810. doi: 10.1007/s40684-021-00343-6. Cited on page/s 136, 137.

[18] Sergei Alyamkin, *et al.* Low-Power Computer Vision: Status, Challenges, and Opportunities. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* **9** (2), 411–421 (6 2019). ISSN 21563365. doi: 10.1109/JETCAS.2019.2911899. Cited on page/s 137.

[19] Rialda Spahic, Mary Ann Lundteigen, and Vidar Hepso. Context-based and image-based subsea pipeline degradation monitoring. *Discover Artificial Intelligence* **3** (1), 17 (5 2023). ISSN 2731-0809. doi: 10.1007/s44163-023-00063-7. URL https://link.springer.com/10.1007/s44163-023-00063-7. Cited on page/s 137.

[20] Dipti Mishra, Satish Kumar Singh, and Rajat Kumar Singh. Deep Architectures for Image Compression: A Critical Review. *Signal Processing* **191** (2 2022). ISSN 01651684. doi: 10.1016/j.sigpro.2021.108346. Cited on page/s 137.

[21] Fangrui Yin. Inspection Robot for Submarine Pipeline Based on Machine Vision. In *Journal of Physics Conference Series* volume 1952. IOP Publishing Ltd (6 2021). doi: 10.1088/1742-6596/1952/2/022034. Cited on page/s 137.

[22] S. Q. Syamsul Amri, A. S. Abdul Ghani, and M. A.S. Kamarul Baharin. Implementation of Underwater Image Enhancement for Corrosion Pipeline Inspection (UIECPI). In *2023 19th IEEE International Colloquium on Signal Processing and Its Applications, CSPA 2023 - Conference Proceedings* pages 195–200. Institute of Electrical and Electronics Engineers Inc. (2023). ISBN 9781665476928. doi: 10.1109/CSPA57446.2023.10087382. Cited on page/s 137.

[23] Mahshad Samnejad, Mahmoud Aboelatta, Cao Vu, and Dung Wood. Asset Inspection Powered by Computer Vision: The Use of Deep Neural Networks for Automating the Detection and Classification of Pipeline External Damage. Technical report (2021). Cited on page/s 137.

[24] Blossom Treesa Bastian, Jaspreeth N, S. Kumar Ranjith, and C. V. Jiji. Visual inspection and characterization of external corrosion in pipelines using deep neural network. *NDT and E International* **107** (10 2019). ISSN 09638695. doi: 10.1016/j.ndteint.2019.102134. Cited on page/s 137.

[25] Amjad Khan, Syed Saad Azhar Ali, Atif Anwer, Syed Hasan Adil, and Fabrice Meriaudeau. Subsea pipeline corrosion estimation by restoring and enhancing degraded underwater images. *IEEE Access* **6**, 40585–40601 (7 2018). ISSN 21693536. doi: 10.1109/ACCESS.2018.2855725. Cited on page/s 138.

[26] George Vachtsevanos, Benjamin Lee, Sehwan Oh, and Michael Balchanos. Resilient Design and Operation of Cyber Physical Systems with Emphasis on Unmanned Autonomous Systems. *Journal of Intelligent and Robotic Systems: Theory and Applications* **91** (1), 59–83 (7 2018). ISSN 15730409. doi: 10.1007/s10846-018-0881-x. URL https://doi.org/10.1007/s10846-018-0881-x. Cited on page/s 138.

[27] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. A Survey on Contrastive Self-Supervised Learning. *Technologies* **9** (1), 2 (12 2020). ISSN 22277080. doi: 10.3390/technologies9010002. Cited on page/s 138.

[28] Yunfeng Diao, Wenming Cheng, Run Du, Yaqing Wang, and Jun Zhang. Vision-based detection of container lock holes using a modified local sliding window method. *Eurasip Journal on Image and Video Processing* **2019** (1) (12 2019). ISSN 16875281. doi: 10.1186/s13640-019-0472-1. Cited on page/s 139.

[29] Damir Tadjiev. Anomaly criteria for general visual inspection of subsea flexible pipes. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE* **4**, 1–9 (2020). doi: 10.1115/OMAE2020-19044. Cited on page/s 140.

[30] William J. Funge. ASCE Pipeline Division Specialty . In *Proceedings of the ASCE Pipeline Division Specialty Conference* (1979). Cited on page/s 140, 141.

[31] D. R. Sarvamangala and Raghavendra V. Kulkarni. Convolutional neural networks in medical image understanding: a survey. *Evolutionary Intelligence* **15** (1) (3 2022). ISSN 18645917. doi: 10.1007/s12065-020-00540-3. Cited on page/s 141, 142.

[32] Min Lin, Qiang Chen, and Shuicheng Yan. Network In Network. In *International Conference on Learning Representations* (12 2013). URL http://arxiv.org/abs/1312.4400. Cited on page/s 142.

[33] Wenqian Liu, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J Radke, and Octavia Camps. Towards Visually Explaining Variational Autoencoders. *arXiv* (2019). doi: https://doi.org/10.48550/arXiv.1911.07389. Cited on page/s 142.

[34] Hossin M and Sulaiman M.N. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process* **5** (2), 01–11 (3 2015). ISSN 2231007X. doi: 10.5121/ijdkp.2015.5201. Cited on page/s 143, 144.

[35] Patrick Perez, Michel Gangnet, and Andrew Blake. Poisson Image Editing. Technical report Microsoft Research UK (2003). Cited on page/s 145.

[36] Inc Heartex. Label Studio: Open Source Data Labeling Platform (2023). URL https://labelstud.io. Cited on page/s 146.

[37] Keras & TensorFlow 2. GlobalAveragePooling2D layer (2023). URL https://keras.io/api/layers/pooling_layers/global_average_pooling2d/. Cited on page/s 147, 149.

[38] Keras & TensorFlow 2. MaxPooling2D layer Keras (2023). Cited on page/s 147, 149.

Chapter 7

New Models for Subsea Pipeline Inspection with UAS

This chapter is based on the following two articles:

- Spahic, Rialda; Hepsø, Vidar; Lundteigen, Mary Ann, *Enhancing Autonomous Systems' Awareness: Conceptual Categorization of Anomalies by Temporal Change During Real-Time Operations*, The Eighteenth International Conference on Autonomic and Autonomous Systems (2022). ISSN: 2308-3913; ISBN: 978-1-61208-966-9

- Spahic, R., Lundteigen, M.A., Hepsø, V. *Context-based and image-based subsea pipeline degradation monitoring.* Springer Nature Discover Artificial Intelligence 3, 17 (2023). DOI: 10.1007/s44163-023-00063-7

All authors contributed to the research conception. Rialda Spahic performed material preparation, literature analysis, and manuscript writing. Mary Ann Lundteigen performed writing reviews and supervision of all prior drafts of the manuscript. Vidar Hepsø contributed to the collection of literature and concept visualisation of the research.

The two articles presented in this chapter discuss suggestions for future UAS inspection of subsea pipelines. The first article proposes a novel categorization of anomalies according to their temporal change. Even though domain-specific anomaly categorization is prevalent in pipeline inspection applications, categorizing changing anomalies that can be tracked with time-series data remains a topic with less research. Similarly, the second article focuses on various factors that contribute to subsea pipeline surface damages, such as the impact of excessive pipeline growth, types of soil surrounding the pipeline, and other geographical properties that contribute to understanding and detecting potential pollution from surface material damage, as well as the opportunities for enhancing pipeline inspection with UAS.

7.1 Enhancing Autonomous Systems' Awareness: Conceptual Categorization of Anomalies by Temporal Change During Real-Time Operations

Abstract

The Unmanned Autonomous Systems (UAS) are anticipated to have a permanent role in offshore operations, enhancing personnel, environmental, and asset safety. These systems can alert onshore operators of hazardous occurrences in the environment, in the form of anomalies in data, during real-time inspections, enabling early prevention of hazardous events. Time series data, collected by sensors that detect environmental phenomena, enables the observation of anomalous data as dynamic instances of the dataset. Recent research characterizes anomalies in terms of their patterns of occurrence in data. However, there is insufficient research on anomalous temporal change patterns. In this section, we examine anomalies in relation to one another and propose a conceptual categorization system for anomalies based on their temporal changes. We demonstrate the categorization through a case study of potentially hazardous occurrences observed by UAS during underwater pipeline inspection. Analyzing anomalies based on their behavior can provide further information about current environmental changes and enable the early discovery of unwanted events, simultaneously minimizing false alarms that overwhelm the systems with low-significance information in real-time.

7.1.1 Introduction

Sensors integrated into Unmanned Autonomous Systems (UAS), such as underwater autonomous vehicles, are reshaping our perception of the world by detecting environmental phenomena and responding to them through inputs such as graphics, motion, pressure, and heat. Underwater UAS, particularly in the offshore industry, are intended to replace operators in remote and potentially dangerous locations by residing on the seabed, collecting the data, and continuously monitoring and inspecting assets and the environment. In crucial situations, real-time data collection and analysis of the environment or assets can provide critical information, signaling us of potentially harmful deviations within the data, known as anomalies. Failure to capture anomalies effectively can have a devastating effect on the environment and result in severe financial loss.

Despite their ample presence in research and industry, anomaly detection methods have not yet matured as they are frequently too specialized or complex to evaluate [1]. Detecting anomalies, particularly for time-series data, is a challenging task that needs real-time processing while learning from analyzed data and making predictions [2]. Most anomaly detection methods are based on statistical samples of some data regions collected over time [3]. When the input data for these data regions changes, it becomes challenging to select the most appropriate strategy for detecting anomalies [3]. More compellingly, it becomes challenging to detect anomalies and capture their changing nature in real-time. The anomalous change detection method searches for unusual discrepancies between measurements taken at the same site at various periods [4]. These discrepancies may be due to harmless changes in atmosphere or sensor equipment. However, they may also be pervasive and potentially indicative of something hazardous evolving at the monitored site, i.e., a deteriorating material of a pipeline surface at the offshore oil and gas platform. Unfortunately, anomaly detection methods can have two significant drawbacks: they can ignore anomalies for the sake of efficiency as tolerable collateral damage [5], or they can overload the system with low-significance data, referred to as false alarms or noise [6]. The ideal outcome of anomaly detection is to alert operators of anomalous occurrences as soon as they are detected while minimizing false alarms [2].

Historically, anomalies have been defined primarily by their pattern of occurrence in data. However, there is insufficient investigation and categorization of anomalies based on how they relate to one another, particularly by the patterns of their temporal change. The time-series data enables the collection and observation of anomalies as dynamic instances of data that alter, evolve, disappear, and reappear. Therefore, this section's contributions is a conceptual categorization of anomalies according to patterns of their temporal change, through an overview of the identification of anomalies during time-series change detection. Analyzing anomalies based on their behavior can provide more information about current environmental changes and allow for the early detection of anomalous, potentially hazardous occurrences in real-time. Consequentially, analyzing anomalies by their behavior can assist in minimizing false alarms by allowing for the more certain elimination of noisy data.

Related Work

Anomaly Characteristics and Categorization Anomalies are instances in a dataset that are unusual in some way and deviate from the dataset's overall or predicted trend [7]. There have been numerous attempts in the literature to categorize anomalies based on their presence in data, the data structures in which they arise, or even application-specific high-level categorization.

Anomalies by Data Structure In a recent review on the nature and categories of anomalies, Foorthius [1] presents an overview of anomaly categories from a data-centric perspective. Because most datasets follow a well-defined, organized format, the author [1] describes the anomalies by examining the data structures that include them: cross-sectional, time-series, time-oriented, sequence, graph, tree, spatial, and spatio-temporal data structures. The author [1] then divides anomalies into univariate, multivariate, and multivariate aggregate anomalies, each of which includes numerical, class, or categorical anomalies and mixed data anomalies.

Anomalies by Occurrence in Data While categorizing anomalies according to the data structure in which they occur simplifies their detection, the literature most often refers to a more general approach to anomaly categorization [8]:

- Global anomaly - one or more independent data points that deviate from the rest of the data. Global anomalies are alternatively referred to as point, and content anomalies [9] [10].
- Collective anomalies - a group of data points that differ from the rest of the data. When observed individually, these points often do not constitute an anomaly. Collective anomalies are alternatively referred to as group or aggregate anomalies.
- Contextual anomalies - anomalies that deviate when an intentionally chosen context is considered, i.e., weather, season, or location. Contextual anomalies are alternatively referred to as conditional anomalies [11].

Anomalies by Data Source According to Erhan et al., [12], sensor systems have become the primary source of data. Therefore, the authors [12] categorize anomalies according to their origins and potential causes (see Table 7.1). Sensor data frequently deviate from predicted behavior. The authors [12] underline the importance of evaluating the performance of anomaly detection systems using physical world data, as opposed to virtual testing with simulators. Since anomalies occur suddenly and are frequently unusual in physical world data, artificially manufacturing them through simulations or data extrapolation can be challenging.

Table 7.1. Anomaly Categorization by Origin, adapted from Erhan et al.[12]

Anomaly origin	Potential cause
Environmental	Unusual events, disasters, weather changes, new objects or compounds
System	System and Hardware limitations, system malfunctions
Communication	Communication and Network loss or delay
Attacks	Malevolent attacks on the physical components, and malevolent interference or attack in network
Spike	Short peak in measured values,
Noise	Increase in the variance in successive data samples
Constant	A constant neutral value reported by sensor
Drift	Off-set in the measurements

Application-Defined and Specific Anomaly Types Ragozin et al.[13] approached forecasting complex time-series within an automated industrial system by *basing anomalies on their distinct dynamic characteristics* to increase the efficiency of information security management within the observed system. The authors[13] developed a method based on structural analysis of multi-component time series and digital signal processing technology for decomposing complex multi-component time series into several essential components for further real-time monitoring of the industrial information system and detecting any component-specific behavior anomaly event or proximity to such event.

Lutz et al.[14] analyzed operational safety-critical anomalies. The authors[14] argue that despite the widely-established benefits of anomaly analysis for operational software, research on anomaly analysis for safety-critical systems has been sparse. Patterns of software anomaly data for operational, safety-critical systems, in particular, are poorly known[14]. The authors[14] describe the findings of two hundred abnormalities on seven spacecraft systems using classification methods. The results of their study demonstrated various classification patterns, including the causal significance of data access and delivery issues, hardware degradation, and unusual incidents. Anomalies frequently revealed hidden software needs critical for the system's robust, accurate operation[14].

Anomalous Change Detection

In a recent review of change detection, Liu et al.[15] classify change detection methods based on their application purpose, data availability, and automation degree. The authors[15] describe anomalous change detection, and time-series change detection as application-specific methods most frequently used in image analysis. By suppressing background and emphasizing alterations, anomalous change detection finds anomalous changes between images. Anomalous change detection is typically focused on detecting minor changes caused by the insertion, deletion, or movement of produced small items and on small stationary objects that exhibit spectrum shifts between images, as with camouflage concealment and deception [15]. The authors [15] argue that the critical point is to examine the image statistics, increase the likelihood of detecting changes induced by human activity, and suppress background in image scene sequences.

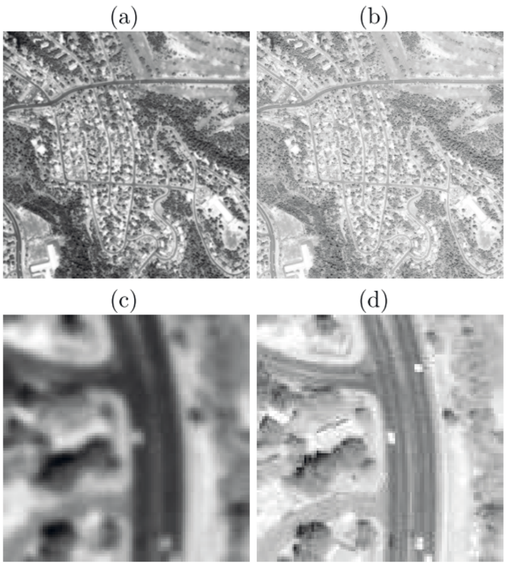

FIGURE 7.1. (a,b) Predictable change in image contrast and brightnes; (c,d) Interesting change with (artificially) added vehicle, adapted from [4]

Theiler et al. [4] employed anomaly detection to identify uncommon changes in images of the same scene captured at various periods and often under varying viewing conditions (see Figure 7.1). The detection of anomalous changes in imaging is of broad general interest and is particularly useful in remote sensing [4]. The authors [4] emphasize that anomalous change is distinct from and more unusual than changes across an entire scene. The authors [4] propose a framework based on a non-flat background distribution stated in terms of data distribution, with anomaly detection treated as a classification problem. The proposed framework identifies anomalous changes capturing meaningful differences between images while avoiding predictable noisy information caused by the camera's focus, contrast, or brightness.

Time-Series Anomaly Detection

Although many organizations collect time-series data, Feremans et al.[16] contend that automatically analyzing them and extracting valuable knowledge, such as a comprehensible model that flags critical anomalies, remains a complex problem, despite decades of effort. After examining various benchmark datasets for time series anomaly detection, the authors[16] discovered that these datasets frequently contain univariate time series with local or global extrema or point anomalies. By contrast, their research concentrated on collective and contextual anomalies, requiring data analysis from multiple sources to detect anomalies successfully. As a result, the authors[16] proposed a method for detecting anomalies in mixed-type time series. The method uses frequent pattern mining methods to create an embedding of mixed-type time series to train a prevalent anomaly detection method, isolation forest. Assuming that the anomalies are infrequent in the data, the isolation forest isolates them by continually splitting the data with low computational costs[17]. Experiments on multiple real-world univariate and multivariate time series and a synthetic mixed-type time series demonstrate that the proposed method outperforms established anomaly detection methods such as MatrixProfile, Pav, Mifpod, and Fpof[16].

Hannon et al.[18] used anomaly detection on streaming data to explain a power-grid system's real-time behavior and provide insight to system operators. The authors examined a real-time anomaly detection followed by a data-driven framework based on the statistical machine learning methods (decision trees and k-nearest neighbors) to enable the remote analysis of individual grid components for monitoring, detecting, and classifying anomalies that generate warnings of possible shortcomings in the system. They[18] concluded that classification of identified anomalies using well-defined probabilistic scores and classification of detected anomalies using interpretable decision trees demonstrates a high level of accuracy, as a result enabling operators to take corrective action to avert cascading blackouts and prevent system failures.

Previous research has established a variety of applications for anomaly detection and a need for a more profound comprehension of anomalies. In a discussion section *Anomalousness: How to measure what you can't define*, Theiler[19] describes anomaly detection as target detection with unknown targets and with the objective to differentiate anomalies (unknown targets with stubbornly undefined attributes) from a background that is generally too cluttered to support an explicit model. Despite the challenges in defining and categorizing anomalies, the outcomes and discussions of previous studies demonstrate a promising direction in application-specific and dynamic-oriented anomaly categorization.

7.1.2 Categorization of Anomalies Based on Their Temporal Changes

After decades of research on anomaly detection, selecting anomalies to investigate and those to disregard as noise continues to be a complex problem, particularly with the pressure of a growing need for autonomous systems. Given the poor camera vision and ambiguous sensor inputs in the subsea environment [20], it is only natural to assume that strange phenomena, such as biological growth or misplaced objects, are frequently misinterpreted. This misinterpretation can further result in the misallocation of resources or the omission of signs indicating a more hazardous occurrence. Using inspiration from prior research on grouping time-series data [21] and integrating time-series and event logs into itemsets [16], we open opportunities to investigate prospects for isolating and analyzing changes in anomalies based on their geospatial context. By combining insights from time-series change detection on dynamic data points [21–23] with application-specific anomalies [14 24], we observe that anomalies can display behavioral patterns such as frequent or reoccurring, disappearing and reappearing, and expanding.

Frequent or Recurring Anomalies Feremans et al. [16] discuss frequent patterns in data, assuming that because anomalous activity infrequently occurs in time series, the frequent patterns represent frequently seen normal behavior. The main advantage of frequent pattern extraction is that the extracted patterns are easily interpretable and aid classifiers and anomaly detection methods in differentiating between normal and anomalous behavior in data. However, it might quickly become problematic if an anomalous event occurs repeatedly or in patterns. Anomalies that reoccur in patterns, hence generating a recurrent pattern in obtained data, present a concern because they can be difficult to spot or even mistaken as part of the normal dataset. Normal data can mask these anomalies, making it particularly difficult to detect when using unsupervised methods.

A practical example, seen on Figure 7.2, is the pipeline with unclear surface material, provided by images collected during a visual inspection of sea bottom infrastructure by an autonomous underwater vehicle. Visually inspecting structures can detect various phenomena, from object detection to material degradation such as corrosion monitoring [25]. However, a less intrusive process, such as biological growth, happens frequently and can readily obscure a more intrusive process, corrosion. Although additional measurements like ultrasonic testing and electromagnetic mapping are used to identify additional information about the corrosion process, the pace of corrosion (spread over time), the exact location, and even plausible causes [25], relying on unsupervised visual inspection of anomalies may not be sufficient.

(a)

(b)

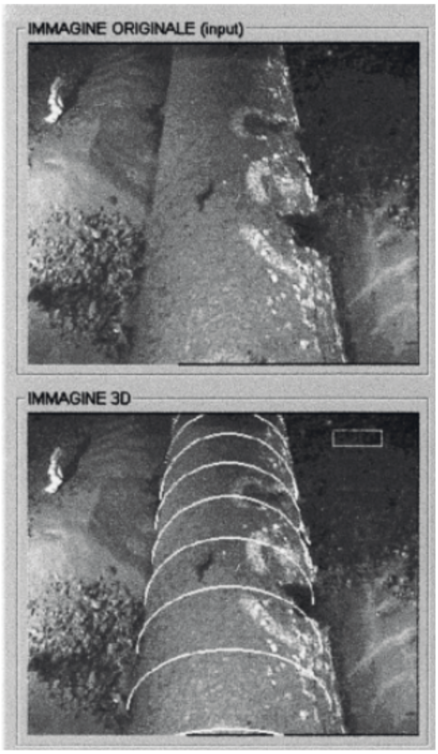

Figure 7.2. (a) Visual inspection of underwater pipeline, images taken by autonomous underwater vehicle, adapted from[20]; (b) 3D scan over the underwater pipeline, adapted from[20]

Disappearing and Reappearing Anomalies Although disappearing anomalies are not usually mentioned in industrial anomaly detection applications, they are a fairly common topic in stock market anomaly detection. During the analysis of the dynamic persistence of anomalies, Marquering et al.[26] highlighted the occurrences of disappearing and reappearing anomalies. Since most seasonal or predictable anomalies are well-known, they should not persist[26]. However, the authors[26] question the persistence of such anomalies as a source of contention. They highlight essential questions on disappearing and reappearing anomalies in data: *Are there still anomalies in recent data? Are they just existent during specific periods, or did they completely vanish? What is the immediate cause of the endurance of the anomaly?* The occurrence of disappearing and reappearing anomalies may be of interest in time-series change detection for various applications.

During a real-time inspection of an underwater pipeline, as depicted in Figure 7.3, recordings of fading unusual events may represent a low-importance environmental phenomenon that does not require comprehensive inspection, thus saving additional resource allocation. However, the persistence of such occurrences may represent something of more profound research interest[26].

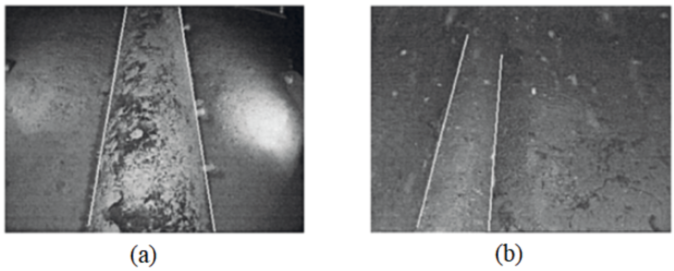

FIGURE 7.3. (a) Visual inspection of underwater pipeline, images taken by autonomous underwater vehicle: Possible material degradation or biological growth?, adapted from[20]; (b) 3D scan over the underwater pipeline, adapted from[20]

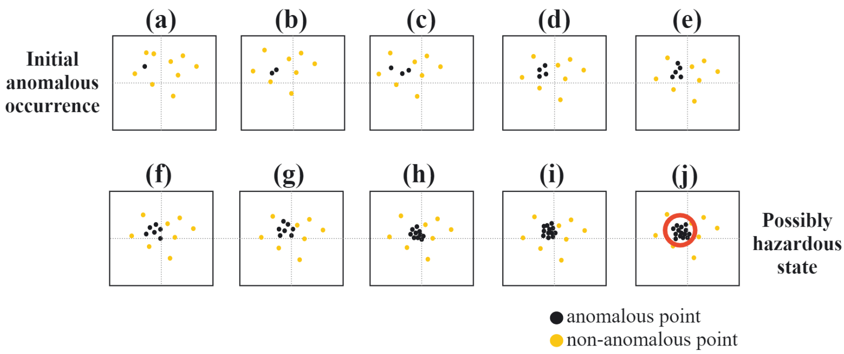

FIGURE 7.4. Anomalies that expand over time

Expanding Anomalies As the environment evolves and changes over time, assuming that anomalous occurrences will exhibit similar changes is natural. Despite anomalies' dynamic and evolving nature being frequently discussed in sensor networks, it is not often discussed in other applications. What appears to be an innocuous anomaly may grow to affect various regions of the inspected structure. The purpose is to identify the onset of the anomaly as fast as feasible while maintaining a low false alarm rate[23]. This detection problem is formulated as a stochastic optimization problem utilizing a delay metric based on the anomaly's worst-case path[23]. In Figure 7.4, we illustrate a point anomaly (Figure 7.4 (a)) expanding into a collective anomaly (Figure 7.4 (b-j)). At an early stage (Figure 7.4 (a)), the detected point anomaly or a smaller collection of anomalies may not yet indicate a high-significance unusual occurrence. However, if unexplored, the anomalous collection may develop into a possibly hazardous state (Figure 7.4 (j)), leaving less time for a reactive response. Detecting anomalies early enables preventative measures. Expanding fractures of the pipeline surface material are a practical example of expanding anomalies during an underwater pipeline inspection.

The proposed conceptual categorization of anomalies according to their temporal changes does not impede their occurrence in data as point, collective, and contextual anomalies. Table 7.2 summarizes the two categories that are intended to complement one another, aiding in our comprehension of unusual events occurring during autonomous operations. Anomalies' behavior is highly dependent on context, not just on their occurrence as a single point or collection of anomalies. The criticality of frequently occurring point and collective anomalies varies by context, as they may be seen as normal and therefore obscure more intrusive processes. This increases the likelihood that the unexposed anomaly may develop into a potentially hazardous event that could have been discovered earlier. Similarly, the context (i.e., seasonal, weather) of disappearing and reappearing anomalies can aid in identifying the cause of their pervasiveness and provide additional reasoning for unanticipated environmental phenomena. Additionally, the point anomaly may expand creating a

Table 7.2. Describing Anomalies by Temporal Change

Anomaly type	Frequent Recurring	Disappearing Reappearing	Expanding
Point	Frequently occurring point anomaly.	Disappearing and reappearing point anomaly may be a sign of pervasive environmental phenomena.	Point anomaly may evolve into a collective anomaly of larger size and impact.
Collective	Frequently occurring collection of anomalies with similar properties (i.e., geospatial context).	Disappearing and reappearing collective anomaly may be a sign of pervasive environmental phenomena.	Collective anomalies may evolve into a more intrusive anomalous occurrence of larger size and impact.
Contextual	Anomalous depending on the context due to a potential risk of being misinterpreted as normal and left unexposed or a frequent anomaly collection obscuring more intrusive processes.	Context (i.e., geospatial, seasonal, weather) aids in determining the anomalousness of the disappearing/reappearing phenomena and finding the causes of their persistence.	Anomalous depending on the context.

collective anomaly of more impactable volume and intrusiveness. Contextual information (e.g., changed material properties due to chemical or temperature variations) can assist in determining the criticality and anomality of observed unanticipated changes. Observing and categorizing anomalies according to their temporal changes adds context to our understanding of how anomalies relate to one another and evolve in a normal and predictable data environment. This knowledge enables the UAS to perceive environmental phenomena and anomalous events in their geospatial and temporal context, improving understanding of the significance and criticality of anomalous occurrences.

7.1.3 Contribution Summary

The research on time-series anomaly detection has been application-oriented and vague. Despite decades of research and categorization approaches, persistent obstacles prevent anomaly detection from maturing and becoming a dependable component of autonomous systems. While an unsupervised and data-driven strategy is common in industry and research, it is insufficient to achieve reliable autonomy. Therefore, this section proposes a fundamentally different perspective of anomalies via a conceptual categorization of anomalies according to their temporal changes. Frequent or recurrent, disappearing and reappearing, and expanding anomalies describe the behavior of anomalies and provide context for their dynamics observed through time-series data analysis. Observing anomalies as they evolve through time enables us to deduce the underlying causes of anomalous occurrences, focusing on more pertinent data from the vast collections of sensor measurements, thus allowing the UAS to react if and when the situation requires it during real-time operations.

7.2 Context-Based and Image-Based Subsea Pipeline Degradation Monitoring

Abstract

This research examines the factors contributing to the exterior material degradation of subsea oil and gas pipelines monitored with autonomous underwater systems (AUS). The AUS have a role of gathering image data that is further analyzed with artificial intelligence data analysis methods. Corrosion and potential ruptures on pipeline surfaces are complex processes involving several competing elements, such as the geographical properties, composition of soil, atmosphere, and marine life, whose effects can result in substantial environmental damage and financial loss. Despite extensive research, corrosion monitoring and prediction remain a persistent challenge in the industry. There is a lack of knowledge map that can enable image analysis using an AUS to recognize ongoing degradation processes and potentially prevent substantial damage. The main contribution of this research is the knowledge map for increased context and risk awareness to improve the reliability of image-based monitoring and inspection by autonomous underwater systems in detecting hazards and early signs of material degradation on subsea pipeline surfaces.

7.2.1 Introduction

Material degradation of a pipeline can result in structural failures that endanger marine life, create environmental hazards, and cause significant financial losses. Understanding the factors contributing to corrosion is essential for understanding the development of corrosion in materials, building anti-corrosion structures, and making risk assessments during monitoring and inspection operations[27]. Several factors contribute to the deterioration of pipeline surface material, and there are multiple factors to consider when identifying the most effective ways to predict deterioration and ultimately prevent substantial damage. The most visible factors are the materials used to construct pipelines and geographical elements such as soil, environment, climate, and marine life. Complex processes affecting material degradation, corrosion, and eventual surface ruptures pose a challenge to industry and are the subject of continuous research.

Autonomous Underwater Systems (AUS), such as autonomous underwater drones and intelligent sensor systems, play an increasingly important role in the monitoring and inspection of remote, potentially chemically contaminated offshore structures, such as pipelines[28]. In the context of autonomous systems, autonomy characterizes self-organizing and self-sufficient systems to achieve a specific task[29]. The industry is increasingly relying on AUS for enhanced safety in remote operations. Missed opportunities in detecting damages at offshore structures, such as ruptures and gas leaks, can lead to catastrophic consequences for employees at connected facilities, the environment and significant financial losses. There is a growth in the usage of artificial intelligence (AI) and machine learning (ML) methods for continuous data and image analysis as the application and interest in AUS for monitoring and inspecting offshore infrastructure grows. However, for reliable data analysis, the AUS requires extensive use of empirical data and causal reasoning[30,31]. There is insufficient labeled data to train ML algorithms for detecting hazardous events. Even more so, the methods lack context in distinguishing significant information from insignificant and hence reliably responding if and when the situation requires it. Commonly, an absence of visibility and contextual knowledge of the operation hinders image-based analysis. Therefore, a knowledge map must be established to determine the most effective and reliable means to plan operations and aid AUS in monitoring and inspecting critical infrastructure.

The main contributions of this section are:

- Analysis of the different factors, geographical properties, soil composition, and marine life in the context of autonomous pipeline monitoring and inspection.
- Mapping of domain knowledge for context-based and risk-informed autonomous monitoring of subsea pipeline material degradation.
- Proposal of a strategy for reconsidering how image-based analysis using AUS is used for safety purposes and overcoming the shortcomings of AUS operations is proposed.

7.2.2 *Motivation and Literature Review*

Pipeline Material Degradation

Pipeline material degradation, such as corrosion, which can lead to material rupture or cracking, has a significant economic and ecological impact, consuming 4% of the gross domestic product of industrialized countries[32]. Due to the significant reliance on pipelines for product transportation, the offshore oil and gas industry is the most impacted. Subsea pipeline degradation is a complex process involving a series of causes and events[33]. Despite extensive research and prevention systems, corrosion remains one of the industry's most significant challenges. Corrosion can cause damage to both the inside and exterior layers of a pipeline and image-based inspection by AUS allows for the detection of exterior corrosion. Hagarova et al.[34] examined the types of corrosion and prevention mechanisms that often involve coatings and cathodic protection. The outer surfaces of pipelines are additionally protected by protective barriers, metal and non-metal coatings, glass fiber, rubber, and epoxide designed to last for the duration of the pipeline's life[33–35]. However, the degradation of the outer surface of the pipeline gains an electrochemical nature. The exterior corrosion highly depends on the soil's chemical composition, the water's salinity in the environment, the existence of currents and atmospheric characteristics, and the presence of microorganisms[34]. Additionally, the surrounding soil influences aeration, affecting water content, oxygen concentration, and other potentially corrosive constituents. While sandy soil acts as a type of protection for pipeline materials, clay-rich soil creates a more corrosive environment. Corrosion damage of gas pipelines is often divided into following categories:

- Sweet corrosion occurs due to CO_2, H_2S, S, H_2O, inorganic salts, chlorides, sand and bacteria in transported products.
- Sour corrosion develops in H_2S environment that becomes corrosive in water-gas environment and can cause cracking in the pipeline wall.
- Microbially influenced corrosion is caused by microorganisms, bacteria, fungi, and other biological growth that produce waste material, such as acids, CO_2, H_2S, which increase the toxicity and promote corrosiveness of the environment.
- Corrosion cracking is induced by mechanical damage that occurs under the component of stress and corrosive environment.

In addition, carcasses of animals and other animal deposit on the pipelines contribute to the increased microbial presence and promote further material degradation.

Geographical Properties and Pipeline Degradation

The physical features of soil significantly contribute to the corrosion propagation of buried metals, which is particularly apparent in pipelines. Wang et al.[27] perfromed the soil corrosivity tests that most commonly analyze oxidation potential, pH factor, water content, and salt saturation. According to the National Association of Corrosion Engineers and the American Society for Testing and Materials, the degradation process of pipeline materials is considerably determined by soil resistivity and corrosivity[27]. Additionally, the authors [27] examined data mining, artificial intelligence, and machine learning approaches to analyze large amounts of sensor-collected soil data and determine soil composition and corrosivity. They observed that this strategy, despite being practical, had significant limitations:

- Soil properties are fixed and cannot be easily changed to fit the conditions on the field. This limitation can be attributed to the need for more context challenges of AI and ML approaches.
- Assessing the soil's corrosivity is a complex task that often involves contradictory factors. During the data analysis with AI and ML methods, the contradictory problems presented by the interaction of multiple factors and the inconsistent effects on soil corrosion need to be sufficiently addressed.
- The lack of specificity in the classification of soil corrosivity contributes to the expansion of uncertainty, necessitating more information from the observations.
- Image data plays a vital role in detecting and analyzing corrosion on subsea pipeline surfaces.

Addressing these limitations would facilitate the prediction of the material degradation levels and improve the monitoring of the rate of material degradation development.

Ohaeri et al. [33] examined climate as another important component of material degradation. They single out the cold environment as a challenge to metallic materials and as one of the primary causes of brittleness, especially in welded areas that lose ductility. The common idea is that materials are less susceptible to corrosion in colder climates, primarily due to the ice covering or permafrost in buried pipelines, which prevents oxygen from accessing surface material [33]. However, high salinity inhibits the freezing process and accelerates corrosion. Similarly, chloride-enriched water and ice create a corrosive environment. Hydrogen is another challenging component contributing to the accumulation of faults and eventual failure in metallic materials. The authors [33] argue that there is not a single factor that contributes to the material degradation of subsea pipelines but rather a series of events that contribute to accelerated material failure, making it vital to observe each event to predict the pipeline lifecycle.

Marine Life at Subsea Anthropogenic Structures

Various attempts have been made to determine the positive and negative effects of pollution and corrosion from offshore structures on the behavioral patterns of opportunistic species[36–41]. The mutual impacts of opportunistic organisms and subsea anthropogenic structures, such as offshore oil platforms, can be summarized in two focus areas:

1. the impact of material degradation and associated pollution from offshore constructions on opportunistic species

2. the impact of opportunistic species on offshore structural materials

Both focus areas provide vital information about the behavioral patterns of opportunistic species in connection to certain pipeline surface materials during degradation (i.e., corrosion, ruptures), and associated pollution. Some species can be the cause of pipeline degradation, while others significantly increase or decrease their appearance when material degradation already occurs[38]. Although the microbial species that cause material degradation, i.e., corrosion, are not visible with image-based inspections[42], many species that appear or disappear under higher saturation of elements in the environment as a result of degradation or pollution are not microorganisms and are visible with imaging equipment. As a result of the elevated components in the environment produced by material degradation of offshore infrastructure, the appearances or disappearances of specific species are frequently classified as either positive indications or negative indicators of pollution[38]. Positive indicators, also known as tolerant opportunistic species, flourish in environments with elevated levels of components produced by material degradation (i.e., increased number of species due to higher saturation of iron in the environment due to corroded material). Negative indicators are species with low tolerance whose disappearance from an area may indicate pollution, or elevated saturation of components produced by material degradation, such as corrosion[38]. Even though positive indicator species thrive in polluted environments, some of them are dependent on the existence of negative indicator species. This dependence leads to a reduction of positive indicators due to the decrease of negative indicators.

Successful observation of opportunistic species around offshore structures is supplemented by information about soil, sediments, seasons, and weather conditions. Su et al.[39] explored the mutual influence of corrosion and microbial communities on buried petroleum pipelines. They argued that soils with distinct microbial populations can have varying effects on the corrosion of buried petroleum pipelines. Their research centered on three distinct types of soil exposed to varying levels of corrosion and petroleum pollution. The authors[39] were able to determine, using electrochemical measurements, that the microbial diversity in soil surrounding corroded pipelines decreased independently of the extent of petroleum contamination. However, electrochemical testing also revealed a more significant concentration of microorganisms that degrade hydrocarbons. Dubiel et al.[42] examined microbial composition in corrosion-surrounded environments and found that such environments considerably modify the microbial composition in the soil and that carbon steel or iron corrosion correlates with sulfate loss in the environment.

Different microbial communities inhabit different soil types, further influencing the observed area's benthic communities. Seasonal context is vital in determining if an increase in larvae in the water contributes to murkiness and invites other species to feed, leading to poor visibility and necessitating varying contextual and sensor inputs for the AUS. Similarly, storms and similar weather conditions can contribute to murkiness in shallower water and sediment deposits. This kind of contextual knowledge can aid in eliminating common issues of AI methods, particularly anomaly detection, pattern recognition and classification, such as biases, an inclination towards efficiency, and a lack of causal and contextual knowledge. While pattern recognition detects patterns or regularities in data, anomaly detection is responsible for detecting data points that do not conform to the data patterns, irregularities, or anomalies[43]. The ability to distinguish between important and irrelevant anomalies is a crucial challenge when using anomaly detection methods to identify potentially hazardous conditions on pipeline surfaces.

AUS and Image-Based Pipeline Monitoring

Image-based monitoring and inspections of structures offshore with mobile cameras that are attached to drones or with stationary cameras attached to structures, can produce extensive collections of image data[44]. Autonomous operations on offshore structures rely on AI for real-time or near-real-time data analysis of extensive image data collections. Detecting material degradation on images of subsea pipelines typically requires computer vision methods, pattern recognition, or anomaly detection. As corrosion and ruptures may be examined through differences in color and texture of the material surface, anomaly detection is a common approach to detect when these changes occur in comparison to the expected appearance of pipeline surface during monitoring operations.

Idris et al.[35] reviewed pipeline inspection using an image-based system for corrosion detection. Through images and videos, a visual examination of pipeline surfaces aids in detecting corrosion by observing changes in texture and color. The following forms of corrosion are classified by appearance and can be detected with image analysis[35,45–50]:

1. Uniform or general corrosion that is evenly distributed across materials.

2. Pitting corrosion, a localised corrosion that leads to small ruptures in metallic materials.

3. Crevice corrosion is one of the most harmful corrosion types that forms inside of ruptures, or spaces and seals.

4. Galvanic corrosion occurs when a metal contacting another conducting, often protective material, results in corrosion, potentially leading to quick deterioration of materials.

5. Erosion-corrosion occurs due to mechanical action, liquids or other particles that can form cavitation.

6. Intergranular corrosion, or stress corrosion, occurs at structural level of the metallic material.

7. Environmentally assisted cracking, including corrosion fatigue, hydrogen damage, and stress-corrosion cracking.

The authors [35] decomposed the problem of incorrect inspection results into underlying causes using a problem tree, as illustrated in Figure Figure 7.5. Each image contains a considerable amount of data, most often noise, making it challenging to select the meaningful data we may be searching for. A substantial amount of sensor-collected data is lost, and retrieving lost data in images is more feasible than other analog signal data. The authors [35] explain the process through a correlation between the lost pixels and their neighbors to retrieve the lost image compression. As Figure Figure 7.5 illustrates, a combination of challenges makes the incorrect image analysis result. The image processing step of image analysis suffers from a lack of information, inexperienced conclusions, wrong data interpretation, challenges derived from artificial intelligence, false data extraction, and undetected defects. Image enhancement layer challenges include poor quality, blur, over-exposure, focus, illumination, environmental constraints, and inappropriate tools and procedures.

Todd et al. [40] performed a review of utilizing remotely operated vehicles (ROV) responsible for collecting the data by observing offshore structures, but not autonomously inspecting the collected images. They found the ROV data collected near offshore anthropogenic structures to be a reliable and readily available information source for researches to further observe and analyze not just offshore structures but also marine life. The authors [40] observed that certain marine animals are taking advantage of the anthropogenic structures due to the growing microhabitat. This result indicated that repeated sightings of specific species can tell us about the microhabitats growing around structures and may signal an ongoing material deterioration or pollution surrounding the structures.

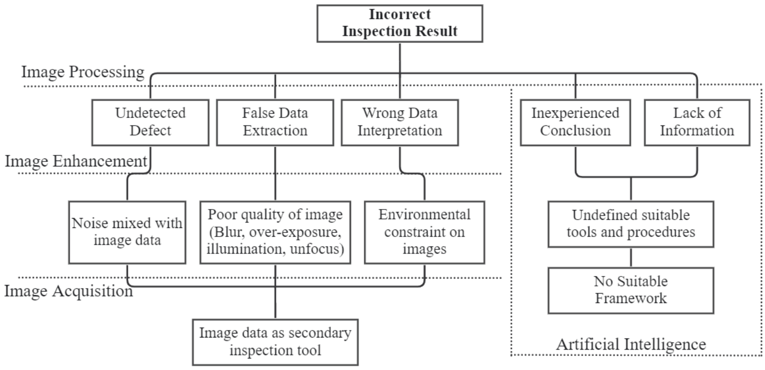

FIGURE 7.5. Image-Based Inspection Problem Analysis[35]

7.2.3 Degradation Probability Under Corrosive Events

Each instance of pipeline deterioration, such as external corrosion, may not be the result of the most common causes of such deterioration. Nevertheless, the likelihood that certain factors contribute to corrosion more than others can be determined by observing multiple risk assessments of corroded pipelines. The chemical composition of the soil and water (i.e., salinity, saturation of CO_2, H_2S), subsea atmospheric characteristics (i.e., currents, harsh environment), and the presence of microorganisms are the most common causes of corrosion[34]. To obtain the level of influence of corrosion-causing events through probability analysis, Yang et al.[51] examined numerous risk assessments, accident reports on corroded pipelines and values assigned by domain experts based on their subject knowledge and experience. Table Table 7.3 describes the probability of corrosion-causing natural factors, excluding human error or faulty sensors. The prior probability of an event is the probability that is assigned before data is considered. Whereas, the posterior probability is obtained with the new event or given the data observations. The prior probability, $P(U, E)$, is used to calculate the posterior probability, $P(U|E)$ described by Equation 7.1, when new data, an observed event or evidence supports the prior hypothesis, where U are all the data variables, and E represents the specific event[51], also listed in Table Table 7.3.

$$P(U \mid E) = \frac{P(E \mid U)}{P(E)} P(U) \tag{7.1}$$

Table Table 7.3 shows prior probability that represents what was initially believed prior to an event, whereas posterior probability is used to revise a prior belief when new information or an event becomes available. Prior and posterior probabilities are also used to estimate the risk of a hazardous occurrence, such as corrosion, by updating the probability of the default state based on previous observations. For machine learning applications, prior and posterior probability are useful in the training phase where posterior probability is updated after each training round. Knowing prior and posterior can enhance contextual understanding of the ongoing operation and facilitate more reliable and confident decision-making for AUS when new events or observations are encountered.

7.2.4 Context-Based AUS Operations

Monitoring, inspection, and intervention operations at offshore structures [28], are characterized by substantial amounts of sensor and image data collection and require intensive work with uncertainties and probabilistic data that can be a task too challenging for human operators. The AUS enables us to turn this data and heavy processing into information ready for interpretation. This data analysis is increasingly reliant on AI methods, such as ML and computer vision, that ease our understanding of ongoing operations and act as decision-support systems. AI connects several traditionally separate disciplines in its lifecycle, data analysis, model building, and software engineering as crucial components of autonomous systems [52]. More specifically, AI systems include problem defi-

Table 7.3. Probability of corrosion caused by common corrosive events, adapted from [51]

Event	Prior Probability	Posterior Probability
High salinity	3.35E-03	1.01E-01
Low temperatures	8.63E-04	2.60E-02
Microorganism presence in the environment	1.30E-03	3.91E-02
Microorganisms on corroded material	2.15E-03	6.47E-02
Anti-corrosive coating failure	6.20E-02	4.73E-01
High CO_2	5.00E-03	1.31E-01
High H_2S	7.15E-03	1.88E-01
Presence of sand in pipeline	5.00E-03	4.40E-02
Internal stress	5.50E-03	4.84E-02
External pipeline stress	2.70E-03	2.38E-02
Deposits or unclean pipeline	1.00E-02	1.07E-02
Harsh subsea environment	1.00E-03	1.19E-02

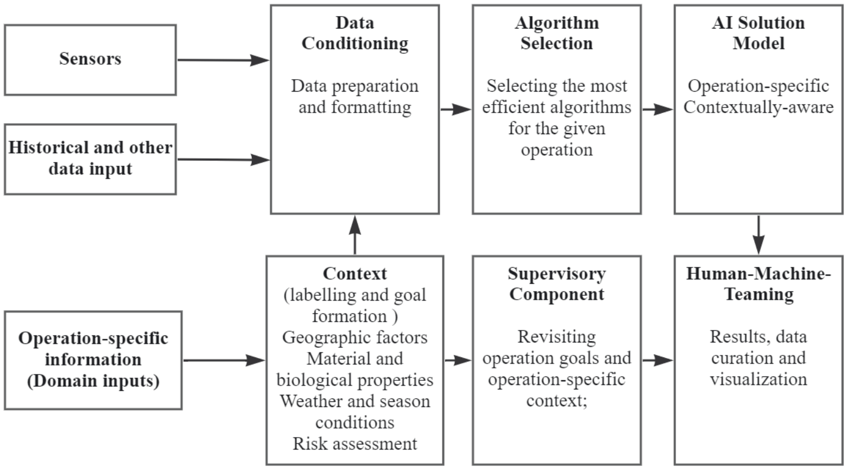

FIGURE 7.6. AI Model Architecture

nition, data collection through sensors or other data inputs, data conditioning, algorithm selection, and solution deployment or delivery as typical steps for developing AI models [52]. Finally, a vital part of the AI model is the Human-Machine Teaming element that represents interactions between humans and the system (i.e., user inputs, results checking, and aided decision-making) . Because of the lack of training data and complexity of the tasks that AI systems are expected to accomplish, the efficiency and accuracy of outputs are often prioritized over interpretability and reliability [53]. Hence, unsupervised and black-box algorithms have become prevalent. These algorithms are often not application specific, challenging to interpret, integrate any risk assessment tasks into, and consequentially cause safety, ethical and moral concerns [53]. Additionally, available training data determines the settings in which classification and anomaly detection methods operate [54]:

1. Supervised: labeled dataset is available to train the model.
2. Semi-supervised: a dataset that does not contain anomalies is used to train the model on distinguishing anomalous from non-anomalous data instances.
3. Unsupervised: there is no available dataset for training the model. The model relies on determining statistical patterns between data instances.

Due to the lack of training data and the difficulty of reliably integrating unsupervised methods, AI approaches may be supplemented with heuristics-based

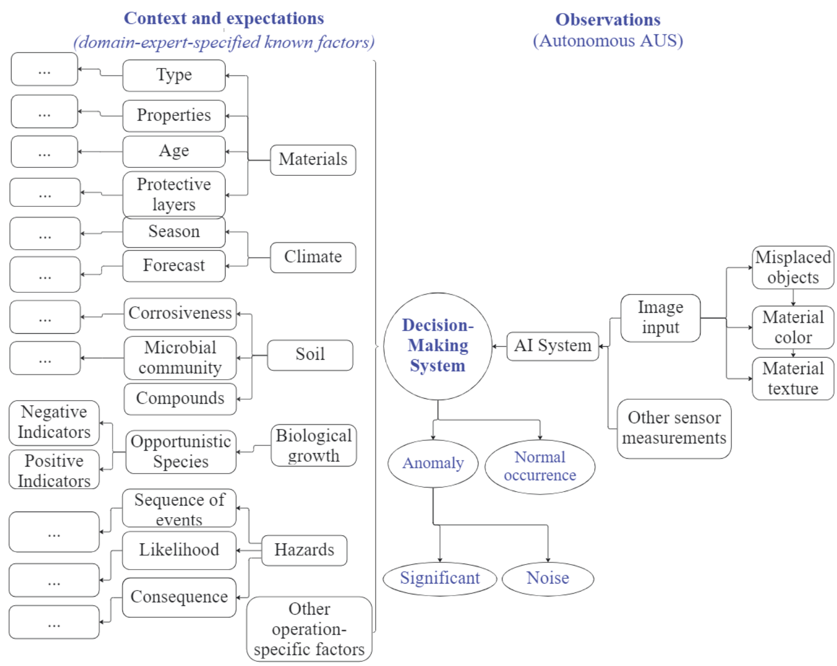

FIGURE 7.7. Knowledge Map for AUS Pipeline Inspection Decision-Making System

methods or heuristic knowledge. Heuristics are described as decision-making methods that employ past experiences to generate quick and efficient solutions to a given problem[55–57], derived from heuristic knowledge that represents the expertise of domain specialists or experts. The systems that take advantage of heuristic knowledge, known as expert systems, serve as decision-support systems and are based on "what-if", "if-else" premises or fuzzy logic[55,56]. The integration of heuristic or empirical knowledge of domain experts, in terms of rules in computing systems, show a great potential in formalizing human knowledge and drawing inferences from observed data for computationally low-cost decision making. However, the integration of heuristics alone into a system may result in biases based solely on past events. Integration of heuristics and analytical tools through AI is therefore one method for avoiding experience bias. Integrating human knowledge by managing heuristic expertise and storing essential skills in dependable and permanent systems can further enhance the interpretability of highly complex systems[57,58]. For subsea pipeline monitoring, the incorporation of domain expertise may help provide operational context with operation-specific information, thereby enhancing the reliability for detection of anomalies or hazards on pipelines. This can be accomplished with risk assessment insights at the training level of the AI life-cycle or during the validation phase of the results, in which the detected anomalies are validated in order to discard noise or discover information about dangerous events. Integration of domain expertise is particularly critical due to recent efforts in AI standardization and functional safety requiring the use and development of more interpretable models that contain operational or application context and carry a high level of reliability. The safety of operations and decision-making systems for AUS cannot be overemphasized. Hence, the interest in integrating risk and context into these systems is increasing in research and industry [59–63]. In Figure Figure 7.6, we propose additional two components to AI models, context and supervisory components as a response to the recent challenges in industry and recommendations for a more standardized approach to AI systems [30,64]. According to [64], a supervisory component or supervisory function, acting as a safe subset of the action space, is advised and expected to be a part of the architecture of AI systems. The AI outputs and decision-making processes are limited by this component, which is regarded as a non-AI component. The proposed context component includes operation-specific knowledge from the domain experts and anticipated risks. The impact of uncertainty on objectives is referred to as risk by [65], and it is typically referred to in terms of risk sources, potential events, likelihood, and consequences. Since AI systems are expected to perform detailed tasks, implementing these systems cannot remain generalized. Context gives us an idea of what kind of setting the AUS operation will take place in. In order to help the AI system decide whether a detected occurrence is significant, Figure Figure 7.7 elaborates on the context and expectations in the AI model for a particular operation. It does this by gathering all the relevant components of the decision-making process. Additionally, Figure Figure 7.7 illustrates the AI system, that includes image inputs and other sensor measurements contributing to decision-making system in detecting significant anomalies from noise, aided by context component. The context includes geographic characteristics like the soil's composition, the climate, the expected marine life and how it might alter in behavior, the material's observed qualities and the several kinds of corrosion that image-based monitoring and inspection makes it feasible to see. On the other hand, the risk assessment picture provides us with information about the risks that we can expect from the operation and lists potential hazards as sources of harm [65,66]. It also shows us how likely it is that each of these things will occur as well as their consequences.

Flage et al.[67] observed the risks through the ways the risk emerges. They describe the emerging or emergent risk as a familiar or unfamiliar risk that becomes apparent in unfamiliar conditions. The authors link the definition of emerging risk to the known/unknown taxonomy derived from the press briefing by United States Secretary of Defense, Donald Rumsfeld[67]:

> There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know.

From this taxonomy, Flage et al.[67] proceeded to address the unknown-knowns and the unknown-unknowns as ambiguous types of risks and link them to the events that lie outside of expectations, known as black swan events. By

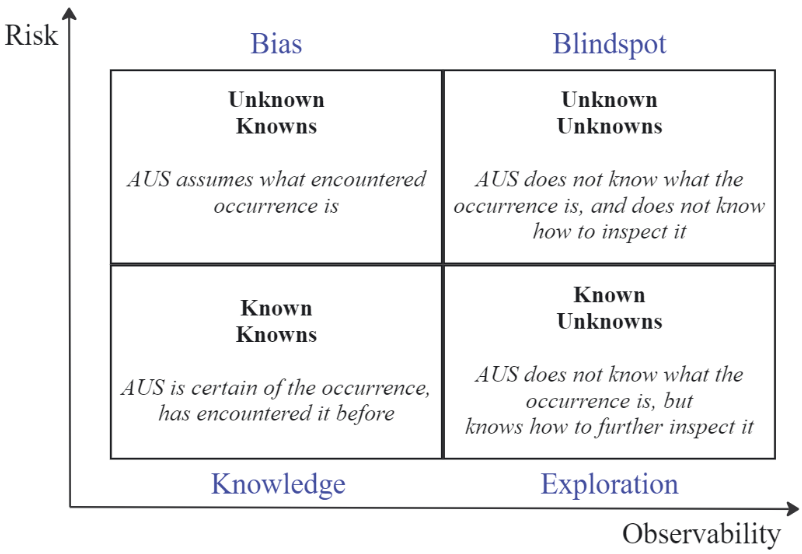

FIGURE 7.8. Rumsfeld Matrix for AUS Exploration

observing the concept of emerging risks, we propose constructing a Rumsfeld matrix for AUS explorations, as shown in Figure 7.8 and demonstrated in Figure 7.9. In the matrix, we place the common challenges from underrepresented data available to the AI model by which AI may derive biased conclusions, known as data biases. An example of a bias in anomaly detection methods is sacrificing anomalies for efficiency and misclassifying them as noise due to a learned experience that there is a higher likelihood of encountering noise rather than a hazardous occurrence[66]. We link the possible situations during the AUS exploration to the emerging risks and identify the following:

1. **(Known-known) Knowledge:** AUS is certain of encountered occurrence because it has learned it from previous experience. The accuracy

is high, confirmed by the supervisory component. Example of a known-known during pipeline inspection: AUS can accurately classify an encountered object, such as biological growth, damage, large boulder, with high confidence.

2. **(Known-unknown) Exploration:** AUS is uncertain of the encountered occurrence, but due to previous experience, knows to inspect the situation further to make the classification. Example of a known-unknown: AUS cannot confidently classify an encountered object, possibly due to the object sharing similar features with more than one classes, such as biological growth and sediment deposit that share the same color and shape.

3. **(Unknown-known) Bias:** AUS assumes the encountered occurrence and proceeds to classify it without further inspection, contributing to unintended bias and potentially problematic conclusions. Example of an unknown-known during pipeline inspection: AUS has not encountered a degraded ship anchor during training phase, but due to high similarity to a known class of biological growth in color and shape, AUS classifies the anchor as biological growth with high confidence, possibly putting the pipe at risk of being damaged due to vicinity of a heavy object.

4. **(Unknown-unknown) Blindspot:** The AUS does not know what the encountered occurrence is, does not know how to proceed in further inspecting it, or experiences challenges in the decision-making process necessitating human interaction. Example of an unknown-unknown: AUS encounters object that is not known and cannot proceed to classify it, potentially requiring human interaction or resulting in an error. This situation may occur if the encountered object is extremely rare and unexpected, or has never appeared in any form during AUS training phase.

Figure 7.9 demonstrates an example of the object classification on the pipeline surface. The image is captured by an underwater drone where the common (known) occurrences, such as biological growth and expected benthic species (Figure 7.9 (a)), are identified, together with texture changes (Figure 7.9 (c)(d)) and sediment deposits (Figure 7.9 (d)). Finally, a sudden material change resembling a material rupture necessitates further inspection (Figure 7.9 (e)). The AUS may need to analyze more images or use more sensor inputs during a subsequent inspection. Determining the right kind of sensors to utilize in a particular situation is crucial when image analysis is insufficient.

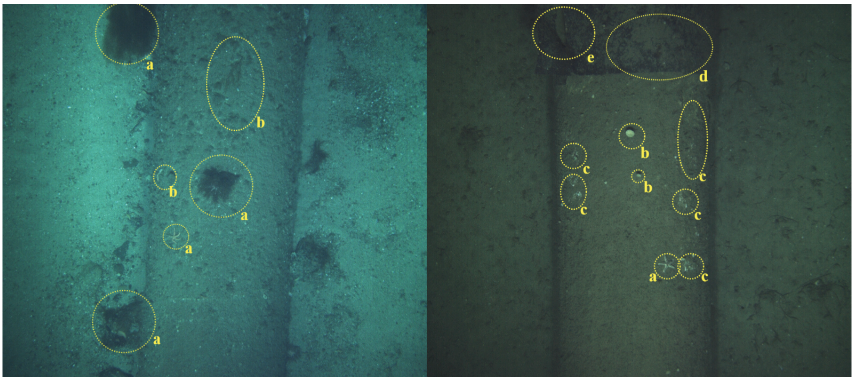

FIGURE 7.9. Subsea Pipeline Images Captured by an Underwater Drone: (a) Known known - Biological growth; (b) Known known - Sediment deposits; (c) Known known - Minor texture/color irregularities; (d) Known known - Major texture/color irregularities; (e) Known Unknown - material rupture. Photo: Equinor

7.2.5 *Rethinking Image-based Monitoring and Inspection*

Widespread use of sensors and imaging equipment for intelligent automation have increased the need for effective and adaptive sensor scheduling, a dynamic sensor control based on environment or operation needs to maximize the efficiency of existing sensors for the intended benefits [68]. Image quality is one of the main drawbacks of image-based analysis. Although many difficulties are camera-related (such as overexposure and focus), the environment can also provide a variety of problems that recur in patterns. During the inspection of offshore pipelines, the oil and gas industry anticipates highly autonomous systems to combine data from multiple sources for the most efficient and reliable data collection [28]. Rather than relying solely on images obtained from visible-spectrum sensors and cameras, it can be beneficial to incorporate data fusion from multiple imaging systems operating in different wavelength ranges, such as infrared cameras. The operation-specific context, however, may assist in identifying the additional types of sensors that may be the most effective under conditions where image-based inspection is problematic. The context of the operation is essential to assessing the validity of conclusions reached by AUS during an image-based pipeline inspection, as was shown in the preceding section. For each factor influencing the objectives of the inspection, such as material degradation, a conditional analysis might be required. Knowing the marine life cycle in such settings is highly beneficial, as a high larval density may lead to a high species density and poor visibility. Not only is it crucial to rely on various sensor inputs in situations when this is an expected pattern, but it is also crucial to avoid mistaking an increase in species presence for an anomalous occurrence. Similarly, in situations when the pipeline's surrounding soil is primarily clay, there is a higher likelihood of corrosion at the points where the soil contacts the pipeline and accumulates on the surface. Similar is true if the encountered occurrence creates ambiguity for an AUS and low confidence in analyzing the situation.

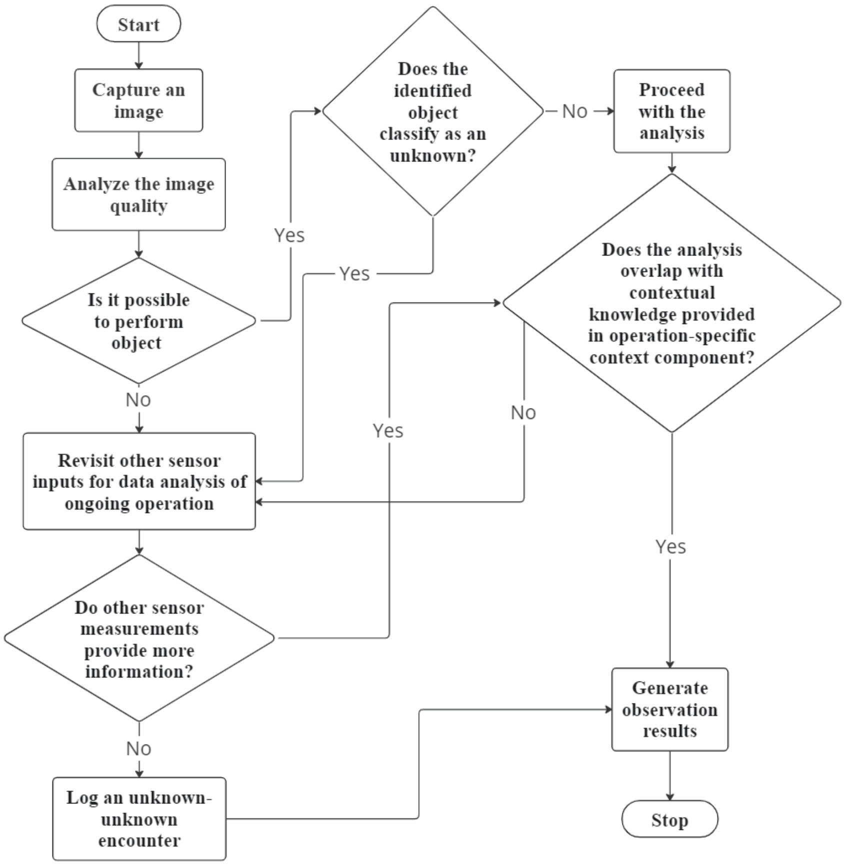

FIGURE 7.10. Image Quality Analysis Flowchart

Figure 7.10 shows a high-level flowchart for entrusting image-based inspection after image quality analysis, illustrating a situation in which an image-based analysis may require additional sensor inputs to classify encountered occurrences during the inspection. An image's quality can be evaluated based on the anticipated conditions, such as exposure, blur, focus, and lighting, after it has been captured by imaging equipment like a drone camera. Object classification begins if the image quality is acceptable and the objects are visible enough for the analysis. Suppose the object is unknown or not identified by the AI model; in that case, the model returns to additional sensor inputs like temperature, pressure, and water content to gather more information. The same sensors are considered if the photos are of poor quality, and object analysis could be more reliable. The AUS needs more inputs to provide observation results if other sensor inputs supply more data to conclude the ongoing observation. Suppose the additional sensor readings are sufficient to support further analysis; in that case, the analyzed output should be updated with the operation-specific contextual information provided in the context component to generate observation results.

7.2.6 Contribution Summary

This research reviewed different factors contributing to subsea pipeline surface material degradation, corrosion, and potential ruptures. The significant factors contributing to the degradation include geographical properties, such as climate and weather variables, soil and water components, microbial communities, and marine species as pollution indicators. Despite extensive study, corrosion continues to be a severe concern to the offshore oil and gas industry, with the potential to cause material failure, which can cause environmental disturbances and substantial financial losses. The employment of autonomous underwater systems for offshore structure inspections, such as intelligent sensor systems and autonomous drones, has consequentially increased the use of AI methods for data analysis in real-time and near-real-time inspections, primarily with image and video captures. These methods include object classification, anomaly detection, and pattern recognition. Although AI methods may evaluate acquired image data effectively, they have flaws that raise concerns about their reliability, such as bias and an efficiency inclination that wastes resources and misdiagnoses pipeline conditions. In this research section, we have examined the factors contributing to the enhanced reliability of image-based pipeline material degradation inspection. We proposed context and supervisory components in AI model architecture and rethinking adaptive sensor scheduling, particularly the image-based inspection, by examining the operation-specific context, emerging risks, and patterned expectations.

It is becoming increasingly important to find context in the operations that AI-dependent systems are required to carry out in order to ensure reliable, intelligible, and ethical outcomes. By analyzing a dataset of underwater images captured by an autonomous drone, we intend to further our research on implementing the AI model context component as a part of image-based anomaly detection and object categorization.

7.3 Conclusions and Key Contributions

This section highlights the key contributions and concludes the chapter and the presented articles.

As the use of UAS becomes more prevalent in the offshore oil and gas industry, the two articles presented identified opportunities to improve the inspection of subsea pipelines. Key findings include the introduction of new temporal anomalies and adaptive sensing for a complex ecosystem of factors that provide early warning signs of potential pollution due to external pipeline damage. These findings are suggestions for future research directions in autonomous pipeline inspection.

7.4 References

[1] Ralph Foorthuis. On the nature and types of anomalies: a review of deviations in data. *International Journal of Data Science and Analytics* **12** (4), 297–331 (2021). ISSN 23644168. doi: 10.1007/s41060-021-00265-1. Cited on page/s 161, 162.

[2] Alexander Lavin and Subutai Ahmad. Evaluating Real-time Anomaly Detection Algorithms - the Numenta Anomaly Benchmark. In *IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015* pages 38–44 Miami, Florida, USA (2015). Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICMLA.2015.141. Cited on page/s 161.

[3] Abed Saif Alghawli. Complex methods detect anomalies in real time based on time series analysis. *Alexandria Engineering Journal* **61** (1), 549–561 (2022). ISSN 11100168. doi: 10.1016/j.aej.2021.06.033. Cited on page/s 161.

[4] James Theiler and Simon Perkins. Proposed framework for anomalous change detection. In *ICML Workshop on Machine Learning Algorithms for Surveillance and Event Detection* pages 7–14 (2006). Cited on page/s 161, 164.

[5] Karima Makhlouf, Sami Zhioua, and Catuscia Palamidessi. On the applicability of ML fairness notions. *arXiv* pages 1–32 (2020). ISSN 23318422. Cited on page/s 161.

[6] R. Sekar et al. Specification-based Anomaly Detection: A New Approach for Detecting Network Intrusions. In *Proceedings of the 9th ACM conference on Computer and communications security - CCS '02* page 265–274 New York, NY, USA (2002). Association for Computing Machinery. ISBN 1581136129. doi: 10.1145/586110.586146. Cited on page/s 161.

[7] Charu C. Aggarwal. An Introduction to Outlier Analysis. In *Outlier Analysis* chapter 1, pages 1–34. Springer, Cham (2017). ISBN 978-3-319-47577-6. doi: 10.1007/978-3-319-47578-3{_}1. Cited on page/s 161.

[8] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A Survey. *ACM Computing Surveys (CSUR)* **14** (1), 1–22 (7 2009). ISSN 15462218. doi: 10.1145/1541880.1541882. Cited on page/s 162.

[9] Alexander Fisch, Idris Eckley, and Paul Fearnhead. Subset Multivariate Collective And Point Anomaly Detection. *Journal of Computational and Graphical Statistics* pages 1–51 (2019). doi: 10.1080/10618600.2021.1987257. Cited on page/s 162.

[10] Michael A. Hayes and Miriam A.M. Capretz. Contextual anomaly detection in big sensor data. In *Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014* pages 64–71. Institute of Electrical and Electronics Engineers Inc. (9 2014). ISBN 9781479950577. doi: 10.1109/BigData.Congress.2014.19. Cited on page/s 162.

[11] Song Xiuyao, Wu Mingxi, Christopher Jermaine, and Sanjay Ranka. Conditional anomaly detection. *IEEE Transactions on Knowledge and Data Engineering* **19** (5), 631–644 (5 2007). ISSN 10414347. doi: 10.1109/TKDE.2007.1009. Cited on page/s 162.

[12] L. Erhan et al. Smart anomaly detection in sensor systems: A multi-perspective review. *Information Fusion* **67** (September 2020), 64–79 (2021). ISSN 15662535. doi: 10.1016/j.inffus.2020.10.001. Cited on page/s 162, 163.

[13] A. N. Ragozin, V. F. Telezhkin, and P. S. Podkorytov. Forecasting complex multi-component time series within systems designed to detect anomalies in dataflows of industrial automated systems. *ACM International Conference Proceeding Series* pages 1–5 (2019). doi: 10.1145/3357613.3357615. Cited on page/s 163.

[14] Robyn R. Lutz and Inés Carmen Mikulski. Empirical analysis of safety-critical anomalies during operations. *IEEE Transactions on Software Engineering* **30** (3), 172–180 (3 2004). ISSN 00985589. doi: 10.1109/TSE.2004.1271171. Cited on page/s 163, 166.

[15] Sicong Liu, Daniele Marinelli, Lorenzo Bruzzone, and Francesca Bovolo. A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges. *IEEE Geoscience and Remote Sensing Magazine* 7 (2), 140–158 (2019). ISSN 21686831. doi: 10.1109/MGRS.2019.2898520. Cited on page/s 163, 164.

[16] Len Feremans et al. Pattern-Based Anomaly Detection in Mixed-Type Time Series. *Machine Learning and Knowledge Discovery in Databases* **11906**, 240–256 (2020). ISSN 16113349. doi: 10.1007/978-3-030-46150-8{_}15. Cited on page/s 165, 166.

[17] Fei Tony Liu, Kai Ming Ting, and Zhi Hua Zhou. Isolation forest. In *Proceedings - IEEE International Conference on Data Mining, ICDM* pages 413–422. IEEE (2008). ISBN 9780769535029. doi: 10.1109/ICDM.2008.17. Cited on page/s 165.

[18] Christopher Hannon, Deepjyoti Deka, Dong Jin, Marc Vuffray, and Andrey Y. Lokhov. Real-time Anomaly Detection and Classification in Streaming PMU Data. In *2021 IEEE Madrid PowerTech* pages 1–6 Madrid (2021). IEEE. ISBN 9781665435970. doi: 10.1109/PowerTech46648.2021.9494800. Cited on page/s 165.

[19] James Theiler. Anomalousness: how to measure what you can't define. *Fourier Transform Spectroscopy and Hyperspectral Imaging and Sounding of the Environment (2015)* page JT1A.2 (2015). doi: 10.1364/FTS.2015.JT1A.2. Cited on page/s 165.

[20] Gian Luca Foresti. Visual inspection of sea bottom structures by an autonomous underwater vehicle. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **31** (5), 691–705 (10 2001). ISSN 10834419. doi: 10.1109/3477.956031. Cited on page/s 166, 167, 168.

[21] Thanawin Rakthanmanon, Eamonn J. Keogh, Stefano Lonardi, and Scott Evans. Time series epenthesis: Clustering time series streams requires ignoring some data. *Proceedings - IEEE International Conference on Data Mining, ICDM* pages 547–556 (2011). ISSN 15504786. doi: 10.1109/ICDM.2011.146. Cited on page/s 166.

[22] Sreelekha Guggilam, Varun Chandola, and Abani Patra. Tracking clusters and anomalies in evolving data streams. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **15** (2), 156–178 (2021). ISSN 1932-1872. doi: 10.1002/SAM.11552. Cited on page/s 166.

[23] Georgias Rovatsos, Venugopal V. Veeravalli, Don Towsley, and Ananthram Swami. Quickest Detection of Growing Dynamic Anomalies in Networks. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* **2020-May**,

8926–8930 (5 2020). ISSN 15206149. doi: 10.1109/ICASSP40776.2020.9053019. Cited on page/s 166, 168, 169.

[24] Teng Wang, Chunsheng Fang, Derek Lin, and S. Felix Wu. Localizing temporal anomalies in large evolving graphs. Society for Industrial and Applied Mathematics Publications (2015). ISBN 9781510811522. doi: 10.1137/1.9781611974010.104. Cited on page/s 166.

[25] Yahya T. Al-Janabi. Monitoring of Downhole Corrosion: An Overview. *Society of Petroleum Engineers - SPE Saudi Arabia Section Technical Symposium and Exhibition 2013* pages 108–118 (5 2013). doi: 10.2118/168065-MS. Cited on page/s 166.

[26] Wessel Marquering, Johan Nisser, and Toni Valla. Disappearing anomalies: A dynamic analysis of the persistence of anomalies. *Applied Financial Economics* **16** (4), 291–302 (2006). ISSN 09603107. doi: 10.1080/09603100500400361. Cited on page/s 167, 168.

[27] Tianyu Wang, Deyu Xu, Lina Qu, Jiangwei Fu, and Zhiliang Li. An extension approach to estimate soil corrosivity for buried pipelines. *International Journal of Pressure Vessels and Piping* **192** (March), 104413 (2021). ISSN 03080161. doi: 10.1016/j.ijpvp.2021.104413. URL https://doi.org/10.1016/j.ijpvp.2021.104413. Cited on page/s 171, 173, 174.

[28] Francesco Scibilia, Knut Sebastian Tungland, Anders Røyrøy, and Marianne Bryhni Asla. Energy industry perspective on the definition of autonomy for mobile robots. In *Communications in Computer and Information Science* volume 1056 CCIS pages 90–101. Springer International Publishing (2019). ISBN 9783030356637. doi: 10.1007/978-3-030-35664-4{_}9. URL http://dx.doi.org/10.1007/978-3-030-35664-4_9. Cited on page/s 172, 179, 185.

[29] Tom Froese, Nathaniel Virgo, and Eduardo Izquierdo. Autonomy: A Review and a Reappraisal. In Fernando e Costa, Luis Mateus Rocha, Ernesto Costa, Inman Harvey, and António Coutinho, editors, *Advances in Artificial Life* pages 455–464 Berlin, Heidelberg (2007). Springer Berlin Heidelberg. ISBN 978-3-540-74913-4. Cited on page/s 172.

[30] Chong Chen, Dazhong Wu, and Ying Liu. Recent advances of AI for engineering service and maintenance. *Autonomous Intelligent Systems* **2** (1), 2–4 (2022). doi: 10.1007/s43684-022-00038-y. Cited on page/s 172, 182.

[31] Ilias Panagiotopoulos and George Dimitrakopoulos. Leveraging on non-causal reasoning techniques for enhancing the cognitive management of highly automated vehicles. *Autonomous Intelligent Systems* **2** (1), 1–13 (2022). doi: 10.1007/s43684-022-00035-1. Cited on page/s 172.

[32] Anupama R. Prasad, Anupama Kunyankandy, and Abraham Joseph. Corrosion Inhibition in Oil and Gas Industry. In *Corrosion Inhibitors in the Oil and Gas Industry* chapter 5, pages 135–150. John Wiley & Sons, Ltd (2020). ISBN 9783527822126. doi: 10.1002/9783527822140.CH5. Cited on page/s 172.

[33] Enyinnaya G. Ohaeri and Jerzy A. Szpunar. An overview on pipeline steel development for cold climate applications. *Journal of Pipeline Science and Engineering* **2** (1), 1–17 (2022). ISSN 26671433. doi: 10.1016/j.jpse.2022.01.003. URL https://doi.org/10.1016/j.jpse.2022.01.003. Cited on page/s 172, 173, 174.

[34] M. Hagarová, J. Cervová, and F. Jaš. Selected types of corrosion degradation of pipelines. *Koroze a Ochrana Materialu* **59** (1), 30–36 (2015). ISSN 18041213. doi: 10.1515/kom-2015-0010. Cited on page/s 173, 178.

[35] Syahril Anuar Idris, Fairul Azni Jafar, Zamberi Jamaludin, and Noraidah Blar. Improvement of Corrosion Detection Using Vision System for Pipeline Inspection. *Applied Mechanics and Materials* **761** (May), 125–131 (2015). doi: 10.4028/www.scientific.net/amm.761.125. Cited on page/s 173, 176, 177, 178.

[36] M. Dubiel, C. H. Hsu, C. C. Chien, F. Mansfeld, and D. K. Newman. Microbial iron respiration can protect steel from corrosion. *Applied and Environmental Microbiology* **68** (3), 1440–1445 (2002). ISSN 00992240. doi: 10.1128/AEM.68.3.1440-1445.2002. Cited on page/s 175.

[37] Michael Redford, Sally Rouse, Peter Hayes, and Thomas A. Wilding. Benthic and Fish Interactions With Pipeline Protective Structures in the North Sea. *Frontiers in Marine Science* **8** (April) (2021). ISSN 22967745. doi: 10.3389/fmars.2021.652630. Cited on page/s 175.

[38] Brage Rygg. Distribution of species along pollution-induced diversity gradients in benthic communities in Norwegian fjords. *Marine Pollution Bulletin* **16** (12), 469–474 (1985). ISSN 0025326X. doi: 10.1016/0025-326X(85)90378-9. Cited on page/s 175.

[39] Hong Su, Shuofu Mi, Xiaowei Peng, and Yejun Han. The mutual influence between corrosion and the surrounding soil microbial communities of buried petroleum pipelines. *RSC Advances* **9** (33), 18930–18940 (2019). ISSN 20462069. doi: 10.1039/c9ra03386f. Cited on page/s 175.

[40] Victoria L.G. Todd, Laura Lazar, Laura D. Williamson, Ingrid T. Peters, Aimee L. Hoover, Sophie E. Cox, Ian B. Todd, Peter I. Macreadie, and Dianne L. McLean. Underwater Visual Records of Marine Megafauna Around Offshore Anthropogenic Structures. *Frontiers in Marine Science* 7 (April) (2020). ISSN 22967745. doi: 10.3389/fmars.2020.00230. Cited on page/s 175, 177.

[41] Virginia Biede, Andrew R. Gates, Simone Pfeifer, Jane E. Collins, Carmen Santos, and Daniel O.B. Jones. Short-Term Response of Deep-Water Benthic Megafauna to Installation of a Pipeline Over a Depth Gradient on the Angolan Slope. *Frontiers in Marine Science* **9** (June), 1–12 (2022). ISSN 22967745. doi: 10.3389/fmars.2022.880453. Cited on page/s 175.

[42] M. Dubiel, C. H. Hsu, C. C. Chien, F. Mansfeld, and D. K. Newman. Microbial iron respiration can protect steel from corrosion. *Applied and Environmental Microbiology* **68** (3), 1440–1445 (2002). ISSN 00992240. doi: 10.1128/AEM.68.3.1440-1445.2002. Cited on page/s 175, 176.

[43] Elena Quatrini, Francesco Costantino, Giulio Di Gravio, and Riccardo Patriarca. Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities. *Journal of Manufacturing Systems* **56**, 117–132 (2020). ISSN 0278-6125. doi: 10.1016/J.JMSY.2020.05.013. Cited on page/s 176.

[44] T Moller, I Nilssen, and T.W Nattkemper. Change detection in marine observatory image streams using Bi-Domain Feature Clustering. In *23rd International Conference on Pattern Recognition (ICPR)* pages 793–798 Cancun, Mexico (12 2016). IEEE Xplore. doi: 10.1109/ICPR.2016.7899732. Cited on page/s 176.

[45] William S. Tait. Controlling Corrosion of Chemical Processing Equipment. *Handbook of Environmental Degradation Of Materials: Third Edition* pages 583–600 (1 2018). doi: 10.1016/B978-0-323-52472-8.00028-9. Cited on page/s 176.

[46] G. S. Frankel. Pitting Corrosion of Metals: A Review of the Critical Factors. *Journal of The Electrochemical Society* **145** (6), 2186–2198 (6 1998). ISSN 0013-4651. doi: 10.1149/1.1838615/XML. URL https://iopscience.iop.org/article/10.1149/1.1838615https://iopscience.iop.org/article/10.1149/1.1838615/meta. Cited on page/s 176.

[47] J. R. Galvele. Tafel's law in pitting corrosion and crevice corrosion susceptibility. *Corrosion Science* **47** (12), 3053–3067 (12 2005). ISSN 0010-938X. doi: 10.1016/J.CORSCI.2005.05.043. Cited on page/s 176.

[48] X. G. Zhang. Galvanic Corrosion. In *Uhlig's Corrosion Handbook: Third Edition* chapter 10, pages 123–143. John Wiley & Sons, Ltd (4 2011). ISBN 9780470080320. doi: 10.1002/9780470872864.CH10. Cited on page/s 176.

[49] S. S. Rajahram, T. J. Harvey, and R. J.K. Wood. Erosion–corrosion resistance of engineering materials in various test conditions. *Wear* **267** (1-4), 244–254 (6 2009). ISSN 0043-1648. doi: 10.1016/J.WEAR.2009.01.052. Cited on page/s 176.

[50] Mary Lyn C. Lim, Robert G. Kelly, and John R. Scully. Overview of Intergranular Corrosion Mechanisms, Phenomenological Observations, and Modeling of

AA5083. *Corrosion* **72** (2), 198–220 (2 2016). ISSN 0010-9312. doi: 10.5006/1818. URL https://meridian.allenpress.com/corrosion/article/72/2/198/199091/Overview-of-Intergranular-Corrosion-Mechanisms. Cited on page/s 176.

[51] Yongsheng Yang, Faisal Khan, Premkumar Thodi, and Rouzbeh Abbassi. Corrosion induced failure analysis of subsea pipelines. *Reliability Engineering and System Safety* **159**, 214–222 (3 2017). ISSN 09518320. doi: 10.1016/j.ress.2016.11.014. Cited on page/s 178, 179.

[52] Lukas Fischer, Lisa Ehrlinger, Verena Geist, Rudolf Ramler, Florian Sobiezky, Werner Zellinger, David Brunner, Mohit Kumar, and Bernhard Moser. AI System Engineering—Key Challenges and Lessons Learned. *Machine Learning and Knowledge Extraction* **3** (1), 56–83 (2020). ISSN 25044990. doi: 10.3390/make3010004. Cited on page/s 179, 180.

[53] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1** (5), 206–215 (2019). ISSN 25225839. doi: 10.1038/s42256-019-0048-x. URL http://dx.doi.org/10.1038/s42256-019-0048-x. Cited on page/s 180.

[54] Markus Goldstein and Seiichi Uchida. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLOS ONE* **11** (4) (2016). doi: 10.1371/JOURNAL.PONE.0152173. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152173. Cited on page/s 180.

[55] Qiuwen Chen and Arthur E Mynett. Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake. *Ecological Modelling* **162**, 55–67 (2003). Cited on page/s 181.

[56] Sajjad Ahmad and Slobodan P Simonovic. Integration of heuristic knowledge with analytical tools for the selection of flood damage reduction measures. *Canadian Journal of Civil Engineering* **28** (2), 208–221 (2001). doi: 10.1139/l00-097. URL https://doi.org/10.1139/l00-097. Cited on page/s 181.

[57] Luiz Fernando de Carvalho Botega and Jonny Carlos da Silva. An artificial intelligence approach to support knowledge management on the selection of creativity and innovation techniques. *Journal of Knowledge Management* **24** (5), 1107–1130 (6 2020). ISSN 17587484. doi: 10.1108/JKM-10-2019-0559. Cited on page/s 181.

[58] Yanqing Duan, John S Edwards, and Yogesh K Dwivedi. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management* **48**, 63–71 (2019). ISSN 0268-4012. doi: https://doi.org/10.1016/j.ijinfomgt.2019.01.021. URL https://www.sciencedirect.com/science/article/pii/S0268401219300581. Cited on page/s 181.

[59] Benjamin J. Smith, Robert Klassert, and Roland Pihlakas. Using soft maximin for risk averse multi-objective decision-making. *Autonomous Agents and Multi-Agent Systems* **37** (1) (2023). ISSN 15737454. doi: 10.1007/s10458-022-09586-2. URL https://doi.org/10.1007/s10458-022-09586-2. Cited on page/s 182.

[60] Michael Fisher, Viviana Mascardi, Kristin Yvonne Rozier, Bernd Holger Schlingloff, Michael Winikoff, and Neil Yorke-Smith. Towards a framework for certification of reliable autonomous systems. *Autonomous Agents and Multi-Agent Systems* **35** (1), 1–65 (2021). ISSN 15737454. doi: 10.1007/s10458-020-09487-2. URL https://doi.org/10.1007/s10458-020-09487-2. Cited on page/s 182.

[61] Georg Hägele and Dirk Söffker. Risk Areas Determination for Autonomous- and Semi-autonomous Aerial Systems Considering Run-Time Technical Reliability Assessment: Requirements, Concept, and Tests. *Journal of Intelligent and Robotic Systems: Theory and Applications* **97** (3-4), 511–529 (2020). ISSN 15730409. doi: 10.1007/s10846-019-01056-4. Cited on page/s 182.

[62] George Vachtsevanos, Benjamin Lee, Sehwan Oh, and Michael Balchanos. Resilient Design and Operation of Cyber Physical Systems with Emphasis on Unmanned Autonomous

Systems. *Journal of Intelligent and Robotic Systems: Theory and Applications* **91** (1), 59–83 (2018). ISSN 15730409. doi: 10.1007/s10846-018-0881-x. Cited on page/s 182.

[63] Robert Bogue. Robots in the offshore oil and gas industries: a review of recent developments. *Industrial Robot* **47** (1), 1–6 (1 2020). ISSN 0143991X. doi: 10.1108/IR-10-2019-0207. Cited on page/s 182.

[64] ISO/IEC. ISO/IEC TR5469:202x(E) Artificial Intelligence - Functional safety and AI systems. Technical report International Electrotechnical Comission (2022). URL https://www.iso.org/standard/81283.html. Cited on page/s 182.

[65] ISO 31000. Risk management — Guidelines, International Organization for Standardization. Technical report International Organization for Standardization (2018). URL https://www.iso.org/obp/ui/iso:std:iso:31000:ed-2:v1:en. Cited on page/s 182.

[66] Rialda Spahic, Hepsø, Vidar, and Mary Ann Lundteigen. Using Risk Analysis for Anomaly Detection for Enhanced Reliability of Unmanned Autonomous Systems. In Maria Chiara Leva, Edoardp Patelli, Luca Podofillini, and Simon Wilson, editors, *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)* pages 273–280 Singapore (2022). Research Publishing, Singapore. doi: 10.3850/978-981-18-5183-4{_}R08-03-390-cd. URL https://rpsonline.com.sg/rps2prod/esrel22-epro/html/toc.html. Cited on page/s 182, 183.

[67] R. Flage and T. Aven. Emerging risk - Conceptual definition and a relation to black swan type of events. *Reliability Engineering and System Safety* **144**, 61–67 (8 2015). ISSN 09518320. doi: 10.1016/j.ress.2015.07.008. Cited on page/s 182, 183.

[68] Dimitris Mourtzis. Advances in Adaptive Scheduling in Industry 4.0. *Frontiers in Manufacturing Technology* **2** (July), 1–29 (2022). doi: 10.3389/fmtec.2022.937889. Cited on page/s 185.

Part III

EPILOGUE

Chapter 8

Conclusion

This thesis examined several aspects of utilizing autonomous underwater systems with artificial intelligence, such as machine learning, anomaly detection, and computer vision, for detecting hazards during a subsea pipeline inspection. During the early stages of research, risk analysis, reliability engineering, resilience engineering, and human-machine teaming, served as a foundation for identifying solutions to well-known challenges and limitations of artificial intelligence. Experiments with sensor and image data collected by autonomous underwater drones demonstrated that the inadequacies of the methods and the imbalances in the collected data undermine the reliability of the results and need to be revised to guarantee the integrity of the inspected offshore structures. Consequently, applying risk analysis to obtain domain-specific hazards and their properties has enabled the detection of hazard warnings among many reported anomalies and noise. The presented contributions have successfully fulfilled main research objectives and answered research questions, while providing directions for extending the research and development of multi-purpose models for AI-based subsea pipeline hazard detection.

8.1 Overview

Chapter 1 introduced the motivation and scope of this thesis, state of the art from the perspective of autonomous systems in subsea pipeline inspection, its implications, the research objectives, and the contributions.

Chapter 2 described the fundamentals of this thesis, namely the role of subsea pipelines in the industry and the complex conditions that expose the pipelines to various types of hazards. This chapter also describes the external failures that can be detected during visual inspection with underwater vehicles and introduces the fundamentals of risk-based analysis for determining the significance of these failures.

Chapter 3 discussed the fundamentals of the experimental methods examined. This chapter describes the method for identifying anomalous data patterns, anomaly detection, the types of anomalies that can be detected with conventional methods, and the necessary data setups for employing anomaly detection in various applications. In addition, computer vision applications are described, including image classification, object detection on images, and object segmentation, as well as the blindspots and challenges associated with their application. Lack of training data, heavy data imbalances contributing to biases and errors in results, and the need to employ complex algorithms that make it difficult to explain results and incorporate risk measures are obstacles shared by anomaly detection and described computer vision methods.

The main contributions of Chapter 4 are identifying concepts from multiple safety-related disciplines to transfer knowledge and address common anomaly detection and machine learning challenges for hazard detection in remote operations. The presented challenges are trust calibration, explainability of algorithms, data biases, and the inadequacy of anomaly detection methods to identify data biases and efficiently differentiate noise from meaningful anomalies. This chapter presents the early concepts of a novel Warning Identification Framework that is anchored in the expectations of risk assessment, reliability engineering, resilience engineering, and human-machine teaming. The steps that configure the Warning Identification Framework rely on the warning identification or lists of expected hazards, hierarchy of warnings or hazard consequences, and orchestration of response methods by the autonomous underwater systems.

The focus of Chapter 5 was on analyzing sensor-collected data to identify the limitations of methods and the potential of the previously proposed Warning Identification Framework. In the first section of Chapter 5, the data collection capabilities of manually and autonomously operated drones were compared. During this experiment, several machine learning and anomaly detection strategies were evaluated to determine the differences in data collection. The outcomes of this analysis provided insight into the potential for addressing issues posed by imbalanced data. Chapter 5's second section analyzed sensor-collected seismic data and compared anomaly detection results with domain expert-specified hazard assessment method results to determine the number of noise and hazards detected by anomaly detection. This analysis demonstrated that incorporating a hazard assessment method can enhance the capability to detect real hazards in a significant number of anomalies and eliminate noise generated by anomaly detection methods.

The main focus of Chapter 6 was exploration of anomalies as risk factors for image-based hazard detection. Analysis of underwater images is challenging due to complex environment that results in hazy and monochromatic images. Additionally, the methods used for the analysis are often complex, challenging to interpret, and, due to computational resources, require resizing of images which may result in loss of information. The image data analyzed in this chapter was collected by an underwater drone that followed a subsea pipeline. Despite high quality images, the pipeline surface was covered in soft sediment and occasional marine growth. Due to the poor visibility of pipeline surface, it was challenging to detect corrosion damage. However, a small number of images displayed ruptures and mechanical damage on the pipeline surface. Due to the high amount of images with pipeline without any damage, and only a few images of damaged pipeline, the dataset was heavily imbalanced, and the image classification methods did not detect any hazards or anomalies. Therefore, by using seamless blending and image manipulation techniques, we were able to generate artificial images of damaged pipelines and balance the training dataset, resulting in improved chances for classification. As an effort to increase explainability of image analysis and classification methods, we have applied localised anomaly detection to highlight distinctive regions on the images that contribute to classification. Finally, we presented the process of expanding the training dataset and enhancing the explainability as a novel methodology for image-based subsea pipeline hazard detection.

Chapter 7 explored future directions for remote operations with autonomous underwater systems for subsea pipeline inspection have been proposed during this research. One of the future directions is a novel categorization of anomalies with time-dependent data. The second proposal for future direction describes dynamic properties that impact the surface degradation of pipeline materials and circumstances under which the visual inspection may necessitate an adaptive sensor approach and modification of traditional AI architecture models.

8.2 Challenges and Lessons Learned

Inspection of subsea pipelines with anomaly detection and classification methods has been studied in the past, but the well-known challenges regarding the reliability of these methods and the lack of risk and hazard insights in the traditional lifecycle of data analysis persist. In addition, the existing literature does not provide extensive applications of these methods under complex conditions, such as monochromatic images, the absence of training data, the interpretation of image analysis methods, and the absence of supervision methods. The stated motivation and objectives of this thesis were to investigate and address the challenges associated with safe AI applications for subsea pipeline inspection in an efficient and effective manner.

Risk and hazard analysis as supervisory components to traditional AI approaches, such as anomaly detection and classification, was one of the primary objectives of this thesis. This has proven to be a novel contribution, and the results of the tested method have been encouraging. This method was only tested on sensor-obtained seismic data containing evidence of seismic tremors due to the difficulty of locating adequate datasets containing relevant data with performed hazard analysis. The most difficult aspect of testing the method of hazard analysis as a supervisor for anomaly detection was locating a dataset with anomalies and research that examined the dataset using hazard analysis and domain experts. Unfortunately, this method has not been tested on images of subsea pipelines and is only applicable to numeric data. Hazard analysis of industry-provided pipeline images necessitates additional knowledge on properties of pipelines and the pipeline environment (e.g., materials, location, sediment types) and the extensive participation of domain experts. However, it was not possible to obtain additional information about pipelines, and the images were anonymized due to industry privacy policies. This made it difficult for domain experts to provide a reliable risk picture.

Existing knowledge on external anomalies as risk factors for manual pipeline inspection could be utilized despite the insufficiency of information on pipeline image data. This knowledge of risk factors enabled the extraction of anomalous images and the observation of existing damage. Due to poor visibility on images, only mechanical damage was detected using hazard detection. Despite the presence of other anomalies, such as large boulders and unknown objects surrounding the pipelines, it was difficult to generate these anomalies from synthetic data. Lack of realism in the images is the primary reason for this challenge. Current image manipulation and color correction techniques are not sophisticated enough to handle complex underwater images with shape and color loss. It is not essential to use these techniques for external pipeline damage. The proposed method was created to detect mechanical damage, but it should be expanded to detect other anomalies or hazards.

8.3 Future Work

Future research will concentrate primarily on pipeline and surrounding environment surveillance with image and video data, supplemented by sensor data from IR cameras and chemical sensors for water and soil analysis. Even when a variety of sensors are employed and sufficient information about the pipeline's environment is available, autonomous hazard detection on subsea pipelines remains a challenging task.

Convolutional Neural Network (CNN) analysis is computationally intensive, particularly due to the requirement for high-quality images. CNN and Isolation Forest anomaly detection have performed admirably, but additional techniques must be evaluated. Lack of explainability in AI is an additional concern with CNN methods, and despite the promising results from Localised Anomaly Detection as an effort to increase explainability, additional approaches must be explored and compared. In addition, the loss of color and/or shape in images presents a formidable obstacle for image-based inspection. It is necessary to implement color correction and shape reconstruction techniques to enhance image clarity and object detection capabilities.

Future work will consist of testing the proposed methods with additional data and detecting multiple anomalies. As discussed in Chapter 7, for future pipeline hazard detection perspectives, the introduction of operation-specific anomalies and the tracking of anomaly development over time are required in order to understand the development of hazards and implement proactive and preventative measures. As flexible and real-time inspection is one of the future goals for remote inspection in industry, it is natural that future research will include video-data and the designing experiments for real-time performance tests. To further generalize the autonomous subsea pipeline hazard detection, two articles in Chapter 7 have focused on future models that could be considered with an access to extensive image, video and other sensor data.

Appendix A

Supporting information for Chapter 5.2

Seismic dataset attributes

1. seismic: result of shift seismic hazard assessment in the mine working obtained by the seismic method (a - lack of hazard, b - low hazard, c - high hazard, d - danger state);
2. seismoacoustic: result of shift seismic hazard assessment in the mine working obtained by the seismoacoustic method;
3. shift: information about type of a shift (W - coal-getting, N -preparation shift);
4. genergy: seismic energy recorded within previous shift by the most active geophone (GMax) out of geophones monitoring the longwall;
5. gpuls: a number of pulses recorded within previous shift by GMax;
6. gdenergy: a deviation of energy recorded within previous shift by GMax from average energy recorded during eight previous shifts;
7. gdpuls: a deviation of a number of pulses recorded within previous shift by GMax from average number of pulses recorded during eight previous shifts;
8. ghazard: result of shift seismic hazard assessment in the mine working obtained by the seismoacoustic method based on registration coming form GMax only;
9. nbumps: the number of seismic bumps recorded within previous shift;
10. nbumps2: the number of seismic bumps (in energy range $[10^2,10^3)$) registered within previous shift;
11. nbumps3: the number of seismic bumps (in energy range $[10^3,10^4)$) registered within previous shift;

12. nbumps4: the number of seismic bumps (in energy range [10^4,10^5)) registered within previous shift;
13. nbumps5: the number of seismic bumps (in energy range [10^5,10^6)) registered within the last shift;
14. nbumps6: the number of seismic bumps (in energy range [10^6,10^7)) registered within previous shift;
15. nbumps7: the number of seismic bumps (in energy range [10^7,10^8)) registered within previous shift;
16. nbumps89: the number of seismic bumps (in energy range [10^8,$10^1$0)) registered within previous shift;
17. energy: total energy of seismic bumps registered within previous shift;
18. maxenergy: the maximum energy of the seismic bumps registered within previous shift;
19. class: the decision attribute - '1' means that high energy seismic bump occurred in the next shift ('hazardous state'), '0' means that no high energy seismic bumps occurred in the next shift ('non-hazardous state') - generated during rule-based classification experiment by [1]

A.1 REFERENCES

[1] Jozef Kabiesz, Beata Sikora, Marek Sikora, and Lukasz Wrobel. Application of rule-based models for seismic hazard prediction in coal mines. *Acta Montanistica Slovaca* **18** (4), 262–277 (2013). Cited on page/s A-2.

ISBN 978-82-326-7190-8 (printed ver.)
ISBN 978-82-326-7189-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology