UNIVERSIDADE FEDERAL DE ALAGOAS

CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

LUCAS GOUVEIA OMENA LOPES

Open-World Learning Applied to Oil Wells Using Autoencoder-Based Clustering

Maceió
2025

Lucas Gouveia Omena Lopes

Open-World Learning Applied to Oil Wells Using Autoencoder-Based Clustering

Doctoral thesis submitted as part of the requirements for obtaining the Doctor's degree in Civil Engineering through the Graduate Program in Civil Engineering at the Technology Center of the Federal University of Alagoas.
Advisor: Prof. Dr. William Wagner Matos Lira
Co-advisor: Prof. Dr. Thales Miranda de Almeida Vieira

Maceió
2025

Catalogação na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

L864o Lopes, Lucas Gouveia Omena.

Open-world learning applied to oil wells using autoencoder-based clustering / Lucas Gouveia Omena Lopes. - 2025.

99 f. : il.

Orientador: William Wagner Matos Lira.

Co-orientador: Thales Miranda de Almeida Vieira.

Tese (Doutorado em Engenharia Civil) – Universidade Federal de Alagoas. Centro de Tecnologia. Programa de Pós-Graduação em Engenharia Civil. Maceió, 2025.

Texto em inglês.

Bibliografia: f. 93-99.

1. Detecção de anomalias. 2. Open-world learning. 3. Detecção de novidades. 4. Representação latente. 5. Monitoramento industrial dinâmico. 5. Poços - Integridade estrutural. I. Título.

CDU: 622.276

OPEN-WORLD LEARNING APLICADO A POÇOS DE PETRÓLEO UTILIZANDO CLUSTERIZAÇÃO BASEADA EM AUTOENCODER

LUCAS GOUVEIA OMENA LOPES

Tese de doutorado submetida à banca examinadora do Programa de Pós-Graduação em Engenharia Civil da Universidade Federal de Alagoas e aprovada no dia 31 do mês de janeiro do ano de 2025.

Banca Examinadora:

Documento assinado digitalmente
gov.br WILLIAM WAGNER MATOS LIRA
Data: 31/01/2025 17:35:15-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. William Wagner Matos Lira
(Orientador – PPGEC/UFAL)

Documento assinado digitalmente
gov.br THALES MIRANDA DE ALMEIDA VIEIRA
Data: 03/02/2025 11:19:07-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Thales Miranda de Almeida Vieira
(Coorientador – UFAL)

Documento assinado digitalmente
gov.br EDUARDO TOLEDO DE LIMA JUNIOR
Data: 01/02/2025 20:48:48-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Eduardo Toledo de Lima Junior
(Avaliador Interno – PPGEC/UFAL)

Documento assinado digitalmente
gov.br JOAO PAULO LIMA SANTOS
Data: 02/02/2025 11:27:59-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. João Paulo Lima Santos
(Avaliador Interno – PPGEC/UFAL)

Documento assinado digitalmente
gov.br SERGIO LIMA NETTO
Data: 01/02/2025 12:43:47-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Sérgio Lima Netto
(Avaliador Externo à Instituição - UFRJ)

Documento assinado digitalmente
gov.br RICARDO EMANUEL VAZ VARGAS
Data: 02/02/2025 10:16:57-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Ricardo Emanuel Vaz Vargas
(Avaliador Externo à Instituição - Petrobras)

Resumo

LOPES, LUCAS G. O.. Open-World Learning Aplicado a Poços de Petróleo Utilizando Clusterização Baseada em Autoencoder. 100f. Tese – Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Alagoas. Maceió-AL, 2025.

Esta tese investiga a aplicação de Open-World Learning no monitoramento de anomalias em poços de petróleo, um aspecto crucial para a segurança e eficiência operacionais. Métodos disponiveis na literatura identificam anomalias conhecidas, mas lidar com anomalias desconhecidas em ambientes dinâmicos segue sendo um desafio. O objetivo deste trabalho é desenvolver uma abordagem de Open-World Learning que integra redução de dimensionalidade com autoencoders, classificadores binários e um método híbrido de agrupamento, visando detectar e organizar anomalias desconhecidas. A hipótese é que essa abordagem pode contribuir para a detecção e classificação de anomalias desconhecidas em dados de produção de poços de petróleo. Caso uma anomalia seja detectada pelo erro de reconstrução do autoencoder, classificadores binários avaliam se pertence a uma classe conhecida. Se não pertencer, um método de agrupamento organiza eventos similares em novas classes, validadas por especialistas humanos. Experimentos com o conjunto de dados 3W e dados reais de poços indicam que os clusters descobertos se alinham a classes conhecidas, com acurácia geral de 81%, superando 95% em certos casos, enquanto classificadores binários atualizados atingem até 99% de acurácia. Os resultados sugerem que a estratégia proposta representa um avanço no monitoramento e na detecção de anomalias, podendo, assim, contribuir significativamente para a segurança e a integridade estrutural de poços de petróleo.

Palavras-chave: Detecção de anomalias, Open-World Learning, Detecção de novidades, Representação latente, Monitoramento industrial dinâmico, Integridade estrutural de poços.

Abstract

LOPES, LUCAS G. O.. Open-World Learning Applied to Oil Wells Using Autoencoder-Based Clustering. 100f. Thesis – Graduate Program in Civil Engineering, Universidade Federal de Alagoas. Maceió-AL, 2025.

This thesis investigates the use of Open-World Learning for anomaly monitoring in oil wells, a critical aspect for operational safety and efficiency. Existing methods identify known anomalies, but handling unknown anomalies in dynamic environments remains a challenge. The objective of this work is to develop an Open-World Learning approach that integrates autoencoder-based dimensionality reduction, binary classifiers, and a hybrid clustering method to detect and organize unknown anomalies. The hypothesis is that this approach can contribute to detecting and classifying unknown anomalies in oil well production data. In the proposed strategy, anomalies are first detected through autoencoder reconstruction errors. Binary classifiers then determine whether the anomaly belongs to a known category. If it does not, a clustering method groups similar events into new categories, which are subsequently validated by human experts. Experiments with the 3W dataset and real-world oil well data indicate that the discovered clusters align with known classes, achieving 81% overall accuracy and exceeding 95% in certain cases, while updated binary classifiers reach up to 99% accuracy. The results suggest that the proposed strategy represents an advancement in monitoring and anomaly detection, thereby significantly contributing to the safety and structural integrity of oil wells.

Keywords: Anomaly detection, Open-World Learning, Novelty detection, Latent representation, Dynamic industrial monitoring, Structural integrity of oil wells.

Acknowledgements

First of all, I thank God, present in all moments of success and difficulty, always illuminating my path and blessing my choices.

To my wife, Luana Marina, whose love and wisdom have helped me reach heights I never thought I could achieve.

To my parents, Artur and Selma, for always placing their love and trust in me, providing the necessary support for my education and development as a person and professional.

To my cousins, uncles, godparents, and grandparents, who form a solid foundation in my life. Even from a distance, they are always present, offering support, affection, and motivation.

To my advisors, William Wagner and Thales Miranda, for their constant availability and attentiveness, sharing their knowledge and contributing to achieving the best results.

To all the professors who were part of my undergraduate and postgraduate journey, especially those who embraced the role of educators by sharing their knowledge in an inspiring way, thereby contributing to my professional training.

To all my friends and coworkers at the Laboratory of Scientific Computing and Visualization (LCCV), who supported me throughout my academic years.

To all my friends who stood by me through both the good and difficult moments of my academic life. A special thanks to my best friends: Edelson, Bela, Chris, Helvio, Josué, Weverton, Heidi, JP, Tiago, Igor, and all the other friends who crossed my path, making this challenging journey a little easier.

To PETROBRAS for supporting the development of projects that contributed to this work by providing data, knowledge, and assistance throughout the research process.

To SED Engenharia and FAPEAL for their support in parts of the thesis results through the Tecnova program.

Contents

List of Figures

List of Tables

1 Introduction

The oil and gas industry plays a vital role in the global energy landscape, supplying approximately 75% of the global energy demand (Zhong et al., 2020). Despite ongoing transitions toward renewable energy, fossil fuels will continue to dominate the energy mix for years to come, underscoring the need for continuous advancements in this sector. Over the decades, the industry has embraced digital transformation, integrating technologies such as artificial intelligence (AI), big data, cloud computing, automation, and the Internet of Things (IoT). These innovations are transforming operational paradigms, improving both efficiency and safety, and supporting real-time monitoring through sophisticated sensor networks that measure critical variables such as pressure, temperature, and flow (Vargas et al., 2019b).

The complexity of oil and gas operations, from exploration to production and distribution, demands meticulous attention to improve efficiency and safety. During production, wells face varying temperature and pressure conditions, which can lead to challenges such as equipment failure, hydrate formation, scale deposition, and other operational disruptions. Monitoring these systems through sensors, such as Permanent Downhole Gauges (PDGs) and Christmas Tree Temperature/Pressure Transducers (TPTs), provides invaluable data for assessing well conditions and detecting anomalies (Huffner et al., 2019). The timely detection and classification of anomalies are crucial for mitigating risks, improving operational safety, and supporting decision-making processes (Venkatasubramanian, 2003).

Machine learning (ML) and deep learning (DL) have been increasingly applied as tools for anomaly detection and predictive maintenance in oil wells. Techniques such as autoencoders, recurrent neural networks, and clustering algorithms have demonstrated their efficacy in identifying deviations from normal operational patterns (Aranha et al., 2024a; Magnusson et al., 2023). Research using the 3W Dataset (Vargas et al., 2019b), the largest labeled dataset of anomalies in oil well data, has laid the foundation for the development of robust models to address challenges such as hydrate formation (Lu et al., 2019; Monday and Odutola, 2021) and scale deposition (Yousefzadeh et al., 2022). Although these advances are significant, existing models often struggle with the detection of previously unseen anomalies, a limitation given the dynamic and evolving nature of oil well operations.

To overcome these limitations, this thesis presents an Open-World Learning (Chen and Liu, 2018) approach that combines anomaly detection with adaptive classification strategies for both known and unknown cases. The central hypothesis of this thesis

is that an Open-World Learning strategy, combining autoencoder-based dimensionality reduction, binary classifiers, and a hybrid clustering method, can enable effective detection, classification, and continuous adaptation to unknown anomalies in oil well production data, thereby improving structural integrity and operational safety. By identifying deviations in well conditions at an early stage, the approach may help mitigate risks associated with wellbore instability, equipment degradation, and pressure anomalies, which are factors in maintaining the long-term integrity of oil wells.

The proposed method first detects anomalies using autoencoder reconstruction error. If an anomaly is detected, the sensor data is transformed into a lower-dimensional representation using the autoencoder's encoder. The encoded output is then processed by binary classifiers—dense models assisted by One-Class Support Vector Machines (OCSVM)—which help distinguish between known and unknown anomaly types. If the data does not belong to a known class, it is treated as an unknown anomaly. To manage these cases, DBSCAN (Deng, 2020) and Mean Shift (Comaniciu and Meer, 2002) are used to identify noise, while a random forest model estimates the number of potential new anomaly classes. Using this estimate, a combination of K-means and Mean Shift clustering is applied to group similar data into potential new anomaly classes. These newly identified classes undergo human validation to refine cluster assignments, register new anomaly types, and update binary classifiers, enabling continuous learning.

Additionally, this work introduces and validates the entire anomaly detection pipeline on a new type of anomaly: the downhole interval control valve (ICV) fault. This dataset has not been previously used to validate Open-World Learning in oil well monitoring, making it a valuable test case for evaluating the adaptability and effectiveness of the proposed approach in real-world scenarios.

The contributions of this work extend beyond individual anomaly types, offering a unified strategy for detecting, classifying, and learning to recognize new anomalies. The methodologies presented can be applied across various operational contexts, including production wells, injection wells, and carbon capture, utilization, and storage (CCUS) systems. By overcoming the limitations of machine learning approaches and introducing a dynamic strategy for continuous learning, this thesis advances anomaly detection and enhances operational safety in the oil and gas industry.

1.1 Related Work

The oil and gas industry has undergone significant advancements with the integration of artificial intelligence (AI) and machine learning (ML) techniques. This thesis builds upon previous research efforts, leveraging methodologies detailed in the following subsections to address anomaly detection and classification challenges in oil well opera-

tions. These studies form the foundation for the development of the Open-World Learning strategy presented in this work.

AI Applications in Oil and Gas Operations

Artificial intelligence applications have been extensively used to optimize operations and improve decision-making in the oil and gas industry. In drilling operations, for instance, Gurina et al. (2020) introduced an anomaly detection methodology for directional drilling, while Anifowose et al. (2019) compared other ML techniques to predict reservoir permeability, emphasizing the importance of AI in characterizing subsurface formations. Recent advances have extended these applications to broader operational contexts. For instance, D'Almeida et al. (2022) and Shu et al. (2022) applied predictive modeling techniques to analyze petroleum reservoirs, showcasing the potential of AI in resource estimation and planning.

The transformative potential of AI is further illustrated in real-time monitoring strategies. For instance, Helmy et al. (2010) explored AI-driven reservoir properties forecasting and Guilherme et al. (2011) enhanced monitoring during the drilling process. Reinforcement Learning (RL) approaches, such as those demonstrated by Hourfar et al. (2019), optimized water-flooding operations, improving resource utilization and recovery efficiency. Neural networks have also been applied to study casing-heading instability, as demonstrated by Salahshoor et al. (2013), contributing to safer well operations under varying pressure and temperature conditions.

The application of ML techniques extends beyond drilling and reservoir management to broader production challenges. Qiao et al. (2023) proposed an improved Bidirectional Gated Recurrent Unit (BiGRU) neural network algorithm with an attention mechanism to classify and identify seven common drilling conditions. Their model achieved an average accuracy of 91.63% and demonstrated strong generalization capabilities, underscoring the potential of neural networks to improve operational safety. Similarly, Pang et al. (2024) developed pre-drilling mud loss prediction solutions to mitigate risks in complex drilling environments, further contributing to efficient production planning.

Anomaly Detection in Oil Well Operations

In the production stage, wells are subjected to various temperature and pressure conditions, which can cause a variety of issues, including equipment failures and operational interruptions. Monitoring strategies, such as PDGs and TPTs, have become critical to tracking well conditions and detecting anomalies (Huffner et al., 2019). These sensors provide data that can be analyzed using artificial intelligence techniques to prevent potential failures and improve efficient production.

Several studies have explored the application of AI for anomaly detection in production strategies. For instance, Monday and Odutola (2021) addressed hydrate formation

detection using supervised and unsupervised learning, while Yousefzadeh et al. (2022) compared ten different intelligent algorithms for scale deposition prediction, reporting accuracies exceeding 90%. Similarly, Gjelsvik et al. (2023) conducted a systematic review of intelligent techniques to address hydrate-related challenges, reflecting the growing reliance on ML in solving flow assurance problems.

Deep learning has emerged as a promissing approach for anomaly detection. Autoencoders have been used to identify anomalies by analyzing reconstruction errors, as highlighted by Aranha et al. (2024a). Other deep learning architectures, such as Long Short-Term Memory (LSTM) networks, have proven effective in forecasting gas production and identifying operational anomalies in mature wells (Loh et al., 2018). Shahani et al. (2022) extended the application of RNN and LSTM models to predict drilling rate indices for rock engineering, demonstrating the versatility of deep learning in diverse industrial contexts.

Pipeline monitoring and leak detection represent another important application area. Li et al. (2023) developed an integrated method that combines singular value decomposition (SVD), variational mode decomposition (VMD) and probabilistic neural networks (PNN), achieving high precision in identifying leak signals. These advancements underscore the adaptability of AI in addressing the complex challenges of oil and gas operations.

Additionally, recent studies have introduced innovative solutions for improving anomaly detection in oil well operations. Aslam et al. (2022) utilized machine learning techniques, including random forest and decision trees, to improve anomaly detection speed and accuracy in offshore oil wells, achieving remarkable performance. Fernandes et al. (2024) conducted a comparative study of one-class classifiers, finding that the Local Outlier Factor (LOF) classifier consistently outperformed others in detecting faults in offshore oil wells. Finally, Gao et al. (2024) presented a comprehensive approach for enhancing anomaly detection in low-quality and imbalanced data, proposing methods to improve both the accuracy and interpretability of detection models.

The Role of the 3W Dataset

The 3W Dataset (Vargas et al., 2019b), the first and largest labeled anomaly dataset for oil wells, has been very important for research in anomaly detection and predictive maintenance. It has facilitated the development of various ML models, including anomaly detectors (Vargas et al., 2019a), event prediction strategies (Li et al., 2020), and visual analytics tools (Soriano-Vargas et al., 2021). Marins et al. (2021b) and Aranha et al. (2024b) employed the dataset to develop fault detection strategies, while Carvalho et al. (2021) used it for classifying multivariate time series data. Unsupervised learning approaches, such as clustering algorithms, have also leveraged the dataset to classify offshore well data into normal and anomalous categories (de Salvo Castro et al., 2021).

Recent studies have expanded the applications of the 3W Dataset. For instance, Chen et al. (2023) introduced a federated learning model for the identification of the oil layer, while Magnusson et al. (2023) explored the use of recurrent neural networks for the detection of anomalies in real time. Gatta et al. (2022) employed convolutional 1D autoencoders to extract features from multivariate time series data, showcasing the dataset's versatility in addressing predictive maintenance challenges.

Novelty Detection and Open-World Learning

Despite significant advancements, ML approaches face limitations when encountering previously unseen anomalies. Most methods rely on labeled datasets and predefined anomaly classes, making them less effective in dynamic environments where new operational conditions emerge. Novelty detection provides a pathway to address this challenge by identifying deviations from known patterns. Techniques such as clustering algorithms (Pimentel et al., 2014) and null space projection (Bodesheim et al., 2013) have been successfully applied to detect and group unknown anomalies for further classification.

Open-World Learning (OWL) extends the principles of novelty detection by enabling strategies to continuously adapt to new anomaly classes. OWL strategies begin with a set of known classes, detect unknown instances, and incrementally learn new classes from dynamic data streams (Parmar et al., 2022; Jafarzadeh et al., 2022; Zhu et al., 2024a). This paradigm offers a comprehensive approach to anomaly detection and classification in evolving environments, addressing the limitations of closed-world ML models.

Recent advances, such as the ORALI strategy (Zhu et al., 2024b) for lithology identification, demonstrate the potential of open-set recognition combined with active learning to address challenges in open-world settings. ORALI utilizes an Open-Set Nearest Neighbor (OSNN) approach to identify and reject instances of unknown classes effectively. Once unknown classes are flagged, expert labeling is employed to accurately classify them, and active learning is used to prioritize and query the most informative instances for labeling. This iterative process enables ORALI to incrementally integrate new classes and adapt to dynamic environments.

Contributions to Anomaly Detection in Oil Wells

Building on these developments, I introduce a novel Open-World Learning strategy designed to detect and classify anomalies in oil wells. The methodology integrates autoencoders for dimensionality reduction with clustering algorithms to identify previously unclassified anomalies. Using the 3W Dataset, this approach dynamically discovers new types of anomalies and continuously adapts to changing operational conditions. The inclusion of a validation step improves the accurate registration of new anomalies and the retraining of the classifiers for future monitoring.

The proposed strategy is validated using novel data on Downhole Interval Control

Valve (ICV) faults, representing a previously unexplored anomaly in oil operations. This validation, along with real-world applications for hydrate formation detection and multivariate anomaly classification, demonstrates the effectiveness of the strategy in addressing practical challenge

1.2 Hypothesis

This thesis hypothesizes that an Open-World Learning strategy integrating autoencoder-based dimensionality reduction, binary classification, and hybrid clustering can effectively detect, classify, and continuously adapt to unknown anomalies in oil well production data. By leveraging autoencoder reconstruction errors for anomaly detection, binary classifiers for distinguishing known and unknown anomalies, and clustering techniques for organizing novel anomaly patterns, the proposed approach enables a dynamic and adaptive monitoring system. The ability to discover and classify previously unseen anomalies is expected to enhance early anomaly detection, support operational decision-making, and contribute to the long-term structural integrity and safety of oil wells. Furthermore, the introduction of a human-in-the-loop validation step ensures the meaningful integration of new anomaly categories, facilitating continuous learning and improving the robustness of the monitoring framework.

1.3 Objectives

The main objective of this thesis is to develop an Open-World Learning strategy for anomaly detection and classification of novelty and anomaly in oil well operations. This strategy addresses the challenge of identifying and classifying previously unknown anomalies, based on modular anomaly detection and novelty clustering research to create a comprehensive strategy capable of adapting to evolving operational conditions. The proposed approach integrates autoencoder-based dimensionality reduction, binary classifiers, and clustering techniques to detect, organize, and classify unexpected events in oil well data. The methodology introduces a human-in-the-loop validation step to incorporate newly discovered anomaly classes, ensuring continuous improvement of the classification strategy.

The development and validation of the Open-World Learning strategy are divided into the following steps, highlighting contributions from each supporting Chapter:

1. **Anomaly Detection:** - Implement and evaluate anomaly detection using autoencoder reconstruction error to identify normal and anomalous data. *(Presented in Chapter 2: "A Strategy to Detect Oilwell Anomalies Using Deep Learning and Decision Diagram Dual Approach")*

2. **Anomaly Classification:** - Develop LSTM-based methods for classifying anomalies. *(Presented in Chapter 3: "Use of LSTM Networks for Anomaly Classification in Oil Wells")*

3. **Open-World Learning:** Design and implement a strategy that combines dimensionality reduction, clustering, and binary classification for anomaly detection and classification in oil wells.

   - Use autoencoder-based latent representations to identify unknown anomalies.

   - Apply clustering techniques to group similar unknown anomalies and define new classes.

   - Incorporate a human validation step to confirm new anomaly classes and retrain binary classifiers for continuous adaptation.

   (Presented in Chapter 4: "Open-World Learning Applied to Oil Wells Using Autoencoder-Based Clustering.")

1.4 Methodology

This thesis proposes an Open-World Learning strategy for the detection and classification of anomalies in oil well operations. The methodology is designed to dynamically handle both known and unknown anomalies, enabling the strategy to adapt to new operational conditions. The process follows a cyclical flow, as illustrated in Figure 1.1, and is divided into six key steps: data preprocessing, anomaly detection, binary classification, clustering of unknown anomalies, human validation, and continuous learning.

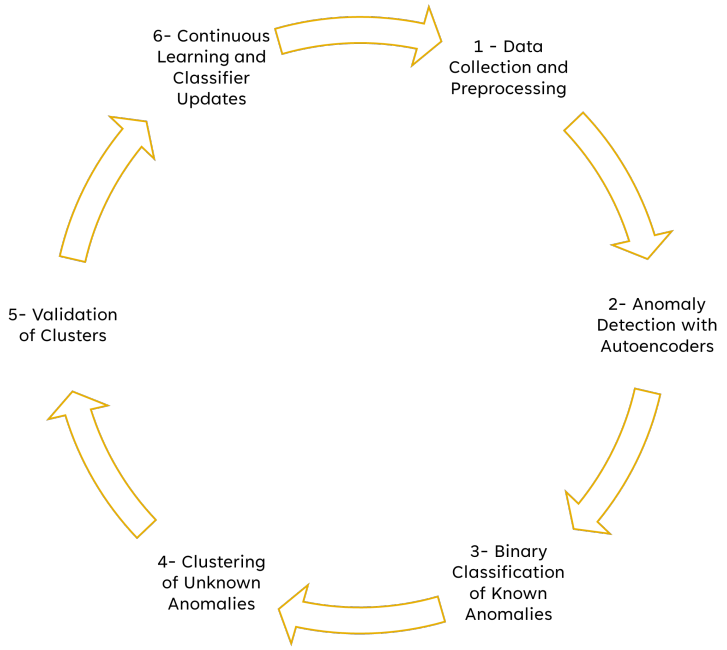
Figure 1.1 represents a cyclic process where data flows through each step iteratively. Starting from data collection and preprocessing, anomalies are detected using autoencoders. Known anomalies are classified by binary classifiers, whereas unknown anomalies are clustered and validated by human operators. The validated clusters are used to update the classifiers, improving the strategy's adaptability.

1.4.1 Step 1: Data Collection and Preprocessing

The strategy begins with the collection and preprocessing of oil well sensor data. The primary dataset is the publicly available 3W Dataset (Vargas et al., 2019b), complemented by proprietary data for specific validation scenarios. The data includes variables such as temperature, pressure, and flow rates, sampled at one-second intervals.

Preprocessing involves: Synchronizing time-series data from multiple sensors; Grouping data based on valve states or operational conditions; Conducting exploratory data analysis to identify patterns and validate data integrity.

Figure 1.1 – Cyclic methodology for Open-World Learning in oil well operations.



Source: Author (2025)

This step improves that the dataset is structured and prepared for downstream anomaly detection and classification tasks. In the context of data analysis, the impact of anomalies in oil well operations can lead to significant changes in loading conditions, impacting also the well structure. Typically, such probabilistic loads exhibit behavior akin to normally-distributed variables, making it feasible to compute probabilities and determine extreme values. Table 1.1 shows the load distribution changes for the 3W Dataset anomalies.

Table 1.1 – Probabilistic distribution for extreme loading: extreme relative variation (%)

	Mean ± Standard Deviation		
ID	P-PDG	P-TPT	T-TPT
1	19.90±16.20	23.20 ± 21.20	-18.50±18.90
2	10.20±10.20	3.40±5.29	-86.80±12.10
3	4.61±6.72	9.70±12.31	-11.60±16.07
4	0.06±0.16	0.99±1.29	-0.61±0.94
5	0.31±2.85	0.72±2.85	-2.62±7.78
6	0.00±3.77	3.55±3.77	-0.09±0.04
7	0.33±1.38	1.34±1.38	-0.44±0.45
8	20.20±18.20	43.51±18.20	-95.10±10.80

There are some insights that is possible to obtain from the loading profile, for instance that the hydrate formation seems to have a larger spread on the loading change during the event. Other conclusion is that each abnormal event have a very unique

signature, making it difficult to create analytical rules to identify and detect each event. Cases with a hard analytical approach are often interesting applications for Deep Learning.

1.4.2 Step 2: Anomaly Detection Autoencoders

An autoencoder is employed to detect anomalies by learning the normal operational patterns of the oil well data. The model compresses input data into a latent representation and reconstructs it. Anomalies are identified based on the reconstruction error: data with high reconstruction error values is flagged as anomalous.

The autoencoder is trained exclusively on normal data, making it effective in identifying deviations that represent anomalies. This step serves as the foundation for both known and unknown anomaly handling.

1.4.3 Step 3: Binary Classification of Known Anomalies

For anomalies previously observed in the dataset, binary classifiers are developed and trained. These classifiers categorize known anomalies into specific types, such as hydrate formation or equipment malfunctions. The classifiers are trained using labeled data from the 3W Dataset and evaluated using standard metrics such as accuracy, precision, and recall.

The anomaly of hydrate formation, for instance, is handled with a specialized DNN binary classifier that is optimized using hyperparameter tuning. This improves good performance in distinguishing hydrate-related anomalies from other operational conditions.

1.4.4 Step 4: Clustering of Unknown Anomalies

For anomalies not recognized by binary classifiers (novelty), clustering techniques are applied to group unknown anomalies into potential new classes. This process utilizes the latent space representations produced by the autoencoder. Techniques such as k-means and mean-shift clustering are used to organize similar unknown anomalies into clusters.

The number of clusters is dynamically estimated based on the characteristics of the latent space, ensuring flexibility in handling diverse and evolving anomalies.

1.4.5 Step 5: Validation of Clusters

Once clusters of unknown anomalies are formed, validation is performed to confirm their classification or adjust cluster assignments. This human-in-the-loop process is essential to improve that newly discovered anomaly classes are meaningful and relevant to operational safety.

Validated clusters are registered as new anomaly classes, and this information is used to update the strategy's classifiers.

1.4.6 Step 6: Continuous Learning and Classifier Updates

The final step in the cycle is the continuous learning process. After human validation, the binary classifiers are re-trained to include the newly discovered anomaly classes. This improves that the strategy evolves and remains effective as new operational conditions and anomalies emerge.

The updated classifiers are then deployed, completing the cycle and preparing the strategy for the next iteration of data collection, anomaly detection, and classification.

The steps outlined above are integrated into a unified cyclic strategy, as shown in Figure 1.1. The iterative nature of the process ensures that the strategy continuously adapts to changes in the oil well environment, providing anomaly detection and classification capabilities.

1.4.7 Validation of the Strategy

The strategy is validated using the 3W Dataset and proprietary data, focusing on key evaluation metrics to improve its effectiveness in detecting and classifying both known and unknown anomalies:

Reconstruction Error: Used as a primary metric to distinguish normal data from anomalous data. A higher reconstruction error indicates deviations from normal operational patterns, enabling the detection of potential anomalies.

Clustering Accuracy: Assesses the alignment between discovered clusters of unknown anomalies and ground-truth classes, reaching a meaningful grouping of previously unseen anomaly types.

Binary Classifier Performance: Evaluated using metrics such as accuracy, precision, recall, and F1-score to measure the effectiveness of classifying known anomalies.

The validation processes, presented in the result section of each Chapter, demonstrates the applicability of the Open-World Learning strategy, improving how it effectively handles evolving operational scenarios in oil well environments. The results highlight the strategy's ability to adapt to new anomaly types while maintaining high classification performance for known anomalies.

1.5 Delimitation

This section outlines the key limitations, assumptions, and scope of the research

presented in this thesis, defining the boundaries and focus of the proposed methodology.

1.5.1 Scope of the Study

This doctoral thesis focuses on the development of an Open-World Learning strategy for anomaly detection and classification in oil well operations, leveraging data from temperature and pressure sensors. The methodology integrates autoencoder-based anomaly detection, binary classification of known anomalies, and clustering techniques to handle previously unseen anomalies. While the proposed strategy has been designed and validated using real and synthetic datasets, certain limitations are inherent to the scope of this research.

1.5.2 Datasets and Validation Approach

The methodology is trained on the 3W Dataset 2.0, validated using 40% of real instances from the same dataset, and tested on a newly introduced downhole interval control valve (ICV) fault. All training, validation, and testing are performed at the sample level. As a result, the effectiveness of the methodology is inherently tied to the characteristics and quality of this dataset, and performance comparisons are made with studies in the same domain. While the strategy is optimized for oil production anomaly detection, applications in other sensor domains, such as manufacturing, healthcare, or transportation, remain unexplored.

1.5.3 Data Representation and Evaluation Levels

This thesis uses two levels of data: instances and samples. The instance level corresponds to an event from the 3W Dataset or ICV fault data, represented by an entire file. The sample level represents a segment of an instance, corresponding to a constrained time series with a fixed length. In other words, a single instance can generate multiple samples. Training, validation, and testing are primarily conducted and evaluated at the sample level. However, instance-level accuracy is also assessed in Chapter 4.

1.5.4 Sensor Data and Limitations

The methodology assumes that the sensor data quality is sufficient for anomaly detection tasks. Issues related to sensor calibration, noise, or missing data are not explicitly addressed, though preprocessing techniques are employed to improve data integrity. Future studies could incorporate more robust preprocessing strategies or adaptive methods to handle such challenges systematically.

This work primarily focuses on pressure downhole gauge (PDG) and temperature and pressure transducer (TPT) sensors, as they are widely used in oil well operations.

Other sensor types, such as flow-rate or vibration sensors, are not explicitly considered. Expanding the methodology to integrate data from a broader range of sensors may further enhance its adaptability and performance. Additionally, while the methodology is validated using publicly available datasets and proprietary Petrobras data, its applicability to land-based wells or other industries requires further study and validation.

1.5.5 Human-in-the-Loop and Open-World Learning Considerations

The Open-World Learning strategy incorporates a human-in-the-loop validation step for clustering unknown anomalies. While this validation is critical for ensuring the meaningful integration of new anomaly classes, it introduces a degree of subjectivity and may not scale seamlessly in real-time operational settings with frequent anomaly occurrences.

The approach allows for the creation of virtual classes with temporary labels, enabling the automatic classification of unknown anomalies before validation is completed. However, bypassing human validation may lead to fragmentation, where a single anomaly is divided into multiple clusters, reducing interpretability. While the validation process is primarily required to assign meaningful labels to newly proposed classes, the proposed approach ensures continuous learning in real-time, even in the absence of immediate validation. Future research should explore automating this process to enhance scalability while maintaining classification accuracy.

1.5.6 Statistical Considerations

Unlike other machine learning studies that emphasize statistical significance testing, this thesis does not include formal statistical significance analyses. However, the robustness of the methodology is illustrated through the presentation of mean and standard deviation values across hundreds of experimental runs, offering insight into the variability and expected range of performance. While this approach provides empirical validation, further studies could conduct statistical hypothesis testing to strengthen confidence in the methodology's effectiveness.

1.5.7 Computational Resources

All GPU-based training experiments were conducted on a system equipped with an Intel Xeon CPU with 2 vCPUs (virtual CPUs), 13GB of RAM, and an NVIDIA T4 GPU. The computational constraints of this hardware setup may have influenced model training efficiency, hyperparameter optimization, and overall performance. Future research could explore the impact of scaling computational resources on model performance, particularly

for real-time or large-scale industrial applications. Chapter 4 presents the main results regarding computational time for different parts of the Open-World Methodology.

1.5.8 Time-Series Data Assumptions

Lastly, the proposed strategy is specifically designed for time-series data. Applications involving non-sequential or mixed data types fall outside the scope of this thesis. Despite these delimitations, the contributions of this research provide a foundation for real-time anomaly detection and classification in oil well operations, with the flexibility to adapt to evolving operational conditions.

1.6 Document Structure

The thesis document is organized into the following Chapters.

- **Chapter 1 - Introduction**: Provides the context of the problem, objectives, and an overview of the methodology.
- **Chapter 2 - Dual Strategy for General Anomaly Detection**: Discusses anomaly detection using autoencoders and decision tree approaches. The results of this methodology are presented in the published paper (Aranha et al., 2024a).
- **Chapter 3 - Anomaly Classification Focused on Binary Classifiers**: Details the development and optimization of binary classifiers for specific anomalies. These results are presented in the article submitted, which is already available as a preprint (Lopes et al., 2024).
- **Chapter 4 - Open-World Learning Strategy**: Presents the complete strategy, integrating anomaly detection, clustering, and human validation. The results have been submitted and have not yet been published.
- **Chapter 5 - Final Considerations**: Summarizes the findings and suggests future research directions.

2 A System to Detect Oilwell Anomalies Using Deep Learning and Decision Diagram Dual Approach

This Chapter presents the development of a strategy for detecting anomalies in oil wells using deep learning techniques combined with a decision diagram approach. The approach applies autoencoders to analyze real-time data and determine whether it corresponds to normal operational conditions.

Autoencoder reconstruction error is leveraged to identify anomalous behaviors, while decision diagrams, specifically decision trees, complement the classification process. This integration enhances anomaly identification by combining unsupervised learning with interpretable decision-making techniques.

The results and methodology presented in this chapter were published as the paper Aranha et al. (2024a) in the SPE Journal.

2.1 Overview

The detection of anomalies plays a critical role in ensuring the safety and efficiency of oil well operations. Given the complexity of oil production systems, sensors such as PDGs and TPTs are commonly used to monitor key variables such as pressure and temperature (Huffner et al., 2019). Anomalous events, whether caused by equipment malfunctions, unexpected conditions, or operational inefficiencies, can disrupt production and lead to significant economic and environmental consequences. Therefore, reliable and efficient anomaly detection systems are essential.

Traditional approaches to anomaly detection in oil wells include rule-based systems, decision trees (DT), and ensemble methods such as Random Forests (RF) (Marins et al., 2021a; Al-Hajri et al., 2020). These methods often struggle with the high dimensionality and sequential nature of time series data generated by oil well sensors. Deep learning techniques, particularly Long Short-Term Memory (LSTM) networks, have emerged as powerful tools for handling sequential data and capturing long-term dependencies, making them highly suitable for analyzing complex datasets such as those collected from oil wells (Malhotra et al., 2015; Machado et al., 2022).

Autoencoders, a type of unsupervised neural network, have been widely used for anomaly detection by learning the normal patterns in data and flagging deviations as

potential anomalies. These methods have proven to be effective in scenarios where labeled anomaly data is scarce, as they rely exclusively on normal data for training (Marchi et al., 2015). Recent works, such as Machado et al. (2022), have demonstrated the application of LSTM-based autoencoders to detect hydrate formation and spurious valve closures in oil wells, leveraging their ability to accurately reconstruct time series data. However, these methods often benefit from complementary techniques to enhance their detection accuracy and robustness.

Rule-based analytical approaches, while simple, can be used to enhance anomaly detection by providing interpretable thresholds or conditions that align with domain knowledge. Figueiredo et al. (2021) compared multiple unsupervised models and highlighted the effectiveness of hybrid approaches that combine machine learning with rule-based systems to detect specific production anomalies. Similarly, Turan and Jaschke (2021) applied DT and other classifiers to oil production anomalies, demonstrating that explainability remains a valuable attribute in operational settings.

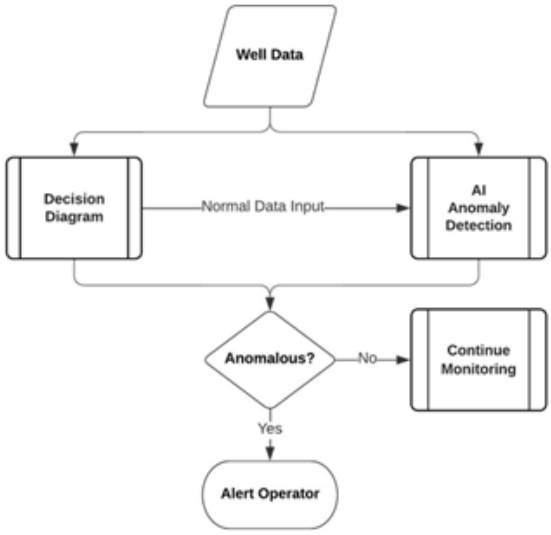
This chapter presents a methodology that integrates LSTM autoencoders with a rule-based analytical approach for real-time anomaly detection in subsea well sensor data. The autoencoder identifies deviations from normal operational patterns using reconstruction error, while the rule-based system adds interpretability and domain-specific insights. The methodology was validated on pressure and temperature data from subsea well sensors, demonstrating its ability to detect both mapped and previously unseen anomalies. This combined approach offers a balance between the robustness of deep learning and the transparency of rule-based systems, making it suitable for deployment in real-time oil well monitoring applications.

2.2 Proposed Methods and Studies

Sensors installed to monitor production in oil wells are located at a few viable positions to monitor the behavior of a well during its operation and regularly transmit data concerning pressure and/or temperature values along the depth. With adequate treatment of these data, followed by pattern identification techniques, it becomes possible to evaluate whether the well operates within normality or not, allowing operators and engineers to be capable of classifying normal and abnormal situations by analyzing the time series. In the anomaly detection system, there is an entry point for sensor data, from which checks are performed with the association of a Decision Diagram and a deep learning network, as illustrated in Figure 2.1. The system can map valve configurations and set its anomaly search mechanisms to suit the current valve status. The methodology is developed in such a way that the system is scalable to monitor simultaneously multiple wells. The anomalies detected by the DD can be classified as mapped and unmapped. For the first case, the

tool has a portfolio of anomalies that contains the attributes of undesirable scenarios in terms of pressure and temperature. The portfolio was designed based on the expected behavior of the well for a given valve configuration. Therefore, the attributes of each scenario are defined by rules and tolerances. As for the second group of anomalies, these are the cases that have not yet been mapped in the system but configure a scenario that does not correspond to the normality of the well in question, being detected due to the behavior patterns that the system attributes to the well for both analytical DD and deep learning approaches.

Figure 2.1 – General methodology of the proposed dual process.



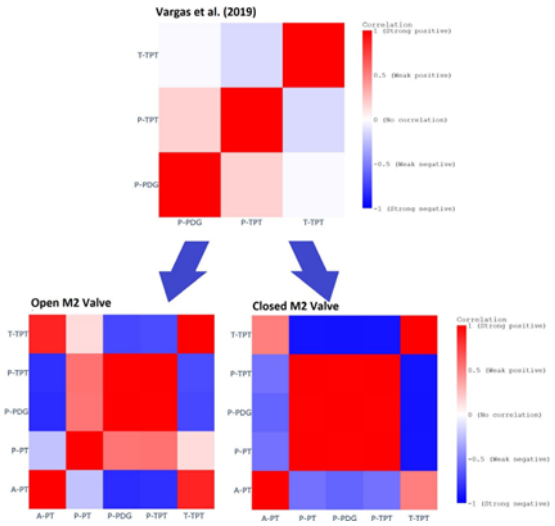
Source: Aranha et al. (2024a)

The anomalies detected by the decision tree can be classified as mapped and unmapped. For the first case, the tool has a portfolio of anomalies that contains the attributes of undesirable scenarios in terms of pressure and temperature. The portfolio was designed based on the expected behavior of the well for a given valve configuration. Therefore, the attributes of each scenario are defined by rules and tolerances. As for the second group of anomalies, these are the cases that have not yet been mapped in the system, but configure a scenario that does not correspond to the normality of the well in question, being detected due to the behavior patterns that the system attributes to the well for both analytical decision tree and deep learning approaches.

Analysis of pressure and temperature sensor data One of the steps in data processing is grouping and synchronizing the pressure and temperature sensor data provided by the oil company. The time series of each sensor are stored in files that contain information collected in different time intervals. The data were synchronized and grouped based on combinations of valve states in production. A first analysis took into account the two main well valve combinations as output, which differ from each other by the open or closed

state of the Master-2 (M2) valve. Graphs and correlation studies are generated to identify behaviors of well operation that are not normal. Figure 3.2 shows observations of the correlation between pressure and temperature values (PT, TPT, and PDG) for both data from the literature provided by (Vargas et al., 2019b) and data from industry provided for the open and closed M2 valve.

Figure 2.2 – Correlation matrices for sensor data.



Source: Aranha et al. (2024a)

The red color indicates a positive correlation (directly related variables), and the blue color a negative correlation (inversely related variables). This study indicates that different valve status configurations should be investigated and taken into account during the development of anomaly detection methodologies. Therefore, it is a qualitative study that indicates the importance of evaluating detection methodologies that take into account different valve states, as they influence the correlation of variables.

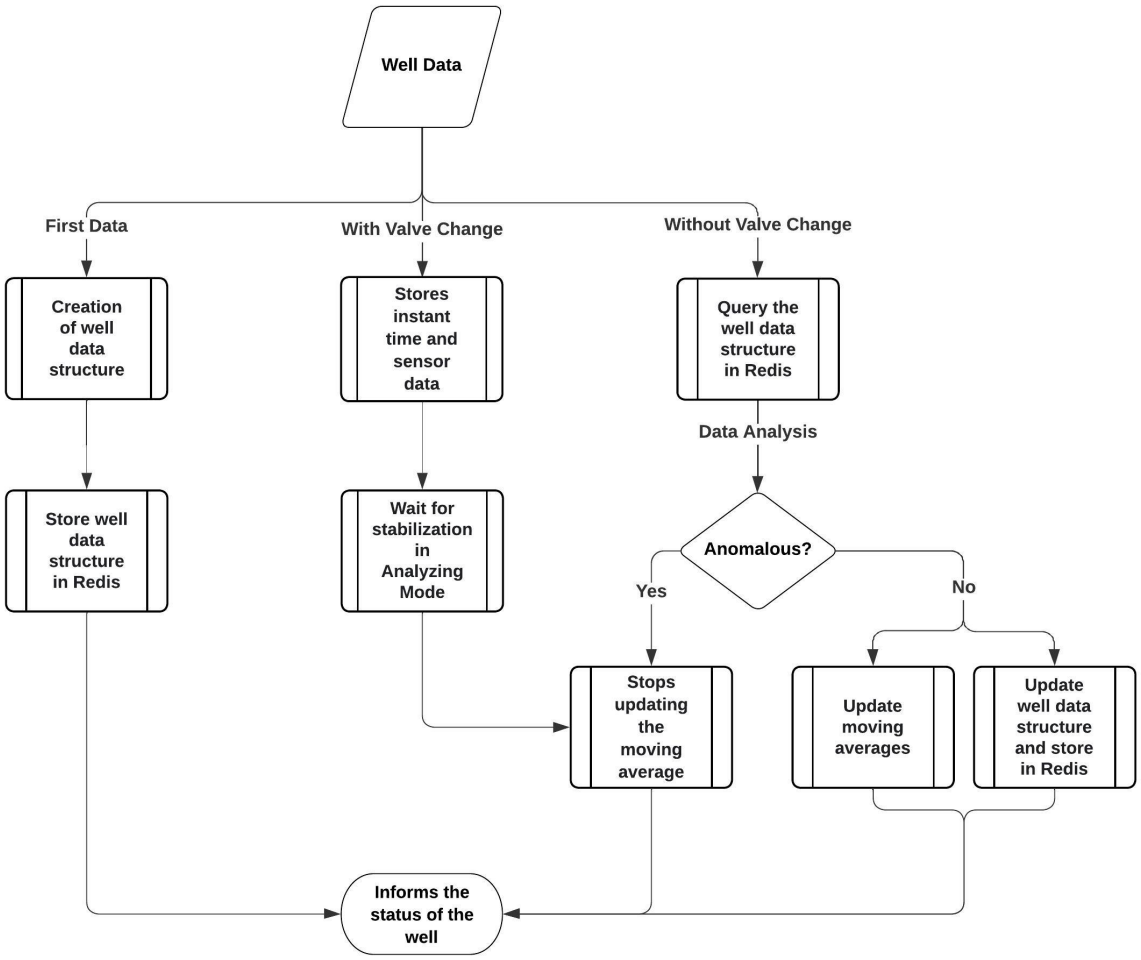
2.2.1 Anomaly Detection Using the Decision Digram

The methodology that defines an action to be taken by the decision tree consists of structuring a set of rules based on layers with nodes. Each node verifies a condition that determines the next node to which the information will be passed until a leaf node. As decisions advance through the internal nodes, the data become more and more specific; therefore, the decision tree can manage the information in such a way as to extract the maximum benefit from them to indicate the occurrence of anomalies.

The first layer of the decision tree refers to the data of the monitored well, given the tool's ability to simultaneously track multiple wells. In the next layer, the type of well data is checked, whether it is the first data, whether valve change has occurred or not. If these are the first well data, only the well data structure will be created and stored in the

database. If there has been any change in the operating valve state, the system goes into standby mode until the pressure and temperature levels normalize. When the valve state remains the same, the well behavior is performed to identify possible anomalies. Then the state of the well is reported. Figure 2.3 displays the structure of the analytical decision tree.

Figure 2.3 – Structure of the Decision Diagram.



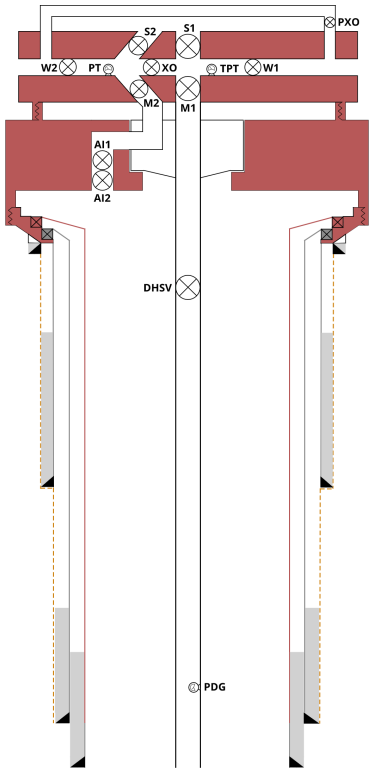
Source: Aranha et al. (2024a)

The flow chart of Figure 2.3 illustrates the decision-making process to monitor and analyze the well data. The initial step involves identifying the type of data received: i.e., first data from the well, data associated with a valve state change, or data without a valve change. For the first data, the system creates a data structure in a well and stores it in the database. If a valve change is detected, the system temporarily halts operations, entering a stabilization phase where pressure and temperature levels are allowed to normalize prior to further analysis. For data without valve changes, the system queries the existing well data structure and performs an analysis to detect potential anomalies. If anomalies are

detected, the system stops updating the moving averages. Otherwise, it updates the moving averages and stores the updated structure in the database. The well status is then reported at the end of each process. This hierarchical workflow ensures efficient handling of well data while prioritizing anomaly detection and stabilization.

Data concerning pressure and temperature sensors PDG, TPT, and PT are first analyzed together with the state of the monitored valves, DHSV, M1, M2, W1, W2, XO, PXO and the choke valve, to develop the proposed methodology. The typical valve and sensor scheme for a subsea oil well is described in Figure 2.4. The data history is collected from wells and production vessels (FPSOs) in several oil basins. Initial studies are conducted to determine the behavior and correlations from the pressure and temperature data for the most common combinations of valve states. An analytical rule-based approach is then developed to detect abnormal variations in the data, serving as a basis for classifying the anomaly based on a known portfolio.

Figure 2.4 – Oil well valves and sensors scheme.



Source: Aranha et al. (2024a)

The response of this analytical system is applied to the training of neural networks, which is also dedicated to anomaly detection. The trained models receive real-time data and monitor valve behavior using load conditions that can indicate abnormal events during oil production based on the current valve scheme. Multiple data sets from the literature (Vargas et al., 2019b) were used to train and validate the proposed methodology.

After validating the joint application of analytical and machine learning methodologies,

an interface is developed using the Node-Red platform coupled with the Python 3.7 code. This tool can receive data from different wells and perform simultaneous monitoring using advanced methods.

2.2.2 Anomaly Detection from Machine Learning

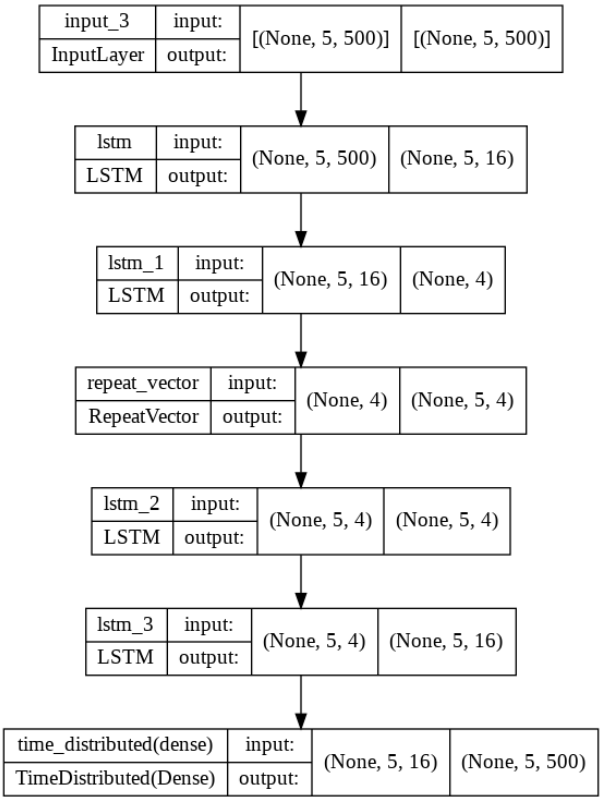
An important challenge for anomaly detection in the oil industry is that anomaly data are scarce, so it is difficult to train an AI model accurately enough to classify anomalies. So, the proposed methodology involves automatically learning what fits as a normal operation state, which happens most of the time, and constantly checking if the new data match the previously trained. Once the system has learned what is considered normal, it can alert the operator when it detects an abnormal state. This means that for each oil well, the network must collect data and learn on its own without requiring operator interference. In this sense, two modules are developed, one for learning and the other for detection. The learning module stores sequential temporal data classified as normal by the DD formulation. The module enters the training mode after acquiring approximately 500 normal data points, which covers roughly 5,000 seconds of real-time production, an amount determined through tests using industry data. Generates three outputs: (1) data normalization parameters, (2) neural network weights, and (3) anomaly tolerance calculated during training. After saving these three outputs, the detection module starts operating, detecting abnormal behaviors in newly acquired data. Networks based on different configurations were tested and the LSTM-based autoencoder proposed by (Larzalere, 2019) had the best results.

The LSTM autoencoder input consists of multivariate time series data capturing temperature and pressure readings from various sensors. Given that the model is an autoencoder, it is trained to reconstruct the input time series after passing through several bottleneck hidden layers, which first encode the original data in a lower-dimension space and then decode the original data. These reconstructed data are compared with the original input, yielding the reconstruction error. This error is then measured against a predefined threshold set during training. Consequently, the LSTM output categorizes the result as Normal or Anomalous.

Figure 2.5 shows the network developed, in which None refers to the batch size that is not predefined on the model constructor, and ReLU refers to the rectified linear activation function. The LSTM features consist of pressure values from the TPT, PT, and PDG sensors, as well as temperature values from TPT and PDG.

Figure 2.5 illustrates the architecture of the developed autoencoder model, designed to process multivariate time series data. The input layer receives sequences with a shape of (no, 5, 500), where none represents the undefined batch size, 5 is the sequence length, and 500 is the dimensionality of the features. The first LSTM layer reduces the

Figure 2.5 – Network based on LSTM architecture.



Source: Aranha et al. (2024a)

feature dimensionality to 16, followed by a second LSTM layer that further condenses the representation to 4 dimensions. The RepeatVector layer reshapes the output to match the sequence length required for reconstruction. The subsequent LSTM layers progressively reconstruct the data, expanding them back to their original feature dimensionality. The final TimeDistributed dense layer outputs the reconstructed data with a shape of (None, 5, 500). This architecture leverages the temporal relationships and patterns in pressure and temperature values from the TPT, PT, and PDG sensors to encode and reconstruct the input data effectively. The ReLU activation function is utilized to introduce non-linearity within the network.

The Adam optimizer (Kingma and Ba, 2014), available in the Keras framework (Chollet, 2015), is used to calculate the network weights. The Adam algorithm improves the classical stochastic gradient descent procedure for optimization. Regarding the error metric, mean absolute error and symmetric mean absolute percentage error (SMAPE) were applied and obtained similar results, leading to the choice of SMAPE as the main error metric, which is given by:

$$\text{SMAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|F_i - A_i|}{\frac{|A_i|+|F_i|}{2}} \tag{2.1}$$

where: n: Total number of observations; F_i: Forecast value; A_i: Actual value; $|.|$: Absolute

value.

In the studies, the number of epochs is 500 with a batch size of 100. Using a similar approach, the deep network training uses 500 samples of data labeled as normal in the DD. After training, the SMAPE value for each prediction is calculated, then approaching this error metric as a normal distribution of values. Assuming that the reconstruction error follows a Gaussian distribution, the tolerance is evaluated by Eq. 2, in terms of the mean of the parameter and its standard deviation, defined as μ and , respectively:

$$tol_{IA} = \mu + k * \sigma, \tag{2.2}$$

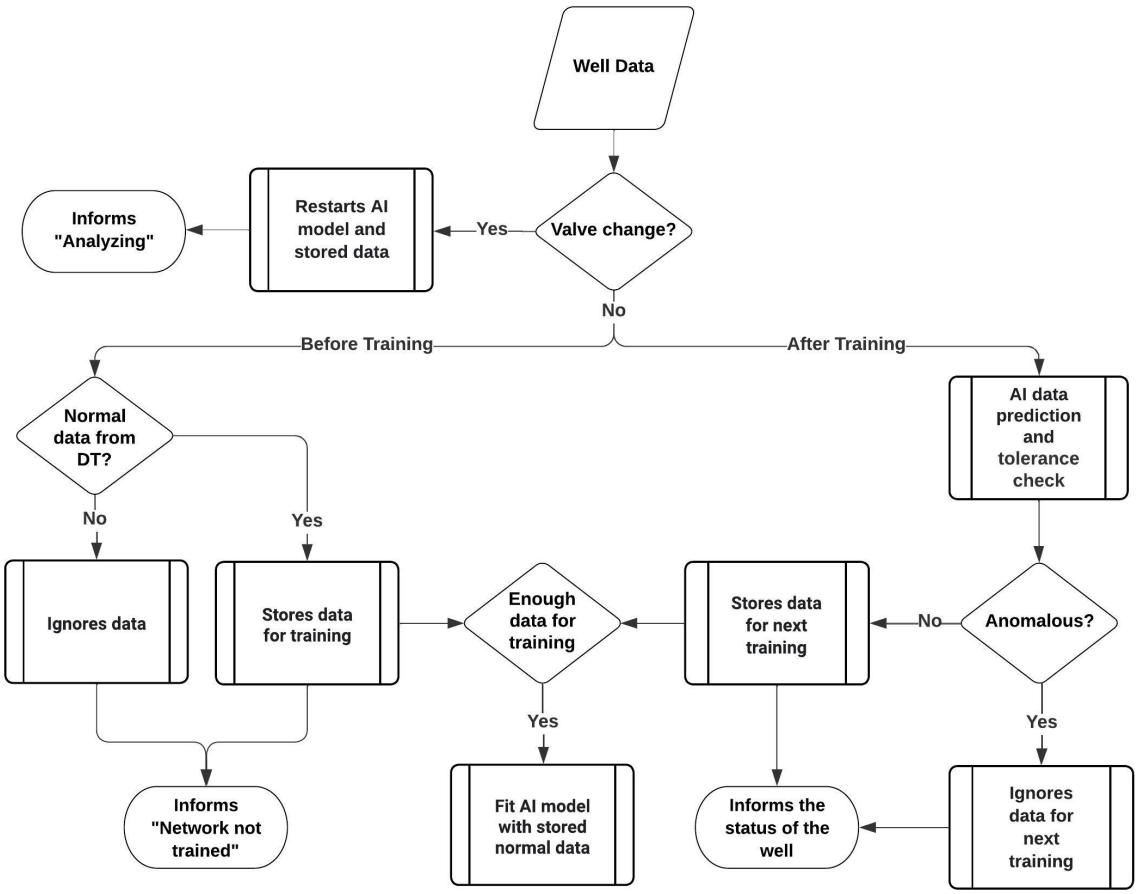
in which k controls the tolerance interval around the mean and can be set for each problem. In this study, k = 3 was adopted according to the database used, which comprises about 99. 7% of the normal data saved. The threshold for occurrence of an anomalous event is based on this tolerance. If a new data SMAPE value is under this tolerance, the new instance is a normal datum. If the SMAPE value exceeds the tolAI value, it can be associated with risk levels defined by the operator, supporting the real-time decision-making process. For the purpose of the system presented here, values exceeding 30%, 70%, and 100% represent low, medium, and high risk anomalies, respectively. Figure 2.6 illustrates the process flow diagram triggered in a fixed time interval for anomaly detection using the developed LSTM-based autoencoder.

Figure 2.6 presents a comprehensive flow diagram detailing the general methodology employed for anomaly detection using the LSTM-based autoencoder. The diagram begins with the collection of well data and branches based on whether a valve state change occurred. For normal data, the system assesses whether there is sufficient data for training. Once trained, the AI model makes predictions and checks for anomalies using the defined tolerance thresholds. The diagram includes provisions for updating the model and storing normal data for future training, as well as categorizing anomalies based on their risk levels. This structured flow highlights the decision-making process at each stage of the anomaly detection pipeline.

By using both literature (Vargas et al., 2019b) and industry data, the autoencoder network was tested to verify its robustness when exposed to time series containing normal, transitional, and anomalous points. It was able to correctly classify 98–99% of the anomalous data, indicating that the autoencoder network is capable of accurately detecting undesirable events in oil production. Figure 2.7 shows an example of an abnormal state using real data from a subsea well, in which can be seen the pressure values gathered from PDG, PT and TPT in the upper graph, the temperature values from PDG and TPT in the central graph, and the anomaly score value calculated by the neural network in the lower graph, compared to the tolerance line (threshold).

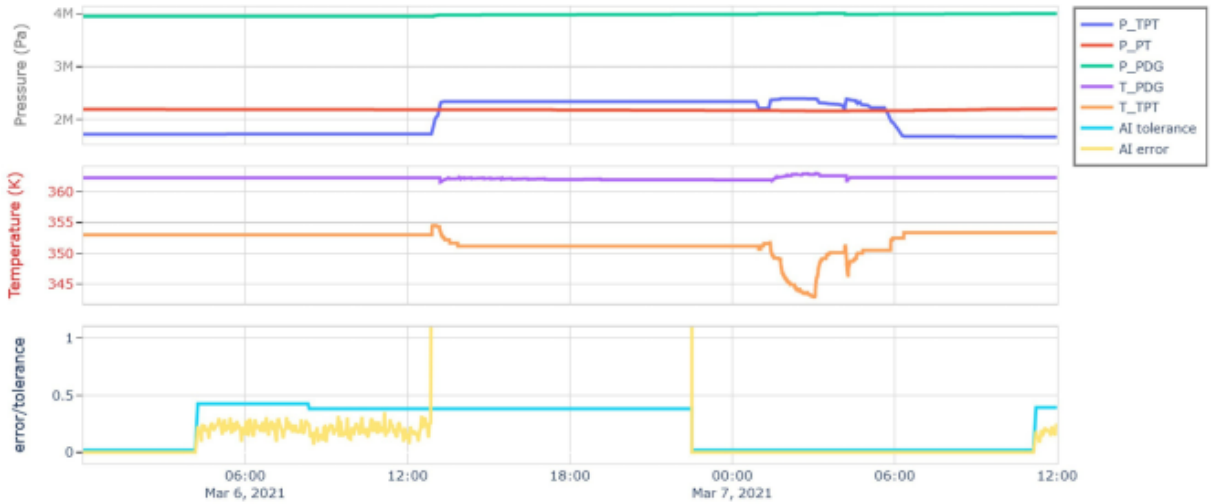
Figure 2.7 illustrates a time series visualization of pressure and temperature readings from

Figure 2.6 – General methodology of anomaly detection using AI



Source: Aranha et al. (2024a)

Figure 2.7 – Illustrative example of anomaly detection using AI.



Source: Aranha et al. (2024a)

different sensors (TPT, PT, and PDG), along with the anomaly detection results provided by the AI system. The upper graph represents pressure measurements, showing distinct trends and fluctuations for each sensor. The central graph represents the temperature readings, reflecting the variations over time. The bottom graph presents the anomaly score calculated by the LSTM-based autoencoder, overlaid with the tolerance line that defines the threshold for anomaly detection. Anomalies are visually apparent where the anomaly score exceeds the tolerance line, highlighting regions of abnormal behavior detected in real-time.

As stated previously, three parameters are obtained after the training: (1) data normalization parameters, (2) weights of neural networks, and (3) anomaly tolerance calculated during training. The first two are configuration parameters relevant to the transformation of input variables and network coefficients. As monitoring is occurring in real time and the conditions of the oil well operation are constantly changing, new training is performed every time 500 new samples are saved. The number of samples required for new training was determined via hyperparameter testing on real data, aiming to have a sufficient amount of data that minimizes the hypersensitivity of the results without penalizing computational resources or causing overfitting. It should be noted that the weights, normalization parameters, and tolerance depend on the valve schematics. If the operator changes the valve configuration, a new network is trained specifically for that new scheme. In other words, each valve configuration has its own deep network associated with it, which evolves over time through the acquisition of new data. It also allows simulation of a look-ahead scenario to predict the system response to a specified action, such as a valve maneuver.

Before training a new network, the data are scaled to be between 0 and 1 using a Min-Max scaler, to ensure that each variable has the same order of magnitude, leading to more stable and expedited training. The procedure is described in Eq. 2.3, in which x and x' refer to the original data point and its scaled value, respectively:

$$x' = \frac{x - \min_x}{\max_x - \min_x} \tag{2.3}$$

This step is crucial to achieve accurate optimization results from the network. However, in some cases, the sensors installed in the wells may transmit data with low variation due to physical conditions or resolution limitations. This scenario can cause the network to become more sensitive, resulting in maximum and minimum values that are close together, so even minor variations may be improperly classified as an anomalous event. To avoid this excessive sensitivity of the network from occasional low variation of the signal, randomly generated noise from a uniform distribution is added before scaling, as depicted in Eq. 2.4:

$$y = x' + \epsilon \tag{2.4}$$

The noise level is set between 0% and 0.5% of the average of the received data. This procedure not only enriches the dataset by introducing variance but also ensures that the model is better equipped to distinguish between significant patterns and minor deviations, reducing the propensity for false positives.

2.3 Metrics Applied for Performance Evaluation

Classification metrics are highlighted and used as a benchmark to evaluate the performance of classification algorithms. Therefore, a set of metrics was used to compare and evaluate the algorithms. Accuracy (ACC) considers all normal and fault samples. ACC can be calculated using Eq. 2.5:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{n_{\mathrm{total}}} \tag{2.5}$$

where: TP: Number of true positives; TN: Number of true negatives; n_{total}: Total number of samples.

When dealing with unbalanced data, accuracy (ACC) is not a suitable measure. Instead, it is recommended to use balanced accuracy ($\mathrm{ACC_b}$) as a more appropriate metric. The calculation for $\mathrm{ACC_b}$ can be estimated using Eq. 2.6:

$$\mathrm{ACC}_b = \frac{\mathrm{REC} + \mathrm{SP}}{2} \tag{2.6}$$

where Recall (REC) indicates the proportion of anomalous data that is correctly detected from all anomalies. Typically, in industrial applications, REC is a prominent metric, as false negatives lead to much more harmful results than false alarms or false positives. The REC is given by Eq. 2.7:

$$\mathrm{REC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2.7}$$

in which FN is the number of false negatives.

Specificity (SP) estimates the ability of the algorithm to predict true negatives over false positives and can be calculated by Eq. 2.8:

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{2.8}$$

where FP is the number of false positives.

The F1-SCORE consists of an important metric, especially in problems containing imbalanced data. It is defined as a harmonic mean between recall (REC) and precision (PR) and can be estimated in Eq. 2.9:

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{REC} \cdot \mathrm{PR}}{\mathrm{REC} + \mathrm{PR}} \tag{2.9}$$

Precision (PR) evaluates how many of the predicted positive samples were actually true positives. It is defined by the ratio of true positives to the sum of true positives and false positives, as defined in Eq. 2.10:

$$\mathrm{PR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2.10}$$

2.4 Results and Discussion

This section presents and analyses the main results for the proposed methods. The case study are first presented along with the main results, than other aspects of the proposed system are discused.

2.4.1 Case Studies

The proposed methodology was integrated into a real-time monitoring system of offshore wells, gathering data from sensors and the valve status of subsea production/injection wells. It has been implemented in more than 20 FPSOs, monitoring more than 250 wells in the Santos and Campos basins, eastern Brazil, all equipped with multiplexed wet Christmas trees. An advantage of this system lies in its compreheniveness in identifying anomalies related to well integrity for hundreds of subsea wells. These anomalies would be impossible for human operators to track. Some use cases are presented to illustrate the efficacy of the system compared to human postclassification (class). In addition, a web-based prototype was developed to provide a user-friendly experience to monitor multiple wells simultaneously.

Three case studies are presented to illustrate the efficacy of the system in real monitoring situations: Well 1 is an oil production well in the presalt area in Santos Basin, ultradeep water scenario, water depth 2041 m, drilled in five phases, directional trajectory, and final depth of 5870 m, equipped with downhole interval completion valves, and producing in three different zones at 5432–5774 m.

Well 2 is an oil production well in the presalt area in the Santos Basin, a depth of water of 2125 m, drilled in four phases with directional trajectories and a final depth of 5552 m, equipped with downhole interval completion valves, and producing in two different zones at 5137-5440 m.

Well 3 is a water-alternating-gas injection well in the presalt area of Santos Basin, at a depth of water of 1946 m, drilled in four phases, a vertical trajectory, and a final depth of

6197 m, equipped with completion valves of the downhole interval and injecting in three different zones at 5437-573 m.

Table 2.1 presents the summarized results for the four classifiers applied to spurious DHSV closure data. RF was used for the results of (Marins et al., 2021a), DT for the model presented in (Turan and Jaschke, 2021), and LSTM refers to the results obtained by (Machado et al., 2022) using the LSTM autoencoder. Finally, DD + LSTM stands for the present study, and the comparison is made in terms of ACC, F1-SCORE.

Table 2.1 – Comparison of metrics between literature and the proposed method: identifying spurious closure events of DHSV from a public database (Vargas et al., 2019b).

Classifiers	**Accuracy (ACC)**	**F1-SCORE**
RF	0.8708	-
DT	0.6000	0.4900
LSTM	0.9922	0.9360
DD + LSTM	0.9894	0.9917

From Table 2.1, the results obtained are in line with other works in the literature. The accuracy achieved is around 98.9%. On the same dataset, (Machado et al., 2022) reported 99.9% accuracy by using the LSTM autoencoder, while (Marins et al., 2021a) presented 87.1% by using RF, and (Turan and Jaschke, 2021) reported 60% with the DT method. Looking at the F1-SCORE values, DD + LSTM performs the best, followed by the LSTM model, both of which are significantly superior to DD.

The coupled method has shown superior performance compared to using the DD alone. To illustrate this, the results obtained using the DD model are compared with those provided by the comprehensive model (DD + LSTM). Table 2.2 presents the average values for accuracy (ACC), balanced accuracy (ACC_b), and F1-SCORE, allowing a direct comparison between the models developed herein.

Table 2.2 – Metrics obtained for the DD model and the complete model (DD + LSTM) for identifying spurious closure events of DHSV from a public database (Vargas et al., 2019b).

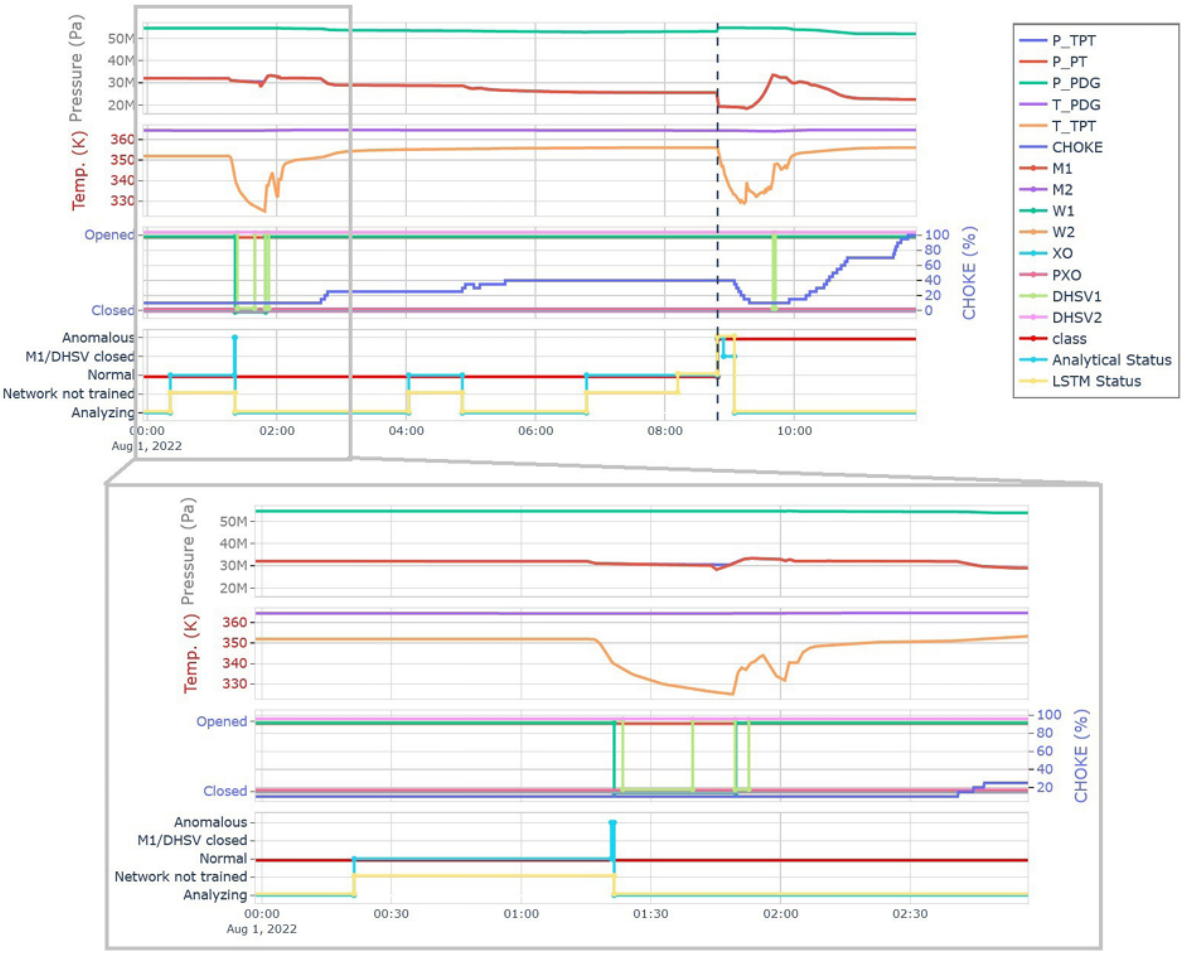
Method	**ACC**	**ACC_b**	**F1-SCORE**
DD	0.9727	0.9615	0.9758
DD + LSTM	0.9894	0.9771	0.9917

Case Studies: The proposed dual system was applied to the three case studies, and the results were compared with human post-classification (class), as presented in Figs. 2.8 through 2.11. It can be seen that the occurrence of anomalies occurs abruptly in all case studies. In these figures, the bottom graphics represent the statuses of the DD and LSTM methods, which can be labeled as follows: Analyzing, for both methods, which is triggered when there is a change of a valve status (e.g., M1 has closed), that will cause temporary

fluctuations in the pressures and temperatures measured by the sensors, so the system awaits some time for these readings to stabilize before resume monitoring; Network not trained, exclusive to the LSTM, reported just after analysis while there is not enough data classified by the DD as normal for the LSTM to train; and Normal / anomalous, for both methods, regarding anomaly detection; and some others exclusive to DD, such as DHSV/M1 closed, reported when according to sensors those valves should be open, but the DD detects that they might actually be closed.

Figure 2.8 – Results for Well 1 for the first anomalous event detected:

PDG, PT, and TPT sensors, valve status, and comparison between model classification (LSTM Status and Analytical Status) and fault classification (class).

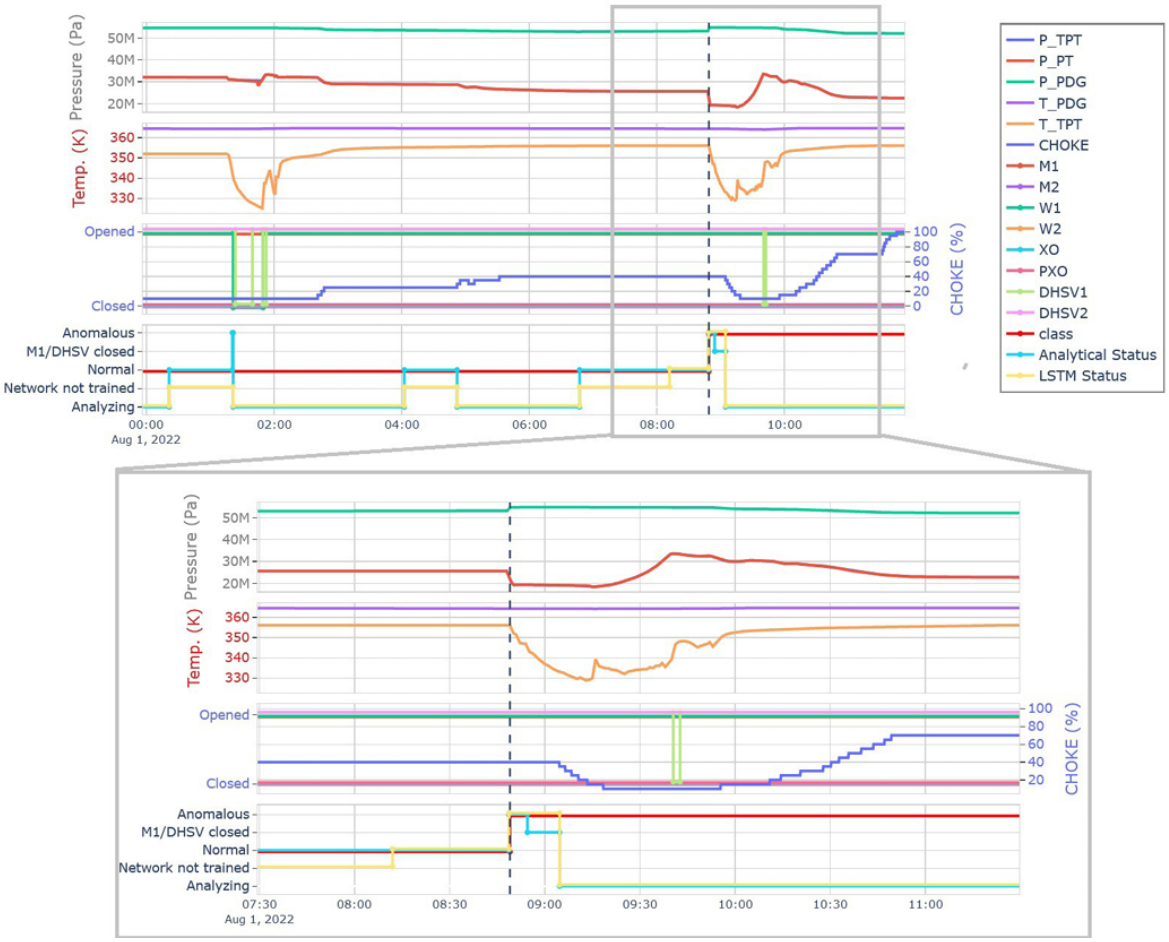


Source: Aranha et al. (2024a)

The data analyzed from well 1 are presented in Figure 2.8 and show data from the PDG, PT, TPT, and valve status sensors along with the results of the analyzes and classifications for a specific time window on 1 August 2022. All valves have a binary open or closed state except for the choke, which reports a partial opening indication. Two anomalous events were identified during the period. The first at 1:16 a.m. (analysis status curve) related

Figure 2.9 – Results for Well 1 for the second anomalous event detected:

PDG, PT, and TPT sensors, valve status, and comparison between model classification (LSTM Status and Analytical Status) and fault classification (class).

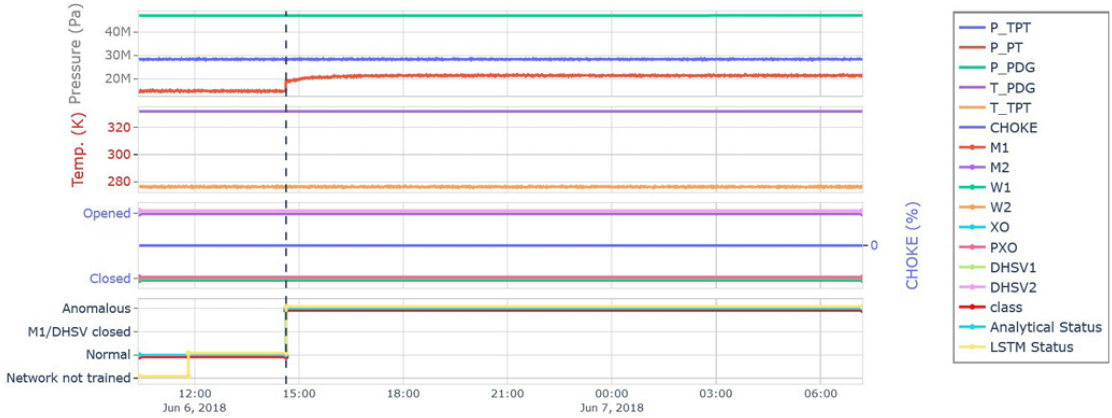


Source: Aranha et al. (2024a)

to spurious closure of the DHSV, corrected by the operator in the sequence with valve cycling performed from 1:39 a.m. to 1:52 a.m. Only the analytical approach captured this anomaly, while the AI did not have time to identify it because the system went into analysis mode due to the change in parameters. Several hours later (Figure 2.9), a second event also occurred at 8:48 am, identified by the model (LSTM curve status and analytical status) and promptly corrected by the operator. In this case, the analytical model correctly classified, after detecting the anomaly, as spurious closure of the DHSV or the M1, and the LSTM model correctly identifies the anomalous event.

The following example refers to Well 2, In this well, the developed system was applied in a retro-analysis to identify an anomaly event at 2:37 p.m. on 6 June 2018 (curves LSTM status and analytical status). In the field, the problem was only detected sometime later in a light workover intervention; the cause was the pressure communication by the gas lift valve with the subsea well closed. Figure 2.10 presents the data from the PDG, PT,

Figure 2.10 – Results for Well 2—PDG, PT, and TPT sensors, valve status, and comparison between model classification (LSTM status and analytical status) and fault classification (class).



Source: Aranha et al. (2024a)

TPT and valve status sensors along with the results of the analyzes and classifications. It is important to note that, in this well, if the developed methodology had already been implemented at that time, the identification and correction of the problem would have been anticipated, avoiding NPT, so bringing enormous gains in oil production.

Figure 2.11 presents the data and results for Well 3, on 18 October 2022. The dual model identified the anomaly at 2:10 am. (curves the LSTM status and analytical status) and correctly classified it in the sequence as spurious closure of DHSV or M1. It is important to note that the operator then closed the well and proceeded with valve cycle and reopening procedures, resuming production in a safe condition. After reopening well at 7:14 a.m., the system starts to analyze and returns after some time to normal condition at 8:29 a.m.

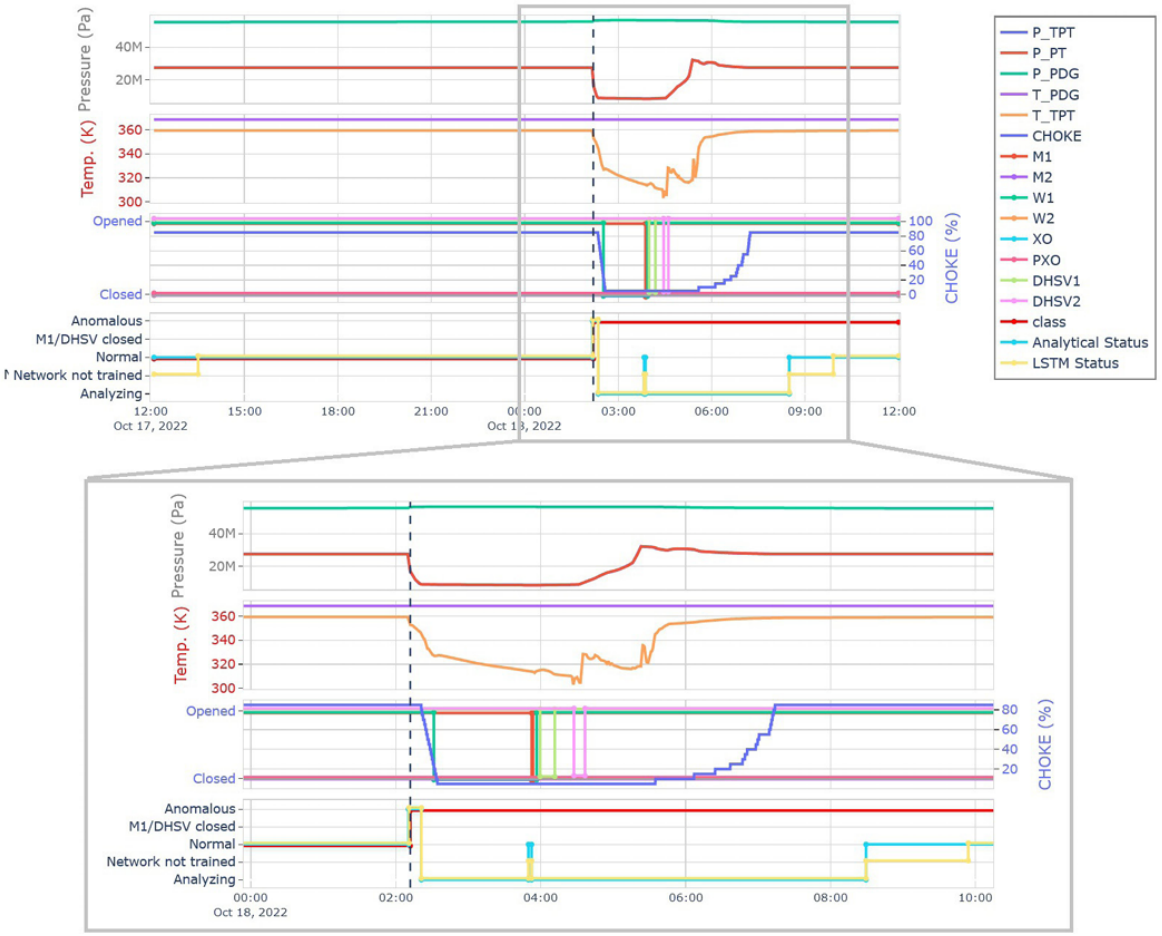
Table 2.3 presents the SMAPE values, balanced accuracy (ACC_b), and F1-SCORE values calculated for the case studies when compared with the human post-classification (class).

Table 2.3 – Comparison of metrics between literature methods and the proposed method for the three case studies.

Case Study	**SMAPE (%)**	**ACC_b**	**F1-SCORE**
Well 1 – first event	39.148	0.9897	0.9987
Well 1 – second event	48.312	0.9988	0.9989
Well 2	33.136	0.9998	0.9998
Well 3	23.390	0.9988	0.9991

The dual model consistently exhibits superior performance across all metrics when contrasted with the standalone DD model. This is further evident in the detection results presented in Table 2.4, where ACC_b and F1-SCORE are compared for the case studies.

Figure 2.11 – Results for Well 3—PDG and TPT sensors, valve status, and comparison between model classification (LSTM status and analytical Status) and fault classification (class).



Source: Aranha et al. (2024a)

Although the DD model on its own already produces impressive results, the combined model surpasses it.

Table 2.4 – Metrics obtained for the DD model and the complete model (DD + LSTM) for the three case studies.

Method	**Case Study**	**ACC_b**	**F1-SCORE**
DD	Well 1 – second event	0.9800	0.9808
	Well 2	0.9990	0.9998
	Well 3	0.9988	0.9983
DD + LSTM	Well 1 – second event	0.9988	0.9973
	Well 2	0.9998	0.9998
	Well 3	0.9995	0.9993

2.4.2 Some Comments on the Proposed System

The proposed system performs well on the task of anomaly detection, presenting detection rates exceeding 90% in the real field production scenarios studied. Changes in well and reservoir conditions over time impose concept drift and data drift in time series, which are challenging factors for machine learning models. The dual approach handles the handling of these aspects, albeit generating the need for recurrent training of LSTM autoencoder, which can result in increased computational cost. The DD part serves the specific and valuable purpose of describing the normal state of the well in different valve statuses. However, on analytic basis, it exhibits lower sensitivity compared to LSTM, resulting in a lower detection rate.

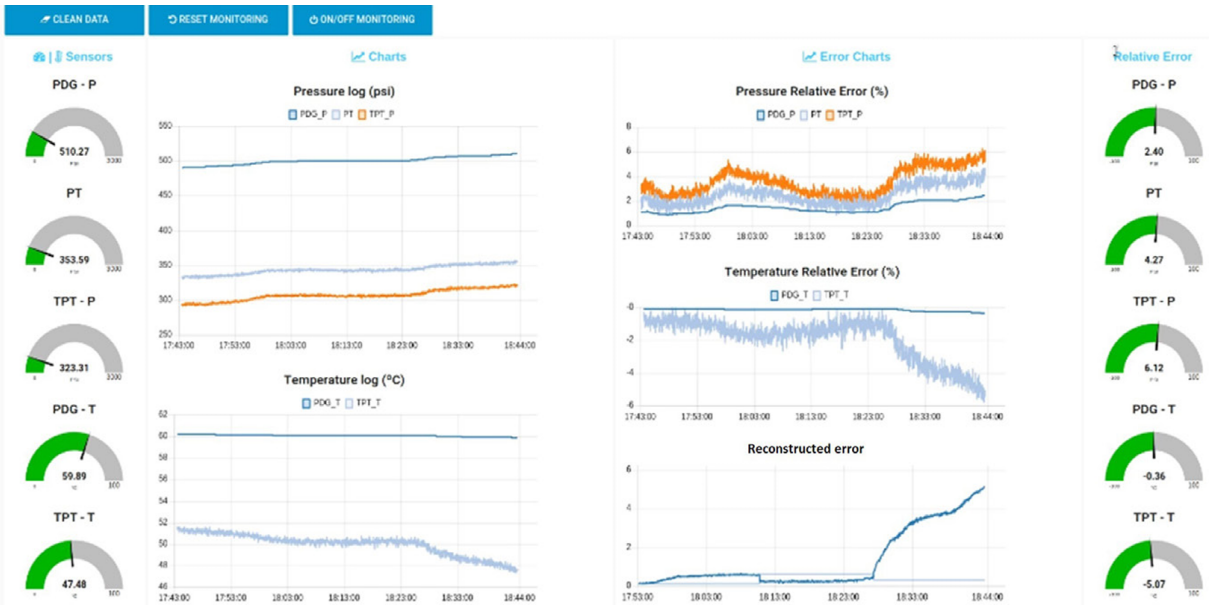
DD is used primarily to assess the system's status under different valve schemes, providing predefined rules and up to a predetermined threshold of 5%. However, in some cases, minor variations can result in false positives. Compared to the model trained via LSTM, the latter exhibits dynamic sensitivity that adapts with each training, enabling it to detect subtle changes that the analytical method may occasionally miss.

2.4.3 Prototype Development

Aiming to provide an agile and accurate monitoring experience, a web-based computational system has been developed that allows the handling of multiple wells simultaneously. From the main dashboard, the operator can set a specific well, whose graphical interface is presented in Figure 3.10. The center of the screen displays two graphs on the left that represent the absolute values of the sensors, as well as two first graphs on the right that display the relative error of the moving average calculated from previous data. The lower right graph shows the reconstructed error and tolerance from the AI approach. The latest absolute values received from the sensors are shown on the left, while the latest calculated relative errors are displayed on the right. At the top of the screen, the well status evaluated by the analytical approach and the neural network is shown, along with some real-time monitoring control buttons.

In addition to searching for anomalies, the system has some functionalities that help prevent false positives. One of them is the detection of frozen sensors, which evaluates the repetition of values from the sensors for a certain period of time. If detected, the frozen sensors are reported on the system output. Another functionality is the identification of invalid data, so that inconsistent data, such as non-numeric or null values for sensors, are not included in the analysis. Changes in valve status can significantly alter well behavior. Therefore, the system has an analysis mode whose objective is to provide adequate time for pressure and temperature levels to stabilize, thus ensuring that planned changes are not mistaken for anomalous events. With these functionalities, the system becomes more robust and is able to provide greater security during monitoring.

Figure 2.12 – Illustrative example of the developed prototype: Parameters and system response during a tubing to annular communication anomalous event.



Source: Aranha et al. (2024a)

2.5 Partial Remarks

In this Chapter, the system for real-time monitoring and anomaly detection is described using production sensor data from oil wells. The system utilizes a combination of a deep learning autoencoder and a rule-based analytic approach to detect unexpected events, with the goal of improving operational safety and reducing the costs associated with NPT and failure repair.

Initial studies were conducted to determine the behavior and correlations of pressure and temperature values for the most common combinations of well valve states. The proposed methodology uses pressure and temperature sensor data to classify the well status via a DD, which is then used to train autoencoders based on deep LSTM networks devoted to anomaly detection. This coupling enables the deep neural network to evolve constantly through the normal data collected by the analytical method. It makes the system versatile enough to adapt itself to variations in the valve scheme while also allowing new anomalies to be added to the cataloged portfolio of the rule-based analytical approach.

A comparison with other approaches using the same public data set is provided, focusing on spurious DHSV closure events. Evaluating metrics such as accuracy, and F1-SCORE indicates that the proposed system either performs comparably or outperforms other machine learning techniques.

The developed system exhibits high accuracy, with true positive detection rates exceeding 90% in the early stages of anomalies identified in both the simulated and actual well

production scenarios. From the latter, it can be highlighted the enhancement of the detection process, typically carried out by humans. For the case studies presented, all three wells had a short transition phase, and the model performed well with a balanced accuracy higher than 98.0%.

The system is implemented in more than 20 FPSOs, monitoring more than 250 production/injection subsea wells, and can be applied in both real-time operation and testing scenarios. The continuous acquisition of data through sensors combined with the adaptability of the detection model ensures robustness over time, along with the changes in well, reservoir, and operation conditions. It is worth mentioning the limited computational infrastructure available at rigsite, which requires expedited models to provide a quick response to the data flow. By observing the performance of the proposed system in massive simultaneous monitoring, it has consistently delivered a prompt response to online well integrity monitoring.

There is still room for future research, such as a rigorous study on how accurately the proposed method can perform early-stage anomaly detection and address persistent challenges related, for example, to valve performance and tightness. Although this chapter focuses on field application, future research could explore the use of other deep learning architectures, such as gated recurrent units or transformer neural networks, which are known for their efficiency in processing sequential data.

3 Anomaly Detection in Oil Wells Using Ensembles of Binary LSTM Classifiers

This Chapter presents the use of long short-term memory (LSTM) networks for anomaly detection in oil wells, utilizing a one-vs-all ensemble of binary classifiers. Operational data from the 3W dataset is used, where each labeled anomaly serves as the positive class, and normal conditions, along with other anomaly types, form the negative class. Validation is conducted using real-world data, with hydrate formation detection as a case study.

The ensemble model achieves an average accuracy of 93%, while hydrate detection specifically achieves 90% accuracy. Confidence intervals are calculated using the bootstrap method to address uncertainty. Benchmarking against multiclass and one-class ensemble models highlights the strengths and weaknesses of the proposed approach. A key advantage of the binary ensemble is its ability to identify data from unknown classes, as such data is unlikely to belong to any positive binary class.

Conversely, multiclass classifiers often constrain data to predefined classes, and one-class ensemble models can present lower accuracy due to the absence of negative samples during training. Using real sensor data from pressure and temperature measurements, the methodology is validated on production datasets and produces results consistent with existing literature, even with limited sensor configurations.

Part of its content is published on the paper Lopes et al. (2024), and the up to date results are submitted to the Petroleum Science and Technology Journal.

3.1 Overview

This Chapter focus on the use of a one-vs-all ensemble (Hafiz and Bhat, 2020) of binary LSTM classifiers for detecting anomalies in the oil industry. A one-vs-all ensemble involves training multiple binary classifiers, each focused on distinguishing one specific class from all others. This method can identify data from unknown classes, as it does not force such data into predefined categories like multiclass classifiers. This approach, validated in real data from the 3W dataset.

While hydrate Lu et al. (2019) formation detection serves as a representative case study, the methodology is versatile enough to extend to multiclass classification of various production anomalies using ensemble strategy. Applications include monitoring production wells, injection wells, and CCUS (Carbon Capture, Utilization, and Storage) wells, as demonstrated using the 3W dataset (Vargas et al., 2019b).

3.2 Proposed Methods and Studies

The methodology adopted in this study focuses on developing a binary LSTM ensemble to detect anomalies in oil wells. 3W dataset anomalies are used as the main case studies to validate the approach with real data, and particularly the hydrate formation anomaly is used to verify results with analytical formulas. The process consists of the steps described Figure 3.1.

Figure 3.1 – Main binary classifier development.

Source: Author (2025)

3.2.1 Exploratory Data Analysis

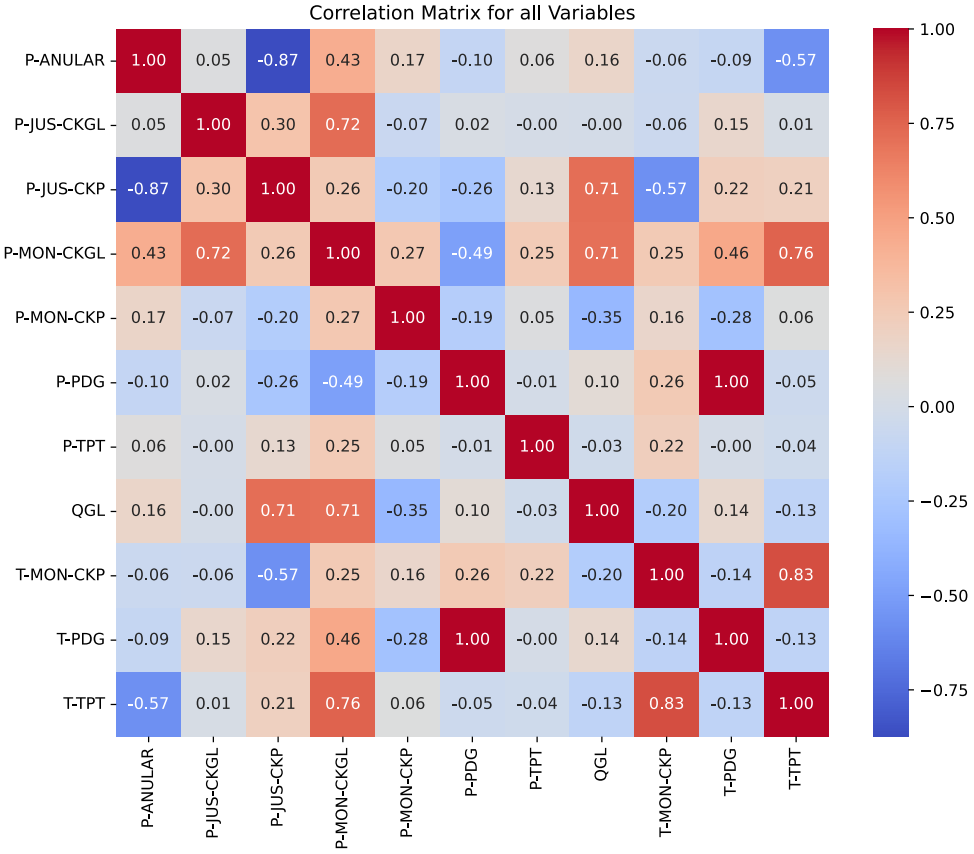
The initial step involved an exploration of the 3W dataset provided by Vargas et al. (2019), consisting of graphical visualization, statistical analysis, and correlation studies among various variables. These authors made available an extensive dataset comprising some of the most common unexpected events in production wells. The Pearson's correlation is used to illustrate the relationships between variables, focusing both on the combined data for all events, as shown in Figure 3.2, and specifically on the data indicative of hydrate formation, as shown in Figure 3.3. In both Figures 3.2 and 3.3, the red colors indicate a direct correlation between the two variables, while the blue colors represent the negative proportionality.

3.2.2 Exploratory Data Analysis

The intensity of the color, as the legend indicates, is related to the proportionality between the variables, from -1 to 1. Comparing both matrices, it is possible to see different behaviors when comparing hydrate, a specific anomaly case, to the correlation behavior of rest of the data. This shows that the signal of a single anomaly is different from the other data, and it indicates to be possible to use a method to specifically identify specific anomaly cases.

The dataset from Vargas et al. (2019) was released with eight different anomaly data and a total of 1984 registered study cases across different anomalies and oil wells. An update was released in 2024 adding new variables, cases and a new class of anomaly, containing 2228 study cases total. The dataset comprises ten folders, each containing several information about the sensors, as described in Table 1. The undesired events presented on the dataset

Figure 3.2 – Correlation for the sensor data.



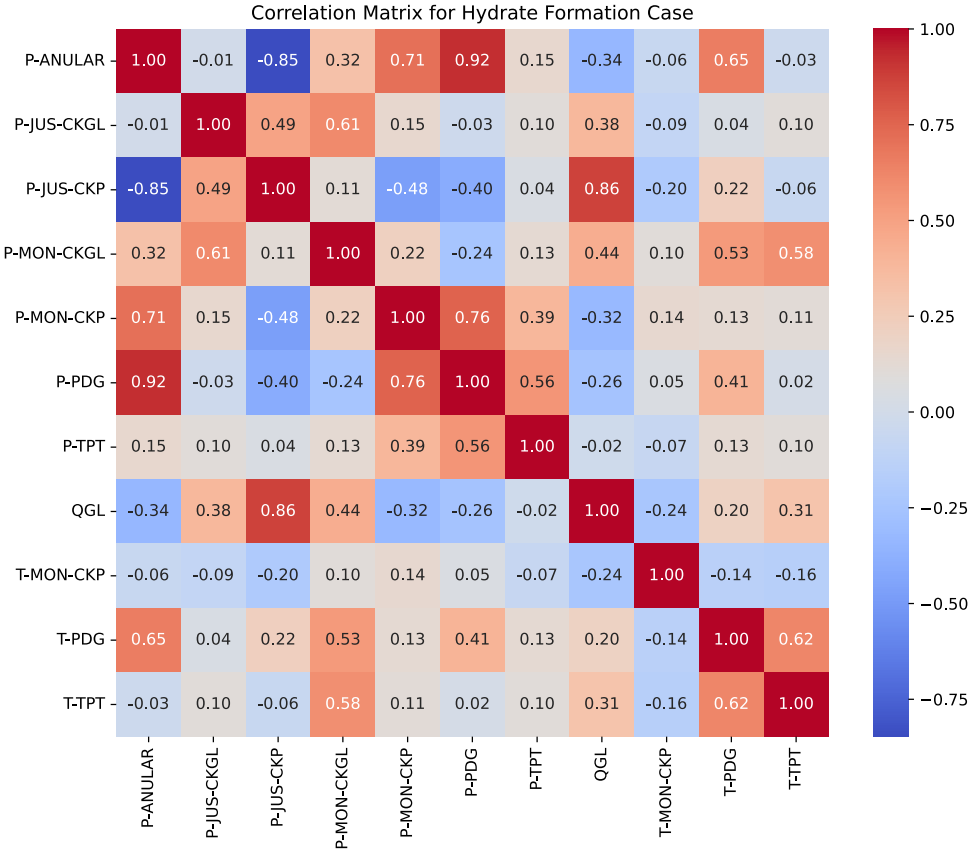
Source: Author (2025)

are: 0 - Normal State, 1 - Abrupt Increase of BSW, 2 - Spurious Closure of DHSV, 3 - Severe Slugging, 4 - Flow Instability, 5 - Rapid Productivity Loss, 6 - Quick Restriction in PCK, 7 - Scaling in PCK, 8 - Hydrate in Production Line and 9 - Hydrate in Service Line. Most sensor data are studied in the correlations, and nine unique anomalies are addressed, as well as data depicting normal operational conditions. Each sensor sample is categorized as a "normal state," "transient state," or "anomalous state." Notably, hydrate formation (Lu et al., 2019) is an important anomaly addressed, because hydrate accumulation can severely impede or even completely block fluid flow.

3.2.3 Binary Deep Neural Network Classifier for Specific Anomaly Detection

The sensor data is processed as numerical time sequences, with each sequence labeled based on the annotation of its final sample. The primary objective of each binary classifier in the ensemble is to determine whether a given data sequence corresponds to a specific anomaly. Sequences ending with a sample labeled as either normal or any anomaly outside the classifier's target class are assigned a label of 0, while sequences concluding with a

Figure 3.3 – Correlation for hydrate formation data



Source: Author (2025)

sample indicative of the target anomaly are assigned a label of 1.

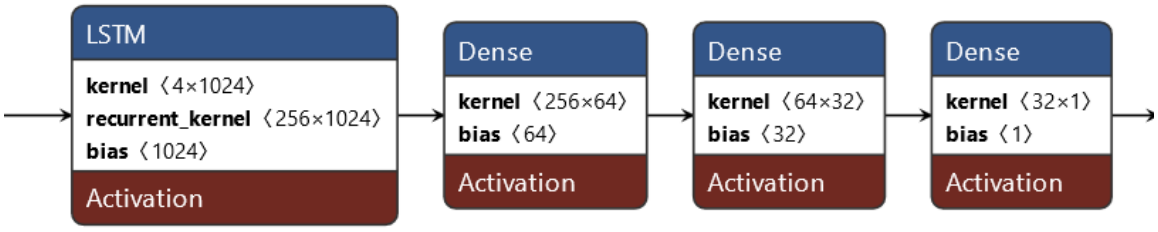
This approach frames the task as a binary one-vs-all classification problem, incorporating multiple temporal input features. 'Multiple input' refers to the inclusion of various sensor readings, such as pressure and temperature, as inputs to the model. 'Binary classifier' signifies that the output consists of two possible classes, while 'one-vs-all' indicates that the target class is treated as the positive class and all other classes as the negative class. Sequences are constructed and labeled based on the annotation of their last data point, with the preceding sequence providing context for anomaly detection.

Various sequence lengths are tested, and hyperparameter tuning is performed to optimize the architecture, which is primarily based on Long Short-Term Memory (LSTM) layers. Data is collected on a second-by-second basis, and the most effective architecture is identified through extensive experimentation. The optimal model architecture, including details about its layers and hyperparameters is depicted in Figure 3.4: LSTM layers and Dense layers (Dropout layers (Srivastava et al., 2014), and Batch Normalization layers are omitted). This figure also specifies the number of neurons and the configuration of each

Table 3.1 – Summary of sensor data points and descriptions.

Name	Description	Unit
P-PDG, P-TPT, P-MON-CKP, P-JUS-CKP, P-JUS-CKGL	Pressures measurements	Pa
T-TPT, T-JUS-CKP, T-MON-CKP, T-PDG	Temperature measurements	°C
QGL, QBS	Flow rates	m^3/s
ESTADO-DHSV, ESTADO-M1, ESTADO-M2, ESTADO-PXO	State of valves	[0, 0.5, or 1]
ABER-CKGL, ABER-CKP	Opening percentage	%

Figure 3.4 – Hyper-parameters for the binary classifier network.



Source: Author (2025)

layer to ensure clarity.

The neural network utilizes data from two key sensors: the temperature and pressure transducer (TPT) and the permanent downhole gauge (PDG). These sensors were chosen due to their frequent use in real-world operational monitoring of wells, as highlighted by Vargas et al. (Petrobras, 2017), and their consistent availability in the dataset. While additional sensors are sometimes present, their data is not always consistently available across all samples in the dataset or in practical operational scenarios.

To enhance the model's applicability to a broader range of real-life operations, this study focuses on pressure and temperature readings provided by TPT and PDG sensors. This decision involves a tradeoff: limiting the input to fewer sensors reduces the amount of data available for identifying anomalous behavior, which may affect the model's performance. However, this simplification increases practicality, as fewer sensors are needed for deploying the models effectively in diverse operational contexts. This balance prioritizes operational

feasibility while maintaining robust anomaly detection capabilities.

3.2.4 Hyper-parameter Optimization

The focus of hyperparameter optimization is to determine the ideal sequence length for the input of the neural network. Input sequences are constructed using n seconds of data samples, ensuring no overlap, resulting in over 200,000 input-output sequence pairs for training and testing. Early experiments showed that the initial architecture could be simplified and optimized by reducing the number of neurons and layers.

Hyperparameter tuning was performed using Keras-Tuner, a tool for optimizing neural networks in the Keras library Chollet, 2016. Random Search was used to explore configurations, evaluated based on validation accuracy. The optimization focused on refining the number of LSTM layers and units (ranging from 32 to 128), dropout rates (0.1 to 0.5), dense layers and units (32 to 64), and learning rates (between 1e-4 and 1e-2). The Adam optimizer (Kingma and Ba, 2014) was chosen for its efficiency, using binary cross-entropy as the loss function. Training was conducted over 25 epochs with early stopping set to 5 steps patience, and a batch size of 64 was used. The final architecture identified during tuning, shown in Figure 3.4, effectively processes TPT and PDG sensor data for binary anomaly classification, balancing simplicity and performance.

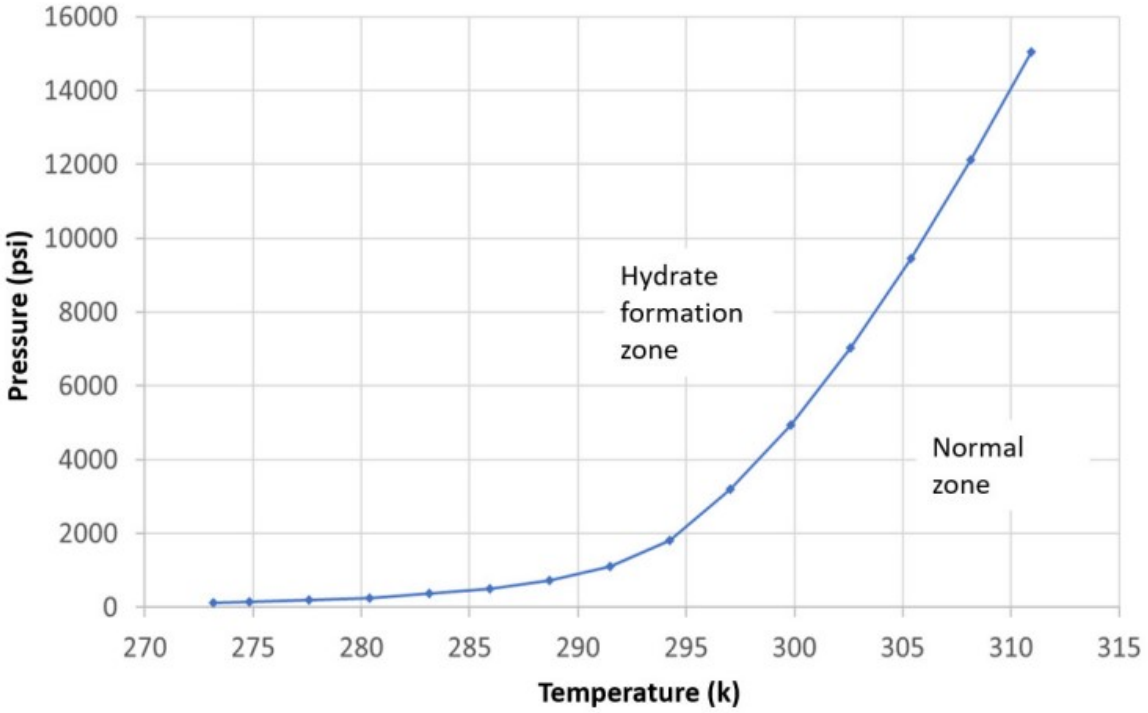
3.2.5 Verification Formulation and Study Case: Pressure and Temperature Analysis (PVT)

Despite developing a binary classifier architecture that could be applied to other anomaly cases, this work uses an analytical formulation to verify the soundness of the models obtained. Many of the 3W dataset anomalies do not have a specific analytical solution or depend on expert interpretation. In this context, to verify our model, it is used the specific anomaly of hydrate formation. Hydrates in oil wells are dependent on specific conditions, notably the presence of water and natural gas under high pressure and low temperature. Since hydrate accumulation can severely impede or even completely block fluid flow, monitoring pressure and temperature in the well is crucial. Preventive measures, as highlighted in the existing literature, represent the most effective strategy for managing this problem.

Safamirzaei et al. (2015) describe a minimum pressure and temperature curve for predicting hydrate occurrences in production lines, as illustrated in Figure 3.5. The Figure shows the curve that represents the analytical hydrate detection formula, where the x-axis contains the temperature (in kelvin) the y-axis is the pressure (psi), and the points the curve divides the normal zone and the hydrate formation zone.

PVT analysis incorporates various formulations that consider the composition of produced

Figure 3.5 – Pressure and temperature curve for hydrate formation



Source: Author (2025)

gas, a key factor in hydrate formation. As Marins (2018) notes, at a fixed pressure, denser gas requires a higher temperature for hydrate formation, shifting the pressure-temperature curve rightward. Modern formulations often include density as an input parameter. To validate hydrate formation, it is applied Motiee's formula (Motiee, 1991)::

$$T = -238.245 + 78.997P - 5.35P^2 + 349.473877\rho - 151.055\rho^2 - 27.60P\rho \tag{3.1}$$

in which T represents the critical temperature threshold for hydrate formation. For a fluid with pressure P and density ρ, temperatures below $T_{hydrate}$ fall within the risk zone for hydrate formation. A confidence value, $C_{hydrate}$, quantifies this risk:

$$C = 100 \cdot (T - T_{hydrate})/T, \tag{3.2}$$

C, ranging from 0% to 100%, indicates the likelihood of hydrate formation. It reaches 100% when the temperature T is below T. Equation 3.1 uses two considerations: The gas considered is methane because it would be the most dangerous case, and the coefficients are designed to convert the values from picoseconds to kPA.

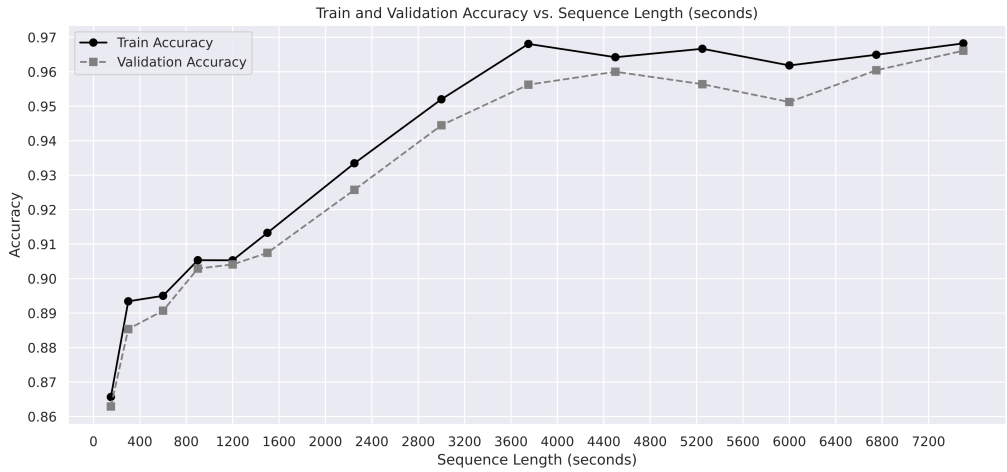
3.3 Results and Discussions

This section details the results obtained from the proposed methodology. The main objective is to demonstrate different aspects of the binary classifiers for the 3W dataset classes and verify the hydrate formation with the analytical formulation. The results demonstrate the efficacy of the deep LSTM neural network and show that the ensemble model can identify correctly the not trained anomaly cases.

3.3.1 Binary Classifier Validation and Literature Benchmark

Figure 3.6 and Table 3.2 present the accuracy achieved with different sequence lengths. The results indicate a clear trend: longer sequences tend to improve model performance. Our findings can be compared to those of (Marins et al., 2021b), who achieved a similar accuracy of 96.9%, albeit with a different setup and sequences of up to 1000 seconds. Similarly, (Brønstad et al., 2021) reported accuracy varying from 89.2% to 100%, using a setup very similar to that of (Marins et al., 2021b).

Figure 3.6 – Training and validation accuracy for the binary LSTM hydrate classifier.



Source: Author (2025)

A key distinction in our approach is that instead of using only normal samples as the negative class, our model classifies all anomalies other than the class of interest as negative, and only the target anomaly as positive. This means that our model must differentiate one anomaly type from a wide range of operational conditions. The validation process relies entirely on real-world cases, reinforcing the model's applicability to real operations. Additionally, it is limited the sensor inputs to TPT and PDG, which are among the most commonly available sensors in operational settings, rather than incorporating all available sensor data. Despite these constraints, our model achieves accuracy comparable to previous works while prioritizing flexibility for real-world applications.

Table 3.2 – Metrics for each sequence length.

Seq Length (s)	**Accuracy**	**Precision**	**Recall**	**F1-Score**
150	0.866	0.863	0.898	0.881
300	0.891	0.883	0.928	0.905
600	0.891	0.891	0.918	0.904
900	0.902	0.885	0.952	0.917
1200	0.898	0.893	0.937	0.915
1500	0.907	0.911	0.935	0.923
2250	0.899	0.923	0.910	0.917
3000	0.930	0.949	0.941	0.945
3750	0.961	0.958	0.985	0.971
4500	0.920	0.895	0.997	0.943
5250	0.962	0.977	0.966	0.971
6000	0.869	0.968	0.832	0.895
6750	0.942	0.922	1.000	0.959
7500	0.966	0.958	0.994	0.975

Even with an accuracy exceeding 96.6%, it is important to consider the findings of Vargas et al. (2019), where hydrate formation detection response times ranged between 30 minutes and 5 hours. This highlights the significance of the 900-second sequence length in our study. This interval allows for detecting transient states within 15 minutes, making the approach viable for practical applications. This conclusion is supported by the fact that validation was conducted solely on unseen real data and utilized only the two most commonly present sensors across study cases.

The validation of the binary ensemble classifier is based on its measured accuracy against real data. As shown in Figure 6 for hydrate detection and Table 3.2 for all anomaly types, a sufficient time window of sensor data leads to high-confidence anomaly detection. Specifically, selecting a 15-minute time window presents 90% accuracy for the hydrate class. Notably, this detection time is still below the minimum response window of 30 minutes, emphasizing the model's potential for timely anomaly identification in operational settings.

3.3.2 Verification of the Binary Classifiers Using Analytical Formulation

A crucial metric in evaluating the performance of the binary classification networks, especially compared to the PVT approach, is their ability to not only improve performance but also minimize false positives and false negatives. Table 3.3 shows the superiority of the binary classifier over the analytical method in this regard.

The binary classifier achieves an accuracy of 90.02%, significantly outperforming the analytical approach, which reaches 76.60%. Additionally, it demonstrates a lower false positive rate (16.8%) compared to the analytical method (18.61%), indicating fewer incorrect hydrate formation predictions. More importantly, the classifier substantially

Table 3.3 – Comparison of Binary Classifier and Analytical Method Metrics

Metric	Binary Classifier	Analytical Method
Accuracy	90.02%	76.60%
False Positive Rate	16.80%	18.61%
False Negative Rate	4.80%	20.10%

reduces the false negative rate to 4.8%, whereas the analytical method records 20.10%, meaning fewer missed hydrate cases.

3.3.3 Evaluation of Model Performance and the Potential of Binary One-vs-All Classifiers

The results in Table 3.4 demonstrate the advantages of the LSTM Binary One-vs-All Classifier compared to LSTM Multiclass, One-Class SVM, and Marins et al. (2021) models. The one-vs-all approach, where each classifier distinguishes a specific anomaly from all other states (including normal conditions), achieves a global accuracy of 0.87, outperforming LSTM Multiclass (0.86) and significantly surpassing One-Class SVM (0.65). The tables also show that the normal state class presents inferior accuracy when compared to the other classes.The initial hydrate detection system employs a dual strategy approach, effectively combining two distinct yet complementary methods. This approach is conceptualized in Figure 7, which illustrates how the components contribute to the overall functionality of the system.

Table 3.4 – All Metrics for the different models: LSTM Binary One-vs-All Classifier, LSTM Multiclass, One-Class SVMs, Marins et al. (2021) Binary (One-vs-One) and Marins et al. (2021) Multiclass.

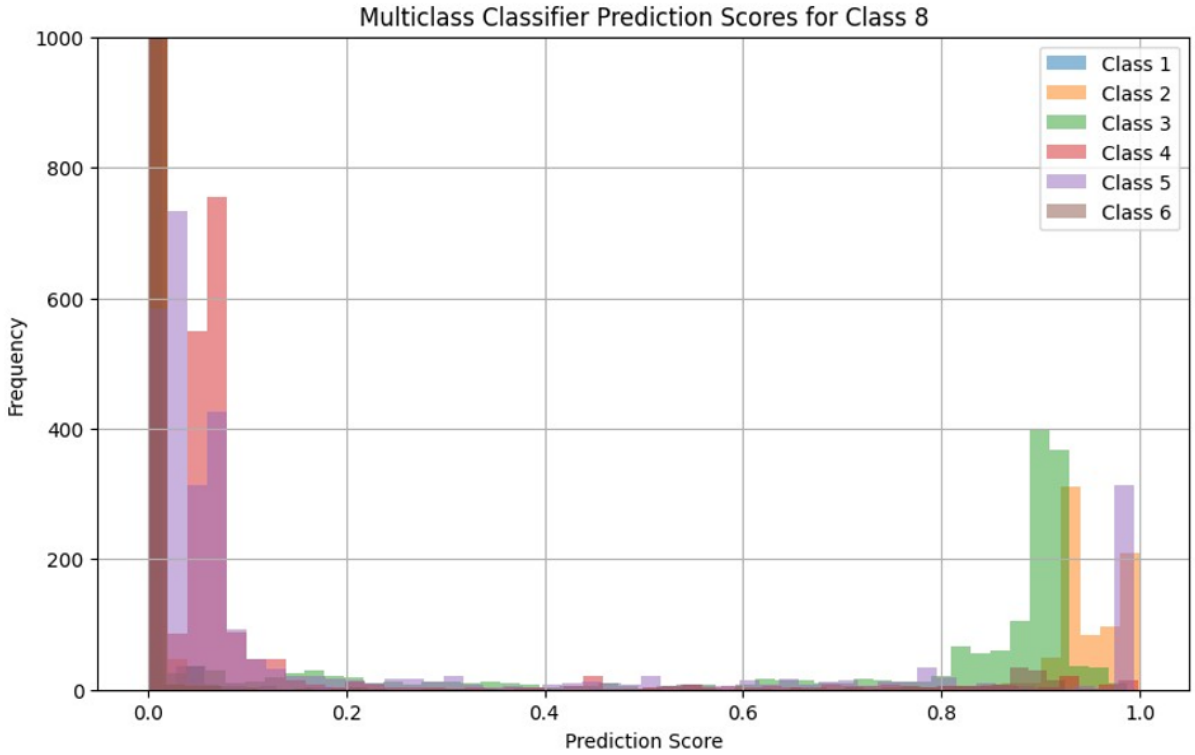
Class	Binary	Multiclass	OCSVM	Marins Binary	Marins Multiclass
Class 0	0.73	0.71	-	-	-
Class 1	0.91 (0.77 $\pm$ 0.07)	0.92	0.63	0.99	0.51
Class 2	0.92 (0.85 $\pm$ 0.04)	0.92	0.57	0.99	0.88
Class 3	0.94 (0.93 $\pm$ 0.04)	0.87	0.67	1.00	0.79
Class 4	0.65 (0.61 $\pm$ 0.02)	0.76	0.64	0.99	0.95
Class 5	0.79 (0.83 $\pm$ 0.04)	0.95	0.69	0.98	0.83
Class 6	0.91 (0.81 $\pm$ 0.11)	0.61	0.63	0.97	0.71
Class 7	0.78 (0.72 $\pm$ 0.03)	0.87	0.65	-	-
Class 8	0.90 (0.70 $\pm$ 0.10)	0.99	0.74	0.99	-
Class 9	0.93 (0.80 $\pm$ 0.06)	0.98	0.63	-	-
Global	0.87 (0.78 $\pm$ 0.05)	0.86	0.65	0.98	0.93

Compared to Marins et al. (2021b) Binary Classifiers, which frames the problem as anomaly vs. normal detection, the one-vs-all method provides finer granularity in classification while maintaining competitive accuracy (0.98 vs. 0.87 globally). Although Marins et al. reach higher scores for some individual anomalies, their binary anomaly detection

approach does not differentiate between different anomaly types, which is essential for operational decision-making.

The one-vs-all ensemble proves more effective than multiclass classification, as it avoids misclassification caused by feature overlaps between anomaly types. Figures 3.7 and 3.8 show that, unlike multiclass classifiers, it does not force unknown data into a predefined category, making it better suited for real-world applications. In the study of both Figures, the multiclass model and the ensemble binary model are trained with 6 classes and the hydrate class (class 8) is used as unknown class. The tested data is the same in both Figures. Figure 3.7 shows that the hydrate class is classified as one of the 6 other classes while Figure 3.8 shows that, for the ensemble binary model, almost all data is classified as none of the classes.

Figure 3.7 – Multiclass trained for 6 classes classifying class 8 in one of the trained classes.



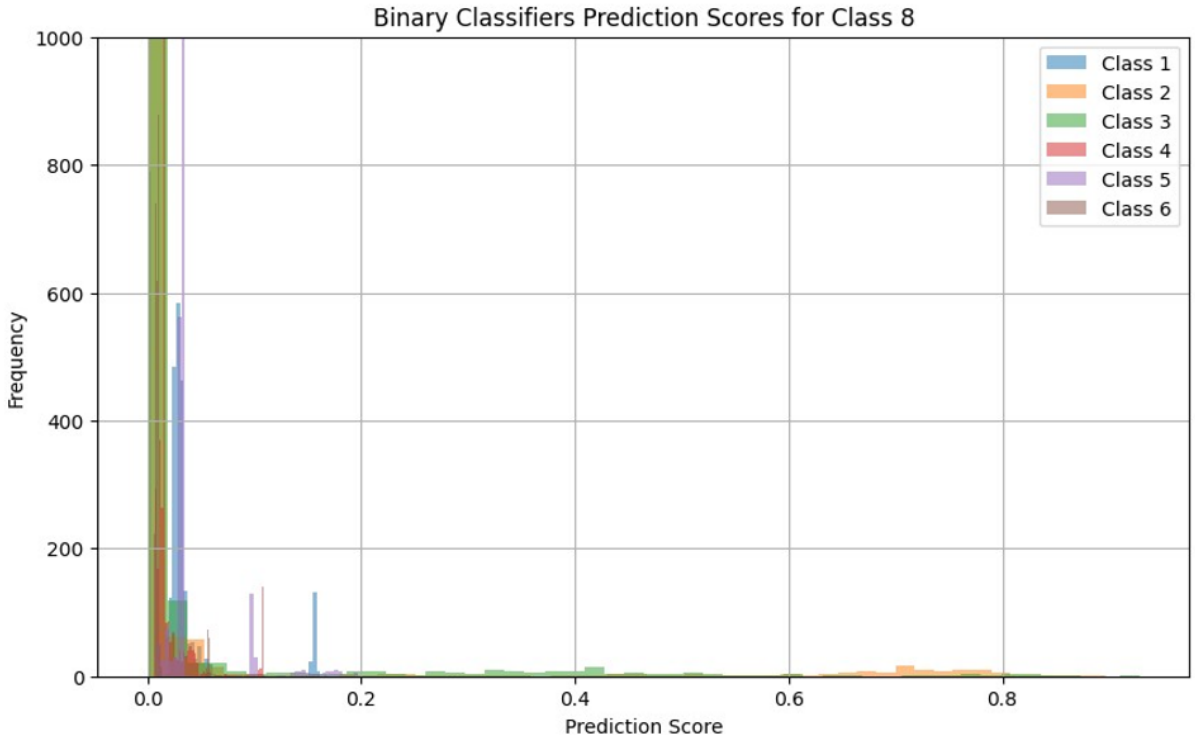
Source: Author (2025)

These results validate the binary one-vs-all strategy as a robust and practical alternative for oil well anomaly detection, offering greater interpretability and flexibility while ensuring strong overall performance.

3.3.4 Applications of the DNN Binary Classifiers

One interesting application proposed strategy is a mock-up real operation monitoring system. Our DNN classifier for the hydrate case can use sensor parameters for pressure and

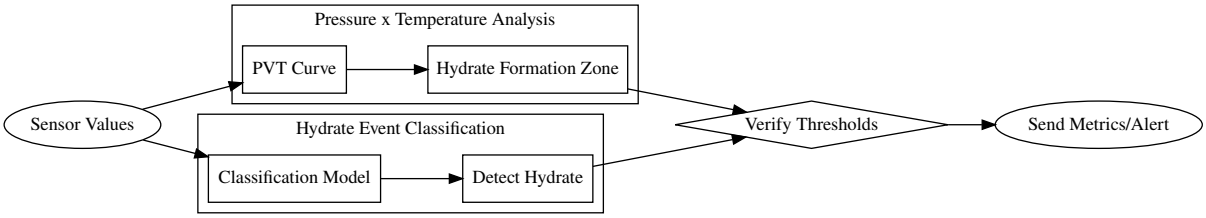
Figure 3.8 – Multiclass trained for 6 classes classifying class 8 in one of the trained classes.



Source: Author (2025)

temperature in real-time analysis. The hydrate detection system employs a dual-strategy approach, effectively combining two distinct yet complementary methods. This approach is conceptualized in Figure 3.9, illustrating how the components contribute to the overall functionality of the system.

Figure 3.9 – Adopted strategy for hydrate detection

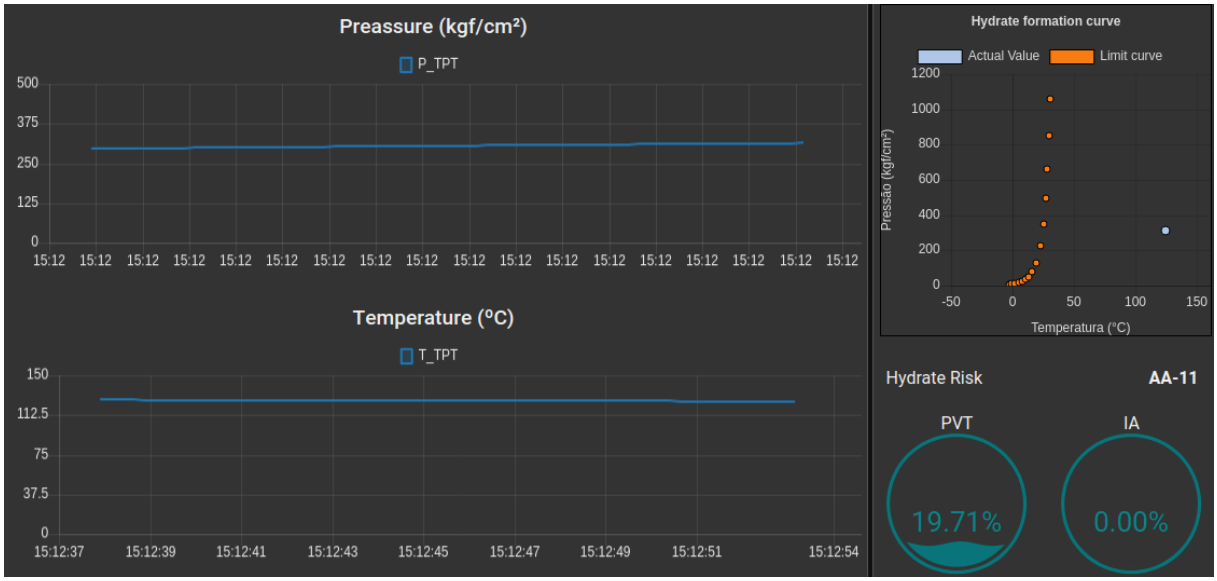


Source: Author (2025)

The first component of the mock-up system is the application of a characteristic pressure-temperature curve derived from PVT analysis, and the second component is the DNN classifier, utilizing the Long Short-Term Memory (LSTM) architecture. Both components are designed to process the sensor data, evaluating the potential risk of hydrate formation. A dashboard developed in Node-RED (Node-RED, 2019) consolidates the outcomes of

methodology, creating a system that can be fed with real-time data, processing it, and alerting operators to potential hydrate events during oil well operations, as shown in Figure 3.10.

Figure 3.10 – Hydrate formation detection system.



Source: Author (2025)

This system also categorizes the likelihood of hydrate formation events, combining PVT analysis with a LSTM binary classifier. By providing operators with a probability assessment of hydrate formation using LSTM deep neural networks and determining if sensor data falls within the danger zone of the pressure-temperature curve, the system offers a comprehensive approach to managing hydrate risks in oil wells.

3.4 Partial Remarks

This study presents the development and evaluation of an LSTM-based one-vs-all ensemble of binary classifiers for oil well anomaly detection. Unlike previous approaches, such as the binary anomaly vs. normal classifier by Marins et al. (2021), our method provides a more refined classification by distinguishing each specific anomaly from all others, including the normal state. The results validate this one-vs-all strategy as an effective alternative to multiclass classification, overcoming feature overlap issues while maintaining strong performance across all anomalies.

The binary one-vs-all ensemble achieves a global accuracy of 0.87, outperforming LSTM Multiclass (0.86) and significantly surpassing One-Class SVM (0.65). This method demonstrates that training separate binary classifiers for each anomaly improves classification accuracy while also allowing for better detection of unknown classes, as opposed to a

forced classification approach used in multiclass models. Additionally, the system relies on TPT and PDG sensor data, ensuring applicability to real-world operations, where sensor availability may be limited.

Benchmarking against analytical formulations further reinforces the superiority of the binary classifiers, with 90.02% accuracy compared to 76.60% for the analytical method, alongside a lower false positive rate (16.8% vs. 18.61%) and a significantly reduced false negative rate (4.8% vs. 20.10%), making it a more reliable and precise tool for real-time anomaly detection.

By validating results on real oil production data, this study confirms that a one-vs-all binary ensemble is a practical and reliable solution for anomaly detection in oil wells. The findings also suggest that implementing an initial anomaly vs. normal classifier before deploying specialized anomaly classifiers could further enhance detection performance, supporting strategies proposed by Aranha et al. (2024).

4 Detection and Classification of Anomalies in Oil Well Production Using an Open-World Learning Strategy

This chapter presents the first Open-World Learning strategy applied to anomaly detection in oil well production data. The proposed strategy detects anomalous behavior, determines whether it belongs to a known labeled anomaly, and, if not, clusters it into newly proposed anomaly classes while continuously learning to classify them.

The approach integrates autoencoder reconstruction error for anomaly detection, autoencoder-based dimensionality reduction to extract latent features, binary classifiers to identify known anomalies, and clustering methods to group similar unseen anomalies. When an anomaly is detected based on reconstruction error, the binary classifiers assess whether it belongs to a known class. If it does not, the clustering method groups similar events into new classes, which are then validated by human experts. This validation step allows for the training of specific binary classifiers for the newly identified classes and updates existing models accordingly.

Experiments on real anomalous oil well production data demonstrate that the discovered clusters align well with ground-truth labels. The clustering methodology achieves an overall sample level accuracy of 81%, exceeding 95% for certain anomalies, while updated binary classifiers reach up to 99% accuracy. These results highlight the proposed method's effectiveness in dynamically adapting to novel anomalies, improving classification accuracy, and enhancing oil well monitoring.

4.1 Overview

The detection of anomalies is essential to maintain safety and efficiency in oil well operations. Advanced AI techniques, such as autoencoders (as discussed in Chapter 2) and clustering algorithms, have been widely used to detect and classify operational anomalies from sensor data, such as pressure, temperature, and flow rates (Vargas et al., 2019b; Aranha et al., 2024a). Although other machine learning methods perform well on known anomalies, they struggle to handle unseen anomaly types that often emerge due to evolving operational conditions. Novelty detection methods, such as clustering (de Salvo Castro et al., 2021) and boundary setting techniques (Pimentel et al., 2014), address this challenge by identifying data points that deviate from known patterns. However, these approaches fail in dynamic systems, such as oil wells, where newly discovered anomalies must be integrated into the

detection process.

To overcome these limitations, this Chapter presents an Open-World Learning (Chen and Liu, 2018) approach that combines anomaly detection with adaptive classification strategies for both known and unknown cases. The proposed method first detects anomalies using autoencoder reconstruction error. If an anomaly is detected, the sensor data is transformed into a lower-dimensional representation using the autoencoder's encoder. The encoded output is then processed by binary classifiers—dense models assisted by One-Class Support Vector Machines (OCSVM)—which help distinguish between known and unknown anomaly types. If the data does not belong to a known class, it is treated as an unknown anomaly. To manage these cases, DBSCAN (Deng, 2020) and Meanshift (Comaniciu and Meer, 2002) are used to identify noise, while a random forest model estimates the number of potential new anomaly classes. Using this estimate, a combination of K-means and Meanshift clustering is applied to group similar data into potential new anomaly classes. These newly identified classes undergo human validation to refine cluster assignments, register new anomaly types, and update binary classifiers, enabling continuous learning.

Additionally, this Chapter introduces and validates the entire anomaly detection pipeline on a new type of anomaly: the downhole interval control valve (ICV) fault. This dataset has not been previously used to validate Open-World Learning in oil well monitoring, making it a valuable test case for evaluating the adaptability and effectiveness of the proposed approach in real-world scenarios.

This Chapter presents the main results of this thesis and advances anomaly detection in the oil and gas sector by providing an adaptive Open-World Learning strategy, ensuring continuous improvement through structured validation, cluster refinement, and incremental classifier updates. This approach bridges the gap between the detection of known anomalies and the discovery of unknown types, enhancing operational safety and efficiency.

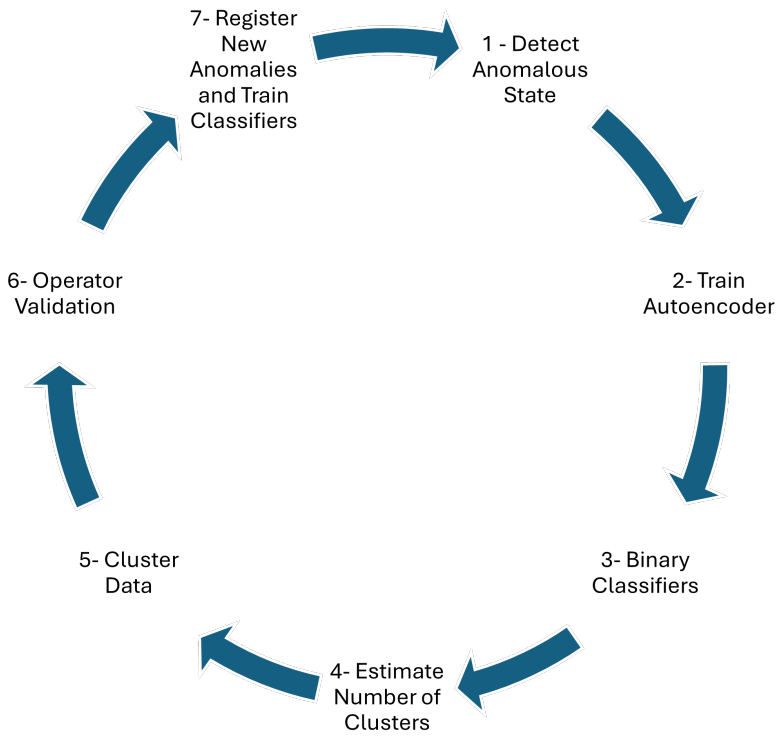
4.2 Proposed Methods and Studies

This Chapter presents a methodology for discovering and classifying unknown anomalies in oil wells through a multistep process that integrates autoencoder-based anomaly detection, binary classifiers for known anomalies, and clustering techniques for unknown anomaly discovery. The approach builds on previous work in oil well anomaly detection while introducing several novel contributions to enhance system adaptability and robustness. Specifically, the methodology incorporates an operator-driven validation system to dynamically register new anomalies, a hybrid K-means and Meanshift clustering approach to group unknown anomalies, and a new application of the encoder's latent space to efficiently represent anomalies for clustering. This system, designed within the Open-World Learning methodology, enables continuous learning and adaptation in oil well monitoring.

The methodology, illustrated in Figure 4.1, operates as an iterative feedback loop where each detected anomaly feeds into the next stage. First, anomalies are detected using an autoencoder model trained exclusively on normal data. The anomaly detection process leverages reconstruction error to identify deviations from the expected operational patterns learned during training.

Once an anomaly is detected, the anomalous data is passed to binary classifiers, which attempt to categorize it into known classes. The binary classifiers operate on the encoder output, derived from an autoencoder trained on both normal and anomalous data, enabling effective feature extraction and dimensionality reduction. The primary classifiers consist of dense deep learning models, supplemented by One-Class Support Vector Machines (OCSVMs) for handling borderline cases and improving generalization. If the anomaly does not match a known class, it is treated as unknown and forwarded for clustering.

Figure 4.1 – Flowchart of the proposed methodology.



Source: Author (2025)

To organize unknown anomalies into potential new classes, the system clusters the encoder's latent representation of the data. First, a combination of Meanshift (Comaniciu and Meer, 2002) and DBSCAN (Deng, 2020) is used to determine whether the data can be grouped or if it consists only of noise. Then, the clustering process is executed using a combined approach of K-means and Meanshift. The number of clusters is estimated using a Random Forest regressor, which learns from previous anomaly distributions to provide an adaptive cluster estimation. Validation results (Subsection 5.4.3) show that

while K-means often produces better clustering results, some anomalies fit better with the Meanshift algorithm due to its non-parametric nature. By overlapping the results of both techniques and selecting the best fit based on clustering quality metrics, the proposed methodology identifies candidates for newly discovered classes that align more closely with ground-truth anomaly patterns.

The clustered data represent candidate new anomaly classes, which undergo operator-driven validation. This human-in-the-loop process allows experts to confirm or refine cluster assignments, ensuring the accuracy and relevance of newly identified anomaly types. Importantly, this validation process forms a closed-loop feedback system in which newly labeled anomalies are incorporated into the training of new binary classifiers, expanding the system's ability to detect them in the future. At the same time, existing binary classifiers are updated with these newly identified negative samples, improving their robustness against false positives. This continuous reintegration of labeled anomalies into the classification pipeline enhances the adaptability of the Open-World Learning system, allowing it to progressively learn and improve its detection of novel anomaly types in real-world oil well monitoring.

4.2.1 Anomaly Detection with Autoencoder Reconstruction Error

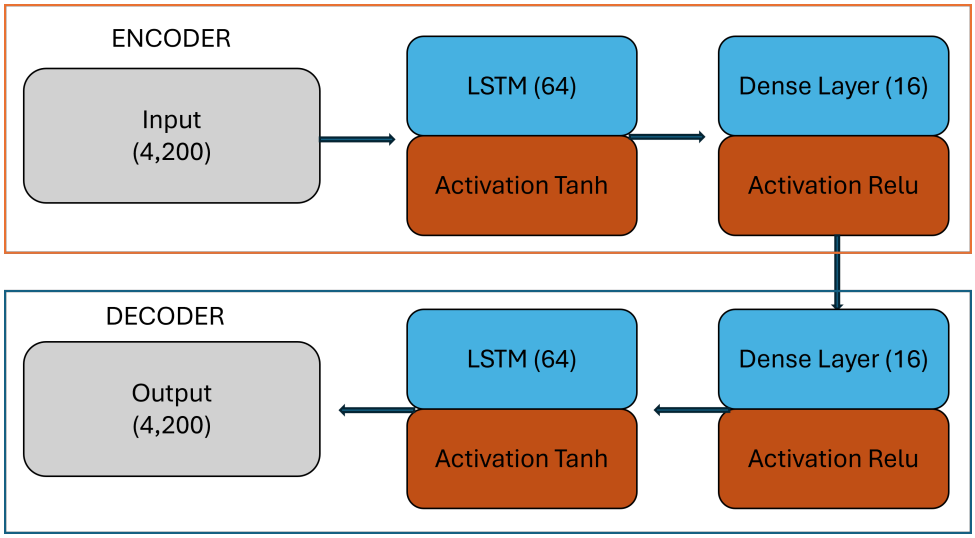
The first stage of the methodology is anomaly detection using an autoencoder (Chen et al., 2018). As in Aranha et al. (2024a), the autoencoder is trained using normal operational data from the 3W dataset (Vargas et al., 2019b), as described in Chapter 2.

Specifically, and inspired by Aranha et al. (2024a), the input data consists of sequences of 200 time steps from four sensors: temperature and pressure readings from the Pressure Downhole Gauge (PDG) and temperature and pressure readings from the Temperature and Pressure Transducer (TPT). These sequences provide a detailed view of well conditions over time, allowing the model to capture the temporal dependencies and patterns associated with normal operations. The inherent parameters of this method are described in Chapter 2.

After the model is trained and the threshold is defined, it is possible to apply the trained autoencoder to detect abnormal behaviors in newly acquired data. Various network configurations are tested, and an LSTM-based network, as proposed by Larzalere (2019), is ultimately selected for its ability to capture temporal dependencies in sequential data. The architecture of this LSTM-based autoencoder is illustrated in Figure 4.2.

Figure 4.2 shows the main layers that compose the trained autoencoder, namely the LSTM layers and the dense layers, the decoder in the first row and the decoder in the second row. Based on the input dimension (4,200), an important hyper parameter that is tuned using greed search from a range between 50% of the original dimension and 1% of this

Figure 4.2 – Autoencoder proposed architecture



Source: Author (2025)

value is the latent space size. The value 16, equivalent to 2% of the original dimension, is obtained after a careful test of the anomaly detection response.

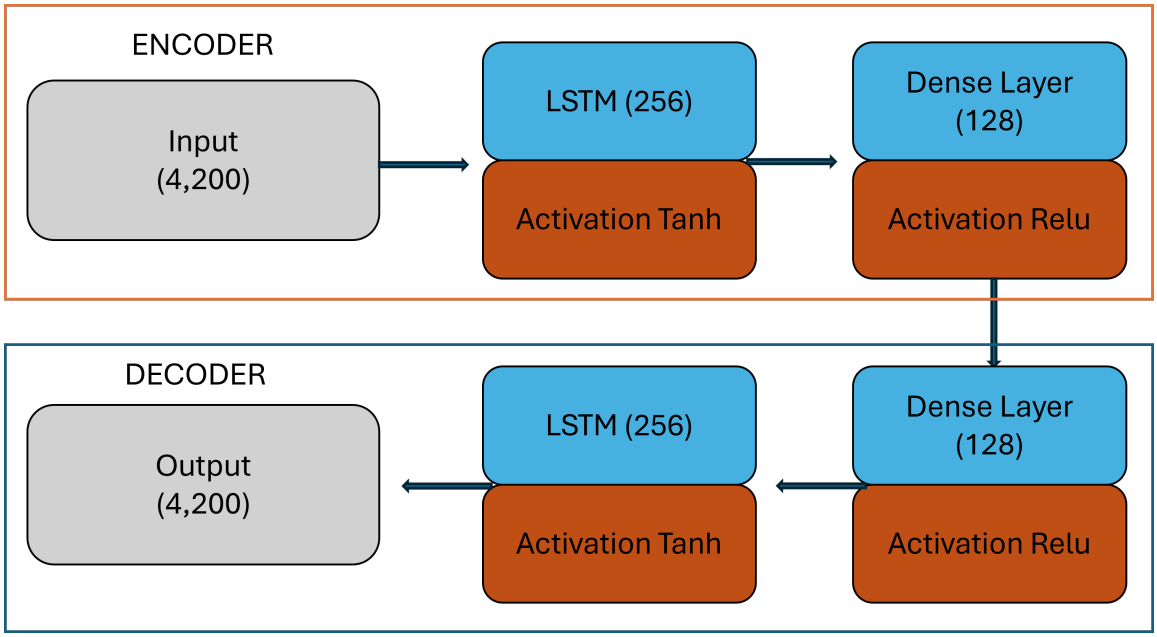
4.2.2 Training an Autoencoder on All Available Data

The second part of the methodology involves training a new autoencoder using all available data, both normal and anomalous, to obtain a meaningful latent representation of the sensor data. The autoencoder is designed to compress the input data into a lower-dimensional latent space that captures essential features representing the operating state of the well. This latent representation, extracted from the encoder, serves as the basis for further analysis, enabling dimensionality reduction while preserving the most informative characteristics of the data (Wang et al., 2019; Marchi et al., 2015).

Obtaining an effective autoencoder for this purpose goes beyond simple hyperparameter optimization tools such as Keras-Tuner (Chollet, 2015), as these tools do not account for the latent space's ability to separate and cluster anomaly classes. Instead, the architecture is chosen based on the tests performed in Lopes et al. (2024), where both binary and multiclass configurations incorporate an LSTM layer and a dense layer. This design is optimized for the characteristics of the 3W dataset, specifically to capture temporal dependencies in sequences. The best architecture is illustrated in Figure 4.3, showing the layers that process sequences of 200 time steps with input from four sensors: temperature and pressure from both the Temperature and Pressure Transducer (TPT) and the Downhole Gauge (PDG). Similar to Figure 4.2, Figure 4.3 depicts the main LSTM and dense layers, with the encoder in the first row and the decoder in the second row.

The model from Lopes et al. (2024) utilized 900-second sequences that focused on hydrate

Figure 4.3 – Autoencoder main architecture



Source: Author (2025)

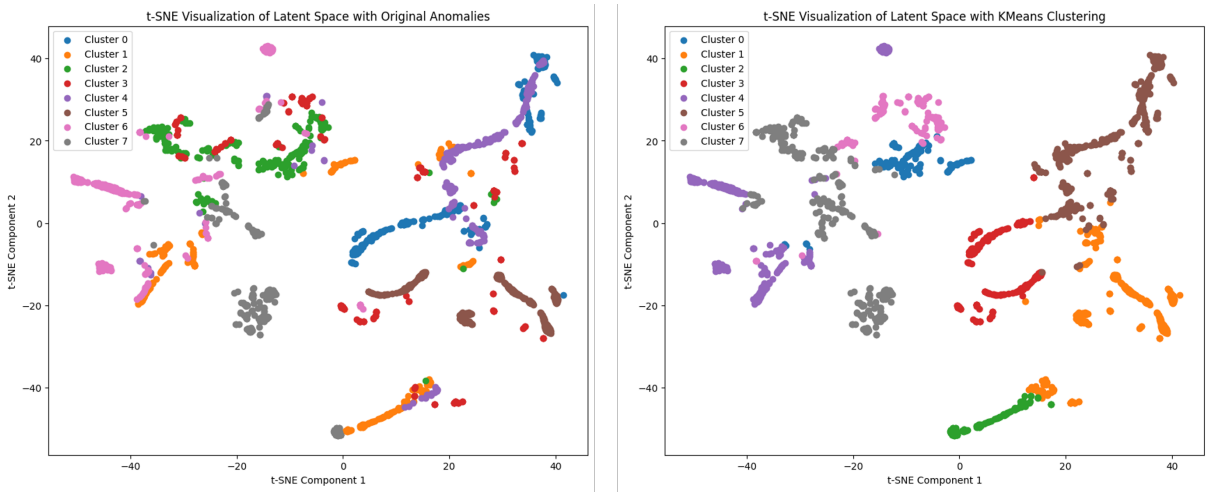
anomalies. However, to encompass a broader range of anomalies in this study, the length of the input sequence is adjusted to 200 seconds, integrating additional sensor data (PDG temperature and pressure) to maintain a comprehensive representation of various anomaly types.

The resulting latent representations from the encoder are then used for clustering. Figure 4.4 compares the original eight labeled anomaly types from the 3W dataset (left) with the clustering results of K-means applied to the latent space (right). In the left visualization, each cluster represents a distinct labeled anomaly from the 3W dataset, while in the right visualization, each cluster is created by K-means clustering. This comparison highlights the clustering potential of the latent space, as clusters derived from K-means exhibit a separation pattern that approximates the distribution of the original labeled anomalies.

In the left plot, each color corresponds to a distinct labeled anomaly from the original dataset, visualized in a t-SNE plot that reflects how different anomaly types are distributed throughout the latent space. On the right, the same t-SNE visualization shows the clustering results of K-means applied to the latent representation, with colors representing the clusters formed by the algorithm. A close correspondence between the colors on the left and right plots indicates that the latent space effectively captures the characteristics of each anomaly class, even in an unsupervised setting.

Autoencoders are valuable for reducing dimensionality and isolating important features in the data, which facilitates the clustering of similar anomaly classes. This approach could be further enhanced by exploring alternative architectures such as Variational Autoencoders

Figure 4.4 – Comparison of the original eight anomalies from the 3W dataset and the K-means clustering of the latent representation of the encoder.



Source: Author (2025)

(VAEs) (Xu et al., 2018) and Deep Cluster Networks (DCNs) (Ren et al., 2022), which offer variations in the structure and distribution of the latent space. By applying the same autoencoder architecture but with VAE or DCN configurations, different latent space representations could be obtained, that could lead to different clustering performance and providing additional insights into anomaly separability in the latent space.

When using other losses metrics to obtain more regular distributions for the latent space using Variational Autoencoders and Deep Clustering Networks, reaching similar accuracy, as shown in Table 4.1.

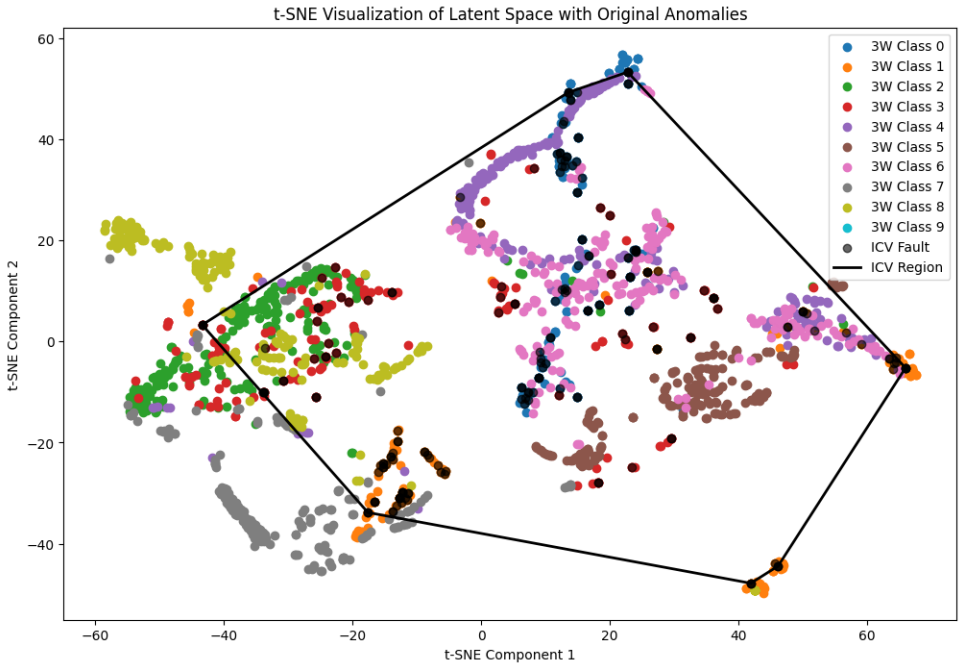
Table 4.1 – Autoencoders, VAE and DCNs Accuracy Comparison

Network	Accuracy
VAE	0.727
DCN	0.725
AE	0.73

Regular autoencoders are slighly better in these experiments, but there are othe benefits than using the Variational Autoencoders and Deep Clustering Networks. Autoencoders have an interesting property that they are mainly weak in reconstructing new unseen data, placing them at the edges of the latent space as presented in Figure 4.8. In other hand, both VAE and DCN changes the latent space in order to accommodate the new data into clusters (DCN) or to regularize and place the data along with the know data (VAE). Both behaviors result in similar problems, as seen in Figure 4.5.

Figure 4.5 presents a t-SNE visualization of the latent space produced by a deep clustering network (DCN), showing how the network organizes known and anomalous data. Different

Figure 4.5 – DCN latent space clustering



Source: Author (2025)

colors represent distinct classes, while the black polygon outlines the ICV (Inter-Class Variability) region. The ICV data are spread in the known clusters, so it is difficult to separate them from the other classes in a single new cluster.

4.2.3 Binary Classifiers for Known Anomalies

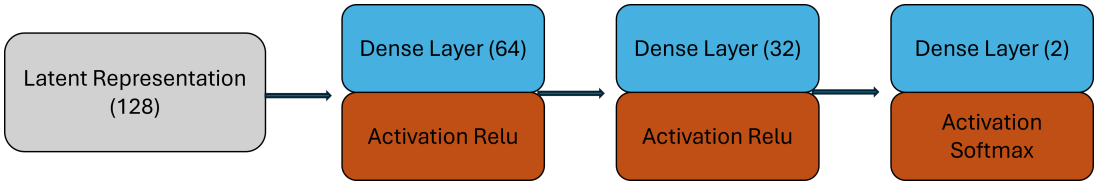
To determine whether a detected anomaly corresponds to a known event, binary classifiers are pre-trained for each known anomaly class using all available labeled data. The autoencoder, including its encoder component, is also pre-trained with this data to capture essential features of both normal and anomalous patterns. This pretraining phase ensures that both the autoencoder and the binary classifiers are fully prepared for evaluation before any real-time detection begins.

During training, the autoencoder learns to reconstruct production operational data—both anomalous and normal—while the encoder captures a compressed latent representation of the input. The extracted latent features serve as input for the binary classifiers, each designed to recognize a specific anomaly type (e.g., hydrate formation or flow instability). The main classifiers use the latent representation from the encoder, which is fed directly into a series of dense layers forming the binary one-vs-all classifier network.

The proposed binary classifier methodology builds directly on Lopes et al. (2024), where binary one-vs-all classifiers were similarly trained for specific anomaly types—such as hydrate anomalies—using LSTM layers for temporal feature extraction followed by dense

layers for classification. In the present work, the binary classifiers use the pre-trained encoder as a feature extractor, with its output passed to a dense layer network, as shown in Figure 4.6.

Figure 4.6 – Binary classifier architecture used to identify specific known anomalies.



Source: Author (2025)

Figure 4.6 illustrates the primary dense layers of the binary classifiers, excluding support layers such as dropout.

Alongside the one-vs-all classifiers, this study also employs auxiliary methods, such as One-Class Support Vector Machines (OCSVMs), to address certain limitations, which are further discussed in the results section. In summary, each classifier outputs a binary decision: 1 if the anomaly matches the specific class it was trained to detect, and 0 for all other classes, including unknown anomalies.

Training the binary classifiers involves labeling each data point as positive (belonging to a specific anomaly class) or negative (representing all other known anomalies, excluding normal data). By excluding normal data from binary classifier training, it is designed that these classifiers focus solely on distinguishing between different types of anomalies, as normal data is already filtered out by the autoencoder reconstruction error during the anomaly detection stage. In cases where the reconstruction error detects a new anomalous state, the data is evaluated by all binary classifiers. If any classifier outputs a positive decision, the anomaly is categorized as the corresponding class. If none of the classifiers identify it as a known anomaly, the data is flagged as an unknown anomaly, leading to further clustering analysis.

Both the autoencoder and the binary classifiers are retrained after unknown anomaly data has been discovered, clustered, and introduced as new classes through the operator validation process. This approach ensures that the models remain up-to-date with the evolving anomaly landscape without requiring frequent retraining, thereby maintaining system stability and efficiency.

4.2.4 Estimating the Number of Anomaly Clusters

The first step in estimating anomaly clusters is to assess whether the data has sufficient structure to form clusters or whether it primarily consists of noise. This is achieved using the Meanshift (Comaniciu and Meer, 2002) and DBSCAN (Deng, 2020) algorithms, with

hyperparameters tuned based on the labeled dataset. DBSCAN is particularly effective for identifying dense clusters and labeling sparse points as noise, while Meanshift clusters data based on density peaks, offering a preliminary check: if these methods detect mostly noise, the data is deemed unsuitable for further clustering; otherwise, the process proceeds to the next stage.

To estimate the number of potential anomaly classes within the detected anomalies, a Random Forest regression model is used. This model, trained on batches of latent representations of known anomalies, predicts the optimal number of clusters based on various clustering metrics. While individual metrics—such as the silhouette score (Shahapure and Nicholas, 2020), Davies-Bouldin index (Davies and Bouldin, 1979), and Calinski-Harabasz score (Caliński and JA, 1974)—each suggest cluster counts, the Random Forest model integrates them to generate a robust prediction, leveraging the unique insights of each metric regarding data structure (Bodesheim et al., 2013; Pimentel et al., 2014).

Each batch is evaluated with these metrics, creating features that serve as input for training the Random Forest model. Key metrics include the silhouette score, Davies-Bouldin index, Calinski-Harabasz score, and density-based measures such as bounding box volume, calculated as the average radius from cluster centers to points in the 128-dimensional latent space. The model's final predicted cluster count combines two values: the average of each metric's optimal cluster count and an entropy-weighted score, refining the estimate:

$$estimated_k = \frac{average_k + entropy_k}{2} \tag{4.1}$$

Once $estimated_k$ is determined, it is used to fit a K-means clustering model on the latent batch data, generating cluster labels based on this estimated number of clusters. Using these labels, additional metrics are computed, including the silhouette score, Davies-Bouldin score, Calinski-Harabasz score, and K-means inertia. These values, along with the number of points in the batch and the bounding box volume, are compiled as feature inputs for training the Random Forest model. The bounding box volume is calculated as the average radius from the center of each cluster to its points in the 128-dimensional latent space.

The final set of input features for each batch is summarized in Table 4.2. This combination of features allows the Random Forest Regressor to integrate multiple clustering metrics, each capturing different aspects of the data structure, resulting in a reliable prediction of the cluster count.

The table above summarizes the key clustering metrics used as input features for the Random Forest Regressor. Each metric captures a distinct characteristic of the clustering structure, providing a comprehensive view of the data attributes. The "Number of Points"

Metric	Description
Number of Points	Number of points in the batch.
Silhouette Score	Cluster cohesion and separation.
Davies-Bouldin Index	Inter-cluster distance ratio.
Calinski-Harabasz Score	Compactness and separation.
Inertia	Squared distances to cluster centers.
Average K	Average optimal cluster number.
Entropy K	Entropy-weighted estimate for K.
Bounding Box Volume	Average cluster radius.

Table 4.2 – Summary of clustering variables used for the Random Forest model.

reflects the size of the dataset batch, serving as a baseline for scale. The "Silhouette Score" and "Davies-Bouldin Index" focus on cluster quality by measuring cohesion, separation, and inter-cluster distance. Similarly, the "Calinski-Harabasz Score" evaluates compactness and separation. The "Inertia" metric quantifies the total squared distance of points to their cluster centers, highlighting overall cluster tightness. Advanced metrics like "Average K" and "Entropy K" provide insights into the optimal cluster count and its weighted distribution, while the "Bounding Box Volume" estimates the spatial extent of clusters. Together, these features ensure a multidimensional approach to predicting the optimal cluster count.

4.2.5 Clustering of Unknown Anomalies

The final stage clusters the latent representations of unknown anomalous data to identify potential new anomaly classes. This process combines the clustering algorithms K-means and Meanshift (dong Qi, 2013) to balance a predefined structure with flexible density-based grouping. K-Means uses the estimated number of clusters provided by the Random Forest model, while Meanshift relies on tuned hyperparameters to detect clusters based on density peaks without a fixed cluster count (Comaniciu and Meer, 2002).

After applying K-means and Meanshift clustering, a third clustering step is performed based on the intersection of the clusters produced by both methods. To further refine these clusters, a distance threshold is introduced. This threshold retains only the points within a specified distance from the cluster center, thereby reducing the impact of outliers. The thresholding process is inspired by the distance-based rejection mechanism of Open-Set Nearest Neighbor (OSNN) (Zhu et al., 2024b). This intersection-based clustering approach improve the resulting clusters to represent the core structure of the data.

The distance threshold is dynamically determined to account for variations in cluster point cloud distributions. The results are inspired by the statistical threshold from Aranha et al. (2024a), and slightly modified after studying the results on the validation experiments presented in Section 5. First, the mean (μ) and standard deviation (σ) of the distances

from the center of the cluster are calculated. The threshold (T) is computed as:

$$T = \mu + k \cdot \sigma \tag{4.2}$$

where k is a dynamically determined scaling factor based on the ratio between the mean and standard deviation:

$$R = \frac{\mu}{\sigma} \tag{4.3}$$

The scaling factor k is dynamically determined as:

$$k = \begin{cases} 2, & \text{if } R \geq 4, \\ 1 + \frac{R-2}{2}, & \text{if } 2 \leq R < 4, \\ 0, & \text{if } R < 2. \end{cases} \tag{4.4}$$

This dynamic adjustment ensures that the threshold adapts to the characteristics of the point cloud distribution. For $R \geq 4$, the distribution is considered highly normal, allowing the threshold to include most points. For $2 \leq R < 4$, k is linearly interpolated between 1 and 2. For $R < 2$, the threshold is equal to the mean. If the threshold removes more than two-thirds of the points, the closest two-thirds points to the cluster center are kept. By maintaining at least two-thirds of the points within the threshold, this approach balances outlier exclusion with good cluster refinement.

Each clustering configuration (from K-means, Meanshift, and the intersection) is then evaluated using clustering metrics, such as silhouette score, Davies-Bouldin index, and Calinski-Harabasz score. A voting mechanism based on these metrics determines the best clustering structure, with the majority vote indicating the final choice. This combined clustering approach enables the strategy to identify distinct patterns within the unknown anomaly data, facilitating the discovery of potential new anomaly classes. Detailed results and comparisons with literature benchmarks are presented in Section 4.3.

4.2.6 New Classes Validation, Continuous Learning and Virtual Classes

The proposed strategy includes a validation process for newly clustered anomalies, which is important for maintaining an accurate and adaptable monitoring system. Validated classes are used to train new binary classifiers, allowing the system to recognize these anomalies in future operations, similar to the methods of Aranha et al. (2024a) and Lopes et al. (2024). To facilitate this validation, algorithms are developed to assist human operators by eliminating problematic points, regrouping points into other clusters, and creating new clusters as needed. Although this human intervention is important at this stage, it could later be replaced by models or automated logic to streamline the validation process.

It is possible to bypass human validation and use the newly proposed classes as they are clustered. However, this can impact the number of new classes, as human operators can identify when separate clusters actually belong to the same anomaly. Without validation, multiple classes may be created for a single anomaly category. Another consequence is that informative names cannot yet be assigned to the proposed clusters, requiring them to be identified by surrogate labels. Therefore, while the proposed method can continuously learn virtual classes, it does so with these limitations, allowing the system to function in real-time while awaiting class validation.

4.2.7 Training and Updating Binary Classifiers

Once new anomaly classes are validated, the system must update its binary classifiers to recognize these classes in future operations. This process consists of two key steps: (1) training new classifiers for each newly validated anomaly class and (2) updating existing binary classifiers to improve their ability to distinguish between similar anomalies.

Newly validated anomaly classes serve as examples for training new binary classifiers and the supporting OCSVM models. Each binary classifier is trained in a one-vs-all fashion, where the new class is treated as the positive class and all other anomaly and normal data serve as negative examples. The autoencoder's latent space representations are used as input features, ensuring that the new classifiers are aligned with the strategy's existing feature extraction pipeline. This enables seamless integration of the new classifiers into the broader anomaly detection system.

Once a new class is introduced, existing classifiers must be retrained to prevent misclassification. For example, if a previously unknown anomaly shares similarities with a known anomaly, the binary classifier for the known anomaly may need adjustment to avoid mislabeling the new class as part of the existing one. To achieve this, previously trained classifiers are fine-tuned using a combination of old and new labeled data, incorporating the newly validated class as a negative example where necessary.

The iterative cycle of anomaly detection, classification, clustering, validation (or adopting virtual classes), and retraining classifiers enables the system to continuously learn and evolve. With each validation round, newly discovered anomalies are incorporated into the monitoring system, improving its robustness and its ability to detect both known and unknown anomalies in a dynamic environment. This ensures that the system remains adaptable to new operational conditions and can progressively refine its anomaly detection capabilities over time.

4.3 Results and Discussions

This section presents a comprehensive evaluation of the proposed methodology, encompassing two distinct analyses: a validation analysis on the 3W dataset and a test analysis on the newly introduced Downhole Interval Control Valve (ICV) fault data. The new data are extracted from nine production oil wells—specifically, anonymized sensor data from wells 1 to 9 available in the supplementary material. In the validation analysis, the 3W dataset anomalies are treated as known classes, allowing us to assess the performance of each stage in the pipeline, from anomaly detection with autoencoders through to clustering and validation.

To provide a clear roadmap for the results and discussions, the following research questions are addressed in each subsection. The following results are structured to respond to specific research questions, facilitating a focused and coherent analysis:

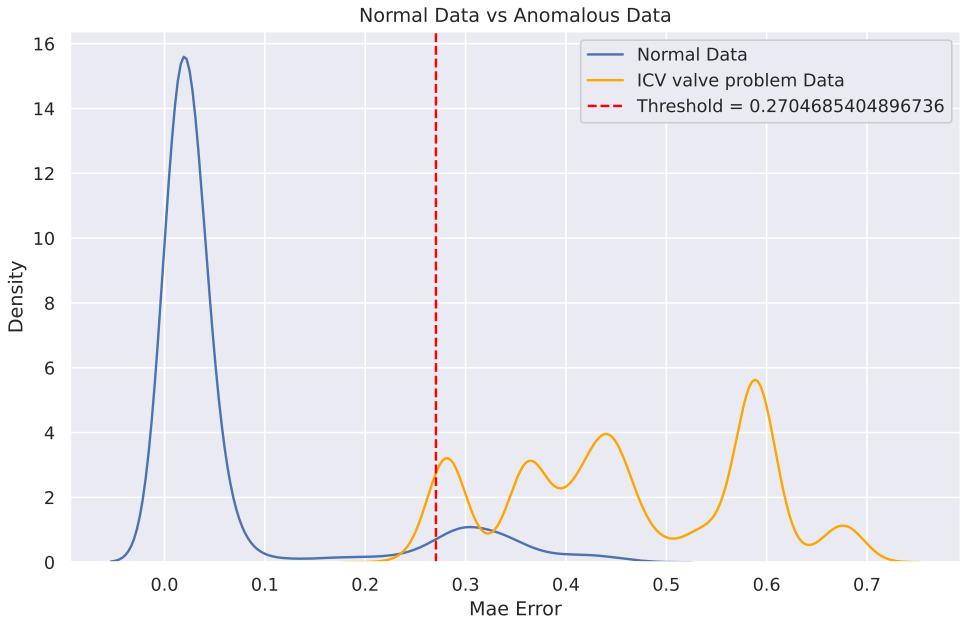
- **Research Question 1:** How effective is the autoencoder reconstruction error in detecting anomalies on new data? (subsection 5.1)
- **Research Question 2:** How well does the latent space representation capture and separate known and unknown anomaly classes? (subsection 5.2)
- **Research Question 3:** Can binary classifiers reliably identify known anomalies and distinguish them from novel ones? (subsection 5.3)
- **Research Question 4:** How effective is the clustering methodology in grouping unknown anomalies and estimating the number of new anomaly classes? (subsection 5.4)
- **Research Question 5:** How well does the system adapt and improve after integrating new anomaly classes into its knowledge base? (subsection 5.5)

These questions are addressed sequentially in the subsections that follow, ensuring a systematic exploration of the performance and adaptability of the proposed methodology. The test analysis with ICV data introduces an entirely new anomaly class, used to evaluate the system's ability to detect, classify, and cluster unknown anomalies within the complete pipeline. This analysis tests the system's adaptability and continuous learning capabilities by incorporating a previously unclassified anomaly type. Performance metrics are reported for both analyses and compared to the relevant literature, as detailed in the subsections that follow.

4.3.1 Autoencoder Anomaly Detection

Autoencoder reconstruction loss is a well-established method for anomaly detection, as demonstrated by Aranha et al. (2024a), which also utilized the 3W Dataset. In this study, the autoencoder model is trained on normal data from both the 3W Datastet and newly acquired real-world data, with a focus on ICV fault detection. Building on the foundation of Aranha et al. (2024a), which presented separate models for each oil well, the proposed approach modifies the autoencoder architecture to create a unified model trained on all available normal data. This single model is then applied to assess both normal and anomalous conditions across multiple wells exhibiting the ICV fault. The current architecture is illustrated in Figure 4.2.

Figure 4.7 – Reconstruction loss threshold applied to the ICV fault data.



Source: Author (2025)

Figure 4.7 presents the data as an interpolated density distribution, where the area under each colored curve sums to one. The x-axis represents the Mean Absolute Error, while the y-axis denotes the density of points corresponding to each error value. The plot demonstrates that normal data predominantly exhibit error values below the probabilistic threshold, whereas ICV fault data present values above this threshold.

After training, an anomaly threshold is determined using the validation set. When applied to reconstruction loss on the normal and ICV fault test data, the threshold effectively distinguishes most normal and abnormal samples, as illustrated in Figure 4.7.

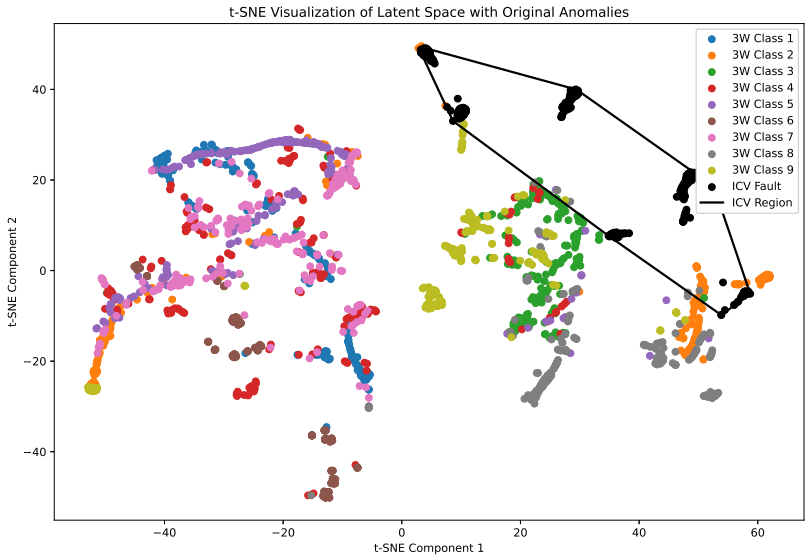
Figure 4.7 visualizes the density of points on the y-axis against the mean squared error of points on the x-axis, providing a smooth interpolation for graphical representation.

Some false positives arise, primarily due to valve opening and closing events, along with a few false negatives. These false positives are one of the key reasons Aranha et al. (2024a) opted to retrain models for specific wells and valve settings. However, within the scope of this Chapter, the results are acceptable, providing a strong initial filter for subsequent processing stages. The presented results, along with those from Aranha et al. (2024a), address Research Question 1, demonstrating that reconstruction error can distinguish normal data from abnormal behavior given a calculated threshold.

4.3.2 ICV Latent Space Representation

A key aspect of the proposed method is reducing the encoder dimension to group similar data together. To illustrate the effectiveness of this approach and address Research Question 2, Figure 4.8 shows the representation of the nine anomaly types from the 3W dataset along with the ICV fault anomaly (highlighted as black points).

Figure 4.8 – Two-dimensional representation of the latent space from nine 3W dataset anomalies and the ICV fault data, highlighting ICV Fault points.



Source: Author (2025)

This Figure illustrates a two-dimensional t-Distributed Stochastic Neighbor Embedding (t-SNE) representation (van der Maaten and Hinton, 2008) of the encoder's latent space. It reveals distinct regions where points corresponding to different anomaly types are concentrated. Anomaly data tend to cluster within a localized region of the latent space. However, it is also evident that anomaly 4, namely Flow Instability, is more widely spread across the latent space. This suggests that this anomaly class is among the most difficult to classify and separate from others.

4.3.3 Binary Classifiers

The binary classifiers are trained on the nine known anomalies of the 3W dataset and tested with ICV fault data. The main advantage of the binary classifier method is its ability to incorporate new knowledge as data are collected, either by training new classifiers or retraining existing ones. Before testing, a validation process using unseen random data is conducted, as shown in Table 4.3.

Table 4.3 – Performance Metrics for Trained Binary Classifiers by Class.

Class	**Exp. Acc.**	**Exp. F1-Scr**	**Mean ± STD Acc.**	**Ins. Mean ± STD Acc.**
1	0.92	0.93	0.82 ± 0.04	1.00
2	0.90	0.90	0.85 ± 0.03	0.98 ± 0.05
3	0.85	0.87	0.89 ± 0.02	1.00
4	0.69	0.68	0.64 ± 0.03	1.00
5	0.82	0.83	0.75 ± 0.03	0.76 ± 0.05
6	0.94	0.95	0.75 ± 0.13	0.34 ± 0.42
7	0.79	0.83	0.80 ± 0.01	1.00
8	0.90	0.91	0.80 ± 0.06	0.94 ± 0.12
9	0.92	0.92	0.81 ± 0.05	1.00

The results in Table 4.3 are presented in five columns. The first column lists the 3W dataset classes. The second column shows the validation accuracy (Exp. Acc.) of the models used in the ICV fault case study experiment. The third column (Exp. F1-Scr) presents the F1-score for the case study, while the fourth column displays the mean and standard deviation obtained from 100 train-validation splits, randomly selected from the 3W dataset, using 40% of the real data for validation. The last column shows both the mean and standard deviation across all time series and at the instance level.

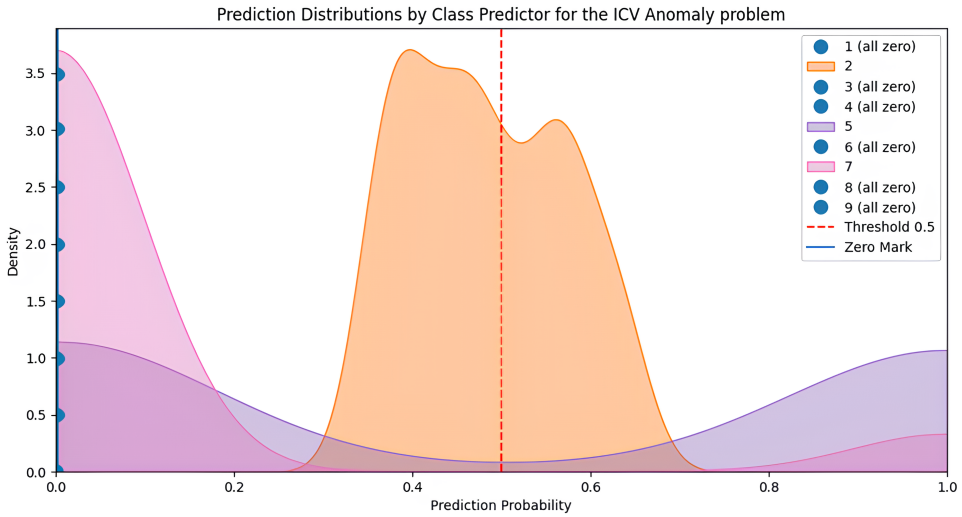
It is important to highlight that all classifiers are trained and validated using the same dataset split. Therefore, Exp. Acc. represents the best validation accuracy observed in the ICV case study, even though some classifiers may show lower Mean Acc. values over multiple training runs. The mean ± standard deviation reflects the overall performance distribution of the models. The best-performing models achieved accuracy values near the upper bound of this range (mean + 2 standard deviations).

Instance-level accuracy is calculated as the ratio of instances with at least one positive detection to the total number of instances for that specific anomaly.

These results partially address Research Question 3. Notably, class 4 (Flow Instability) exhibits the weakest performance metrics among all classifiers. This suggests that the anomaly shares similarities with other classes, leading to misclassifications. Class 6 has the worst instance level accuracy, leading to most of the instance files to be misclassified for that anomaly.

The evaluation of ICV fault data using these binary classifiers is shown in Figure 4.9.

Figure 4.9 – Prediction Distributions by Class Predictor for the ICV Anomaly Problem



Source: Author (2025)

Figure 4.9 presents density interpolation curves for the predictions of each of the nine binary classifiers. In this figure, each curve represents the distribution of prediction outputs for a binary classifier trained on one of the known anomaly classes from the 3W dataset. The x-axis represents the classifier's output, which corresponds to the probability that the input data belong to the anomaly class the classifier represents. If the output exceeds 0.5, it indicates that the data point has been classified as belonging to that specific class. The y-axis represents the density of these predictions, where the area under each curve sums to one. This visualization provides insight into the distribution and concentration of prediction scores across different thresholds, highlighting how well each classifier distinguishes the ICV anomaly from known classes. The threshold line at 0.5, marked in red, represents the decision boundary for classification into a given class.

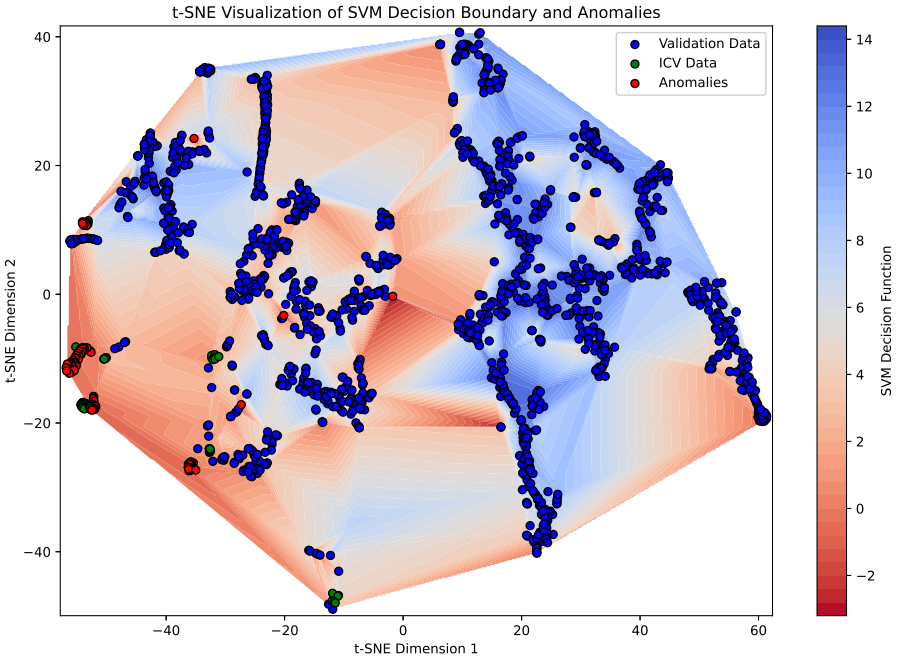
Ideally, to correctly treat ICV fault data as a new anomaly, all points should fall to the left of the red line (below 0.5), as they theoretically do not belong to any of the known classes. However, when using solely the binary classifiers with the 0.5 threshold, many points are unexpectedly classified as belonging to a combination of two classes: Class 2 - Spurious Closure of DHSV and Class 5 - Rapid Productivity Loss. This misclassification is reasonable, as these anomalies share similarities with ICV faults. This pattern is also observed in Figure 4.8, where the ICV fault data region overlaps with points from Classes 2 and 5.

The class intersections highlight the limitations of binary classifiers. When one or more known anomaly classes are too similar to the new data, the predictions may incorrectly assign the data to those classes. To mitigate this issue, two combined processes are used: introducing an auxiliary classifier with a non-class detection mechanism and applying a

multi-classifier consensus approach with a One-Class Support Vector Machine (OCSVM). The multi-classifier consensus also ensures a single definitive classification when multiple classes are detected.

A two-stage approach combining One-Class Support Vector Machines (OCSVM) and a pre-trained autoencoder is employed to tackle the challenge of detecting samples that belong to none of the known classes (non-class) in the absence of explicit "none" examples. The OCSVM is trained on the latent representations generated by the autoencoder to learn a decision boundary that encapsulates the distribution of the positive class, effectively identifying samples that deviate significantly from this boundary. Additionally, the autoencoder provides the reconstruction error as an auxiliary signal to quantify how well a sample conforms to the known data distribution. These two complementary signals are then integrated into a new classifier, which is utilized alongside binary dense classifiers to enhance the detection capability.

Figure 4.10 – OCSVM decision boundary, separating "known" classes and the non-class



Source: Author (2025)

Figure 4.10 illustrates the t-SNE visualization of the SVM decision boundary, highlighting its role in separating "known" classes from non-class samples. The red-to-blue gradient represents the SVM decision function values: blue regions correspond to areas confidently classified as "known" data, while red regions indicate non-class samples. The plot overlays three categories of data: validation data (blue points), ICV data (green points), and non-class samples (red points).

In this case, a strategy such as the multiclassifier consensus approach can be applied. If all classifiers are uncertain or produce low confidence scores, the input can be classified as "none of the classes." The classifier outputs are normalized, and a threshold is applied to the maximum probability across all classifiers. If this probability is below the threshold, the sample is classified as "none"; otherwise, it is assigned to the class with the highest probability value.

OCSVMs serve as auxiliary classifiers, complementing this strategy by modeling the distribution of each individual class. Unlike binary classifiers, which rely on comparisons between positive and negative samples, OCSVMs define the regions of feature space occupied by the positive class. This makes them particularly effective at identifying outliers or novel inputs that deviate significantly from the known class distributions.

The Weighted Voting Scheme provides a structured approach for classifying samples and detecting those that do not belong to any predefined class. This scheme integrates the outputs of binary classifiers and auxiliary OCSVMs through weighted contributions. Each classifier produces a normalized score that reflects its confidence in assigning the sample to a given class. Weights are assigned to both the binary classifier scores and OCSVM scores based on their performance during training and validation.

The final score for each class i is calculated as:

$$S_i = w_i^{(b)} s_i^{(b)} + w_i^{(o)} s_i^{(o)}, \tag{4.5}$$

where $w_i^{(b)}$ and $w_i^{(o)}$ are the weights of the binary classifier and OCSVM scores, respectively, and $s_i^{(b)}$ and $s_i^{(o)}$ are the normalized scores for the i -th class of the binary classifier and OCSVM.

The predicted class $\hat{y}$ is then determined as follows:

$$\hat{y} = \begin{cases} \arg\max_i S_i, & \text{if } \max_i S_i \geq \tau, \\ \text{“none of the classes”}, & \text{otherwise.} \end{cases} \tag{4.6}$$

Here, τ is a threshold calibrated through validation experiments to ensure that samples with low confidence across all classes are correctly identified as belonging to none of the classes. The threshold τ is calibrated through validation experiments to optimize the classification results through greed search, where certain classes are randomly hidden to simulate novel class detection, and its value may vary accordingly. For the ICV detection fault case study, the value is set to 0.65.

By tuning the weights $w_i^{(b)}$ and $w_i^{(o)}$ and the threshold τ, the Weighted Voting Scheme effectively integrates binary classifiers and OCSVMs, providing a flexible and reliable

classification framework. $w_i^{(b)}$ is related to the autoencoder reconstruction error classifier for the "no class". For the autoencoder, the probability of anomaly is defined by:

$$predict_{AE} = \begin{cases} \frac{r_{error}}{tol_{IA}}, & \text{if } r_{error} < tol_{IA} \\ 1.0, & \text{otherwise.} \end{cases} \tag{4.7}$$

In this equation, r_{error} represents the reconstruction error and tol_{IA} is defined in Equation 8 from Chapter 2. The results of the combined binary classifiers and OCSVM, namely accuracies and weights, are described in Table 4.4.

Table 4.4 – Updated Classifier Performance Metrics

Class	**Accuracy**	$w_i^{(b)}$	$w_i^{(o)}$
1	0.92	0.65	0.35
2	0.87	0.54	0.46
3	0.91	0.75	0.25
4	0.68	0.52	0.48
5	0.82	0.58	0.42
6	0.93	0.77	0.23
7	0.79	0.62	0.38
8	0.90	0.53	0.47
9	0.90	0.54	0.46
Class	**Accuracy**	$w_i^{(AE)}$	$w_i^{(o)}$
None	0.55	0.26	0.74

The table shows that the weights assigned to the binary classifiers are generally higher than those for the OCSVMs. Additionally, the combined accuracy is often lower than that of the binary classifiers alone. However, the combined approach is necessary to reduce false positives and effectively detect non-class samples. The application of the combined binary classifier approach resulted in 48% of ICV fault data being classified into one of Classes 2 and 5. This means that for the next steps in the proposed Open-World Learning pipeline, 52% of the anomalous data are classified as unknown and will be used in the clustering process.

4.3.4 Clustering of Unknown Anomalies

To address Research Question 4, after the binary classifiers fail to assign an anomaly to any known class, the combined random forest and clustering algorithm, composed of K-means and Meanshift, is applied to the latent representations of the unknown anomalies. Similar to the validation process for binary classifiers, this step is validated using the 3W dataset and tested with the ICV fault data.

4.3.4.1 Random Forest Estimation of Number of clusters

The random forest algorithm provides a method for estimating the number of clusters within a given point cloud. Table 4.5 presents clustering performance metrics, demonstrating the model's accuracy and reliability in estimating the number of unknown anomaly classes.

Table 4.5 – Clustering Performance Metrics

Metric	**Value**
Accuracy	98.3%
Precision	0.98
Recall	0.86
F1-Score	0.87

Table 4.5 shows that the model performs with high accuracy across all metrics. In cases where the estimated cluster count deviates from the actual count, the average error margin remains small (1.12 clusters). This accuracy provides a good foundation for the subsequent clustering process, helping the cluster process of unknown anomalies. The methodology allows the system to dynamically estimate the number of unknown anomaly clusters, thereby improving its ability to adapt to new anomaly types as they emerge.

4.3.4.2 K-means Clustering Experimentation

To evaluate the model's ability to generalize to different unknown anomaly combinations, scenarios are simulated where only six of the nine known anomalies are used for training, treating the remaining three as unknown. An example of this setup involves treating anomaly classes 3, 6, and 8 as unknown and applying clustering.

Figure 4.11 illustrates the original data distribution per class for this sample, while Figure 4.12 displays the results of clustering using K-means. The performance metrics for this specific case are summarized in Table 4.6. For this selection of three unknown anomalies, the clustering results closely align with the original classes, as reflected in the metrics.

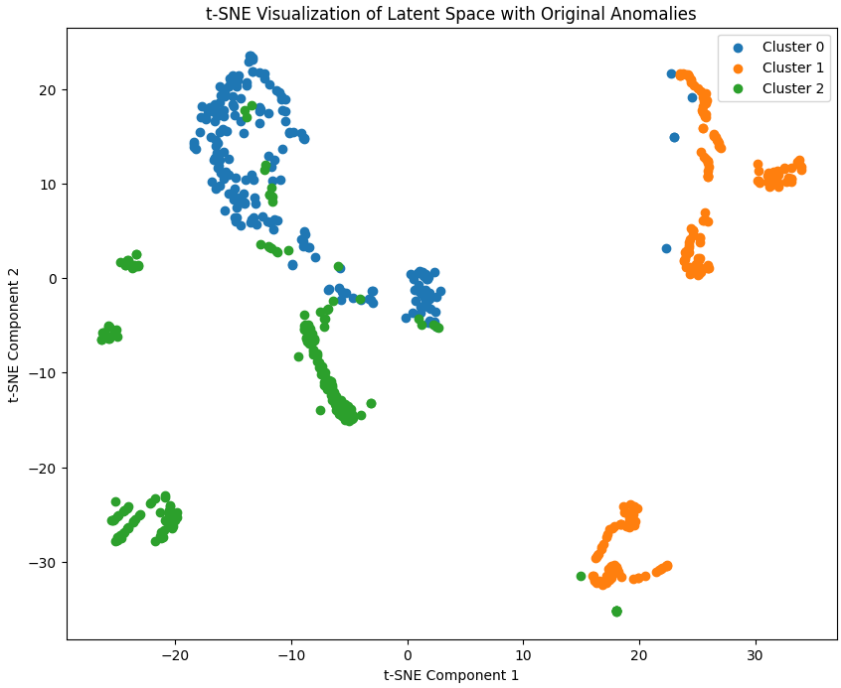
Table 4.6 – Per-Class Metrics of a sample set of 3 unknown anomalies

Class	**Accuracy**	**Precision**	**Recall**	**F1 Score**
3	0.62	0.83	0.62	0.71
6	1.00	0.94	1.00	0.97
8	0.83	0.70	0.83	0.76

The selection of the 3 remaining anomalies is done, to obtain a more accurate result, by using all possible combinations of trios from 9 elements, totaling 168 different groups of clusters. Table 4.7 shows the best result using K-means clustering method, and Figure 4.13 shows the normalized confusion matrix.

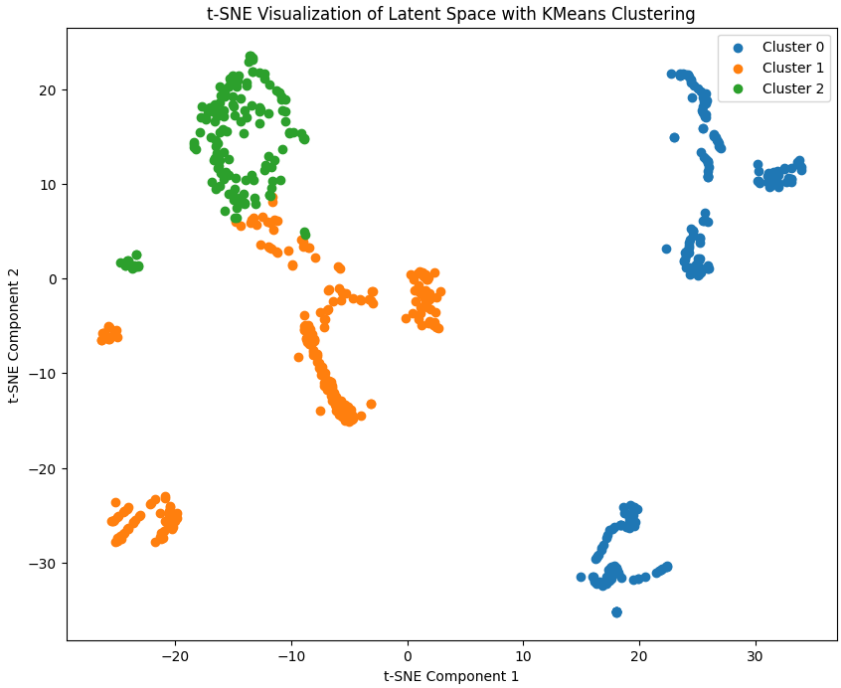
Similar to the binary classifiers in Table 4.3, Table 4.7 presents accuracy and F1-score values from the main experiment, along with the mean and standard deviation obtained from

Figure 4.11 – Original data from the anomalies 3, 6 and 8 (represented respectively as clusters 0, 1 and 2)



Source: Author (2025)

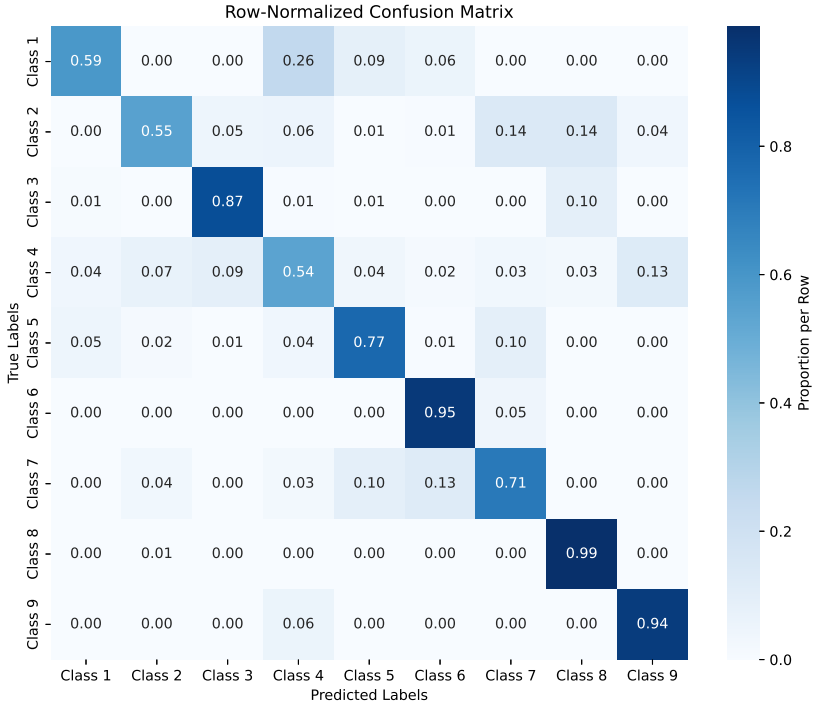
Figure 4.12 – K-means clusters for the anomalies 3, 6 and 8.



Source: Author (2025)

100 different random validation splits. The results highlight the classification challenges associated with class 4, as reflected in both the table and the confusion matrix.

Figure 4.13 – K-means Clustering Confusion Matrix



Source: Author (2025)

Table 4.7 – K-means Clustering Metrics Comparison.

Class	Exp. Acc.	Exp. F1-Scr	Mean ± STD Acc.	Ins. Mean Acc.
1	0.71	0.72	0.57 ± 0.16	0.95
2	0.44	0.52	0.49 ± 0.13	0.46
3	0.88	0.73	0.72 ± 0.19	0.92
4	0.38	0.41	0.48 ± 0.13	0.67
5	0.74	0.72	0.77 ± 0.07	0.60
6	0.99	0.71	0.93 ± 0.10	0.93
7	0.81	0.71	0.66 ± 0.18	0.87
8	0.93	0.75	0.92 ± 0.04	0.88
9	0.72	0.70	0.81 ± 0.15	0.73
Global	0.73	0.67	0.71 ± 0.15	0.78

4.3.4.3 Combined Clustering with Distance Filter

Table 4.8 presents the clustering results when a distance filter is applied, excluding data points that are far from the cluster center. Table 4.9 displays the best results for the combined clustering approach using K-means, Meanshift, and the distance filter. Similar to Table 4.7, the columns from both Tables are described as follows: Exp. Acc. represents the validation accuracy in the ICV case study, Exp. F1-Scr represents the validation F1-score in the ICV case study, Mean ± STD Acc. denotes the mean and standard

deviation of accuracy, and Ins. Mean Acc. represents the mean instance-level accuracy. Both tables compare accuracy and F1-score for the main experiment, along with the mean and standard deviation of accuracy computed across 100 randomly selected validation splits.

Table 4.8 – K-means Clustering Performance After Applying Distance Filter.

Class	**Exp. Acc.**	**Exp. F1-Scr**	**Mean ± STD Acc.**	**Ins. Mean Acc.**
1	0.61	0.68	0.58 ± 0.11	0.95
2	0.56	0.71	0.60 ± 0.19	0.46
3	0.89	0.83	0.70 ± 0.20	0.93
4	0.52	0.62	0.58 ± 0.13	0.67
5	0.78	0.76	0.83 ± 0.10	0.71
6	0.95	0.95	0.96 ± 0.07	0.99
7	0.71	0.76	0.74 ± 0.18	0.84
8	0.98	0.93	0.98 ± 0.02	0.99
9	0.94	0.83	0.94 ± 0.08	0.89
Global	0.77	0.76	0.77 ± 0.15	0.82

Table 4.9 – Performance of the Combined Clustering Approach (Distance Filtering + K-means + Meanshift).

Class	**Exp. Acc.**	**Exp. F1-Scr**	**Mean ± STD Acc.**	**Ins. Mean Acc.**
1	0.80	0.73	0.60 ± 0.10	0.96
2	0.61	0.69	0.64 ± 0.16	0.51
3	0.91	0.75	0.73 ± 0.19	0.92
4	0.52	0.57	0.60 ± 0.12	0.68
5	0.80	0.71	0.84 ± 0.09	0.74
6	0.99	0.94	0.98 ± 0.04	0.98
7	0.71	0.73	0.77 ± 0.16	0.85
8	0.99	0.90	0.98 ± 0.02	1.0
9	0.94	0.83	0.96 ± 0.06	0.92
Global	0.81	0.74	0.79 ± 0.15	0.84

The clustering results across the nine anomaly classes demonstrate consistent performance, with the highest accuracy observed for class 6. Conversely, class 4 exhibits the lowest accuracy (52%), a trend also observed in binary classifiers and pure K-means clustering. The overall accuracy of 81% highlights the model's ability to correctly cluster most unknown anomalies. While certain classes show strong clustering performance, others achieve moderate to low accuracy, reflecting the inherent complexity of certain anomaly types.

To benchmark the clustering performance against supervised approaches, the results of the proposed method are compared to those reported in Marins et al. (2021b) and to the supervised multi-class models of Lopes et al. (2024). Table 4.10 presents a comparison between accuracy results from different methods and works. Although the clustering

method is unsupervised, the results demonstrate competitive performance relative to fully supervised methods, particularly for classes 8 and 9, which achieved accuracy rates close to those of the supervised approaches. However, for certain anomaly classes, such as class 2 and class 4, supervised methods performed better. It is important to highlight that the these are results with different goals, so accuracy values is only a first hint when selecting a strategy between the presented from Table 4.10.

Table 4.10 – Comparison of Clustering Accuracy with Supervised Methods.

Columns Description: *Lopes et al. (2024); **Marins et al. (2021b); Proposed 1: Unsupervised Clustering of the Latent Space; Proposed 2: Binary Supervised Classifiers.

Class	**Literature 1***	**Literature 2****	Proposed 1	Proposed 2
1	0.92		0.79	0.92
2	0.92	0.89	0.61	0.90
3	0.87	0.79	0.90	0.85
4	0.76	0.95	0.52	0.69
5	0.95	0.83	0.79	0.82
6	0.61	0.71	0.99	0.94
7	0.87	-	0.70	0.79
8	0.99	-	0.98	0.90
9	0.98	-	0.94	0.92
Global	0.86	0.93	0.81	0.92

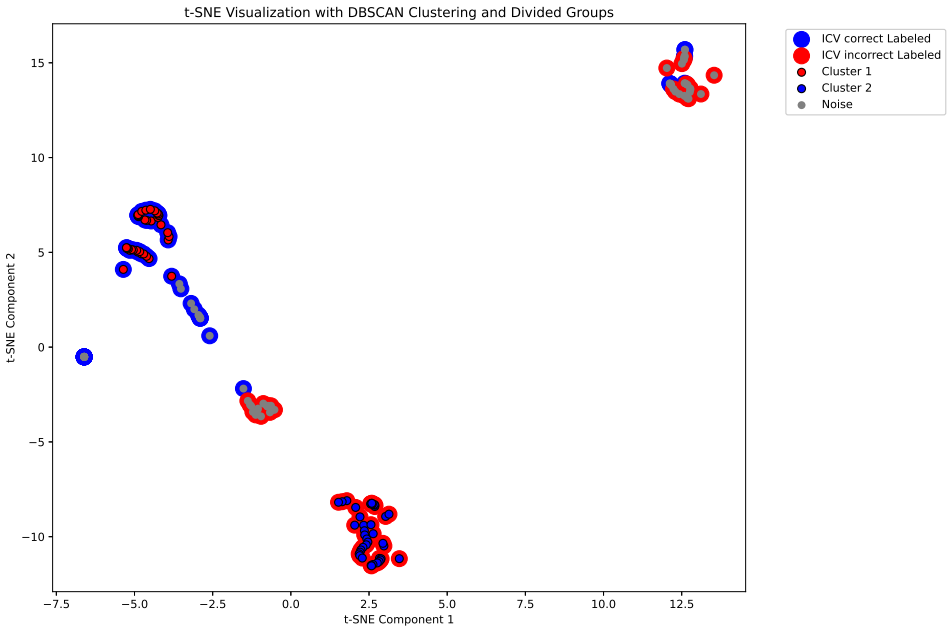
When comparing clustering performance with other supervised methods, the proposed approach performed competitively for some classes. It is important to highlight that a direct accuracy comparison is not possible since the classifiers have distinct objectives; the comparison only provides a general indication of overall performance. The global accuracy of 81% is not far behind the results achieved by Marins et al. (2021b), which obtained a global accuracy of 93% in a fully supervised scenario, and Lopes et al. (2024), which achieved 83%. However, the proposed method has the advantage of autonomously discovering new anomalies without requiring extensive labeled data. When compared to binary classifiers, the performance of the proposed method is similar to approaches found in the literature.

4.3.5 Testing the combined clustering in the ICV Fault Data

The first step in analyzing the remaining 52% of ICV data, which is classified as negative by all trained binary classifiers (indicating it is unseen data), is to determine whether it can be grouped into a distinct cluster. When applying DBSCAN and Meanshift to all ICV data, including the portion classified as other anomalies, both algorithms identify two distinct clusters along with some noise, as illustrated in Figures 4.14 and 4.15.

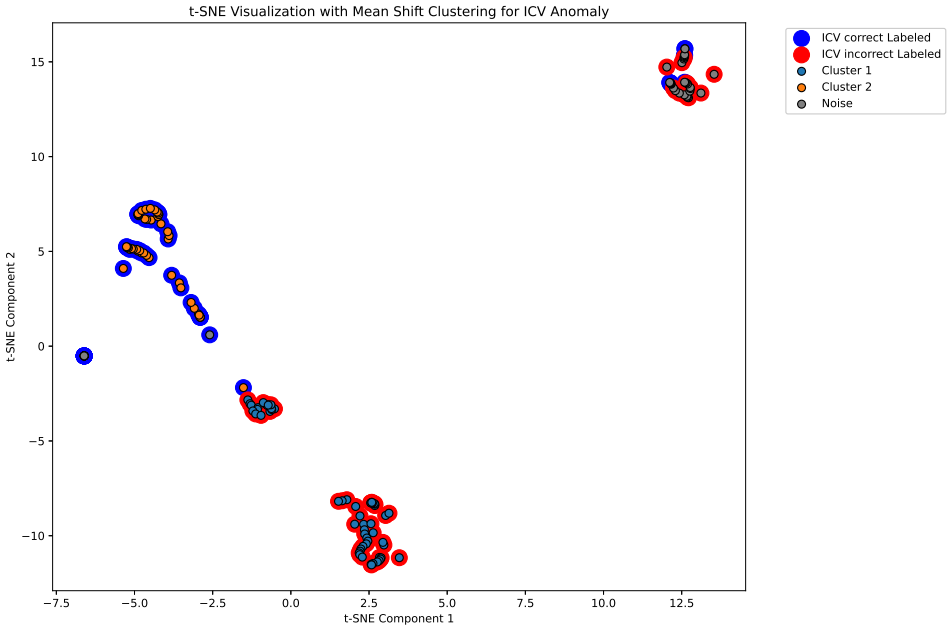
The existence of two clusters is logical, as part of the data exhibit behavior similar to known anomalies, leading them to be classified into existing anomaly classes. The figures

Figure 4.14 – DBSCAN clustering of ICV Fault Data



Source: Author (2025)

Figure 4.15 – Meanshift clustering of ICV Fault Data



Source: Author (2025)

indicate that the misclassified region corresponds to one of the clusters for both Meanshift and DBSCAN.

Since at least one cluster exists in the region classified as unknown, the Random Forest regressor also estimated one cluster as the probable number of classes. In this particular case, it is unnecessary to apply the combined algorithm to determine that the 52% of ICV

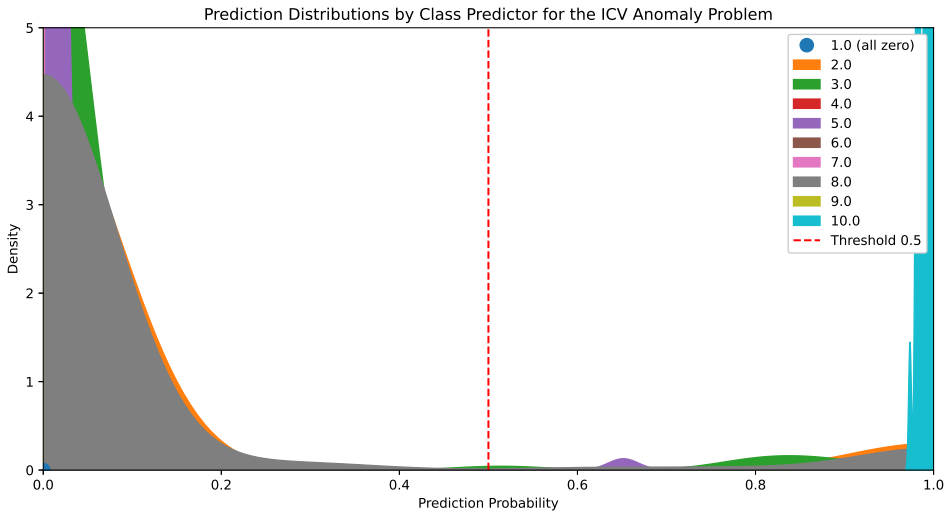
data classified as unknown belong to a single class. Given that only one cluster is detected, there is no need to separate the data further.

4.3.6 Validation and Train New Binary Classifiers

Table 4.11 – Performance Metrics for Trained Binary Classifiers by Class

Class	Accuracy	Precision	Recall	F1 Score
1	0.93	0.89	0.99	0.93
2	0.89	0.91	0.88	0.89
3	0.85	0.89	0.82	0.85
4	0.67	0.66	0.73	0.69
5	0.77	0.86	0.67	0.75
6	0.94	0.92	0.98	0.95
7	0.82	0.75	0.95	0.84
8	0.90	0.86	0.96	0.91
9	0.93	0.93	0.94	0.93
10	0.99	0.98	0.98	0.98

Figure 4.16 – New Prediction Distributions by Class Predictor for the ICV Anomaly problem



Source: Author (2025)

The newly introduced 10th classifier is trained exclusively on the data previously classified as unknown. This dataset now serves as positive samples for the new class and negative samples for all other classes. The new validation metrics for the binary classifiers are presented in Table 4.11.

To assess the improvement in the system's knowledge, Figure 4.16 illustrates the same scenario as Figure 4.9, but with 10 classifiers, including the newly learned ICV Fault

Anomaly. Despite a few false positives, which could be considered noise, the majority of the previously misclassified data is now correctly classified as an ICV Fault anomaly. This result effectively addresses Research Question 5.

4.3.7 Training and Execution Times

Table 4.12 – Training and Execution Times for Different Model Components

Process	**Time**
Train autoencoder for anomaly detection	$\approx$ 1.37 min
Train binary classifiers	$\approx$ 1.02 min per classifier
Train encoder (autoencoder)	$\approx$ 1.24 min
Detect anomaly	6 ms per batch of 32 sequences
Classify anomaly	7 ms per batch of 32 sequence/classifier
Noise detector (DBSCAN + MeanShift)	$\approx$ 2 sec per 200 points
Cluster data	$\approx$ 3 sec per 200 points
Retrain classifiers	$\approx$ 1.02 min per classifier + new classifiers

All GPU-based training experiments were conducted on a system equipped with an Intel Xeon CPU with 2 vCPUs, 13GB of RAM, and an NVIDIA T4 GPU. The computational constraints of this hardware setup may have influenced model training efficiency, hyperparameter optimization, and overall performance. Table 4.12 presents the training and execution times for several stages of the proposed methodology, including autoencoder training, binary classifier training, anomaly detection, and clustering processes.

4.4 Partial Remarks

This Chapter presented a system for detecting, classifying, and clustering unknown anomalies in oil wells, applying Open-World Learning principles to a domain where labeled anomaly data are often scarce. The proposed approach integrates autoencoder-based anomaly detection, binary classifiers for known anomalies, and a hybrid K-means and Meanshift clustering method for unknown anomaly discovery. This enables the system to autonomously detect, classify, and continuously adapt to previously unseen anomalies—an essential capability for dynamic and complex operational environments such as oil well monitoring.

The key contributions of this Chapter include the application of latent space representations for anomaly clustering, the hybrid clustering strategy combining K-means and Meanshift, and the integration of an operator-driven validation system that enables continuous learning.

The autoencoder successfully detected anomalous states based on reconstruction errors, achieving 95% accuracy for the unseen ICV fault data. Binary classifiers demonstrated high

accuracy across most known anomaly classes, with 90% and 92% accuracy for anomalies in Classes 3 and 8, respectively. For unclassifiable anomalies, the Random Forest regressor reliably estimated the number of clusters with 98.3% accuracy. The combined K-means and Meanshift clustering method aligned well with ground-truth labels, achieving a global accuracy of 81%, approaching the performance of supervised methods in many cases.

Testing the pipeline on the ICV fault dataset confirmed the adaptability of the approach. The system successfully identified the ICV fault as a new anomaly class and integrated it into the model, leading to significant improvements in classification accuracy for this fault type. This demonstrates the system's ability to dynamically learn and incorporate new anomalies without requiring extensive labeled datasets.

Despite strong performance, supervised methods still achieved higher accuracy for certain challenging classes, particularly Class 4, where supervised models reached 95% accuracy, compared to 69% for binary classifiers and 52% for clustering. However, the Open-World Learning approach offers a key advantage: the ability to autonomously discover and classify new anomalies without relying on predefined labels, a critical requirement for real-time industrial applications.

Future work can focus on automating the operator-driven validation step, potentially replacing it with machine learning models capable of handling point elimination, cluster adjustment, and autonomous validation of new classes. Another key improvement will focus on enhancing feature extraction and developing specialized architectures to better differentiate classes with similar behavior, particularly Flow Instability (Class 4), which has consistently struggled with categorization and grouping. Additional enhancements will aim to distinguish anomalies caused by sensor interference from those arising from genuine operational issues, thereby reducing false positives and further refining the model's accuracy for improved predictive maintenance and operational safety.

5 Final Remarks

This Thesis presents a comprehensive set of contributions that aim to advance the field of anomaly detection and classification in oil well monitoring. By integrating machine learning techniques, rule-based approaches, and Open-World Learning principles, this work addresses challenges in ensuring operational safety, efficiency, and adaptability in highly dynamic oil production environments. The proposed strategy and models span anomaly detection for known faults, early-stage fault detection, and adaptive systems capable of identifying and incorporating new, previously unseen anomalies. Together, these contributions provide a strong foundation for the development of intelligent and scalable monitoring systems for the oil and gas industry.

Chapter 2 introduced a system that combines LSTM-based autoencoders with a rule-based analytical approach to detect general anomalies in oil wells. This system effectively monitored the behavior of pressure and temperature sensors, adapted to various valve configurations, and identified faults such as spurious closures of the DHSV. Using the autoencoder reconstruction error and the interpretability of rule-based logic, the system achieved high precision and adaptability, consistently detecting anomalies in the early stage with a true positive detection rate exceeding 90% on sample level. Moreover, its practical implementation across more than 20 FPSOs, monitoring over 250 subsea wells in real-time, demonstrates its robustness, scalability, and ability to function within the computational constraints of the rigsite infrastructure. This study also highlighted the importance of using hybrid approaches to integrate human expertise and domain-specific knowledge into machine learning systems.

Chapter 3 presents the study of binary classifiers capable of detecting anomalies in oil well operations, with the detection of hydrate formation serving as a study case for verification and validation. Binary classifiers based on LSTM-based Deep Neural Networks (DNN) demonstrated their effectiveness in identifying hydrate-related anomalies with precision exceeding 90% 90% on sample level. This achievement ensured timely detection within a 15-minute window, within the 30-minute response time required to mitigate the risks of hydrate formation, which carry significant economic and safety implications. By reducing the false negative rate, the detection model validated the reliability of this binary classification approach compared to traditional analytical methods. In addition, the study underscored the versatility of binary classifiers in multiclass anomaly detection contexts, showing that variations in pressure and temperature in different faults provided distinct characteristics for training models capable of accurately classifying various types of anomalies. This work presented a method using a reduced number of sensors (only TPT data), emphasizing its broad applicability in diverse oil well setups. Furthermore,

the study proposed a staged detection system, where an initial anomaly detection model distinguishes normal from abnormal states before forwarding data to the appropriate binary classifier, enhancing overall accuracy and efficiency of the system.

Chapter 4 represents a major contribution to the field of oil well monitoring by introducing an Open-World Learning (OWL) strategy for anomaly detection, anomaly and novelty classification, and novelty clustering. This methodology addresses a critical limitation of traditional machine learning approaches: the reliance on fixed, labeled datasets that cannot adapt to new, unseen anomaly types. The OWL strategy combines autoencoder-based anomaly detection, binary classification for known anomalies, and clustering methods (k-means and MeanShift) to identify and organize previously unknown anomalies. A human-in-the-loop validation step is integrated into the pipeline, allowing operators to confirm or adjust clusters, register new anomaly classes, and incrementally update the model.

A key achievement of this study is the successful validation of the OWL pipeline on new data involving Downhole Interval Control Valve (ICV) faults. These faults, previously unrepresented in 3W dataset, were identified as distinct classes and integrated into the system. The clustering methodology demonstrated robust performance, achieving an overall clustering accuracy of 80.8% and nearly 95% for certain events, while the binary classifiers achieved high accuracy for most known anomaly classes. This shows the system's ability to dynamically learn and adapt to new operational conditions without the need for extensive retraining. The OWL strategy thus fills the gap between detecting known faults and discovering new ones, providing a truly adaptive monitoring system for complex real-time operational environments.

This Thesis not only delivers practical methodology, but also highlights implications for the oil and gas industry:

- Real-Time Application: The methodologies have been implemented and validated in real-world operations, showcasing their scalability and robustness in monitoring over 250 subsea wells.
- Dynamic Adaptation: The OWL strategy enables systems to evolve continuously, addressing the challenge of identifying new types of anomalies in a highly dynamic field.
- Cost and Safety Improvements: Early-stage fault detection, as demonstrated in the hydrate formation and ICV Fault cases, reduces nonproductive time and minimizes risks to equipment and personnel.

5.1 Limitations and Future Work

While the proposed methodologies exhibit promising results, certain limitations remain. Along with all cited limitations described in Section 1.5, early-stage anomaly detection could benefit from further refinement, particularly for anomalies with subtle signatures. Automating the operator-driven validation step in the OWL pipeline remains an open challenge, and future work could explore the use of advanced clustering algorithms or reinforcement learning models to handle this process autonomously. Additionally, distinguishing between operational anomalies caused by sensor interference (e.g. human interventions) and genuine faults is an area that requires further exploration to reduce false positives.

Future research can focus on expanding the detection capabilities to encompass more uncharted anomalies and exploring advanced architectures such as transformers and gated recurrent units to enhance the efficiency of sequential data processing. Finally, integrating predictive maintenance systems with the proposed methodologies could unlock new possibilities for proactive decision-making and operational optimization.

5.2 Final Considerations

This Thesis represents an advancement in anomaly detection and classification for oil well monitoring, offering a suite of robust, scalable, and adaptive solutions. By bridging rule-based methods with state-of-the-art machine learning techniques and introducing Open-World Learning principles, this work lays the foundation for monitoring systems capable of addressing the evolving challenges of the oil and gas industry. These contributions provide a roadmap for safer and more efficient operations, ensuring that oil well monitoring systems can adapt and thrive in increasingly complex environments.

Bibliography

Al-Hajri, N. M., Al-Ghamdi, A., Tariq, Z., et al. (2020). Scale-prediction/inhibition design using machine-learning techniques and probabilistic approach. *SPE Prod Oper*, 35(4):987–1009.

Anifowose, F., Abdulraheem, A., and Al-Shuhail, A. (2019). A parametric study of machine learning techniques in petroleum reservoir permeability prediction by integrating seismic attributes and wireline data. *Journal of Petroleum Science and Engineering*, 176.

Aranha, P., Gouveia Omena Lopes, L., Paranhos Sobrinho, E., Oliveira, I., de Araújo, J., Santos, B., Lima Junior, E., Silva, T., Vieira, T., Lira, W., Policarpo, N., and Pinto, M. A. (2024a). A system to detect oilwell anomalies using deep learning and decision diagram dual approach. *SPE Journal*, 29:1540–1553.

Aranha, P. E., Policarpo, N. A., and Sampaio, M. A. (2024b). Unsupervised machine learning model for predicting anomalies in subsurface safety valves and application in offshore wells during oil production. *Journal of Petroleum Exploration and Production Technology*, 14:567–581.

Aslam, N., Khan, I., Alansari, A., Alrammah, M., Alghwairy, A., Alqahtani, R., Alqahtani, R., Almushikes, M., and Hashim, M. (2022). Anomaly detection using explainable random forest for the prediction of undesirable events in oil wells. *Applied Computational Intelligence and Soft Computing*, 2022:1558381.

Bodesheim, P., Freytag, A., Rodner, E., Kemmler, M., and Denzler, J. (2013). Kernel null space methods for novelty detection. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3374–3381.

Brønstad, C., Netto, S. L., and Ramos, A. L. (2021). Data-driven detection and identification of undesirable events in subsea oil wells. In *Proceedings of the International Conference on Advances in Condition-Based Monitoring*, pages 1–6, Athens, Greece. IARIA.

Caliński, T. and JA, H. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3:1–27.

Carvalho, B., Vargas, R., Salgado, R., et al. (2021). Flow instability detection in offshore oil wells with multivariate time series machine learning classifiers.

Chen, B., Zeng, X., Zhang, W., Cao, S., and Zhou, J. (2023). Knowledge sharing-based multi-block federated learning for few-shot oil layer identification. *Energy*.

Chen, Z. and Liu, B. (2018). *Open-World Learning*, pages 77–89. Springer International Publishing, Cham.

Chen, Z., Yeo, C. K., Lee, B. S., and Lau, C. T. (2018). Autoencoder-based network anomaly detection. In *2018 Wireless Telecommunications Symposium (WTS)*, pages 1–5.

Chollet, F. (2015). Keras. https://github.com/fchollet/keras.

Chollet, F. (2016). Building powerful image classification models using very little data. https://github.com/fchollet/keras.

Comaniciu, D. and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.

D'Almeida, A., Bergiante, N., Ferreira, G., Leta, F., Lima, C., and Lima, G. (2022). Digital transformation: a review on artificial intelligence techniques in drilling and production applications. *The International Journal of Advanced Manufacturing Technology*, 119:1–30.

Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.

de Salvo Castro, A., de Jesus Rocha Santos, M., Leta, F., et al. (2021). Unsupervised methods to classify real data from offshore wells. *American Journal of Operations Research*, 11(5):1–14.

Deng, D. (2020). Dbscan clustering algorithm based on density. In *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, pages 949–953.

dong Qi, X. (2013). K-means clustering algorithm combined with mean-shift and minimum spanning tree. *Computer Engineering*.

Fernandes, W., Komati, K., and de Souza Gazolli, K. A. (2024). Anomaly detection in oil-producing wells: a comparative study of one-class classifiers in a multivariate time series dataset. *Journal of Petroleum Exploration and Production Technology*, 14(2):343–363.

Figueiredo, I., Carvalho, T., and Silva, W. (2021). Detecting interesting and anomalous patterns in multivariate time-series data in an offshore platform using unsupervised learning. In *Offshore Technology Conference Proceedings*, number OTC-31297-MS.

Gao, B., Kong, X., Li, S., Chen, Y., Zhang, X., Liu, Z., and Lv, W. (2024). Enhancing anomaly detection accuracy and interpretability in low-quality and class imbalanced data: A comprehensive approach. *Applied Energy*, 353:1151–1166.

Gatta, F., Giampaolo, F., Chiaro, D., and Piccialli, F. (2022). Predictive maintenance for offshore oil wells by means of deep learning features extraction. *Expert Systems*.

Gjelsvik, E. L., Fossen, M., and Tndel, K. (2023). Current overview and way forward for the use of machine learning in the field of petroleum gas hydrates. *Fuel*, 334, Part 2.

Guilherme, I. R., Marana, A. N., Papa, J. P., Chiachia, G., Afonso, L. C., Miura, K., Ferreira, M. V., and Torres, F. (2011). Petroleum well drilling monitoring through cutting image analysis and artificial intelligence techniques. *Engineering Applications of Artificial Intelligence*, 24(1):201 – 207.

Gurina, E., Klyuchnikov, N., Zaytsev, A., Romanenkova, E., Antipova, K., Simon, I., Makarov, V., and Koroteev, D. (2020). Application of machine learning to accidents detection at directional drilling. *Journal of Petroleum Science and Engineering*, 184.

Hafiz, A. M. and Bhat, G. M. (2020). Multiclass classification with an ensemble of binary classification deep networks. *CoRR*, abs/2007.01192.

Helmy, T., Fatai, A., and Faisal, K. (2010). Hybrid computational models for the characterization of oil and gas reservoirs. *Expert Systems with Applications*, 37(7):5353 – 5363.

Hourfar, F., Bidgoly, H. J., Moshiri, B., Salahshoor, K., and Elkamel, A. (2019). A reinforcement learning approach for waterflooding optimization in petroleum reservoirs. *Engineering Applications of Artificial Intelligence*, 77:98 – 116.

Huffner, L. N., Trierweiler, J. O., and Farenzena, M. (2019). Are complex black-box models for permanent downhole gauge pressure estimation necessary. *J Pet Sci Eng*, 173:715–732.

Jafarzadeh, M., Dhamija, A. R., Cruz, S., Li, C., Ahmad, T., and Boult, T. E. (2022). A review of open-world learning and steps toward open-world learning without labels.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint, last revised 30 January 2017.

Larzalere, B. (2019). Lstm autoencoder for anomaly detection. `https://towardsdatascience.com/lstm-autoencoder-for-anomaly-detection-e1f4f2ee7ccf`. Accessed: 2019-10-30.

Li, H., Chu, L., Lu, J., Liu, Q., Li, F., and Zhang, K. (2023). Svd-vmd algorithm and its application in leak detection of natural gas pipeline. *Petroleum Science and Technology*, 41(2):230–255.

Li, Y., Ge, T., and Chen, C. (2020). Data stream event prediction based on timing knowledge and state transitions. *PVLDB*, 13(10):1779–1792.

Loh, K., Shoeibi Omrani, P., and Linden, R. (2018). Deep learning and data assimilation for real-time production prediction in natural gas wells.

Lopes, L. G. O., Vieira, T. M. d. A., de Lima Junior, E. T., and Lira, W. W. M. (2024). Use of lstm networks for anomaly classification in oil wells emphasizing hydrate detection. *SSRN Electronic Journal*.

Lu, Z., Zhai, G., Zuo, Y., Wang, Q., Fan, D., Tang, S., Hu, D., Liu, H., Wang, T., Zhu, Y., and Xiao, R. (2019). The geological process for gas hydrate formation in the qilian mountain permafrost. *Petroleum Science and Technology*, 37(13):15661581.

Machado, A. P. F., Vargas, R. E. V., Ciarelli, P. M., and Munaro, C. J. (2022). Improving performance of one-class classifiers applied to anomaly detection in oil wells. *Journal of Petroleum Science and Engineering*, 218.

Magnusson, L., Olsson, J., and Tran, C. (2023). Recurrent neural networks for oil well event prediction. *IEEE Intelligent Systems*.

Malhotra, P., Vig, L., Shroff, G., et al. (2015). Long short-term memory networks for anomaly detection in time series. In *23rd European Symposium on Artificial Neural Networks, ESANN 2015*, Bruges, Belgium.

Marchi, E., Vesperini, F., Eyben, F., Squartini, S., and Schuller, B. (2015). A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional lstm neural networks. *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*.

Marins, M. A. (2018). Machine learning techniques applied to hydrate failure detection on production lines. "http://www.pee.ufrj.br/index.php/en/producoes-academica/dissertacoes-de-mestrado/2018/2016033300-machine-learning-techniques-applied-to-hydrate-failure-detection-on-production-lines/file".

Marins, M. A., Barros, B. D., Santos, I. H., Barrionuevo, D. C., Vargas, R. E., Prego, T. M., Lima, A. A., Campos, M. L., Silva, E. A., and Netto, S. L. (2021a). Fault detection and classification in oil wells and production/service lines using random forest. *Journal of Petroleum Science and Engineering*, 197.

Marins, M. A., Barros, B. D., Santos, I. H., et al. (2021b). Fault detection and classification in oil wells and production/service lines using random forest. *Journal of Petroleum Science and Engineering*.

Monday, C. U. and Odutola, T. O. (2021). Application of machine learning in gas-hydrate formation and trendline prediction. In *SPE Symposium: Artificial Intelligence - Towards a Resilient and Efficient Energy Industry*.

Motiee, M. (1991). Estimate possibility of hydrates. *Hydrocarbon Processing*, 70:98–99.

Node-RED (2019). Node-red: Documentation.

Pang, H.-W., Wang, H.-Q., Xiao, Y.-T., Jin, Y., Lu, Y.-H., Fan, Y.-D., and Nie, Z. (2024). Machine learning for carbonate formation drilling: Mud loss prediction using seismic attributes and mud loss records. *Petroleum Science*, 21(2):1241–1256.

Parmar, J., Chouhan, S. S., Raychoudhury, V., and Rathore, S. S. (2022). Open-world machine learning: Applications, challenges, and opportunities.

Pimentel, M. A., Clifton, D. A., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99:215–249.

Qiao, Y., Xu, H.-M., Zhou, W.-J., Peng, B., Hu, B., and Guo, X. (2023). A bigru joint optimized attention network for recognition of drilling conditions. *Petroleum Science*, 20(6):3624–3637.

Ren, Y., Pu, J., Yang, Z., Xu, J., Li, G., Pu, X., Yu, P. S., and He, L. (2022). Deep clustering: A comprehensive survey.

Salahshoor, K., Zakeri, S., and Haghighat Sefat, M. (2013). Stabilization of gas-lift oil wells by a nonlinear model predictive control scheme based on adaptive neural network models. *Engineering Applications of Artificial Intelligence*, 26(8):1902 – 1910.

Shahani, N. M., Kamran, M., Zheng, X., and Liu, C. (2022). Predictive modeling of drilling rate index using machine learning approaches: Lstm, simple rnn, and rfa. *Petroleum Science and Technology*, 40(5):534–555.

Shahapure, K. R. and Nicholas, C. (2020). Cluster quality analysis using silhouette score. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 747–748.

Shu, X., Bao, T., Yangtao, L., Jian, G., and Zhang, K. (2022). Vae-talstm: a temporal attention and variational autoencoder-based long short-term memory framework for dam displacement prediction. *Engineering with Computers*, 38:1–16.

Soriano-Vargas, A., Werneck, R., Moura, R., et al. (2021). A visual analytics approach to anomaly detection in hydrocarbon reservoir time series data. *Journal of Petroleum Science and Engineering*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958.

Turan, E. M. and Jaschke, J. (2021). Classification of undesirable events in oil well operation. In *23rd International Conference on Process Control (PC)*, Strbske Pleso, Slovakia.

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11):2579–2605.

Vargas, R. E. V., Munaro, C. J., and Ciarelli, P. M. (2019a). A methodology for generating datasets for development of anomaly detectors in oil wells based on artificial intelligence techniques. *I Congresso Brasileiro em Engenharia de Sistemas em Processos*.

Vargas, R. E. V., Munaro, C. J., Ciarelli, P. M., Medeiros, A. G., do Amaral, B. G., Barrionuevo, D. C., de Araújo, J. C. D., Ribeiro, J. L., and aes, L. P. M. (2019b). A realistic and public dataset with rare undesirable real events in oil wells. *Journal of Petroleum Science and Engineering*, 181:106223.

Venkatasubramanian, V. (2003). A review of process fault detection and diagnosis part i: Quantitative model-based methods. *Computers & Chemical Engineering*, 27:293 – 311.

Wang, Y., Pan, Z., Yuan, X., Yang, C., and Gui, W. (2019). A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network. *ISA Transactions*, 96.

Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., Liu, Y., Zhao, Y., Pei, D., Feng, Y., Chen, J., Wang, Z., and Qiao, H. (2018). Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 187–196, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Yousefzadeh, R., Bemani, A., Alireza, K., and Ahmadi, M. (2022). An insight into the prediction of scale precipitation in harsh conditions using different machine learning algorithms. *SPE Production & Operations*, pages 1–19.

Zhong, Z., Sun, A. Y., Wang, Y., and Ren, B. (2020). Predicting field production rates for waterflooding using a machine learning-based proxy model. *Journal of Petroleum Science and Engineering*, 194.

Zhu, F., Ma, S., Cheng, Z., Zhang, X.-Y., Zhang, Z., and Liu, C.-L. (2024a). Open-world machine learning: A review and new outlooks.

Zhu, X., Zhang, H., Ren, Q., Rui, J., Zhang, L., and Zhang, D. (2024b). Orali: Open-set recognition and active learning for unknown lithology identification. *Engineering Applications of Artificial Intelligence*, 133:108623.