UNIVERSITÀ
DEGLI STUDI
FIRENZE

Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione (DINFO)
Corso di Dottorato in Ingegneria dell'Informazione
Curriculum: Ingegneria Informatica

Application of Artificial Intelligence for Anomaly Detection in Energy Generation Systems

Candidate
Piero Danti

Supervisors
Prof. Fabio Schoen

Dr. Alessandro Innocenti

PhD Coordinator
Prof. Fabio Schoen

ciclo XXXVI, 2020-2023

Università degli Studi di Firenze, Dipartimento di Ingegneria dell'Informazione (DINFO).

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Information Engineering. Copyright © 2024 by Piero Danti.

To my mom, dad, and Natalia,
who may not fully grasp my work,
but always support me wholeheartedly.

Abstract

In recent years, two major subjects have emerged in the scientific community: artificial intelligence and sustainability. These topics converge in developing data-driven techniques for detecting malfunctions or inefficiencies in energy generation systems. Extensive research exists on applications for large-scale generators, however, smaller generators, particularly micro-cogenerators, have received minor attention. This study focuses on the YANMAR's best-selling micro-cogenerator model in Europe. However, findings from this investigation can be extended to a diverse range of sizes and models. Dealing with real data, this thesis explores various facets related to the challenge of anomaly detection in industrial components, thus giving rise to multiple contributions. In particular, it addresses anomaly detection in energy systems by tailoring a general deep learning technique, the autoencoder, to handle time series data. The study presents a general methodology to determine the optimal dataset size for training an autoencoder. An application of the algorithm to predict cogenerator faults is proposed after estimating the dataset size that offers the best compromise for the YANMAR micro-cogenerator under analysis. False positives are reduced through a frequency-based technique. Additionally, a failure root cause analysis is conducted to identify features associated with abnormalities. The proposed approach is validated by predicting various fault types several weeks prior, potentially preventing breakdowns and inefficiencies. Furthermore, the study applies a technique to quantify the algorithm's detection confidence. This technique enables the development of condition-based maintenance strategies tailored to the fault's uncertainty level. Finally, a method is proposed to retrain the model, ensuring its performance despite variations due to the system's aging and seasonal operations which may lead to erroneous abnormality detections. By addressing these challenges, this research significantly advances autoencoder-based methodologies and anomaly detection in micro-cogenerators. Rooted in real-world industrial data, the developed procedures offer practical solutions to enhance the reliability and efficiency of energy systems. This work attempts to fill the gap between artificial intelligence techniques and their application in industrial contexts, providing valuable insights for future research and practical implementations.

Keywords: anomaly detection, autoencoder, energy systems, cogeneration unit, time-series, artificial intelligence, deep learning, train-set size, false pos-

itives reduction, heat exchanger, fouling, uncertainty quantification, Monte Carlo dropout, domain adaptation.

Contents

Chapter 1

Introduction

1.1 Problem Statement

In recent years, the field of Artificial Intelligence (AI) has experienced significant growth, driven by advancements in hardware capabilities, increased computational power, and the availability of vast amounts of data. This growth has been further fueled by the development of Internet of Things (IoT) infrastructures, leading to a digital transition across various industries. One area that has witnessed substantial advancements is maintenance, specifically in keeping machinery and equipment in optimal condition through repairs, problem correction, and periodic adjustments. Traditional maintenance strategies, such as Time-Based Maintenance (TBM), have evolved into more modern approaches, such as Condition-Based Maintenance (CBM), due to the opportunities presented by advancements in AI. In the context of industrial energy plants, maintenance strategies have traditionally relied on TBM, where equipment or systems are serviced at predetermined intervals, regardless of their actual condition. This usually involves regular inspection, cleaning, and replacing worn or consumable parts. Ranges are often based on manufacturer recommendations, industry standards, or previous experience. Although TBM aims to prevent equipment failure and reduce unscheduled repairs, if maintenance intervals are temporally too tight, interventions may result in redundancy leading to an increment of costs, on the other hand, if they are spread out over time, faults can occur damaging devices and causing plant shutdowns.
Nonetheless, in the era of big data and Industry 4.0, other maintenance

strategies are considered more practical, such as CBM and Predictive Maintenance (PM); both approaches focus on monitoring the actual condition of the equipment using various sensors, and then performing maintenance only when necessary. Ahmad et al. [2] gives an overview of industrial applications of TBM and CBM. With the advent of AI and cloud computing, a growing number of facility managers strive to transition towards a data-driven maintenance approach. The main reason is customer loyalty: indeed, a machine that seldom breaks and proves to be reliable generates invaluable benefits to the brand reputation.
Recently, another topic to be taken into consideration, which in the past was often overlooked, is the increasing costs of energy joined with the European will to become the first climate-neutral continent that poses attention to generators' efficiency. A maintenance intervention must be planned not only to prevent a failure but also to mitigate machines' wearing and loss of performance.
Anomaly detection, intended as the identification of rare items, events, or observations that deviate significantly from the majority of the data and do not conform to a well-defined notion of normal behavior, plays a critical role in industrial energy plants, particularly in detecting abnormalities in machine behavior that could lead to energy inefficiency, increased consumption, or costly breakdowns. Machine learning and deep learning models have arisen as the most promising methodologies for addressing these issues [99]. Deep learning, a powerful machine learning technique inspired by the human brain's structure and function, has emerged as a leading approach. Utilizing artificial neural networks with multiple hidden layers, deep learning can extract features from data and encode them into different levels of abstraction, outperforming traditional machine learning methods in various applications. However, the success of deep learning heavily relies on extensive datasets [34]. For machines operating at lower power levels, maintenance or interventions are typically carried out when the presence of a problem is certain, necessitating not only anomaly detection but also identification of the sensors involved and confidence levels for the anomalies. This further reinforces the importance of CBM [1].

1.1.1 Root Cause Analysis and Failure Identification

Root cause analysis and component failure identification play a key role in the successful implementation of an anomaly detection routine. When

anomalies occur within a system, merely addressing the surface-level symptoms might provide temporary relief but will not prevent future occurrences. By delving into root cause analysis, one can uncover the underlying factors that led to the anomaly, thus enabling the development of targeted and effective interventions. Identifying the specific component that failed is crucial as it facilitates the response process. Instead of conducting a general survey and potentially wasting time and resources, knowing the exact component allows technicians to arrive on-site equipped with the appropriate tools and spare parts, ready to tackle the problem head-on. This approach enhances efficiency, minimizes downtime, and fosters proactive maintenance strategies, ultimately resulting in more reliable and optimized system performance.

1.1.2 Uncertainty Quantification

Another aspect that has central importance in practical applications is having a robust uncertainty quantification of the anomaly detection algorithm output. Anomaly detection algorithms can provide valuable insights into potential issues within a system, but blindly relying on their output can lead to costly and unnecessary maintenance actions. By quantifying uncertainty, the algorithm gains a measure of confidence in its predictions; in this way, it is possible to set a threshold tailored to the specific importance of the system at hand. Less crucial systems might demand higher confidence levels before initiating maintenance actions, ensuring that interventions are targeted and justified. In contrast, critical systems might accept lower certainty thresholds, facilitating proactive maintenance measures. Uncertainty quantification fosters a risk-based approach, reducing the chances of unnecessary downtime or premature interventions, while also enabling more informed decision-making to optimize the overall system reliability and performance.

1.1.3 Domain Adaptation

Small facilities are often managed by insufficient staff who cannot closely monitor and track all the installations for which they are responsible. This is because the cost savings from efficiency improvement or basic maintenance do not justify a significant allocation of human resources. Therefore, it is crucial to automate the detection of failures or deteriorations as much as possible while minimizing false positives (detected warning signals for anomalies that are not really present). Indeed, an algorithm that produces

many false positives not only becomes unreliable but also consumes a considerable amount of time for an operator who must oversee multiple plants to do the anomaly check. For this reason, addressing the problem of domain adaptation is essential to prevent non-dangerous trends in the machine (e.g., due to seasonal variations in the system) from being flagged as anomalies merely because the algorithm has not encountered them during the training phase.

1.1.4 Objectives

The proposed doctoral thesis aims to combine YANMAR's expertise and data in the domain of energy generators and internal combustion engines with signal filtering, data analysis, and machine learning techniques to develop a methodology for detailed diagnostic activities. The primary objective encloses four critical aspects: firstly, identifying potential anomalous states in the machinery under analysis; secondly, determining the root cause of these faults; thirdly, quantifying the algorithm's output confidence, and fourthly experimenting with the capability of the algorithm to adapt to new patterns when acquiring online data.
As a final consideration, a detail that adds significance to the already interesting objectives described is the fact that the work in question aims to operate with real data rather than toy datasets, as can occur in pure research. The greatest challenges to be faced in this framework include:

- obtaining availability and collaboration from domain experts, a task that may seem straightforward but is not guaranteed;
- finding relatively clean, non-noisy data to study the behavior of the relevant CHP units;
- retrieving comprehensive and statistically representative data, especially in the context of energy generators where an essential consideration involves accounting for the seasonality of data patterns;
- identifying CHP units where domain experts can confirm the normal behavior of the machine for characterization using data-driven techniques;
- CHP units where domain experts are aware of the presence of an anomaly and its resolution;

- when identifying an anomaly, being able to understand when the anomaly began. The difficulty in establishing an exact moment of anomaly onset poses a significant challenge in evaluating results using metrics commonly employed in the literature.

1.2 Contributions

This work attempts to explore and develop four essential topics. Firstly, it aims to establish a versatile methodology applicable to many variants of autoencoders. The primary objective is to determine the optimal dataset size required for effectively training an autoencoder, the chosen model for addressing the problems tackled in this thesis, enabling it to reconstruct specific inputs without encountering overfitting issues. The second topic deals with a practical anomaly detection and root cause analysis challenge applied to a YANMAR Combined Heat and Power (CHP) system. This particular system plays a crucial role in meeting the energy needs of the facility where it is installed, making efficient anomaly detection vital for its smooth operation. The third topic assumes a paramount role in enhancing CBM strategies. It focuses on incorporating critical information related to the level of uncertainty associated with each anomaly detection, further refining maintenance decisions, and overall system reliability. Lastly, a partial retraining strategy is used to cope with false positives generated by temporal domain shift.
Through these comprehensive investigations, this work seeks to contribute significantly to the fields of autoencoder-based methodologies and anomaly detection in CHP systems, utilizing real-world data from the industrial domain with all the associated challenges and complexities.
Chapter 2 addresses a comprehensive state-of-the-art investigation, covering various topics related to anomaly detection. It spans from its application in the energy domain to the subset of algorithms known as novelty detection, extending toward the utilization of time-series data. Furthermore, the chapter delves into an in-depth exploration of deep learning techniques, particularly focusing on the autoencoder. Lastly, there is a sector-specific analysis pertaining to the concept of predictive uncertainty quantification and domain adaptation, aimed at mitigating the intrinsic temporal variations within the data. Chapter 3, on the other hand, provides an overview of the fundamental techniques employed in this thesis, specifically going through the mathemat-

ical theory of certain cornerstones upon which the work is built. Chapter 4 introduces the subject of the study, the YANMAR micro CHP, from a purely engineering and plant-oriented perspective. Chapter 5 aims to fill a technical gap present in the literature: a great number of works propose the application of autoencoders to a variety of problems even though the procedure to size the train-set is not provided. Researchers are prone to exploit the whole available data to maximize their model's performance; unluckily this behavior may lead to unpleasant drawbacks:

- implementing an autoencoder that needs a lot of data to be trained is against the industry trend to foster plug-and-play products. Indeed, it is challenging to sell a product that will be operative months after the installation;
- using a big train-set may lead to training the autoencoder with data that are far in the past and so less representative;
- using a small train-set may lead the researcher to neglect some deep learning techniques, like autoencoders, due to poor performance and underfitting.

In summary, Chapter 5 contributes to the literature by presenting:

- a methodology to define an optimal train-set size by considering a pretuned autoencoder. The methodology has been validated over three real case studies demonstrating that, when dealing with a 20 kW_e CHP unit, it is acceptable to collect 6 weeks of data to capture the system's normal behavior;
- a general procedure to define the retrain frequency. In particular, it has been explained how a CHP exhibits daily patterns and, according to the computational resources availability, it can be useful to retrain the algorithm every 24 hours;
- a cross-validation routine to evaluate the methodology outcomes respecting the constraint of not shuffling time-series data. This routine produces an optimized splitting of data to assess the autoencoder's reconstruction performance when varying the train-set size.

Chapter 6 tackles an untreated problem in the literature: anomaly detection and root cause analysis applied to micro-CHP units. In particular, this chapter aims to fill the technical gap by introducing a model to detect abnormal

behaviors, developing a post-processing technique to reduce false positives detection, and devising a method to address which unit components led to the abnormal behavior. The main outcomes of this part of the work are:

- a deep learning architecture trained with normal data capable of detecting CHP's anomalies;
- a frequency-based filtering procedure that improves algorithm robustness by masking unreliable detections;
- a root cause analysis technique to highlight which measurements of the unit deviate from the healthy status.

Chapter 7 investigates an unexplored aspect in the current literature, namely the utilization of Bayesian quantification for anomaly detection in micro-CHP units. Specifically, an existing autoencoder architecture is enhanced to accommodate the incorporation of Monte Carlo Dropout (MCD) layers. The challenge lies in striking a balance that preserves the detection performance of the model despite the regularization introduced by the dropout layers, while also enabling the quantification of anomaly severity both from a system-level perspective and from the individual contribution of each signal.
Chapter 8 addresses the challenge of domain adaptation to prevent false positives, and it explores whether fine-tuning a pre-trained autoencoder can restore model performance without complete retraining. Results indicate successful restoration of performance using a targeted fine-tuning approach, reducing the training data requirement from 6 weeks to just 1 week. The study's significance lies in demonstrating the importance of fine-tuning for adapting models to changing conditions and improving anomaly detection in micro-CHP systems, enhancing operational efficiency and safety.

Chapter 2

Literature Review

This chapter provides an overview of state-of-the-art in anomaly detection from various perspectives. In the first analysis, this chapter focuses on literature that applies data-driven maintenance strategies to energy systems (Section 2.1). This review reveals that the specific problem addressed in this thesis has not been previously investigated. Later on, a brief overview of anomaly detection (Section 2.2) is provided to frame the problem. In order to select a suitable technique, the second part of the analysis delves into anomaly detection strategies, particularly focusing on novelty detection problems (Section 2.3). Additionally, the review focuses on works within the domain of multivariate time-series (Section 2.4). Given the superior performance of deep learning techniques in the context of time-series data, a comprehensive review of works focusing on deep learning and autoencoders has been conducted (Section 2.5), as they represent the primary deep learning novelty detection algorithm. Moreover, a brief research has been done to select works that investigate the problem of quantifying uncertainty when solving a data-driven task (Section 2.6). In the end, Section 2.7 reviews works dealing with deep learning domain adaptation to keep machine learning models updated with the current status of real data.

2.1 Anomaly Detection Applied to the Energy Domain

This introductory paragraph aims to demonstrate that the problem addressed in this thesis is not present in the existing literature. Additionally, it seeks to illustrate how machine learning and deep learning algorithms have been applied to the field of energy generation in recent years, albeit with a notable omission of micro-cogenerators.
More in-depth, anomaly detection in the context of small to medium-scale energy generators remains an area with limited research, with the exception of wind turbines, photovoltaics (PV), and building energy consumption, where relevant literature is available. Bellanco et al. [9] reviewed possible faults and diagnoses for heat pumps, indeed many heat pumps installed in buildings are not performing well, leading to lower efficiency and an increase in energy consumption. Lee et al. [58] started from a Fault Detection and Diagnosis (FDD) model trained on a water chiller's actual data and transferred it to a group of 100 refrigerators optimizing energy use reaching conservation of 17.3% of the total. Using a Principal Component Analysis (PCA) based approach, Beghi et al. [8] detected and diagnosed typical chiller faults such as reduced evaporator water flow, refrigerant leakage, condensate flow reduction, and compressor efficiency decrease. Han et al. [46] compared a Support Vector Machines (SVM) with a Least-Squares SVM (LS-SVM) to detect anomalies in refrigeration systems finding out that the LS-SVM performs slightly better in terms of accuracy but needs a remarkably lower time (about 37% less) for training and inference. Yu et al. [110] utilized PCA to discern failures from the normal operation of a sewage source heat pump system ensuring a healthy and efficient operation.
Several recent studies have shown much interest in predicting wind turbine anomalies to improve condition monitoring and power generation efficiency while reducing maintenance costs. Feng et al. [33] introduce an innovative unsupervised anomaly detection framework, incorporating physical-statistical feature fusion and Graph Neural Networks (GNNs) to tackle the challenges posed by massive field data, particularly the lack of labeled SCADA data. Their approach effectively captures latent nonlinear correlations and temporal dependence, significantly enhancing anomaly detection performance. In another study, Matsui et al. [72] address the critical issue of blade damage caused by lightning strikes, leading to secondary damage due to continued

rotation. To mitigate downtime and enhance wind turbine availability, they employ a machine learning model along with SCADA data for assessing the soundness of blades post-lightning strikes. This timely evaluation allows for quicker resumption of operations and increased uptime. Furthermore, Urmeneta et al. [102] contribute to performance-based maintenance strategies for wind turbines by offering a generic methodology for system-level performance assessment. Through the detection of critical periods of low performance, leveraging multiple machine learning methods and models on SCADA data, their approach provides valuable insights for analysts, particularly in scenarios where knowledge about variables impacting performance is limited. This holistic approach holds significant promise for the wind energy sector, supporting asset profitability and improving overall maintenance efficiency.

Moreover, the field of failure detection in PVs is quite investigated. Many researchers have explored various methods and techniques to detect anomalies and defects in photovoltaic systems to ensure reliable and efficient operation. From deep learning-based approaches like SeMaCNN [56] and attention classification-and-segmentation networks [50] to model-based anomaly detection algorithms [30] and unsupervised monitoring procedures [47], the efforts to develop accurate and efficient detection methods have been significant. Additionally, some studies have focused on specific types of anomalies, such as micro-crack detection [50] and cell-level anomalies segmentation [80], utilizing advanced techniques like Convolutional Neural Networks (CNN) and weakly supervised learning. With the increasing adoption of PV systems and the growing importance of green energy technologies, reliable failure detection mechanisms play a crucial role in ensuring the long-term sustainability and performance of photovoltaic installations.

Numerous researchers have shown significant interest in the field of anomaly detection in building energy consumption. Copiaco et al. [19] propose an innovative deep anomaly detection approach that utilizes two-dimensional (2D) energy time-series images, enabling a supervised Deep Transfer Learning (DTL) approach to detect anomalies effectively. They combine physical-statistical feature fusion and GNNs to capture latent nonlinear correlations, resulting in superior performance compared to state-of-the-art methods. Himeur et al. [49] present a comprehensive review of existing anomaly detection frameworks for building energy consumption based on artificial intelligence. Their survey classifies algorithms based on different modules

and parameters, including machine learning algorithms, feature extraction approaches, anomaly detection levels, computing platforms, and application scenarios. They also discuss domain-specific challenges and future research directions to promote wider applications and effectiveness in anomaly detection technology. Wanasundara et al. [104] focus on detecting thermal anomalies in buildings using frequency and temporal domain analysis. They develop a Fourier transform-based method to identify places with thermal anomalies, enabling the identification of poorly-performing areas within building environments using non-permanent sensors. Capozzoli et al. [13] propose an automated load pattern learning and anomaly detection methodology for enhancing energy management in smart buildings. Their approach characterizes energy consumption patterns over time identifying infrequent and unexpected energy patterns, and provides timely alerts based on anomaly detection outcomes. Fan et al. [32] investigate the potential of autoencoders in unsupervised anomaly detection in building energy data. They propose an autoencoder-based ensemble method, comparing different autoencoder types and training schemes to provide advanced tools for anomaly detection and performance benchmarking. Qu et al. [86] develop a combined genetic optimization with AdaBoost ensemble model for anomaly detection in buildings' electricity consumption. Their approach uses synthetic samples, dimension reduction, and an ensemble deep learning network based on AdaBoost to effectively detect anomalous electricity consumption, outperforming other detection methods in sensitivity and the Area Under the Curve (AUC). Araya et al. [5] developed an ensemble learning framework to reduce energy waste in smart buildings by detecting abnormal consumption: this framework led to 8% false positives reduction and 15% true positives increase. [112] et al. explore the potential of Generative Pre-trained Transformers (GPT) in automated data mining for building energy management. They demonstrate that GPT-4 can automatically solve various data mining tasks in this domain, overcoming the barriers of practical applications and revealing future research directions.

Certain specific devices have collected increased attention in recent years, likely due to substantial investments directed toward renewable energies, coupled with intensified focus on consumption reduction. This trend, however, is also propelled by the adoption of micro-cogenerators, which represent a proven, cost-effective, and energy-efficient solution for supplying both electricity and heat. In conclusion, a limited quantity of works treat the anomaly

detection problem applied to particular devices but a total lack of attention toward micro-cogeneration units has been documented.

2.2 Anomaly Detection, a First Overview

The first step that must be taken before framing the problem addressed in this thesis is to describe the various types of existing anomalies and the taxonomy of methods found in the literature to solve them. In general, the concept of anomaly detection encompasses two different aspects: outlier detection and novelty detection. Outlier detection aims to identify an observation that appears inconsistent with the rest of the observations, whereas novelty detection refers to determining a novelty as something different from what had been observed and learned before.
Moreover, when delving into the domain of anomaly detection, it is crucial to discern between various methodological approaches. Anomalies can be detected through both unsupervised and supervised techniques. Unsupervised methods operate without labeled data, relying solely on the inherent patterns within the dataset to identify outliers or novelties. On the other hand, supervised methods utilize labeled data to train the model, allowing it to distinguish between normal and abnormal instances based on the provided examples. Between these two techniques lies the subset of semi-supervised algorithms, wherein only partial examples are provided. Typically, these algorithms are constrained to learn solely the normal behavior of the object under analysis [11], [88], [92].
Additionally, the detection methods can be categorized as univariate or multivariate. Univariate approaches assess anomalies based on a single variable, whereas multivariate techniques consider the relationships and dependencies between multiple variables, enabling a more elaborated understanding of complex anomalies that might involve multiple interrelated factors.
Furthermore, the context in which anomalies are identified plays a significant role. Point anomalies refer to individual data points that deviate from the norm, whereas contextual anomalies are identified concerning the surrounding context, taking into account the relationships and dependencies between data points within a specific context or environment.
Lastly, the temporal aspect of anomaly detection distinguishes between static and sequential methods. Static approaches analyze the dataset at a specific point in time, making independent assessments of anomalies. In contrast,

sequential methods consider the temporal order of data points, detecting anomalies by analyzing the sequential patterns and trends, providing a dynamic perspective on anomaly detection. In this case, data must be shaped as time-series.
Understanding these diverse dimensions of anomaly detection methods is fundamental for choosing the appropriate approach tailored to the specific problem under consideration, ensuring accurate identification and effective mitigation of anomalies in various applications.
Before diving into the discussion, it is important to clearly define the problem explored in the thesis. This step is crucial for focusing the literature review. While a general, but non-exhaustive, overview of the anomaly detection landscape has been given, the aim is to narrow down the focus to specific areas relevant to this thesis. Figure 2.1 illustrates a tree diagram representing the aforementioned anomaly detection context. In particular, the dark gray areas represent the thesis's scope, while the light gray ones with dashed borders indicate the areas not covered by the thesis.

2.3 Strategies for Anomaly Detection

The first step in the process of selecting the appropriate strategy is to assess the type of dataset available and, particularly, the quality of the available labels. A supervised learning approach can be pursued in the case of well-documented data with annotated instances of anomalous and normal behavior. By training a model on this labeled data, higher performance may be achieved in anomaly detection. Supervised learning allows the model to learn patterns from the labeled examples and subsequently identify similar anomalies in new instances.
However, not all situations provide such labeled data. When tackling a problem of anomaly detection in an industrial context, one of the main issues researchers experience is the lack of labeled data. It is even more difficult to have a balanced dataset containing acquisition describing the normal behavior of the machine under analysis and acquisition catching abnormal functioning. A well-known workaround is to build an algorithm able to learn the normal behavior of a system, for instance, in the early years following the commissioning of the object under analysis, and then detect when the operating condition is deviating from the statistical distribution acknowledged as normal. These techniques take the name of semi-supervised algorithms and

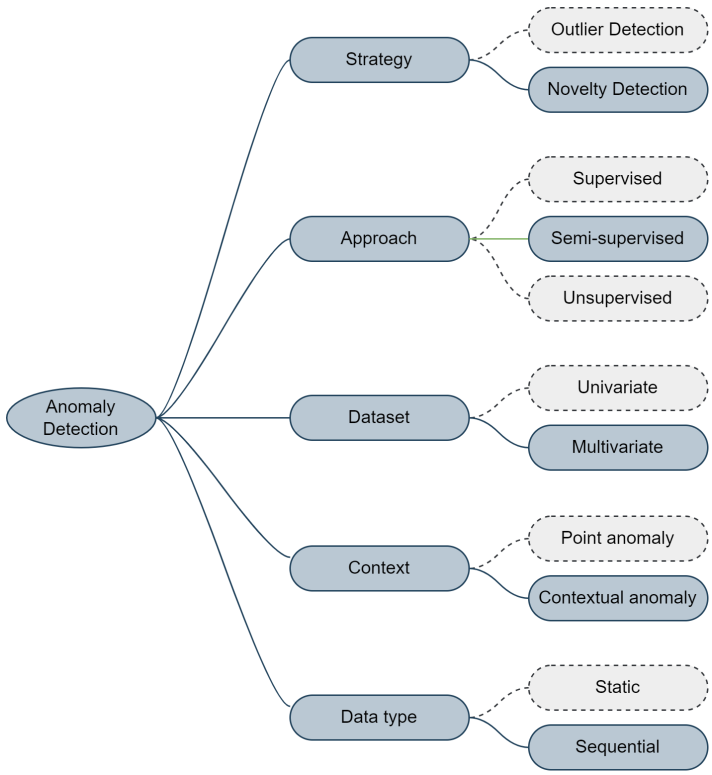

Figure 2.1: Anomaly detection context and frame of the problem under analysis.

the task slightly changes the definition from anomaly detection to novelty detection.
The notion of novelty detection refers to the identification of unusual or anomalous patterns within large amounts of typical data. An anomaly is a deviation from the expected behavior of a pattern and it is often referred to as a novelty. Detection of novelty is meant to identify abnormal behavior within the system that differs from what is normal [73].
Pimentel et al. [83] group novelty detection in five branches:

- Probabilistic novelty detection involves determining the likelihood of new data by estimating the probability distribution of normal data. This allows for setting a threshold to identify abnormal data and determine if it is from the same distribution as the normal data. Gaussian Mixture Model (GMM) is a common choice, Oluwasegun et al. [79] exploit a GMM to detect anomalies in control element drive mechanism of nuclear power plant reaching values of 85% of precision and 94% of recall.

- Distance-based novelty detection calculates the similarity between two data points by means of a specific distance metric. One popular method is the k-Nearest Neighbors (k-NN) algorithm, which supposes that normal samples will be closely surrounded by other normal samples in the train-set, while novel data points will be farther away from them. k-NN found a large application in rolling element bearings [90] [109], gears [61] and motors [38]; Ding et al. [29] gather a collection of datasets and test different benchmark algorithms proving that k-NN achieves a competing overall performance. Some recognized flaw of k-NN is that setting an optimal neighborhood parameter is not trivial.

- Reconstruction-based novelty detection methods can independently model the data, and when new data is presented, the difference between the new data and the output of the model can be used to determine its novelty. This difference is called the reconstruction error and it is related to the novelty score. AutoEncoder (AE), together with its several modifications, belongs to this group and nowadays is the most used technique for anomaly detection tasks. Sohaib et al. [95], Zhou et al. [116] and Sun et al. [98] leaned on AEs to diagnose bearing faults. Lu et al. [67] individuated the faulty component of rotary machinery by means of an AE-based health state classificator and they reported

a classification rate ranging between 91.8% and 95.6% depending on the operating condition.

- Domain-based methods calculate a boundary depending on the layout of the train-set. These approaches are not affected by the distribution of the target class, because they focus on the boundary of the target class, rather than the density of the class. Unknown data is classified by its location in relation to the pre-calculated boundary. Like in classical SVM, the modified SVM developed for novelty detection, referred to as One-Class SVM (OCSVM) in literature, draws the classification boundary by considering only the points that are in its neighborhood, these points are called support vectors. Only support vectors are taken into account when setting the novelty boundary, other training data are neglected. Tan et al. [100] and Saari et al. [88] apply an OCSVM respectively to a naval propulsion plant and to a windmill bearing.

- Information-theoretic methods evaluate the amount of information in a dataset using measures such as entropy and relative entropy. These methods are based on the assumption that novel data will significantly change the amount of information in the normal dataset. Keogh et al. [51] discern suspect sub-sequence inside various types of time-series introducing a novel algorithm called HOT SAX. He et al. [48] use a Local-Search heuristic Algorithm (LSA) for outlier detection tasks: LSA limitation is represented by the fact that the number k of expected outlier must be given as input to the algorithm. In industrial applications, this number is usually unpredictable.

2.4 Anomaly Detection for Sequential Data

When data is organized sequentially and includes information about the timing of measurements, it is referred to as time-series data. In this case, an additional challenge must be addressed, indeed it is recommended to employ an algorithm capable of retaining samples' temporal correlation. Time-series data refer to data that are collected over a period of time, and they are often used in the context of forecasting or prediction tasks. Using deep learning algorithms with time-series data can be useful because it allows for the modeling of complex temporal dependencies and patterns in the data. These algorithms, such as Recurrent Neural Networks (RNN) can learn to identify

patterns in the data and use them to make predictions about future values. Anomaly detection in time-series data is a crucial task in various domains, including High-Performance Computing (HPC) systems, smart manufacturing, marine autonomous systems, and more. In recent years, deep learning approaches, especially autoencoders, have gained popularity for their ability to capture complex patterns and represent data in a compressed form. This section focuses on the use of anomaly detection applied to time-series data, particularly leveraging deep learning techniques.

One common strategy is to use a deep neural network to build a digital twin of the examined system and then to identify data that deviate significantly from the output of the digital twin; when this deviation overpasses a pre-determined threshold, abnormal behavior is detected. Bahlawan et al. [6] presented a study adopting a particular type of RNN, a Nonlinear Auto-Regressive network with eXogenous inputs (NARX), to create a digital twin of a heavy-duty gas turbine. They reported successful simulations of the dynamic behavior of the power generation system during start-up; the developed model could be an efficient tool for the diagnosis of gas turbine faults.

A state-of-the-art approach is to use AE with RNN layers and in particular, Long-Short Term Memory (LSTM) layers [77], [70], [69]. Till et al [101] design a Temporal Convolutional Network AutoEncoder (TCN-AE) where dilated 1D-convolutional layers substitute the fully connected layers; the proposed algorithm outperforms several benchmark techniques in detecting anomalies in medical time-series. Kiranyaz et al. [55] review the major applications of 1D-CNNs including fault detection and condition monitoring: they show how 1D-CNNs achieve optimal results by being characterized by a fast training phase.

Borghesi et al. [12] investigate anomaly detection and anticipation in High-Performance Computing (HPC) systems. The authors propose using a deep learning model trained on labeled data extracted from a service monitoring tool to detect anomalies and predict their occurrence in HPC systems.

Kieu et al. [52] propose two ensemble-based outlier detection solutions for time-series data using recurrent autoencoders. The methods aim to improve overall detection quality by reducing the effects of overfitting to outliers.

Netti et al. [75] present a machine learning approach for online fault classification in HPC systems. The proposed method uses machine learning classifiers to detect and classify faults in real-time, facilitating timely correc-

tive actions.
Alfeo et al. [3] suggest the use of an autoencoder in the design of an anomaly detector for smart manufacturing. The autoencoder is employed to learn representations of multivariate time-series data for anomaly detection.
Anderlini et al. [4] develop a remote fault detection system for underwater gliders using time-series data. They use feedforward deep neural networks and autoencoders for online anomaly detection in glider operations.
Shu et al. [93] present a dam anomaly assessment system based on a sequential Variational AutoEncoder (VAE) and evidence theory. The proposed method is designed to detect anomalies in dam sensor data. Zheng et al. [115] introduce Deeppipe, a semi-supervised learning approach for operating condition recognition of multi-product pipelines. The method leverages autoencoders and semi-supervised learning techniques for anomaly detection.
Yu et al. [111] analyze different RNN autoencoder variants for time-series classification and machine prognostics. The study investigates various architectures and their performance for time-series anomaly detection.
Ou et al. [81] propose a deep sequence multi-distribution adversarial model for bearing abnormal condition detection using autoencoders. The model combines VAEs and LSTM predictors for robust prediction.
Chen et al. [17] present SeqVL, a neural network model that integrates unsupervised anomaly detection and time-series prediction under one framework. The model merges VAEs and LSTM for improved detection and prediction.
von Schleinitz et al. [103] propose VASP, a VAE-based selective prediction framework for robust time-series prediction in the presence of anomalies. The method is applied to motorsport data.
Li et al. [60] present a clustering-based anomaly detection method for multivariate time-series data.

2.5 Autoencoders for Sequential Novelty Detection

Among deep learning algorithms, this work focuses on a particular type of semi-supervised neural network: the autoencoder.
The autoencoder is part of a larger family of representation learning methods capable of learning features from unlabeled data automatically. These methods are designed to map the input data to an internal latent represen-

tation, which is then used to produce an output similar to the input data Several methods exist that can automatically learn features from data, including the autoencoder. An internal latent representation is generated from the input data, which is then used to create output that is similar to the input data [65].
Although many papers exploit autoencoder-based models, there are no best practices on how to tune the hyper-parameters and how to size the train-set. Hyperparameter tuning in deep learning algorithms refers to the process of systematically optimizing the hyperparameters, which are external configuration settings not learned from the data. These parameters significantly impact the performance of the model but cannot be directly learned during training. The challenge lies in finding the ideal combination of hyperparameters, such as learning rate, batch size, and network architecture, to achieve optimal model performance. To date, various advanced techniques such as Bayesian Optimization, Neural Architecture Search, Reinforcement Learning, as well as more traditional approaches like Random Search, Grid Search, and Genetic Algorithms, have been developed for hyper-parameter optimization. These algorithms rely on a specific metric to optimize, such as accuracy or F1-score in anomaly detection. Unfortunately, industrial datasets often lack reliable labels, making it challenging to rely on existing labels, which can even prove counterproductive for the algorithm. Furthermore, when employing autoencoders, it is not feasible to solely minimize reconstruction error, as it could lead to creating a structure that simply replicates inputs to outputs, rendering the model ineffective for anomaly detection purposes. However, the adoption of manual fine-tuning remains widespread in industrial domains due to its relatively lower cost and the challenges associated with using unreliable labels or minimizing reconstruction error in autoencoders.
After a brief overview of the most notable works employing techniques based on autoencoders for anomaly detection, this section will also investigate the presence of research addressing the challenge of determining the appropriate trainset size.

2.5.1 Overview of Autoencoder Applications

Qian et al [84] offer an exhaustive review of autoencoder-based applications in industrial processes monitoring. The study introduces the encoder-decoder framework and emphasizes its importance in unsupervised feature

extraction. The review focuses on the characteristics of industrial data and explores state-of-the-art fault detection and fault diagnosis strategies.
Liu et al. [64] examine unsupervised deep learning strategies applied to time-series collected by IoT devices suggesting that the autoencoder and its variants are widely used in the literature. A Probabilistic Deep AutoEncoder (PDAE) is suggested by Lin et al. [62], in which non-parametric estimated distributions are utilized to construct the uncertainty intervals of measurements in the first layer of neural networks. Deep autoencoder structures are formulated based on the non-parametric estimated uncertainty intervals of the measurements. Outliers can be accurately detected with the proposed PDAE model trained on 20000 samples.
Deng et al. [27] develop a Stacked AutoEncoder (SAE) joined with a softmax layer to classify anomalies in industrial processes able to reach an accuracy greater than 99% in the three case studies tested.
Sarajcev et al [91] exploit a Stacked Denoising AutoEncoder (SDAE) together with a voting ensemble classifier for power system transient stability assessment. The dataset used consists of 3.8 GB signals generated by electromechanical transient simulations and, to deal with class imbalance, a stratified shuffle split is used. Even if the dataset has remarkable dimensions, the possibility of gradually reducing it and tracking the performance trend is not investigated.
González-Muñiz et al. compare various autoencoder structures, including VAEs and Deep AutoEncoders (DAEs), applied to rotating machines, hydraulic systems, and body motion systems. They conclude that VAEs perform better than DAEs because they are capable of learning the probability density function of healthy states [40]. A subsection dedicated to the three different datasets used is present but no mention has been made related to their dimension.
Gokhale et al. present a new deep learning framework for gene selection using an SAE for cancer classification, obtaining a model accuracy ranging from 90% to 100% on ten different datasets [39]. For the experiments the authors collected ten benchmark datasets used in the literature and then applied a data augmentation technique but, due to the nature of the problem, a sensitivity analysis on the influence of the trainset size could not be provided: gene expression datasets tend to be generally poorly populated.
To determine health indicators for bearing performance degradation assessment, Xu et al. develop an SAE model and show that the proposed method

has superior denoising performance to other benchmark models [106]. In this work, two open-source datasets are used and no specific correlation between the size of the training dataset and the proposed models' performance is emphasized.
Zhou et al. propose a novel contrastive autoencoder for anomaly detection in multivariate time-series data and emphasize its capacity to model the normal data pattern [117]. The authors have examined three datasets of varying sizes, and it seems that the nature of these datasets does not significantly impact the results of the four models compared. Furthermore, regardless of the dataset used, the ranking of the best-performing algorithms remains consistent throughout the tests.
Luo et al. propose a Multi-mode Non-gaussian VAE (MNVAE) for anomaly detection of electromechanical equipment that achieves a 90% of accuracy without any false positive [68]. Six methods are compared over two different datasets whose dimension is detailed but not described in relation to the model performance.
Shen et al. develop a Hybrid Robust Convolutional AutoEncoder (HRCAE) for unsupervised anomaly detection of machine tools under noises that obtain better performance compared to other benchmark techniques [107]. In this case, three datasets of 315 samplings are collected but, because the size does not vary, the analysis does not allow extrapolation of any information on the influence of train-set size.
Castangia et al. train a VAE to reconstruct the consumption profile of a washing machine, a dishwasher, and a dryer obtaining a better classification score with respect to the One-Class Support Vector Machines (OC-SVM) used as a reference for the comparison [15]. To run their experiments, the dataset used consisted of 8 months of recorded power consumption acquired at 1 Hz and a classical k-fold validation technique with fixed trainset size has been applied.
A method for detecting and diagnosing anomalies has been developed by Zhang et al. for wind turbines using Long Short-Term Memory-based Stacked Denoising AutoEncoders (LSTM-SDAE) and extreme gradient boosting obtaining an accuracy of 91% [113]. Due to the confidentiality of the dataset, no information is divulged about the train-set size.
Qu et al. introduce an approach to predict anomalies by means of Echo State Network (ESN) and Denoising AutoEncoder (DAE) yielding an accuracy of 83%, 4 points greater than the autoencoder accuracy [85]. The authors di-

vide the available data dedicating the 70% for training and the remaining 30% for testing and they explain this decision by asserting that it is the same ratio of normal and abnormal samples in the dataset.
Borghesi et al. [11] propose an automated semi-supervised approach for anomaly detection in high-performance computing systems based on autoencoder neural networks, demonstrating superior accuracy compared to current techniques, with an increase of approximately 12% in detection accuracy. The research presents a section that explores the impact of reducing the size of the train-set on detection accuracy. Results indicate variations in accuracy levels among different components, with certain components requiring larger train-sets to achieve satisfactory results. On average, a train-set size of approximately 500 examples, corresponding to about 42 hours of continuous monitoring, can yield adequate accuracy. In a subsequent study [12], the same research group applies a fusion of supervised and semi-supervised techniques, both based on autoencoders, to anticipate the high-performance computing system automatically annotated labels by approximately 45 minutes. Given the specific field of application, this time interval is regarded as remarkably interesting.
Zhou et al. [118] tackle weakly-supervised anomaly detection with limited labeled abnormal data. The model, consisting of a feature encoder built upon an autoencoder and an anomaly score generator, has been tested on eight datasets of different sizes originating from various fields of application achieving superior performance over the four comparison algorithms.
Lately, federated learning has sparked a new wave of interest as researchers and industry experts recognize its potential to revolutionize machine learning by enabling collaborative model training across distributed devices without compromising data privacy; Liu et al. [63] introduce five benchmark algorithms for time-series anomaly detection under four federated learning frameworks.

2.5.2 Train-set size definition

If the use of autoencoders in anomaly detection is widely discussed, less attention has been given to studying the impact of the train-set size on reconstruction performance, particularly when the available data consists of temporal frames collected from field sensors. Bengio [10] offers practical recommendations for handling hyperparameters in deep learning algorithms, specifically in the context of backpropagation and gradient-based optimiza-

tion. It emphasizes the importance of adjusting hyperparameters to achieve better results and provides insights into training and debugging large-scale neural networks. While it is acknowledged that increasing the amount of data used for training a neural network can solve potential overfitting issues, there is a lack of specific techniques or guidelines to determine the optimal size of the train-set.

Dawson et al. [26] evaluate the performance of nine CNNs across four architectures for carbonate core classification using deep learning. Transfer learning and fine-tuning are employed on three geological datasets of varying sizes, showcasing the potential of deep learning in automated carbonate classification. The Inception-v3 architecture achieves 92% accuracy on the larger dataset, while the VGG19 architecture is suitable for smaller datasets. Dataset size significantly impacts model performance, with smaller datasets prone to overfitting even with transfer learning.

Radiuk et al. [87] investigate the impact of batch size on the performance of CNNs. MNIST and CIFAR-10 datasets are used to analyze the influence of batch size on image recognition accuracy. Results confirm that larger batch sizes yield higher accuracy, but the hypothesis regarding the specific type of batch size value's impact on CNN performance is not supported. The optimal batch size, depending on computational resources, is determined to be 200 or greater.

Gütter et al [44] examine the influence of train-set size on the robustness of a deep learning strategy for performing detection in satellite imagery, specifically against omission noise in crowdsourced datasets. The experiment involves introducing controlled levels of omission noise to a dataset and training the model on subsets of varying sizes. The results indicate that the train-set size has a notable impact on the model's robustness against label uncertainty. Larger train-sets generally lead to improved robustness, resulting in better performance and increased tolerance to label noise. These findings support the assumption that train-set size positively influences the model's robustness.

Gulamali et al. [43] focuses on evaluating the lowest dataset dimension required for training computer vision autoencoders, specifically examining the point at which classifiers face difficulties in class differentiation. The findings offer a valuable tool for estimating sample sizes in fully connected networks for computer vision deep learning applications.

El-Nouby et al. [31] challenge the necessity of large-scale datasets like Im-

ageNet for self-supervised pre-training in computer vision. By focusing on DAEs, the research demonstrates their superior robustness to the type and size of pre-training data compared to other self-supervised methods. Results show competitive performance on various classification tasks, surpassing supervised ImageNet pre-training in certain scenarios, indicating that large-scale datasets may not be essential for effective self-supervised pre-training with DAEs.
Masters et al. [71] examine the impact of mini-batch size on deep neural network training and generalization performance. Experimental results across CIFAR-10, CIFAR-100, and ImageNet datasets demonstrate that smaller mini-batch sizes lead to better stability and performance. The optimal mini-batch sizes range from 2 to 32, contradicting recent recommendations for larger batch sizes. Additionally, the study highlights the challenges and decreased optimization potential associated with larger batch sizes, supporting the notion that smaller batch sizes offer advantages in terms of learning rate range and performance.
As mentioned earlier, the issue of training dataset size is predominantly addressed in the context of images, with specific sensitivity studies conducted for individual cases. However, to the author's knowledge, there is only one paper [43] that offers a methodology, albeit applied to a computer vision problem and therefore not transferable to time-series problems.

2.6 Evaluating Uncertainty in Detection Algorithms

Frequently, energy generators are operated with low productivity or, in worst cases, some negative trends appear weeks before a real breakdown of the machine; this leads to higher energy consumption and unplanned reparations that could be easily avoided through ad-hoc maintenance intervention. In contexts where machines are highly expensive and production cannot be interrupted, every detection is carefully considered to avoid costly breakdowns or performance losses. Conversely, for energy devices with productions below 100 kW, maintenance or interventions are typically carried out when the presence of a problem is certain. In this regard, it is important not only to provide anomaly detection but also a confidence level for the anomaly itself, and where possible, to point the sensors involved in the problem. This reinforces the concept of CBM.

Regarding adding a confidence level to the anomaly prediction, despite the success of standard deep learning methods in solving various real-world problems, they cannot provide information about the reliability of their predictions [1]. Two main solutions have been proposed in recent years to introduce the information of uncertainty quantification: the VAE [53] and the Monte Carlo dropout (MCD) [35].
VAEs are a type of generative model that learn a low-dimensional representation of the input data by encoding it into a latent space and then decoding it back to the original input space. In this process, VAEs minimize a reconstruction loss between the original input and its reconstruction, as well as a regularization term that encourages the latent space to follow a prior distribution. The resulting model can be used to encode-decode test samples and perform anomaly detection by measuring the reconstruction error of these samples [97], [15], [40], [21].
MCD, on the other hand, is a dropout-based technique that uses dropout [96] during inference to estimate the model's uncertainty as detailed in Section 3.3. Dropout randomly drops out units from the neural network during training, which acts as a regularization technique. During inference, dropout is applied multiple times with different dropout masks, and the resulting predictions are averaged to estimate the model's uncertainty. In anomaly detection, MCD can be used to estimate the uncertainty of the model's prediction for each test sample returning a rate of how many anomalies have been detected during the multiple inferences [14], [89], [59].
One key difference between VAEs and MCD is that VAEs are generative models, while MCD is a technique for uncertainty estimation. VAEs can be used not only for anomaly detection but also for tasks such as data generation. However, MCD is a simpler technique that can be easily applied to any existing neural network architecture. Another difference is that VAEs require a prior distribution over the latent space, while MCDs do not. The choice of prior distribution can have a significant impact on the quality of the VAE's latent space representation, and finding an appropriate prior can be challenging. [54] Finally, VAEs tend to be computationally more expensive than MCDs, as they require training in a full generative model. MCD, on the other hand, only requires running the inference multiple times with different dropout masks.

2.7 Adapting Models in Changing Scenarios

In the context of predictive maintenance in complex machinery, the concept of model retraining holds central significance. Ensuring that machine learning models are aligned with the real-world behavior of the systems they monitor is essential to avoid an excess of false positives, which can lead to desensitization among maintenance technicians. This phenomenon, where technicians might ignore alerts due to their big volume, can result in delayed responses to genuine issues, leading to costly breakdowns and downtime. Therefore, this thesis explores the key role of systematic model retraining in maintaining the efficacy of AI systems.

In the domain of neural networks and machine learning models, the concept of lifelong learning presents a major challenge. This challenge arises from the need to continuously acquire new information from non-stationary data distributions. Unfortunately, this continuous acquisition often leads to problems like catastrophic forgetting or interference, especially for cutting-edge deep learning models. These models are designed to learn from fixed collections of training data, and they don't take into account scenarios where information becomes available gradually over time.

To address this challenge, Parisi et al. [82] review available neural network techniques that aim to relieve this problem. Despite meaningful advancements in domain-specific learning with neural networks, the development of robust lifelong learning capabilities for autonomous agents and robots requires extensive research efforts.

Cossu et al. [20] discusses the importance of continuous learning throughout the lifetime of a machine learning model to ensure robustness against data distribution drifts. It mentions that advances in continual learning with RNNs have the potential to be applied in various domains, such as natural language processing and robotics, where incoming data is non-stationary. However, the existing research on continual learning is currently fragmented, with application-specific approaches and diverse evaluation protocols and datasets.

Hadsell et al. [45] discuss the relevance of continual learning in artificial intelligence research, which aims to enable machine learning models to learn sequentially from a continuous stream of correlated data. It emphasizes that the ability of humans to adapt to changing tasks and learn incrementally is currently missing in modern machine learning methods. Continual learning is seen as a crucial attribute of human-level artificial general intelligence.

The article introduces biologically inspired approaches to continual learning and suggests that developing neural network models with sequential learning capabilities could improve data efficiency, facilitate continual adaptation to changing target specifications, and enhance knowledge transfer between tasks.

Losing et al. [66] explore the significance of incremental and online learning in the context of big data and data streams. It examines eight popular incremental methods, representing different algorithm classes, to understand their key properties. The evaluation focuses on online classification error, convergence behavior, and model complexity. Additionally, the paper addresses hyperparameter optimization, often overlooked in the context of incremental learning, and tests its robustness based on limited examples. The extensive evaluation provides insights into the performance of these methods, aiding the selection of the most appropriate method for specific applications.

Zhang et al [114] introduce Adaptive Online Incremental Learning (AOIL) to tackle the challenges of online incremental learning, including concept drift, catastrophic forgetting, and learning latent representations. AOIL uses an auto-encoder with a memory module to detect concept drift and adjust model parameters. It partitions features into common and private aspects to avoid catastrophic forgetting. A self-attention mechanism is employed to fuse extracted features, enhancing latent representation learning. The addition of a de-noising auto-encoder improves the algorithm's robustness. Extensive experiments show that AOIL outperforms other state-of-the-art methods on various datasets.

Gori et al [42] discuss the application of a continual learning approach to an RNN to monitor the status of a remarkable amount of sensors in a turbomachinery application. Turbo-machinery prototypes are equipped with a large number of sensors, and manually checking the health of each sensor is time-consuming. The authors propose an approach where an RNN is continuously trained on a daily basis to create a virtual sensor based on other sensors' data. The predicted signal is then compared to the real signal to detect potential anomalies in the sensor readings. Kullback-Leibler (KL) divergence is used to estimate the overlap between input distributions available during training and those observed during testing, providing a measure of confidence in the predictions.

Xiaolan et al. [105] present an evolving anomaly detection method for network streaming data to address network security challenges. The proposed

method incrementally updates clusters and detects outliers based on local and global density thresholds. A buffer is used to store temporary outliers to avoid misclassifying normal samples. The algorithm is validated on three datasets outperforming other anomaly detection methods regarding detection accuracy, false-positive rate, and computational cost.

Chapter 3

Background

This chapter presents an overview of foundational concepts pertinent to subsequent discussions. Specifically, Section 3.1 outlines the mathematical notation of the autoencoder and the hyperparameters that define it. Section 3.2 introduces the Fourier transform, which will be utilized as a filtering technique to mitigate false positives. Lastly, Section 3.3 elucidates the concept of Monte Carlo dropout, which will subsequently be employed for uncertainty quantification.

3.1 Autoencoder

An autoencoder is a neural network that is trained to attempt to copy its input to its output. Internally, it has a hidden layer h, often named latent space, that describes a code used to represent the input. The network may be viewed as consisting of two parts: an encoder function $h = f(x)$ and a decoder that produces a reconstruction $r = g(h)$. Usually, input information is reduced by means of a funneled shape of the encoder that does not allow a perfect copy of the information and forces the model to coarsely copy inputs. This constraint theoretically leads the autoencoder to learn useful properties of the data. Indeed, traditionally, autoencoders were used for dimensionality reduction or feature learning. Since autoencoders are the coupling of two feed-forward networks they can be trained using back-propagation [41]. Encoder and decoder are artificial neural networks that may consist of one or more layers interconnected by means of the latent layer (Figure 3.1);

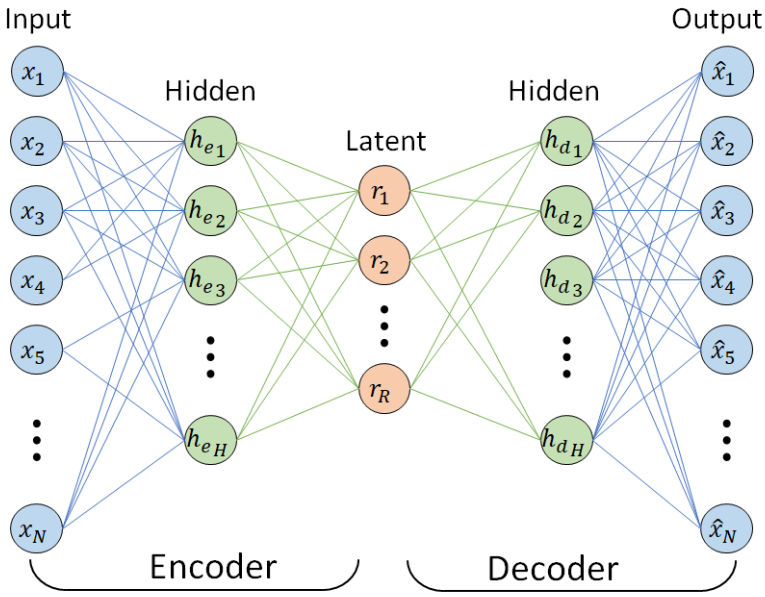

Figure 3.1: Autoencoder structure: encoder-decoder.

usually, the two neural networks share the same structure in terms of number and type of hidden layers (Section 3.1.1), number of nodes in each layer (Section 3.1.2), type of activation function in each node (Section 3.1.3). Furthermore, the dimension of the latent layer may influence the autoencoder's performance and its applicability. Increasing the structure complexity, the autoencoder will be able to catch more sophisticated behaviors of the system under analysis. The encoder and decoder are trained together to minimize a loss function (Section 3.1.4), typically a quantitative measure of the error that the decoder introduces trying to reconstruct the features inputted into the encoder.
In this work, the following mathematical notations are used:

- input (or features') vector:

$$X = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_N \end{bmatrix}^T \tag{3.1}$$

  where N is the number of features;

- output (or reconstructions') vector:

$$\hat{X} = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & \dots & \hat{x}_N \end{bmatrix}^T \tag{3.2}$$

- weights' matrix:

$$W^k = \begin{bmatrix} w_{11}^k & w_{12}^k & w_{13}^k & \dots & w_{1n_l}^k \\ w_{21}^k & w_{22}^k & w_{23}^k & \dots & w_{2n_l}^k \\ & & \dots\dots\dots\dots & & \\ w_{n_k1}^k & w_{n_k2}^k & w_{n_k3}^k & \dots & w_{n_kn_l}^k \end{bmatrix} \tag{3.3}$$

  where w_{uv}^k is the weight of the edge linking the u-th node of the layer l (whose dimension is n_l) with the v-th node of the layer k (whose dimension is n_k);

- biases' vector:

$$B^k = \begin{bmatrix} b_1^k & b_2^k & b_3^k & \dots & b_{n_k}^k \end{bmatrix}^T \tag{3.4}$$

  where b_v^k is the bias of the v-th node of the layer k;

- note that subscript i is used to refer to the N features.

3.1.1 Hidden layers

Hidden layers' typology is usually chosen in accordance with the data type available to train the model whilst the latent layer is usually built upon fully-connected nodes. In the literature, three types of data are the most frequent:

- tabular data: the best practice suggests using fully connected layers both for hidden and latent layers;

- images: 2-D Convolutional Neural Network (CNN) is the state-of-the-art technology for hidden layers [57];

- sequential/time-series data: typically the implementation of LSTM hidden layers [70] or 1D-CNN hidden layers [55] outperforms fully connected layers; another option is represented by the TCN that provides simplicity, autoregressive prediction, residual blocks, and long memory, all of which make them more appealing [101].

In their conclusion, Bai et al. [7] propose that 1D-CNN could be a natural starting point for sequence modeling tasks, since the common association between sequence modeling and recurrent networks may need to be reconsidered. In this work, an autoencoder with convolutional hidden layers and a fully connected latent layer is considered to deal with time-series data.

3.1.2 Dimension of layers

Similarly to the number of layers, the number of nodes of each layer affects the complexity of the autoencoder's structure; obviously the number of input and output nodes is not a hyper-parameter but must be equal to the number of features used to train the model.
Conversely, the dimension of the latent layer affects the autoencoder's behavior not only in terms of performance but also in terms of the field of application. In the literature, the following differentiation is explained:

- Contractive Autoencoder: autoencoders risk learning the identity function if the latent layer has as many neurons as input and output. Latent layers must therefore have fewer neurons than input and output layers in order to satisfy this constraint [94]. Hence, an autoencoder where input and output layers have the highest dimension and the latent layer has the lowest dimension is named a contractive autoencoder;
- Sparse Autoencoder: if the latent layer's dimension is higher than the input/output dimensions, the autoencoder is defined as sparse and, imposing a sparsity constraint on the latent units (also on the hidden units, if present), it will be still capable of extracting valuable patterns from data [76].

In anomaly detection problems, Contractive Autoencoders are typically favored for capturing the most relevant information embedded in the data. However, it is worth noting that employing Sparse Autoencoders with appropriate regularization can also yield promising results. Experimentation with various architectures and hyperparameter combinations is common practice in order to determine the most effective model structure.

3.1.3 Activation functions

Activation Functions make the decision of whether or not to pass a signal to the next layer. They take in the weighted sum of inputs plus a bias. They

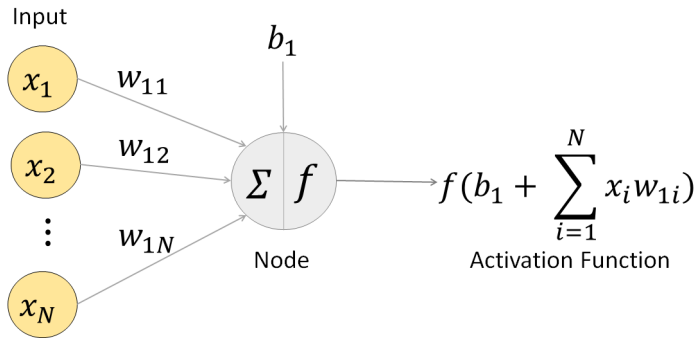

Figure 3.2: Node structure: inputs, node and activation function.

are usually applied to both hidden and output layers and, due to their non-linearity, make neural networks capable of learning complex functions from data. Figure 3.2 shows the foundation element of a neural network: inputs x_i are linearly combined with a bias b and inputted to the node where the activation function f introduces its non-linearity.
The most used activation functions are listed below:

- linear, $f(x) = x$
- sigmoid, $f(x) = \frac{1}{1+e^{-x}}$
- tanh, $f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$
- ReLu, $f(x) = max(0, x)$
- LeakyReLu,

$$f(\alpha, x) = \begin{cases} \alpha x, & if\ x \leq 0 \\ x, & if\ x > 0 \end{cases} \quad for\ \alpha > 0$$

3.1.4 Loss functions

The main purpose of the autoencoder is to reassemble its input as accurately as possible. Training of the model achieves this goal by utilizing a specific loss function, named reconstruction loss function, that penalizes the model for generating outputs different from the inputs. As a loss function, it is

usually used the Mean Squared Error (MSE) between input and output (Equation 3.5).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{3.5}$$

3.2 Fourier Transform

The Fourier Transform is a fundamental mathematical concept widely used in various scientific and engineering domains to analyze signals and functions. Proposed by Joseph Fourier in the early 19^{th} century, this transformative tool revolutionized the understanding of time-varying phenomena.
The primary goal of the Fourier Transform (FT) is to decompose a complex signal or function into its constituent frequency components. This process enables the representation of the signal in the frequency domain, uncovering the contribution of each frequency to the overall signal. Expressing the signal in frequency terms through the Fourier Transform facilitates the analysis and comprehension of its spectral characteristics, such as dominant frequencies, periodicities, and amplitude variations.
The mathematical representation of the Fourier Transform involves integration over an infinite range. However, in practice, fast algorithms such as the Fast Fourier Transform (FFT) have been developed, significantly reducing the computational complexity for practical implementations.
The Fourier Transform finds extensive applications in signal processing, telecommunications, audio and image analysis, as well as in fields like quantum mechanics and medical imaging. In signal processing, for instance, it is utilized for filtering, noise reduction, and modulation, while in quantum mechanics, it helps analyze wavefunctions and understand the behavior of quantum systems. In conclusion, the Fourier Transform is a fundamental tool that provides a powerful bridge between the time and frequency domains, enabling profound insights into the inner workings of signals and functions. Its widespread applications across scientific and engineering disciplines have made it an indispensable technique for understanding and manipulating complex phenomena in the modern world.
As above mentioned, the FT is widely used in signal processing to analyze and extract frequency domain information from monitored measurements. The idea is to decompose the original signal into the sum of innumerable

sine and cosine waves with different frequencies in order to analyze the components of the original measurement [28].
The FT of a generic function $f(x)$ is defined as:

$$F(\omega) = \int_{-\infty}^{+\infty} f(x)\, e^{-i\omega x} dx \tag{3.6}$$

The inverse Fourier transform is described by:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega)\, e^{-i\omega x} d\omega \tag{3.7}$$

When dealing with real data, signals are discrete, and, thus, the Discrete Fourier Transform (DFT) is needed. The DFT of a, known as the spectrum of a, is:

$$A_k = \sum_{n=0}^{N-1} e^{-i\frac{2\pi}{N}kn} a_n \tag{3.8}$$

The sequence A_k is the DFT of the sequence a_n; the inverse DFT of the sequence A_k is the sequence a_n:

$$a_n = \frac{1}{N} \sum_{k=0}^{N-1} e^{i\frac{2\pi}{N}kn} A_k \tag{3.9}$$

Since computing the DFT of a signal is highly time-consuming, the Fast Fourier Transform (FFT) is needed: FFT is a fast algorithm for computing the DFT. Indeed, calculating the DFT of an N-point sequence using Equation 3.8 has a $O(N^2)$ complexity whilst the FFT algorithm computes the DFT with a $O(N \log N)$ complexity as explained by Cooley et al. [18].

3.3 Monte Carlo Dropout

This section briefly introduces dropout emphasizing its role in mitigating overfitting and how it can be used to introduce the concept of uncertainty quantification.
Dropout is a regularization method that selectively deactivates neurons in a neural network during training. Each neuron has a dropout rate denoted by p, representing the probability of being ignored during a training step. Typically, p is set between 0 (no dropout) and 0.5 (approximately 50% of neurons deactivated), adjusted based on network type, layer size, and the

severity of overfitting observed during training.
The primary motivation behind dropout lies in its role as a regularization technique to mitigate overfitting. When dealing with limited data or complex networks, models may memorize the training data, leading to a poor generalization of unseen data. Dropout counteracts overfitting by forcing neurons to distribute their weights more evenly across the network and reducing sensitivity to input changes, thereby promoting better generalization.
It is important to emphasize that dropout is exclusively utilized during the training phase (Figure 3.4) and is not applied during inference (Figure 3.3) time. During inference, the entire network, with all trained neurons and connections, is used to make predictions.
MCD, proposed by Gal and Ghahramani in 2016 [35], establishes a compelling connection between regular dropout and Bayesian approximations of probabilistic models, specifically Gaussian processes. It capitalizes on the random sampling nature of dropout to generate a multitude of different networks, each with different neurons dropped out. These networks can be regarded as Monte Carlo samples from the space of potential models, yielding valuable insights into uncertainty estimation.
By interpreting dropout as a Bayesian approximation, MCD enables the assessment of model uncertainty. The diversity in the dropout ensemble provides a basis for reasoning about uncertainty and often contributes to enhanced performance. Indeed, when using MCD, dropout is active also during inference by randomly deactivating a certain percentage of nodes. Repeating the inference process numerous times the percentage of detection can be calculated and used as a degree of certainty.
In summary, dropout serves as a powerful regularization technique to prevent overfitting and improve generalization. On the other hand, MCD leverages dropout's stochastic nature to approximate Bayesian inference and estimate model uncertainty effectively. These techniques are of major importance in enhancing the performance and reliability of neural networks, particularly in scenarios with limited data or complex architectures.

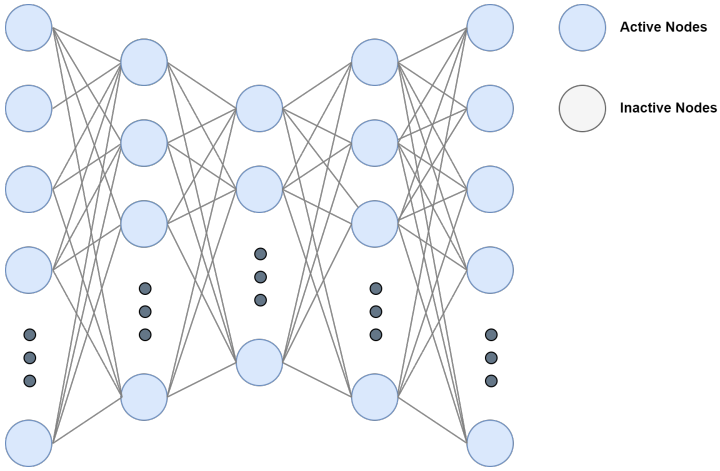

Figure 3.3: Autoencoder without dropout.

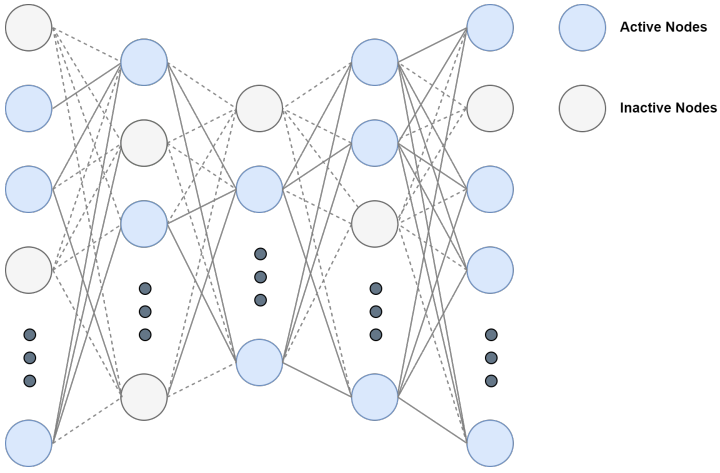

Figure 3.4: Autoencoder with dropout ($p = 0.3$).

Chapter 4

Case Study

This chapter outlines the model of the main characters of each chapter of this work. Although in subsequent chapters, the subject of study may vary, the model and specifications remain consistent. Section 4.1 encloses all the considerations related to the 20 kW_e YANMAR CHP, thus preventing the need for their repetition in each subsequent chapter. Section 4.2 also elucidates the steps involved in acquiring and processing the available data for each considered machine.

4.1 20 kW YANMAR CHP

The energy system considered is a YANMAR CHP, whose datasheet is available at [108], with a rated electrical power of 20 kW_e.
CHP is an energy device supplied by natural gas that produces electricity and thermal energy at high efficiencies. The kind of unit under analysis is designed to satisfy the electrical and thermal needs of the customers' energy plants where it is installed. Figure 4.1 offers an illustrative representation of the plant layout, emphasizing the interconnections of the generators. The red lines denote the piping system responsible for conveying hot water from the micro-CHP unit and the gas boiler to meet the thermal load, while the blue lines indicate the return flow of water after heat exchange with the customer's facility. The electrical connections, drawn in green, represent the wiring network that provides power supply to the facility.
Cogeneration is a process that generates electricity and heat simultaneously.

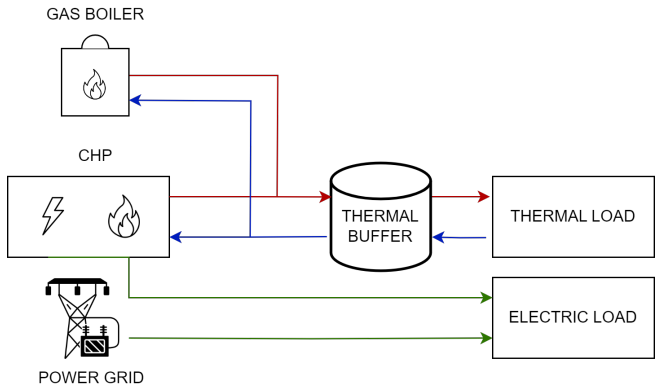

Figure 4.1: Case study plant layout.

It is a highly efficient way to produce energy because it recovers the waste heat that is produced during the electricity generation process and uses it to provide heat or hot water. This results in higher overall energy efficiency compared to generating electricity and heat separately. CHPs are often used in industrial and commercial settings, but can also be used in residential buildings. The CHP unit has at its core an internal combustion engine (Figure 4.3) that converts fuel energy (natural gas in this case) into electrical energy. The heat removed from the engine to cool it as well as the heat present in the high-temperature exhaust gas is recovered and transmitted to the carrier fluid (water) that is feeding the plant. Indeed, the CHP is part of a more complex plant (Figure 4.2) composed of a buffer tank and a heating boiler. The 20 kW_e CHP's engine, backed up by the grid, meets the electric load and the recovered heat is used to charge the buffer tank. The CHP's hot circuit is connected in parallel with the backup boiler to cover the thermal request.
Each of the cases under consideration shares identical plant layouts, thereby resulting in a highly similar dataset with the exception of two specific aspects. The dataset related to the CHP systems could exhibit varying acquisition times ranging from 1 minute to 15 minutes. Furthermore, there could be disparities in the volume of available data for each plant, contingent upon factors such as commissioning dates, potential connection issues, plant shutdowns prompted by internal or external factors, and the governing data management policies.

Generally, these datasets present multiple features from different elements installed in the plant's layout; since this work focuses on learning the normal functioning of only 1 device of the plant, the CHP unit, a technical analysis has been performed and only 16 features have been considered remarkable and detailed below:

- the *active power* is the electric power generated by the CHP system used to meet the electrical demand. By monitoring the electric power produced, it is possible to ensure that the system is operating at optimal efficiency and identify any potential issues affecting its performance.
- the *maximum power* the CHP is allowed to produce (set by operator).
- *number of starts of the CHP in the time interval* is an index of how the CHP is managed.
- *engine inlet temperature* is the fluid temperature at the intake manifold just before entering the engine. It is an important parameter to consider when operating the CHP, as it can affect the performance and efficiency of the engine. Furthermore, the spread between the engine inlet and engine outlet temperature should not have a too high value. Indeed, when the spread increases remarkably, serious damages can be reported to some engine components and in particular to the cylinder head.
- *engine outlet temperature* is the temperature of the cooling fluid after cooling down the engine. Measuring and monitoring the engine outlet temperature is an important aspect of engine operation, as it can help ensure that the engine runs efficiently and effectively. As mentioned above, it is also important to calculate and monitor the spread temperature between the engine's inlet and outlet manifolds.
- *engine cooling circuit pressure* is the pressure of the system that helps to regulate the temperature of the engine by circulating a coolant fluid through passages in the engine block and cylinder head. The coolant absorbs heat from the engine and then flows through a radiator, where it releases the heat into the air. The circulation of the coolant is driven by a pump, and the pressure of the coolant in the system is a result of the resistance to flow caused by the various components in the circuit,

including the radiator, hoses, and water pump. The pressure in the engine cooling circuit is an important parameter to monitor, as it can affect the performance and efficiency of the cooling system. If the pressure is too low, it can result in poor circulation and insufficient cooling, which can lead to overheating and damage to the engine. If the pressure is too high, it can cause leaks or even ruptures in the system, which can also result in engine damage.

- *oil temperature* is the temperature of the lubricating oil that is used to reduce friction and wear in the engine. In general, it is desirable to maintain the oil temperature within a certain range in order to ensure that the oil is able to effectively lubricate the engine and prevent wear. If the oil temperature gets too high, loses its effectiveness, which can lead to increased wear and potentially cause damage to the engine. On the other hand, if the oil becomes too cold, several issues can arise due to its changed viscosity and reduced effectiveness.

- *CHP cabin temperature* refers to the temperature inside the enclosure or control room of the CHP system. This temperature is influenced by the amount of heat generated by the CHP system and the surrounding environment. The CHP cabin temperature is vital for ensuring the safety and comfort of the operator, as well as for maintaining the proper functioning of the CHP system.

- *ambient temperature* is the outdoor temperature.

- *exhaust gas temperature* is the exhaust gas temperature when expelled from the combustion chamber, measured at the outtake manifold. It will vary depending on the operating conditions of the engine, such as the load, speed, and air-fuel ratio. It is an important indicator of how combustion is occurring in the combustion chamber (injection, stoichiometry, combustion complementation, etc.). Monitoring the exhaust gas temperature can help to identify problems with internal heat exchangers.

- *heat exchanger exhaust temperature* is measured by a probe onto the pipe carrying the operating fluid, just after the heat exchanger (the device used to transfer heat from the cooling fluid, such as water, to the operating fluid in the plant). It is related to the heat exchanger and

provides information on the state of the element and the heat transfer process between the fluids involved.

- *heating circuit pump rate* is the measured value in revolutions per minute (rpm) of the pump managing the heating circuit. Is the circulation pump that provides the main water flow. It is placed at the beginning of the circuit that brings cold water from the thermal buffer to the engine, recovering its heat for cogeneration and thus preventing overheating and damage.

- *CHP pump rate* indicates the CHP pump motor rotation rate. The CHP pump circulates water through the system recovering heat from the engine and from the exhaust gases and transferring it to the heat exchanger.

- *lambda sensor value* is the measurement provided by the lambda sensor, also called the oxygen sensor. It measures the amount of oxygen that has not been burned in the exhaust pipe of a combustion engine. This information is used to adjust the mixture of air and fuel in the engine to ensure it is operating efficiently. The lambda sensor helps to determine if the air-fuel ratio is lean or rich.

- *gas mixer position* is the opening degree of the gas mixer valve, the valve that controls the flow of fuel into the combustion chamber of an internal combustion engine. It is typically located in the intake manifold and is used to adjust the air-fuel ratio in the engine. This is important because the optimal ratio can vary depending on the load and operating conditions of the engine. By adjusting the gas mixer valve, the engine can be made to run more efficiently and with fewer emissions.

- *throttle position* is the opening degree of the throttle valve, the valve that controls the flow of air and fuel into the combustion chamber of the internal combustion engine. By regulating the flow of air and fuel into the engine, this valve also controls the power output of the engine.

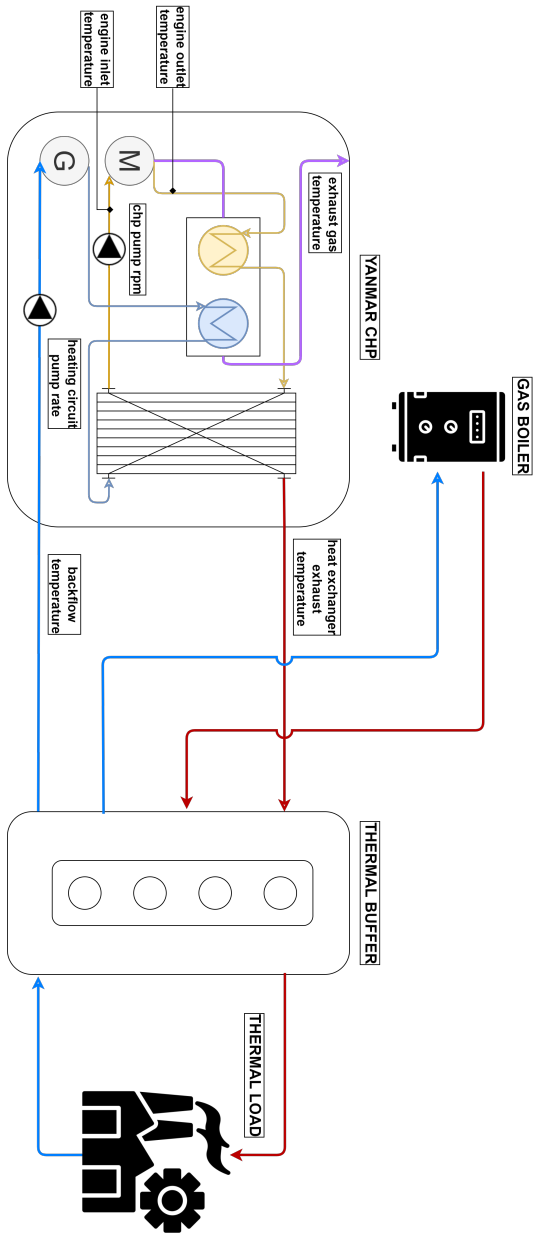

Figure 4.2: Plant layout with YANMAR CHP, gas boiler, and thermal buffer.

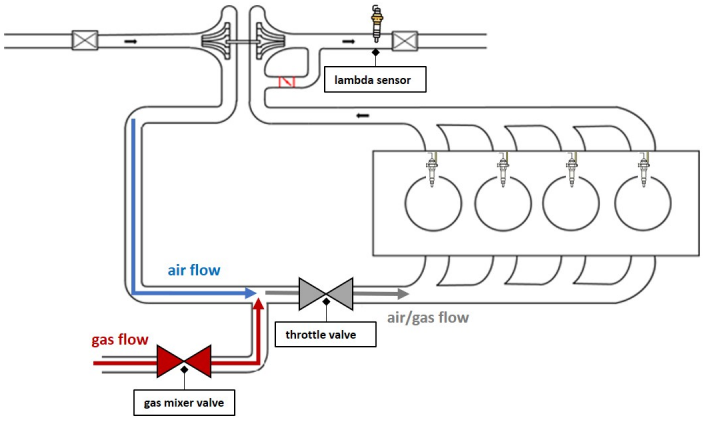

Figure 4.3: Details of the gas engine driving the CHP.

4.2 Data Preprocessing

For every YANMAR CHP, data are acquired by the local Programmable Logic Controller (PLC) and pushed into the cloud: in particular, data is sampled and hosted on a MariaDB database [74]. The acquisition time T_s is set by the customer in accordance with the connectivity tariff available; if data are transmitted by means of a Global System Mobile (GSM) sim card then the sampling frequency cannot be increased in a steady-state functioning and T_s must be set to 15 minutes. Otherwise, if the plant guarantees a LAN connection, then T_s can be set to 1 minute. Furthermore, the PLC does not rank the data transmission as a high-priority task, and values are pushed with variable frequency and on average every T_s minutes implying that the real acquisition period is highly irregular.
A series of pre-processing operations have been done on the original data. At first, the 16 features are queried and re-sampled to meet the hypothesis of having a value every 1 or 15 minutes.
A binary value denoting whether the *lambda sensor is reliable* or not is introduced to allow for tracking of unreliable measurements. Indeed, when the cogenerator is shutting down, the lambda sensor may produce inconsistent values close to overflow not reflecting reality. In this situation, the binary value is set to 1 to indicate that the measurement is not reliable, while the

value of the lambda sensor is set to 0.
The *CHP pump rate* is encoded to a boolean value. Both pumps have, in fact, an embedded inverter that controls the motor's rpm but, while the *heating circuit pump* is actually operated in the full range of speeds indicated by the datasheet, the *CHP pump* is controlled with a two-state logic: it starts to run before the CHP switches on and shuts down right after the CHP turns off and the engine is cooled down.
In the end, all sensors' values are normalized according to engineering experience; minimum and maximum values of each signal have been set a-priori.

Chapter 5

A Methodology to Determine the Optimal Train-set Size

This chapter proposes a procedure to determine the optimal train-set size to minimize the reconstruction error of an autoencoder with a pre-defined structure and hyper-parameters that will be trained to encode the normal behavior of energy generation systems. This procedure exploits the outcome of learning curves, a powerful tool to track algorithm performance while the train-set dimension varies. Afterward, the procedure is applied to three real case studies where two types of autoencoders are trained to learn the normal behavior of a YANMAR CHP unit with the scope of detecting incoming anomalies. In the end, the outcomes of the procedure are explained and, under the constraint of a daily re-training frequency, 6 weeks are identified as the optimal train-set size for both autoencoders. This chapter is structured as follows: the proposed methodology is explained in Section 5.1, and the three real case studies are described (Section 5.2). Results (Section 5.3) are illustrated and conclusions (Section 5.4) are drawn. [1]

[1] Part of the contents of this chapter has been published as "A methodology to determine the optimal train-set size for autoencoders applied to energy systems" in *Advanced Engineering Informatics* [24].

5.1 Proposed Methodology

In this section, the procedure to size the optimal train-set size for a pre-tuned autoencoder is explained; this procedure exploits a powerful machine learning tool: learning curves. In order to produce a model as generalizable as possible, a cross-validation routine is applied. The described procedure is a further development of the methodology proposed by Giola et al. [37]. The main difference lies in the fact that the present work is aimed at utilizing semi-supervised models where only partial information about the machine's state is available, whereas Giola et al. focus on supervised regression models for energy load prediction where the data is well labeled. Additionally, in this work, the autoencoder operates in the domain of multivariate data rather than univariate. Finally, a procedure based on the frequency behavior of the specific CHP system is also proposed to determine the optimal period between retraining intervals.
In this work, a multivariate time-series with N features is analyzed; each observation of the time-series is $X \in \mathbb{R}^N$. At the time $t_k \in \mathbb{R}_0^+$, $X(t_k)$ represents the vector of observations of the input variables; the first acquisition of the available dataset is indicated with t_0. A list of length q of eligible train-set size $p = [p_1, p_2, ..., p_q]$ has to be pre-set in order to show how the algorithm is able to generalize when is trained with different sizes of train-set. The test-set size d is strongly related to the nature of the problem and the seasonality of the dataset; in this work, the test-set size is set to 1 day accordingly to the procedure explained in Section 5.3.1. As shown in Figure 5.1, the testing period has been assumed to be immediately after the train-set; this is not a constrain but just a hypothesis that does not lead to a lack of generality, as long as the time order is retained the test period can be shifted forward.
The proposed methodology can be categorized among Out-Of-Sample (OOS) approaches: OOS methodologies have traditionally been used to assess predictive performance for time-varying data. Basically, the out-of-sample method keeps the last part of the time-series to be used for testing. These approaches do not fully exploit the total amount of data but retain their temporal order allowing for to management of dependencies between observations and temporal correlations between consecutive values in the time-series [16].
Given d as a constant dimension of the test-set, then, for every train-set size

p_j starting at time t_i, the analyzed train-sets are:

$$X(t_i) \leq X \leq X(t_i + p_j) \tag{5.1}$$

and the test-sets:

$$X(t_i + p_j + 1) \leq X \leq X(t_i + p_j + 1 + d) \tag{5.2}$$

In order to evaluate the algorithm's performance without being influenced by the train-set local characteristic, for each p_j train-set size, this scheme is applied over multiple test periods (Figure 5.1). Obviously, as the methodology is designed, each time the train-set is moved, the test-set slides.

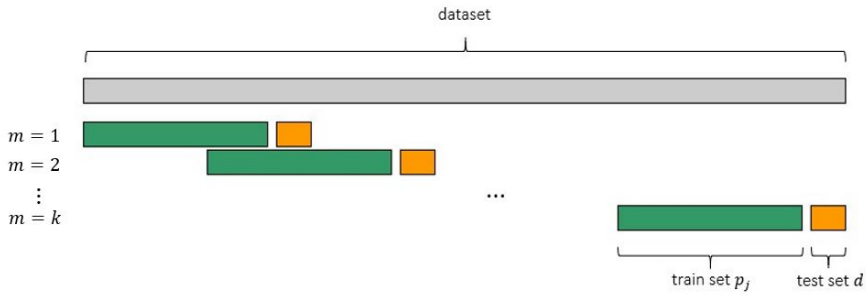

Figure 5.1: Scheme of the proposed methodology.

The number k of the tested periods must be chosen in order to be statistically representative of the dataset's temporal behavior. In particular, k depends on the dimension of the biggest train-set and on the time-series seasonality and trend. The selection of k is based on a general rule of thumb developed in the present research. This rule suggests that if we denote s as the number of acquisitions present in the maximum seasonality of interest within a particular dataset, and n as the available data points for analysis, then a value of k consistent with Equation 5.3 can be considered.

$$k = 10 \cdot \frac{n}{s} \tag{5.3}$$

In the present work, the CHP maximum seasonality of interest is acknowledged as 1 year and the available acquisitions correspond to 1 year for the

first two datasets whilst 6 months for the third one. Consequently, $\frac{n}{s} = 1$ and $k = 10$ for the first two case studies have been considered, while $\frac{n}{s} = 0.5$ and $k = 5$ for the third one. This rule serves as a practical guideline but, in the case of different applications, may be adjusted according to experience and domain knowledge.
In order to quantify the train-set and test-set reconstruction errors, many metrics can be calculated: the most common metrics are Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), or Root Mean Squared Percentage Error (RMSPE) and they all have different shortcomings and merits. In this work, the selected metric is the RMSE: being the reconstruction error a mean of the single scaled feature's reconstruction errors, RMSE is not fully explicative but, on the other side, gives a robust and intuitive interpretation of the error. Indeed, both RMSPE and MAPE present high values when a feature is close to zero invalidating the graphical visualization of learning curves. The formula for calculating the training RMSE of the m-th testing period is expressed by Equation 5.4.

$$RMSE_{m,train} = \sqrt{\frac{1}{p_j}\sum_{i=1}^{p_j}\frac{1}{N}\sum_{f=1}^{N}\left(x_{i,f}^{m} - \hat{x}_{i,f}^{m}\right)^2} \tag{5.4}$$

where $x_{i,f}^{m}$ is the actual value of the feature f and $\hat{x}_{i,f}^{m}$ is the forecast value of the same feature. The training procedure is repeated k times for each train-set-size p_j obtaining k RMSE test-set's reconstruction errors and k train-set's reconstruction errors. The procedure to calculate the testing RMSE is analogous with the difference in that it takes into consideration the d time-steps of the test-set (Equation 5.5)

$$RMSE_{m,test} = \sqrt{\frac{1}{d}\sum_{i=1}^{d}\frac{1}{N}\sum_{f=1}^{N}\left(x_{i,f}^{m} - \hat{x}_{i,f}^{m}\right)^2} \tag{5.5}$$

The q average RMSE errors e_k for training (Equation 5.6) and testing (Equation 5.7) represent two points in the learning curves graph. In addition, to show the dispersion of the data, the error variance for both the training and testing curves is represented with a colored gradient.

$$e_{k,train} = \frac{1}{k}\sum_{m=1}^{k} RMSE_{m,train}. \tag{5.6}$$

$$e_{k,test} = \frac{1}{k} \sum_{m=1}^{k} RMSE_{m,test}. \tag{5.7}$$

The flowchart in Figure 5.2 explains the methodology where the parameters have been set in accordance with Section 5.3.1.

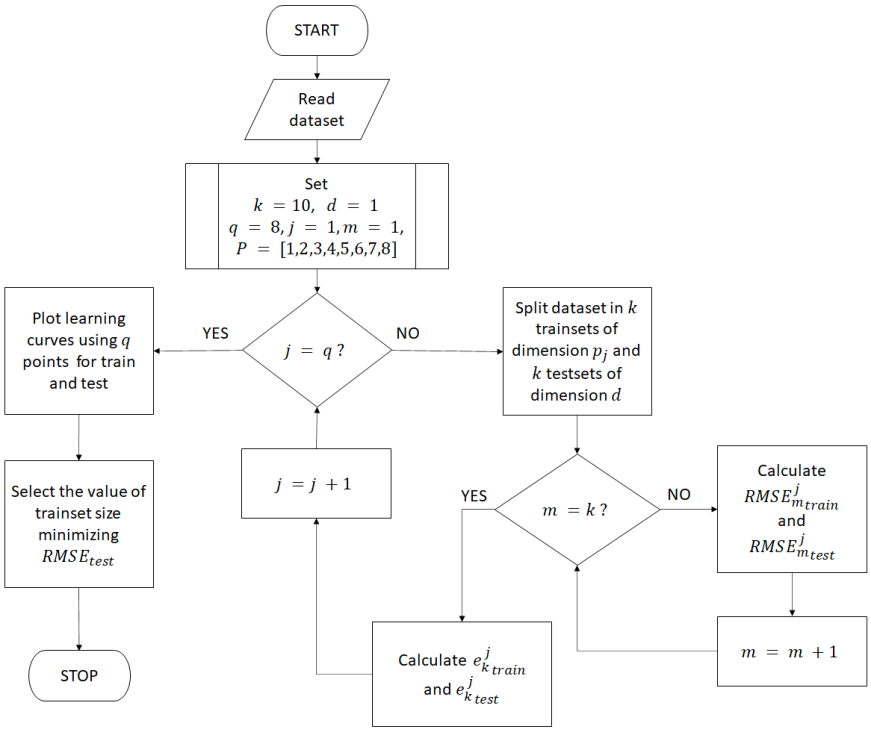

Figure 5.2: Flowchart of the proposed methodology.

In this section, a general methodology is proposed but the reader must be aware that there are some dataset traits that strongly influence the methodology outcome:

- *autoencoder hyper-parameters*: the optimal train-set size depends on the architecture of the model. Hence, the hyper-parameters must be set before running the proposed methodology.

- *retraining frequency*: the quality of the reconstruction depends on how far in the future the autoencoder is applied without retraining. This means that also the test-set size must be fixed (Section 5.3.1) and the interval between two retrainings must be equal to the test-set size.

- *number of features*: the quantity of data needed to train the model is directly influenced by the order of magnitude of the number of features populating the dataset. When the number of not highly correlated features increases, it implies that the relationships among those features are more complex. As a result, the model requires a larger quantity of data to accurately capture and learn these complex relationships. The absence of sufficient data may result in underfitting and the model may fail to generalize well to unobserved data.

- *sample time*: autoencoder's performance is clearly weakened if trained with low-frequency data, indeed, increasing the interval among two observations, may cause information loss. On the other side, the high sampling frequency may lead to redundant samples causing a slowdown in the autoencoder's training without improving reconstruction performance. In the context of the thesis, the dataset is sampled at intervals of either 1 or 15 minutes. It is important to note that these sampling intervals should not impact the methodology.

5.2 Case Studies

The proposed methodology has been applied to define the optimal train-set size of two types of autoencoder, the Vanilla AutoEncoder (AE, Section 5.2.2) and the Convolutional AutoEncoder (ConvAE, Section 5.2.3), subsequently employed for a problem of sequential novelty detection. The units under analysis (20 kW_e YANMAR CHPs described in Chapter 4) are located in Germany and satisfy the electrical and thermal needs of three different plants: the YANMAR facility where CHPs are produced, a manufacturing company, and a telecommunication company.

5.2.1 Dataset

The three cases considered have identical plant layouts and, thus, a very similar dataset except for two details. The first CHP has a dataset with an

acquisition time of 1 minute, while the other two have 15 minutes. The first and second CHPs have one year of data available, while the third has only six months. These datasets present multiple features from different elements installed in the plant's layout; since this work focuses on learning the normal functioning of only 1 device of the plant, the CHP unit, as described in Sections 4.1 and 4.2, only 16 features plus 1 manipulated, acquired with a sample time of 1 minute, have been considered remarkable for the current work:

- the *active power* produced by the CHP;
- the *maximum power*;
- *number of starts of the CHP in the time interval*;
- *engine inlet temperature*;
- *engine outlet temperature*;
- *engine cooling circuit pressure*;
- *oil temperature*;
- *CHP cabin temperature*;
- *ambient temperature*;
- *exhaust gas temperature*;
- *heat exchanger exhaust temperature*;
- *heating circuit pump rate*;
- *CHP pump rate*;
- *lambda sensor value*;
- *lambda sensor is reliable*;
- *gas mixer position*;
- *throttle position.*

5.2.2 Vanilla Autoencoder

AE is the simplest form of autoencoder, it is a three-layer neural network where interconnections are made by means of fully connected nodes: each node of layer l is connected to every node in the layers $l-1$ and $l+1$. This algorithm considers each sample as independent, which leads to faster training disregarding the time correlation among acquisitions. The structure and the hyper-parameters selected for this work are reported in Table 5.1:

hyper-parameters	**AE**
input dimension	17
latent dimension	70
output dimension	17
activation function	*relu*
batch size	32
learning rate	10^{-4}
l1 regularization	0.0
dropout	0.2
optimizer	*Adam*
loss function	MSE

Table 5.1: Vanilla Autoencoder hyper-parameters.

5.2.3 Convolutional Autoencoder

ConvAE presents a more complicated shape: it is not fed with the row samples but, at each step, a sliding window of the input vector is provided. Inputs are treated by means of a 1D-convolutional layer characterized by a certain number of filters; each filter has a predefined dimension (*kernel size*). The latent space is made by fully-connected nodes and its output is inputted to a transposed 1D-convolutional layer that produces an output vector of the same dimension as the input. The convolutional layers and the small size of the bottleneck allow for neglecting the use of dropout and regularization. A complete list of the ConvAE hyper-parameters can be found in Table 5.2

hyper-parameters	**ConvAE**
sliding window	10
input dimension	(10,17)
latent dimension	3
output dimension	(10,17)
activation function	*relu*
batch size	32
learning rate	10^{-4}
l1 regularization	0.0
dropout	0.0
padding	*same*
strides	1
filters number	10
kernel size	5
optimizer	*Adam*
loss function	MSE

Table 5.2: Convolutional Autoencoder hyper-parameters.

5.3 Results

This section reports the results obtained running the methodology proposed in Section 5.1 to estimate the optimal dimension of the train-set for two types of autoencoder described in Section 5.2.2 and Section 5.2.3. The methodology has been applied to three different case studies.

5.3.1 Case study 1: YANMAR facility

In the first instance, the methodology parameters have to be explicated: a test-set of 1 day has been set, as motivated later in this Section, the number of periods of test k has been fixed to 10 (according to Equation 5.3) and 8 train-set sizes have been explored (from 1 week up to 8 weeks). A monitoring duration of up to 8 weeks may be regarded as an acceptable upper limit for customers before an anomaly detection service becomes operational. The whole dataset is composed of 17 features recorded for 1 year and describes the behavior of a 20 kW_e YANMAR CHP.
For the sake of precision, a sum-up of the methodology's parameters is re-

ported below:

- $d = 1$, test-set size expressed in days;
- $k = 10$, number of periods tested for each train-set size;
- $p = [1, 2, 3, 4, 5, 6, 7, 8]$, the list of train-set sizes expressed in weeks.

Figure 5.3 reports the real splits applied to the case study dataset according to the proposed methodology described by Equations 5.1 and 5.2.

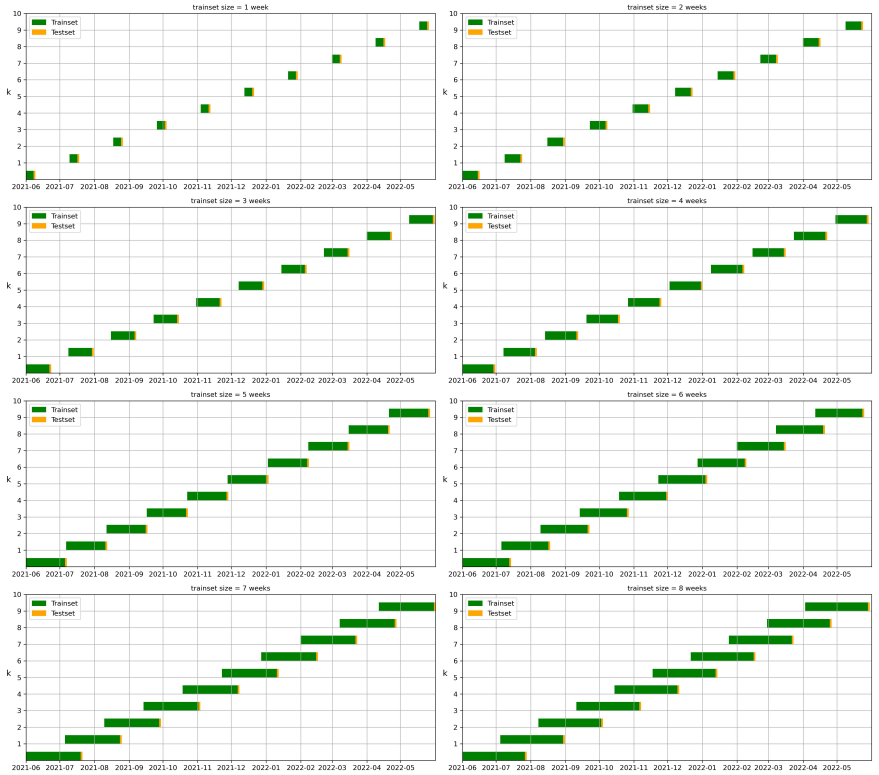

Figure 5.3: Methodology applied to a real case study: YANMAR facility.

Retraining frequency

Over time, the behavior of the system may change (aging, degradation, etc.). If the algorithm has been trained with data too far back in the past, it may not be able to capture the current behavior of the system. Therefore, it is essential to provide periodic re-training with more recent data that contains information about the unit's current operation. Usually, the system's signals are characterized by a periodicity due either to their nature or to the plant's management logic. To maximize the deep learning model's performance, it is important to have a statistical similarity between the train-set and the test-set so that the model is able to infer reconstructed features that are subject to the same external condition (i.d. period of the year, management logic, cost of raw materials, etc...) of the input features.
In this section, a frequency domain-based procedure to set the retraining frequency, and consequently the test-set size, is explained. The FFT (Section 3.2) is applied to all features and Fourier components are calculated. Frequencies corresponding to a period smaller than 8 hours have been filtered because considered not physically interesting for the energy system under analysis. Then, the most representative Fourier components of each feature are considered and the most representative frequency is selected.
Figure 5.4 shows the frequency components of the features describing YANMAR CHP under analysis; the orange, green, and yellow shaded areas represent the Fourier components relative to 1 week, 1 day, and 12 hours. Experience with YANMAR CHPs suggests that the 12-hour pattern is linked to the day/night cycle, the 24-hour pattern is linked to the daily operation, and the week pattern is generated by the different loads requested at the week-end. Concluding, a re-training period of 1-day is then chosen for three main operating reasons:

- retraining a deep learning algorithm every 12 hours may be computationally expensive considering that, in some cases, the anomaly detection routine runs locally on on-field devices;
- a daily pattern is considered the most significant from an operating point of view and it is present in all YANMAR CHPs;
- waiting 1 week before feeding the model with new data may lead the algorithm to detect some seasonal trends as anomalies.

It is important to emphasize that 1 day is the ideal period after which, if no

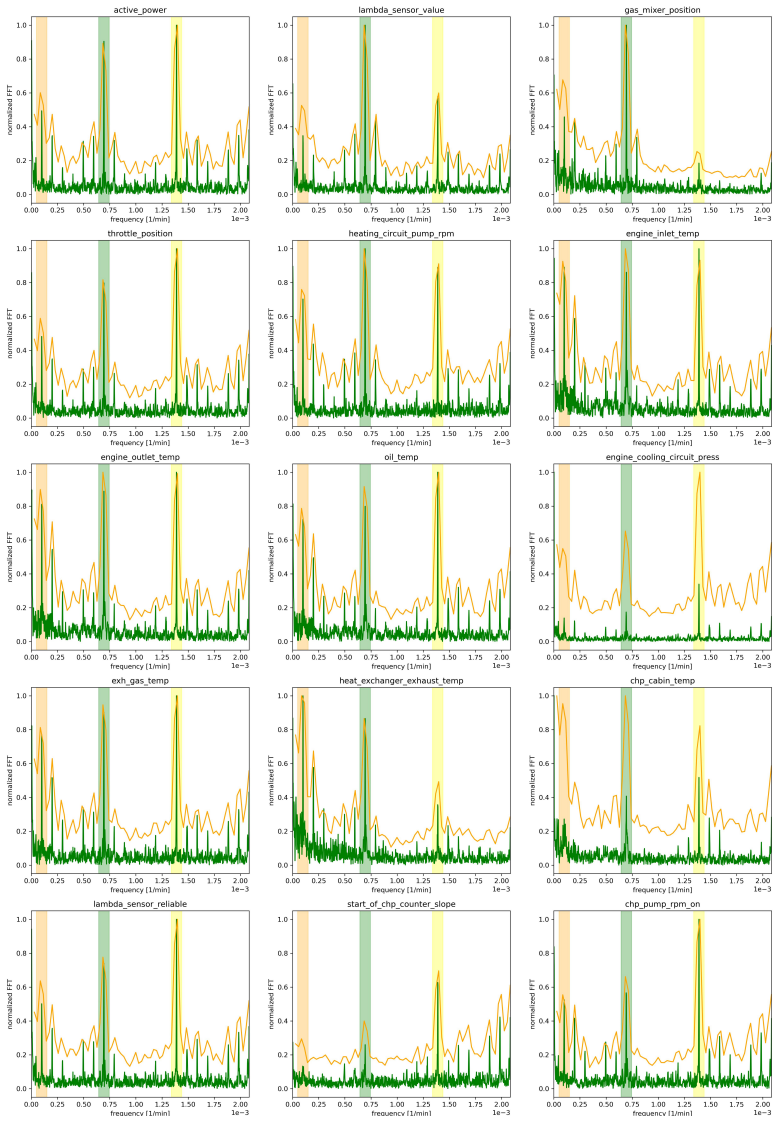

Figure 5.4: Features' FFT components.

anomalies have occurred, it is highly beneficial to retrain the autoencoder. However, before retraining, it is imperative to ensure that no anomalies took place on the preceding day. This necessitates prompt confirmation of the CHP system's normality by YANMAR maintenance technicians during the development of the anomaly detection routine. As of today, it has not been possible to do this; therefore, in Chapters 6 and 7, the autoencoder is not retrained daily. Instead, in Chapter 8, an approach to retraining is proposed after the data has been carefully selected to meet certain criteria.

Optimal train-set size

Figure 5.5 shows the learning curves built for the AE: red and green curves report the training RMSE averaged over the k folds whilst the red and green areas, similarly, represent the RMSE variance. In this case, a train-set of 6 weeks should be considered optimal; indeed, increasing the train-set size, the test score decreases reaching its minimum at 6 weeks and then starts to increase again. Also, the tiniest RMSE variance happens to be at 6 weeks.

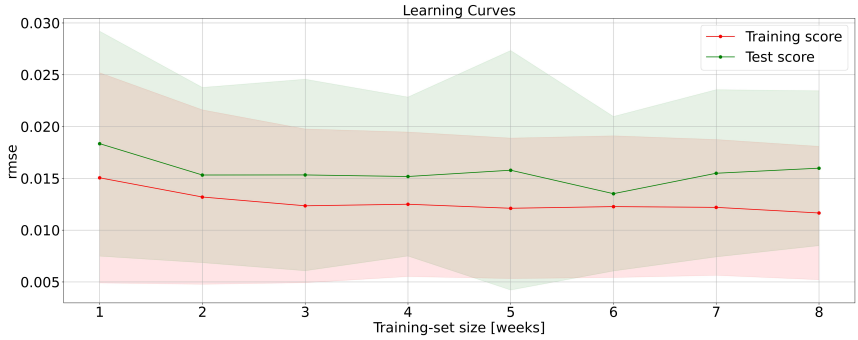

Figure 5.5: Case study 1. Learning curves for Vanilla Autoencoder.

In the same way, Figure 5.6 shows the learning curves built for the ConvAE: also in this case, 6 weeks appear to be the optimal train-set size even if the average RMSE is a bit higher. At first glance, analyzing the learning curves, the reader may conclude that the difference, in terms of RMSE, between a train-set size of 2 weeks and a train-set size of 6 weeks is not re-

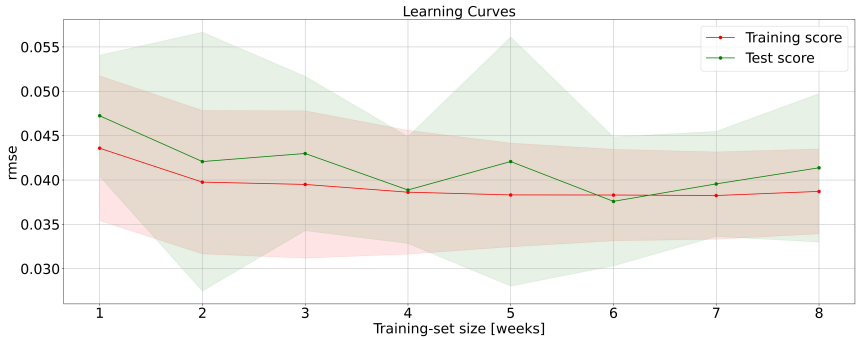

Figure 5.6: Case study 1. Learning curves for Convolutional Autoencoder.

markable. For sake of clarity, Figure 5.7 and Figure 5.8 depict the actual and the reconstructed values of three of the most significant features of the same test-set for the two structures of autoencoder discussed in this work. Both AE and ConvAE, drastically improve their performance when increasing the dimension of the train-set, in particular when the signal is not in a steady condition. From a practical point of view, some consideration must be done: if the minimum test-set RMSE does not match the customer's expectation then the structure of the autoencoder can be changed in order to improve performance.
On the other hand, if 6 weeks are considered too much, then the autoencoder can be simplified. Figures 5.5 and 5.6 demonstrate that the AE is more sensible to the train-set size while the ConvAE reaches an RMSE value close to the minimum with only 2 weeks of training. Increasing the train-set size produce an RMSE variance reduction for both autoencoders even if a strange behavior is noticed when 5 weeks of training are used: algorithms' performance drastically worsens in term of RMSE variance.

5.3.2 Case study 2: Manufacturing Company

To maintain conciseness, the hyper-parameters and retraining frequency considerations pertaining to the methodology, as discussed in the previous case study, are omitted as they remain consistent. Figures 5.9 and 5.10 depict the learning curves for the models built on data acquired from the manufacturing

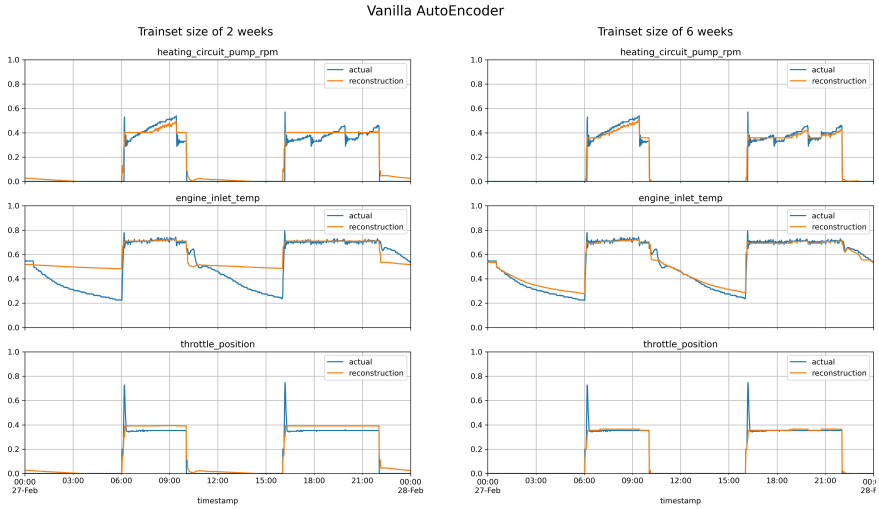

Figure 5.7: Case study 1. AE reconstructions comparison, 2 weeks vs 6 weeks of train-set size.

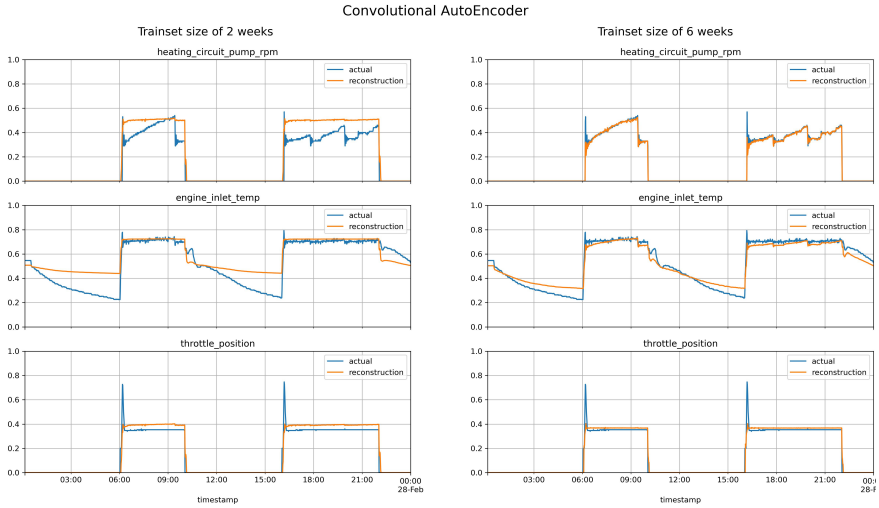

Figure 5.8: Case study 1. ConvAE reconstructions comparison, 2 weeks vs 6 weeks of train-set size.

company and, also in this case, 6 weeks appear to be the optimal train-set size. Figures 5.11 and 5.12 underscore how a marginal reduction in the av-

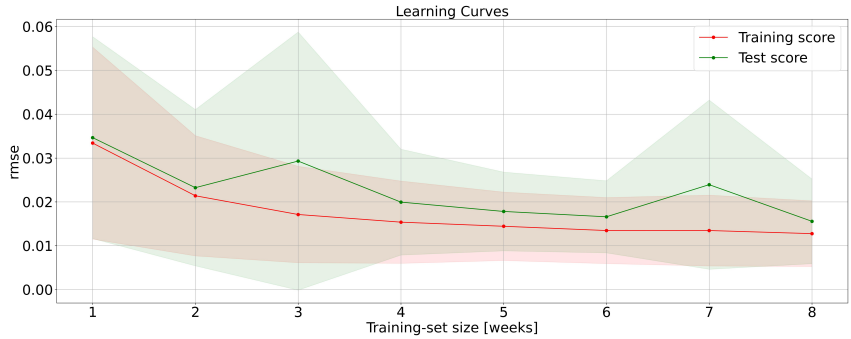

Figure 5.9: Case study 2. Learning curves for Vanilla Autoencoder.

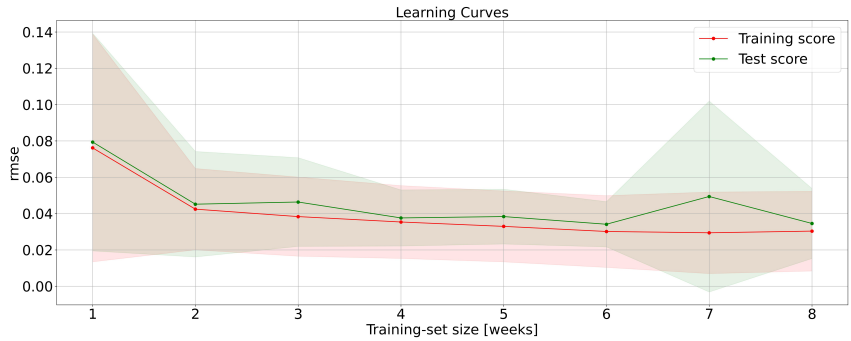

Figure 5.10: Case study 2. Learning curves for Convolutional Autoencoder.

erage error within the learning curves, results in an improved reconstruction of input signals.

5.3.3 Case study 3: Telecommunication Company

Similarly, in this case, the hyper-parameters of the methodology and the considerations remain largely unchanged, with the only difference being at-

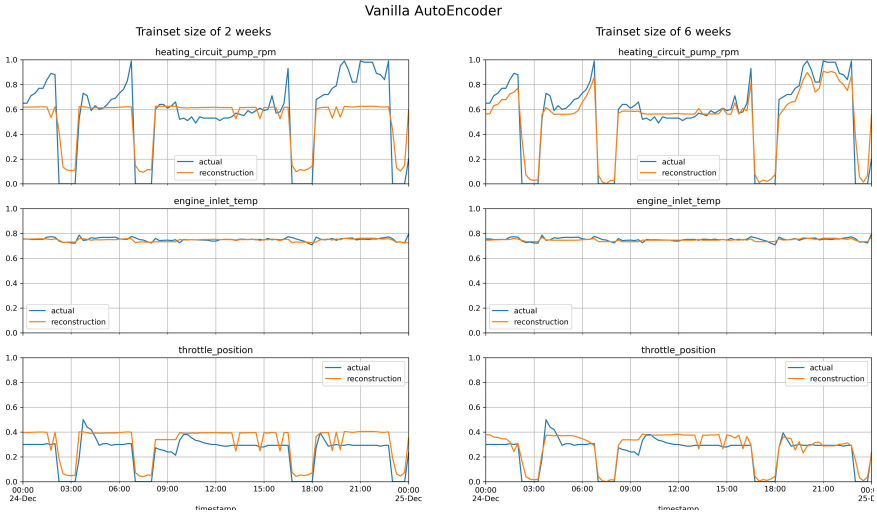

Figure 5.11: Case study 2. AE reconstructions comparison, 2 weeks vs 6 weeks of train-set size.

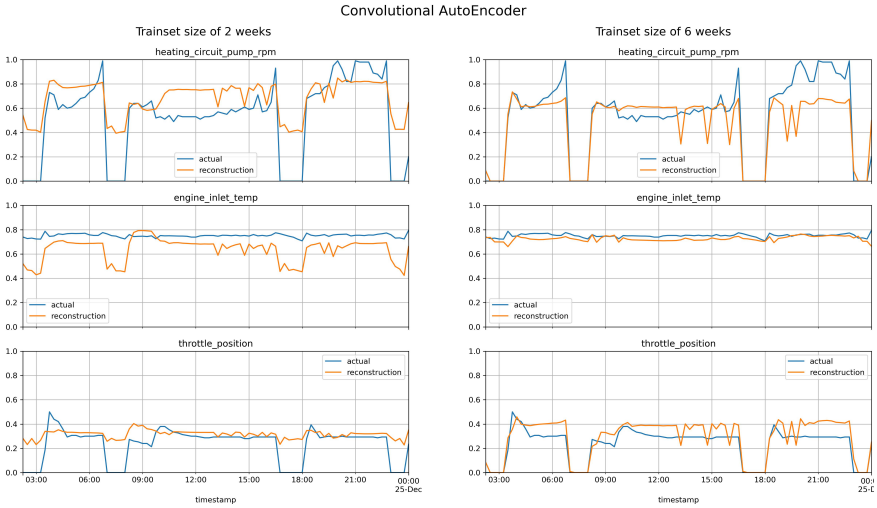

Figure 5.12: Case study 2. ConvAE reconstructions comparison, 2 weeks vs 6 weeks of train-set size.

tributed to the limited availability of data. In this particular case, only six months of data are obtainable therefore, by employing Equation 5.3, the value of k used is determined to be 5, and different splits are utilized (Figure 5.13).

Figure 5.13: Methodology applied to a real case study: telecommunication company.

Figures 5.14 and 5.15 report the learning curves for the two autoencoders trained using data from the telecommunication company. In this case, the learning curves indicate that a slightly larger train-set might be beneficial, particularly for the Vanilla AE. The graph suggests that exploring a dimension greater than 8 weeks could have been advantageous, whereas, for the ConvAE, 7 weeks seems to be the more appropriate choice. However, due

to the limitation of only having 6 months of data, applying the proposed methodology with a larger train-set exceeding 8 weeks within a 24-week dataset could lead to significant overlapping in the methodology's dataset split. Further data acquisition would be necessary to augment the dataset to confirm an optimal train-set size greater than 8 weeks.
Nevertheless, taking into account the test score variance and the stabilization of the test score average beyond a train-set size exceeding 5 weeks, it can be deemed reasonable, albeit not universally optimal, to opt for a train-set size of 6 weeks. For consistency with previous cases, a comparison is provided

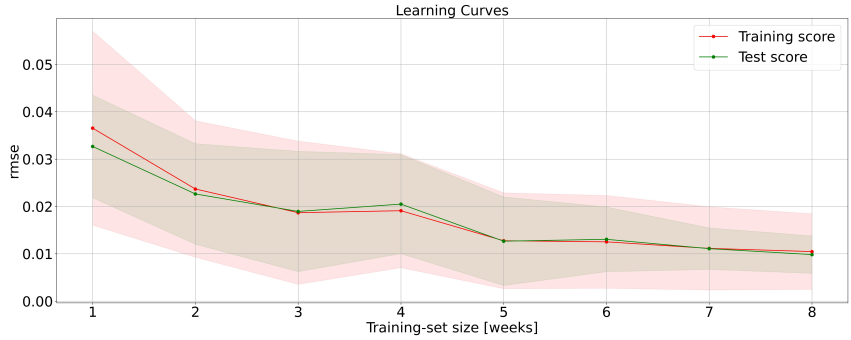

Figure 5.14: Case study 3. Learning curves for Vanilla Autoencoder.

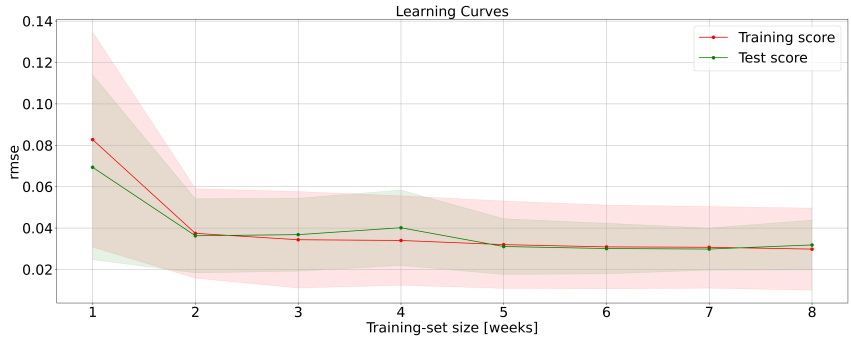

Figure 5.15: Case study 3. Learning curves for Convolutional Autoencoder.

between signal reconstruction achieved by a model trained for 2 weeks (Figure 5.16) and one trained for 6 weeks (Figure 5.17). In this case, as well, Figures 5.16 and 5.17 emphasize how a small decrease in the average feature reconstruction error significantly enhances the reconstruction quality of the two presented autoencoders.

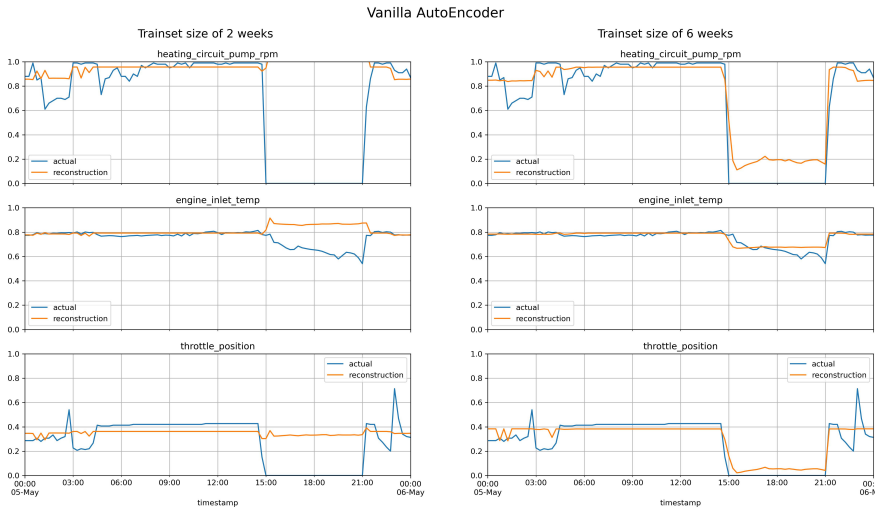

Figure 5.16: Case study 3. AE reconstructions comparison, 2 weeks vs 6 weeks of train-set size.

5.4 Chapter Conclusions

This work tries to fill the gap between academic research and industrial applications with regard to the utilization of autoencoders in the energy production domain. The problem has been explained and then a methodology to select an optimal dataset dimension to train an autoencoder has been proposed. Three particular case studies have been detailed and a procedure to set the test period during which is not helpful to retrain the model has been described. In the end, results have been shown taking into account two structures of autoencoder: AE and ConvAE.
The proposed methodology, designed to retain temporal correlation among

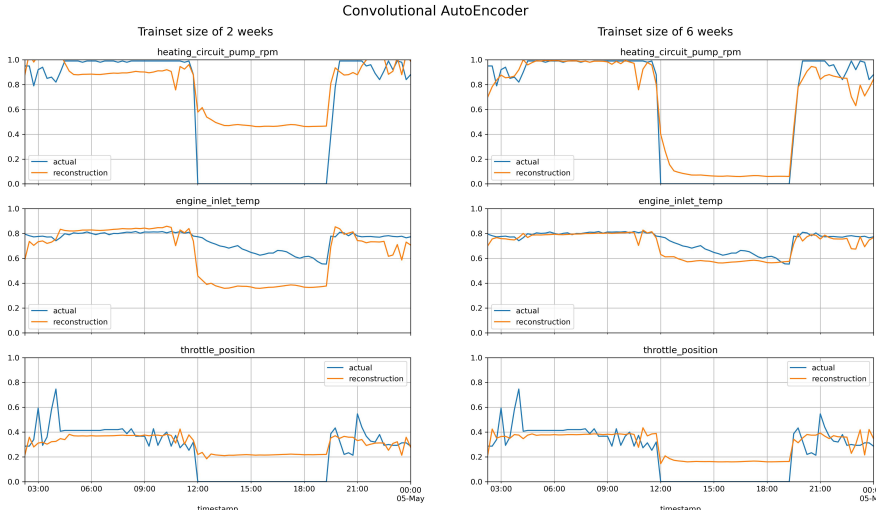

Figure 5.17: Case study 3. ConvAE reconstructions comparison, 2 weeks vs 6 weeks of train-set size.

samples, has been applied to learn the normal functioning of three YANMAR CHPs where the datasets of interest include the behavior of 17 features. For these case studies, it has been demonstrated how 6 weeks can be considered a good selection for the train-set size when dealing with 20 kW_e YANMAR CHPs, and in two case studies out of three, this dimension minimizes the reconstruction error in the test-set. FFT transform has been used to calibrate the test-set dimension and 1 day has been individuated as the best choice. The outcome of this work is a tool to help researchers in designing autoencoder models to be more usable by industries; indeed the proposed methodology gives also information on how much the autoencoder performance will decrease if the customer needs the model to be deployed before the optimal train-set size quantity of data is available.

Chapter 6

Anomaly Detection and Root Cause Analysis

This chapter presents a semi-supervised deep learning solution to detect anomalies of a 20 kW_e YANMAR micro-cogeneration unit installed in the energy plant of a facility school. The dataset considered consists of 13 features acquired every 15 minutes. An autoencoder with 1D-convolutional layers has been designed and, after being trained to learn the normal behavior of the CHP, is employed to report abnormal operations. In consideration of the fact that autoencoders have the tendency to yield false positives, an FFT-based technique has been applied to filter spurious detections and improve the algorithm's robustness. As the last contribution, a naive methodology to address the root cause of the anomalies has been explained and its effectiveness has been proved in two real malfunctionings of the CHP. Section 6.1 presents the real application and details of the available dataset. Section 6.2 treats the proposed methodology. Section 6.3 defines two algorithms to benchmark the work's outcomes. In the end, results obtained with two different test-sets are discussed (Section 6.4), and conclusions are drawn (Section 6.5). [1]

[1] Part of the contents of this chapter has been submitted as "Anomaly Detection and Root Cause Analysis using Convolutional Autoencoders: a Real Case Study" to *Energy* [25].

6.1 Dataset

The target unit of this chapter, described in Chapter 4, is a YANMAR co-generation system that satisfies the energy needs of a school in Germany. As mentioned in previous chapters, the dataset contains information about different parts of the plant but, since the main focus of the study is to understand the normal functioning of the CHP unit, only 17 features were selected. Unluckily, for this particular case study, some sensor was not installed or was producing unreliable measurements, so 4 features were discarded. Thus 13 signals, sampled every 15 minutes, have been used for this study:

- the *active power*;
- *engine inlet temperature*;
- *engine outlet temperature*;
- *oil temperature*;
- *CHP cabin temperature*;
- *exhaust gas temperature*;
- *heat exchanger exhaust temperature*;
- *heating circuit pump rate*;
- *CHP pump rate*;
- *lambda sensor value*;
- *lambda sensor value is reliable*;
- *gas mixer position*;
- *throttle position.*

Specifically, in the case of this system, four other measurements describing the CHP behavior have been excluded:

- *maximum power* recorded inconsistent values;
- The counter responsible for tracking the *number of CHP starts within the specified time interval* became static at zero;

- The *engine cooling circuit pressure* remained unconnected;
- The sensor designated for measuring *ambient temperature* was erroneously positioned.

6.2 Proposed Methodology

This section will first present the data collection steps and pre-process operations that were necessary in order to maximize available information and to neglect disturbance. This will be followed by a description of the construction of the model, the core of the work, and a detailed overview of the proposed methodology (a flowchart is reported in Figure 6.1). The proposed methodology is based on two pillars: the autoencoder and the FT. A brief introduction to these two techniques has been given in Chapter 3, providing some hints of the instruments used, which do not claim to be exhaustive but only supportive for the reader.

6.2.1 Data collection and pre-processing

In Chapter 5, it has been demonstrated how a train-set size of 6 weeks is a suitable trade-off between having enough data to describe CHP's dynamic and, in the meanwhile, not complicating too much the autoencoder structure. In this work, the outcome obtained in the previous chapter has been deemed applicable to this case study as well, even though the signals are slightly fewer: this assumption is made on the premise that such a result could be attributed to the seasonal dynamics of the machine rather than the number of features.
Consequently, the first challenge of this work has been to find a batch of data containing 6 weeks of normal data: this activity joined the activity reports of the CHP maintenance engineers, the error codes generated by the onboard controller of the machine, and a deep analysis of each univariate describing the CHP's behavior. Every YANMAR CHP has a diagnostic onboard system that at each acquisition reports the CHP status. Unluckily faulty statuses are created utilizing coarse thresholds and, consequently, the presence of an alarm is a sufficient condition to say that the CHP has an anomaly but it is not a necessary condition. Many faults or degradation trends go unnoticed by the diagnostic system. The analyzed dataset presents the following fault' codes with the respective message:

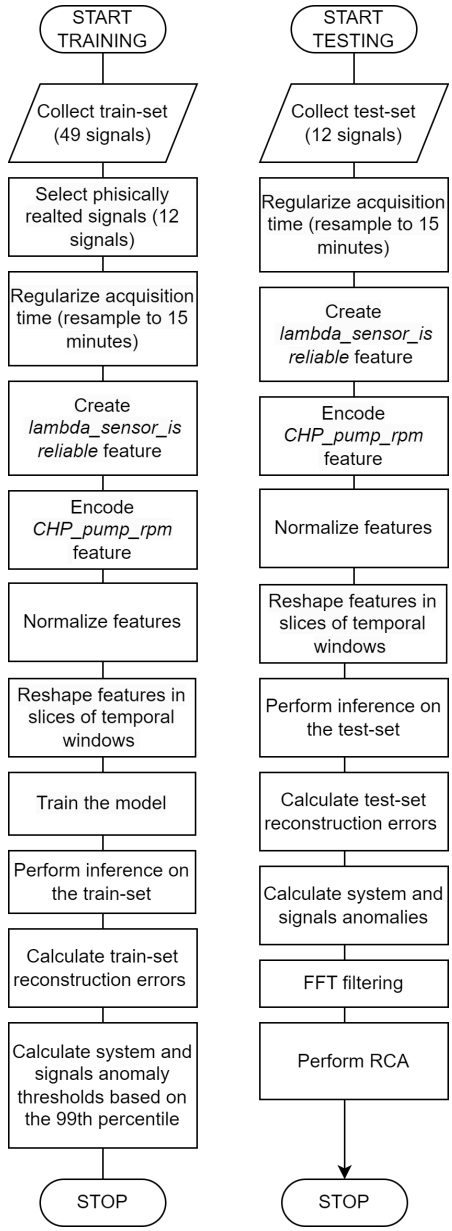

Figure 6.1: Flowchart of the proposed methodology.

- 0 - CHP stopped but ready to start;
- 2 - CHP stopped. Maintenance needed;
- 11 - Shutdown: generator protection;
- 12 - Low hydraulic pressure;
- 16 - Shutdown: water pressure too low;
- 18 - Shutdown: overheat interior;
- 19 - Shutdown: overheat engine oil.

Furthermore, the CHP can operate in four different operating modes. If the anomaly detection model is trained on a specific mode then the test-sets must be selected accordingly. Five operation modes are present:

- Mode 0, the CHP is turned off;
- Mode 1, the CHP is ready to charge an electric car;
- Mode 2, the CHP is optimized for summer operation;
- Mode 3, the CHP is optimized for heat production;
- Mode 4, the CHP is optimized for power production.

However, the CHP under analysis mainly operates to optimize heat production; as depicted in Fig. 6.2, except for Mode 0 when the plant is idle, the CHP primarily functions in either Mode 2 or Mode 3. When the CHP operates at maximum capacity, it is consistently configured to operate in Mode 3, which may be attributed to the high demand for thermal power within the facility. However, at partial load, the CHP primarily operates in Mode 2. Finally, in order to retain the temporal behavior of the CHP, samples are manipulated into sequences of 10 acquisitions by means of a sliding window. The normalized frames used to train the neural network and create a baseline of normality are shown in Figure 6.3, with the note that only the *lambda sensor is reliable* signal has been omitted as it is considered of limited interest in defining the CHP behavior.

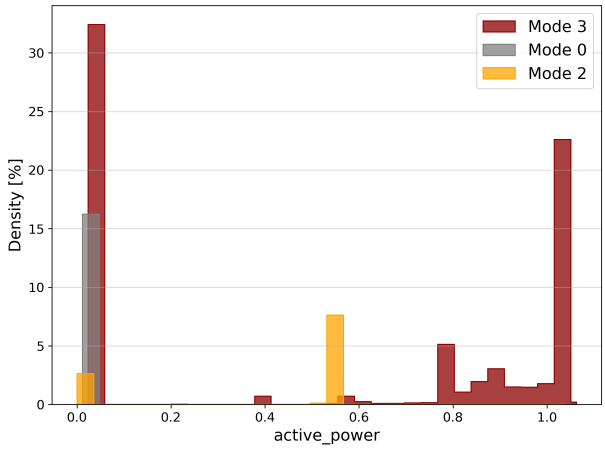

Figure 6.2: Distribution of the active power in relation to the CHP operating mode.

6.2.2 ConvAE: an ad-hoc structure of autoencoder

A tailored structure of autoencoder hereafter referred to as ConvAE, is presented in this subsection. The ConvAE proved to perform well in mimicking the CHP's normal behavior. It presents a straightforward but unprecedented shape (Figure 6.4): the input layer is provided with the manipulated acquisition, indeed, at each iteration, a sequence of 10 elements is inputted and then treated using a 10 filters 1D-convolutional layer; all filters have the same fixed dimension of 5 (*kernel size*). Results of the convolution are processed by a ReLu activation function ($f(x) = max(0, x)$) that introduces nonlinearities and then are compressed to a 3-node latent space.
The latent space consists of dense nodes and produces an output that is directly given as input to a transposed 1D-convolutional layer. Like in the encoder part, the output of the convolution passes through a ReLu activation function and then outcomes an array of the same size as the input. The autoencoder is trained using a learning rate of 10^{-4}, a batch size of 32, and the Adam optimizer to minimize an MSE loss function. The reduced size of the bottleneck and the 1D-convolutional layers permit ignoring the use of any kind of regularization such as dropout. A validation-set has been extracted randomly from the train-set and the training history is reported in Figure 6.5. The hyper-parameters used to build the ConvAE are de-

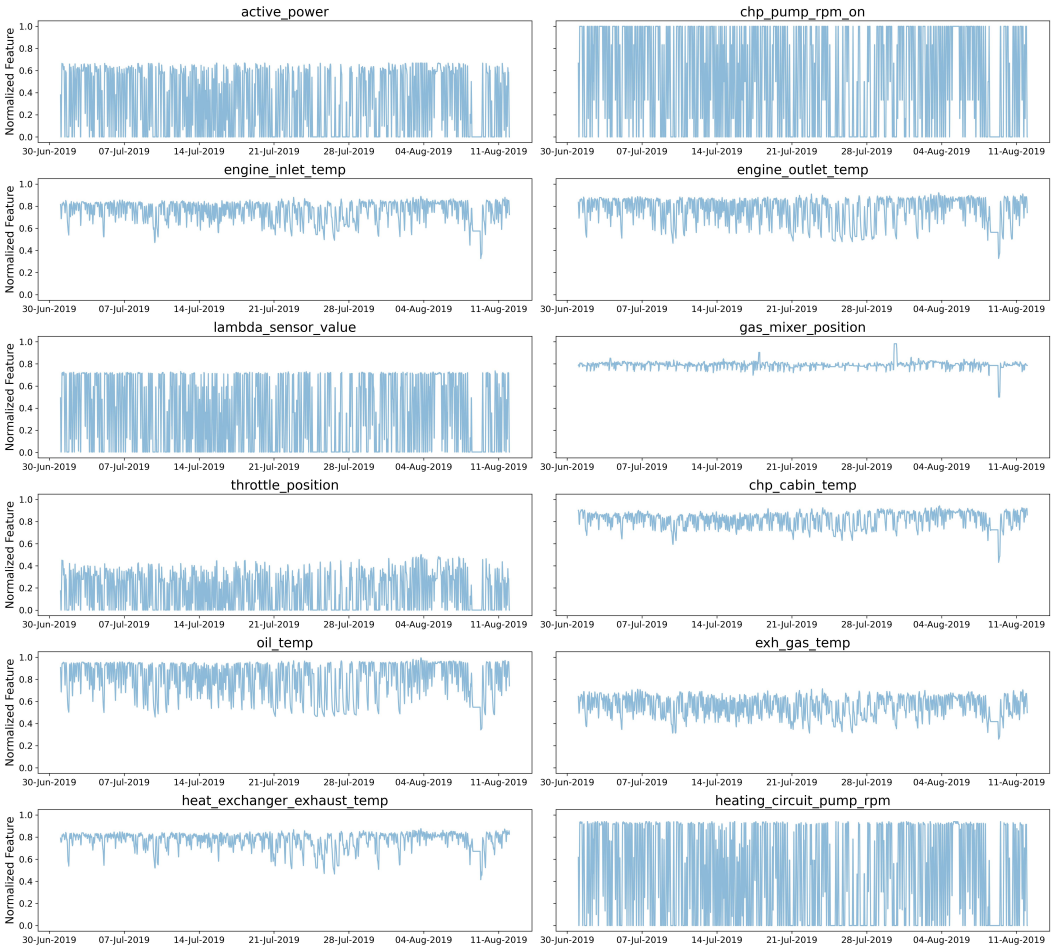

Figure 6.3: Time-series of the normalized features used to train the ConvAE.

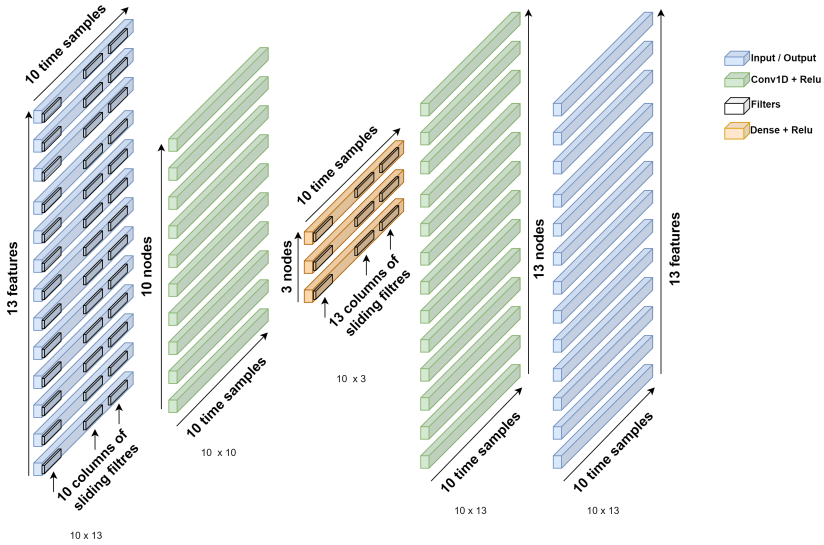

Figure 6.4: Convolutional Autoencoder architecture.

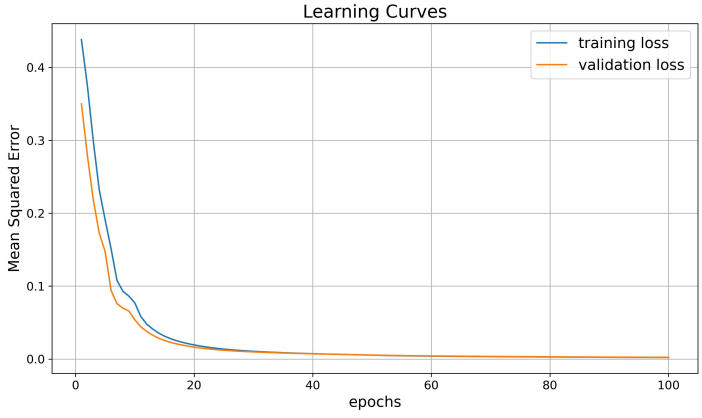

Figure 6.5: Train and validation loss history during the training phase.

tailed in Table 6.1. During the training phase, a threshold is calculated to

hyper-parameters	**ConvAE**
sliding window	10
input dimension	(10,13)
latent dimension	3
output dimension	(10,13)
activation function	*relu*
batch size	32
learning rate	10^{-4}
l1 regularization	0.0
dropout	0.0
padding	*same*
strides	1
filters number	10
kernel size	5
optimizer	*Adam*
loss function	MSE
validation split	0.1

Table 6.1: Convolutional Autoencoder hyper-parameters.

discriminate between normal and abnormal observations. This threshold is calculated on the 99^{th} percentile of the training reconstruction errors.
The percentile value was chosen after conducting a sensitivity study (Figure 6.6). A period was selected during which the CHP exhibited normal behavior, which clearly does not coincide with either the train-set or the test periods. This period ranges from July 15, 2020, to August 15, 2020. The choice was constrained by the available data, as a longer period where the machine's behavior was completely certain to be normal could not be found. Ten percentile values ranging from 90% to 99% were tested, showing a consistent trend of improving accuracy. Eventually, by selecting a percentile from 97% to 99%, every false positive was eliminated. Given that, a problem encountered in using autoencoders for anomaly detection is the high number of false positives, a more cautious approach was adopted and the value of 99% was chosen.
To be precise, it is important to note that it would have been scientifically more accurate to conduct a sensitivity analysis based on metrics such as F1-

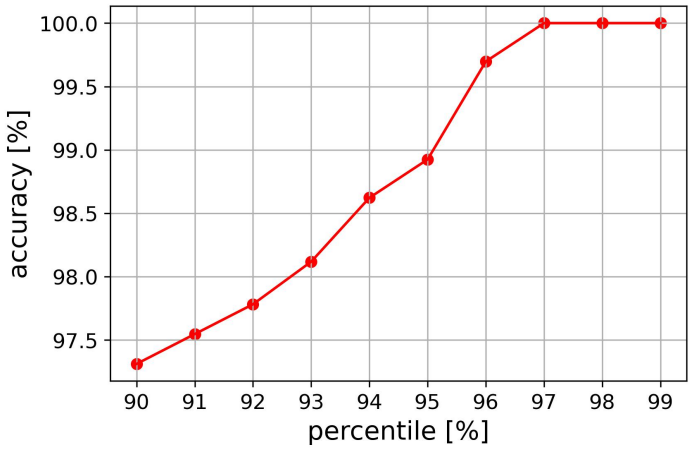

Figure 6.6: Sensitivity analysis for selecting percentile threshold, ConvAE.

score. Unfortunately, this was not possible because it would have required an additional dataset with known anomalies. Therefore, it was decided to choose the percentile in a way that maximizes accuracy, which in this context means minimizing false positives in a period of normality of the machine under analysis.

Since this work deals with time-series and observations are treated by means of a fixed-size sliding window, an observed sub-sequence is considered anomalous if all the 10 samples composing the sub-sequence itself have a reconstruction error greater than the pre-calculated threshold. Furthermore, since each sample is a multivariate consisting of a vector of 13 elements (as the number of inputs), the reconstruction error for each sample is calculated as the average error made reconstructing the 13 features in that time instant.

For the sake of clarity, the mathematical formulation for the reconstruction error *recon_error* is expressed below:

$$recon\_error = \frac{1}{W}\sum_{t=1}^{W}\frac{1}{N}\sum_{i=1}^{N}(x_i(t) - \hat{x}_i(t))^2 \tag{6.1}$$

where $x_i(t)$ is the i^{th} feature at time t, $\hat{x}_i(t)$ is the i^{th} reconstruction at time t, N is the number of features and W is the sliding window size.

Figure 6.7 shows the reconstruction error for the train-set; blue scatter points

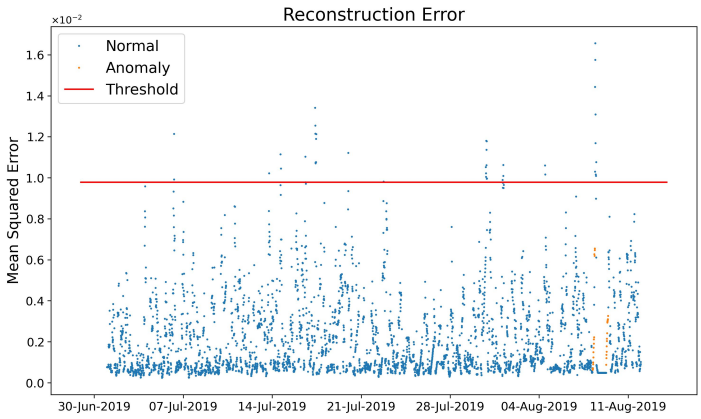

Figure 6.7: ConvAE reconstruction error of the train-set.

represent the normal samples according to the alarms reported by the internal controller of the machine. Due to the fact that the dataset comes from a real case study, irrelevant orange points are present (abnormal samples) and they can be considered perturbations. Indeed, in the industrial context, having a remarkable quantity of normal subsequent acquisitions is a strong hypothesis that seldom can be satisfied.

6.2.3 Post-processing

As already stated, autoencoders (but also other machine learning models) tend to produce false positives. Therefore, it was considered to post-process autoencoder detections using frequency-based techniques. False positives occur when the model detects transient conditions that deviate from its average normal behavior. For autoencoders, this also happens when the model is well-trained for variations related to seasonal or physiological changes in operating parameters. By using frequency-based techniques such as, for example, Kalman and moving average filters, the number of false positives can be reduced and the overall performance of the model can be improved. In this work, an FFT-based technique is proposed: the main idea, based on the domain knowledge of CHP functioning, is that detection can be considered reliable if it remains active for at least 12 hours.

The anomaly detection signal is transformed to the frequency domain and a low-pass filter is built to neglect components corresponding to high frequencies (i.e. with a period below 12 hours). In the second step, the anomaly detection signal is reconstructed back to the temporal domain by means of IFFT (Inverse FFT). Due to the low-pass filter, the generated detection signal continuously ranges between 0 and 1, thus a threshold (empirically set to 0.6) is used to discretize the signal and produce a boolean anomaly detection output. Figure 6.8 shows an example of the post-processing technique used: the raw detection signal is depicted in dark red and then, after being filtered and reconstructed, is reported in black on the left side. The discretization threshold is horizontally drawn in red (left side) and the results of its application are shown on the right side where the discretized anomaly detection signal is depicted in black.
This procedure outputs an anomaly only when the raw anomaly detection persists for a remarkable interval of time, indeed, the dark red signal appears to be more unstable while, on the other hand, the discretized black signal gives an indication of a reliable status of abnormality.

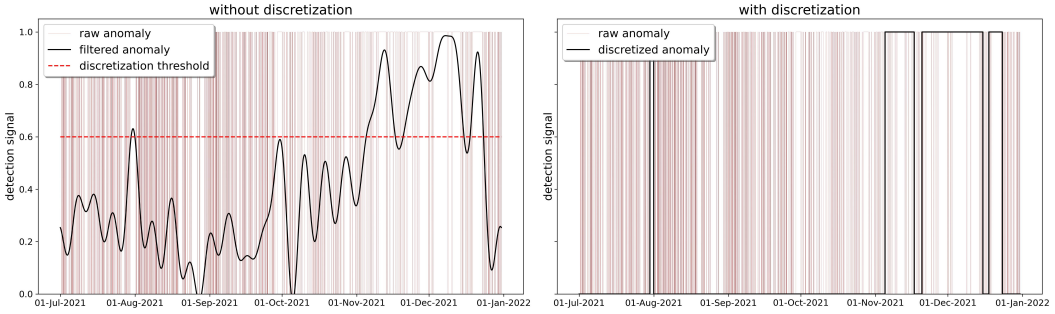

Figure 6.8: Post-processing technique explained.

6.2.4 Root Cause Analysis

Extracting the reconstruction error of every feature reconstructed by the autoencoder can help to understand which measurements led to an anomaly. An autoencoder is a particular version of a neural network that is trained to reconstruct its input data, and the reconstruction error is the difference

between the original input and the output of the autoencoder. By analyzing the reconstruction error for each feature, it is possible to identify which features have the largest error, indicating that they are the most different from the expected or normal data. This can help to pinpoint which measurements led to the anomaly and identify the source of the problem.
In literature, it is a common approach to build digital twins of a system in order to predict the behavior of a measurement and, in case the real value diverges from the inferred one, an anomaly is suspected. The proposed approach can be thought of as an ensemble of N models reproducing the behavior of N different sensors: clearly, building different models is more expensive both from research and computation perspectives. Furthermore, since the sensors are installed in different parts of the same machinery, there is a remarkable interconnection among measurement behaviors that justifies the exploitation of one single autoencoder with N outputs.
Considering each output of the model as an isolated reconstruction, a $99th$ percentile threshold is pre-calculated during training for each feature i, and each reconstruction error is calculated as follows:

$$recon\_error_i = \frac{1}{W} \sum_{t=1}^{W} (x_i(t) - \hat{x}_i(t))^2 \tag{6.2}$$

Extrapolating diverging behaviors of each feature is convenient for detecting components' faults that do not cause an evident malfunctioning of the machine. Indeed, if the machine is stable and does not start to operate in a way that affects many measurements, averaging the reconstruction errors leads to a mask on the local fault. Creating a general signal of the machinery anomaly gives a hint of the global behavior of the CHP; meanwhile, generating an individual signal of anomaly gives a deeper understanding of the CHP's components' life status and can be treated as a fault precursor to performing CBM.

6.3 Benchmark models

To objectively evaluate the results of the proposed method, two additional technologies were employed for the same task. A vanilla autoencoder with fully connected nodes, belonging to the same family as ConvAE but lacking the specific capability to retain temporal information, and a OneClass SVM, which is widely recognized as one of the benchmark techniques for out-

lier and anomaly detection problems. To ensure consistency in both model construction and assessing model stability in terms of false positives, the same preprocessing and postprocessing techniques described in the previous sections were applied.

6.3.1 AE: vanilla structure of autoencoder

The Autoencoder (AE) utilized in this study, which is the same one described in Section 5.2.2 with a different input dimension, is a basic structure consisting of three layers. The selected configuration and hyper-parameters for the AE model are described in Table 6.2: Coherently with Subsection 6.2.2,

hyper-parameters	**AE**
input dimension	13
latent dimension	70
output dimension	13
activation function	*relu*
batch size	32
learning rate	10^{-4}
l1 regularization	0.0
dropout	0.2
optimizer	*Adam*
loss function	MSE

Table 6.2: Vanilla Autoencoder hyper-parameters.

a sensitivity analysis (Figure 6.9) was carried out to define an optimal value for the percentile of the training reconstruction errors used to calculate the threshold discerning between normal and faulty acquisition during predictions. As for ConvAE, the value has been chosen equal to 99%.

6.3.2 One-Class SVM

The OCSVM [78] is also a semi-supervised algorithm that encloses normal data points within a hyperplane during training. During testing, any data points falling within the hyperplane are classified as normal, while those outside the hyperplane are identified as anomalies. Due to this operating principle, it is not straightforward to determine the specific feature that

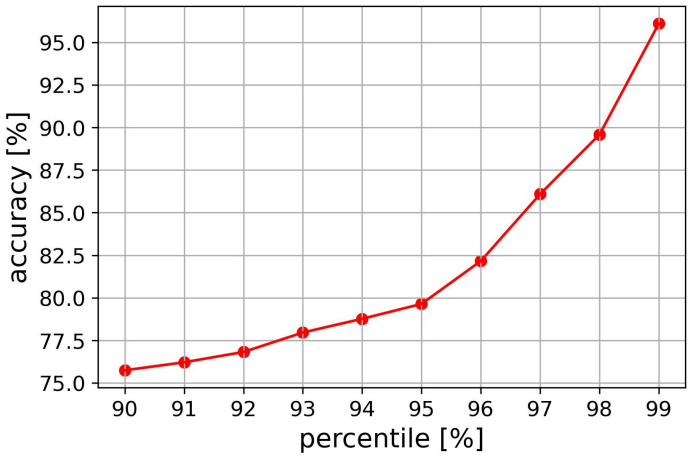

Figure 6.9: Sensitivity analysis for selecting percentile threshold, Vanilla AE.

caused the anomaly and perform a Root Cause Analysis (RCA). The problem formulation is as follows [92]:

$$\begin{aligned} \min_{w,\xi,\rho} \quad & \frac{1}{2}||w||^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - \rho \\ \text{s.t.} \quad & w \cdot \Phi(X_i) \geq \rho - \xi_i, \\ & \xi_i \geq 0. \end{aligned} \tag{6.3}$$

The aim is to enclose all normal points into a hyperplane, allowing for a fraction, denoted by $\nu\%$, of the data to be incorrectly separated. In the above formula, w represents the coefficients of the hyperplane and ρ the bias in the feature space, $\Phi()$ denotes a non-linear transformation (associated with the kernel used in the dual formulation), ξ_i represents the classification error for the i_{th} example, and $\nu \in (0, 1)$ is a hyper-parameter. When $\nu = 0$, the penalty term for violating the separation constraint becomes infinite, causing the separator to enclose all observations with a classification error $\xi_i \approx 0$ for every example i. This only happens by generating a very wide region containing the entire dataset. On the other hand, if $\nu > 0$, it is possible to obtain a good surface capable of correctly classifying many examples without the requirement of including all of them. In a somewhat informal sense, ν can be

associated with the tolerable fraction of classification errors on the train-set. Once the system has been trained on the train-set, the decision function in Equation 6.4 will be employed for the test example. It will produce negative outputs in case of anomalies and positive outputs otherwise.

$$f(x) = w \cdot \Phi(x) - \rho \tag{6.4}$$

The hyper-parameters to be tuned for optimizing the performance of the OCSVM are the ν parameter, the kernel, and the γ parameter.The value of ν was found to be irrelevant in the problem under consideration (Figure 6.10), likely due to the problem's complexity, which generates a highly nonlinear boundary between normal and novel instances. It was set to 1% to maintain consistency with the 99% threshold chosen for both the ConvAE and the AE. The kernel, on the other hand, was selected as the Radial Basis Function (RBF) kernel, described in Equation 6.5, to model the high nonlinearity present in the dataset.

$$K(x, y) = \Phi(x) \cdot \Phi(y) = \exp\left(-\gamma ||x - y||^2\right)) \tag{6.5}$$

The γ parameter was calculated using the procedure proposed by Ghafoori et al. [36].

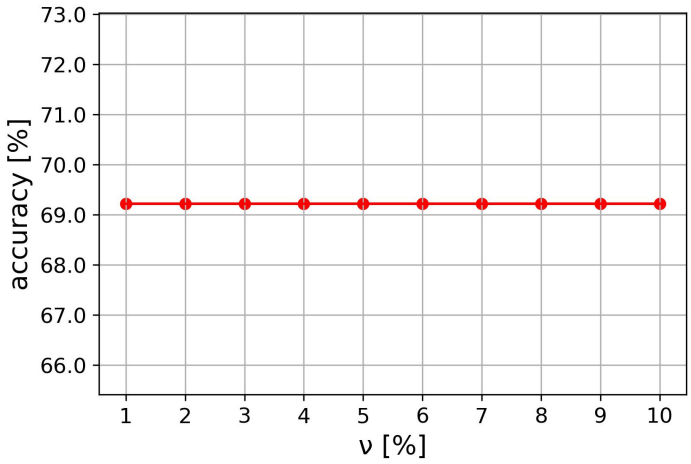

Figure 6.10: Sensitivity analysis for selecting ν, OCSVM.

For the sake of precision, Table 6.3 provide a brief sum-up:

hyper-parameters	OCSVM
kernel	RBF
ν	0.01
γ	5.1

Table 6.3: One-Class SVM hyper-parameters.

6.4 Results

In Sections 6.1 and 6.2.1 the features composing the dataset have been explained and the methodology to choose the train-set has been briefly treated. Thereafter, two test-sets have been selected to stress the model performance in both normal functioning and in the presence of an a-priori known anomaly. Table 6.4 reports the dates and details of each dataset split. In the first test-set (referred to as the *Anomalous Test-set* in Table 6.4) a component failure is present, and the heat exchanger starts to present fouling affecting CHP's performance. The second test-set (referred to as the *Healthy Test-set* in Table 6.4) was recorded immediately after the heat exchanger was replaced, and the CHP was serviced by specialized personnel. Therefore, no anomaly is present and the unit runs in a healthy state. Fig. 6.11 depicts the CHP

	From	To	CHP status
Train-set	1 Jul 2019	12 Aug 2019	HE Healthy
Anomalous Test-set	1 Jul 2021	31 Dec 2021	HE Faulty
Healthy Test-set	15 Jul 2022	7 Sep 2022	HE Healthy

Table 6.4: Datasets details: dates and Heat Exchanger (HE) status.

behavior in terms of a-dimensionalized produced electric power (orange scatter), alarms reported by the onboard diagnostic system (red scatter), and the human detection made by the plant supervisor (purple vertical line). Grey-shaded areas emphasize the dataset split used for this work detailed in Table 6.4. The observations presented in Fig. 6.2 are supported by the findings in Fig. 6.11, which indicate that the CHP system primarily operates either at full load (between 80% and 100% of the rated power) or at 50% partial load. It is worth noting that the training period for the model corresponds to a period where the CHP management logic is slightly distinct from the two test-sets. Specifically, the range of powers covered during the train-

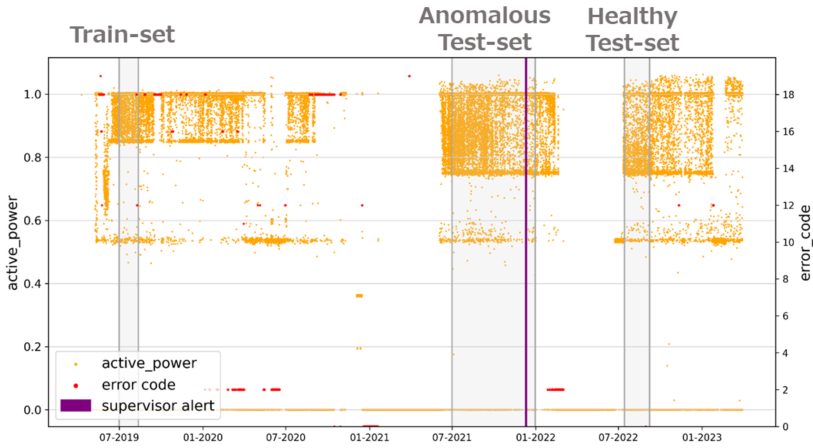

Figure 6.11: Active power describing CHP behavior.

ing phase is comparatively smaller than that observed after a plant shutdown period of approximately 5 months, where the power range has increased. It is important to remark that the errors reported by the onboard CHP system (red dots in the picture) are not correlated to the heat exchanger fouling acknowledged by the plant supervisor. Furthermore, these errors are of minor interest in terms of research as trivial faults can be recognized more easily due to their disruptive consequences often leading to a machine shutdown. On the other hand, the heat exchanger deterioration studied in this work is very interesting due to its slow degradation trend (low negative) over time which is not easily detectable by naive thresholds or by the human naked eye. Figure 6.12 shows the results of the ConvAE predictions on the *Anomalous Test-set*. Specifically, each measured input signal is plotted over time (blue curve). The same plot also shows the two predictions of ConvAE in terms of anomaly: in red the anomaly detected on the specific signal and in black the system ensemble anomaly calculated as described in Section 6.2.2.
In general, various scenarios that may arise are considered. The first and most critical scenario for system safety occurs when an overall anomaly occurs, as observed, for example, in the December 2021 period. It is evident that, during such anomalies, certain individual signals not directly related to fouling do not display anomalous states. The global anomaly represents an average of different signal anomalies, and it is not necessary for all of them

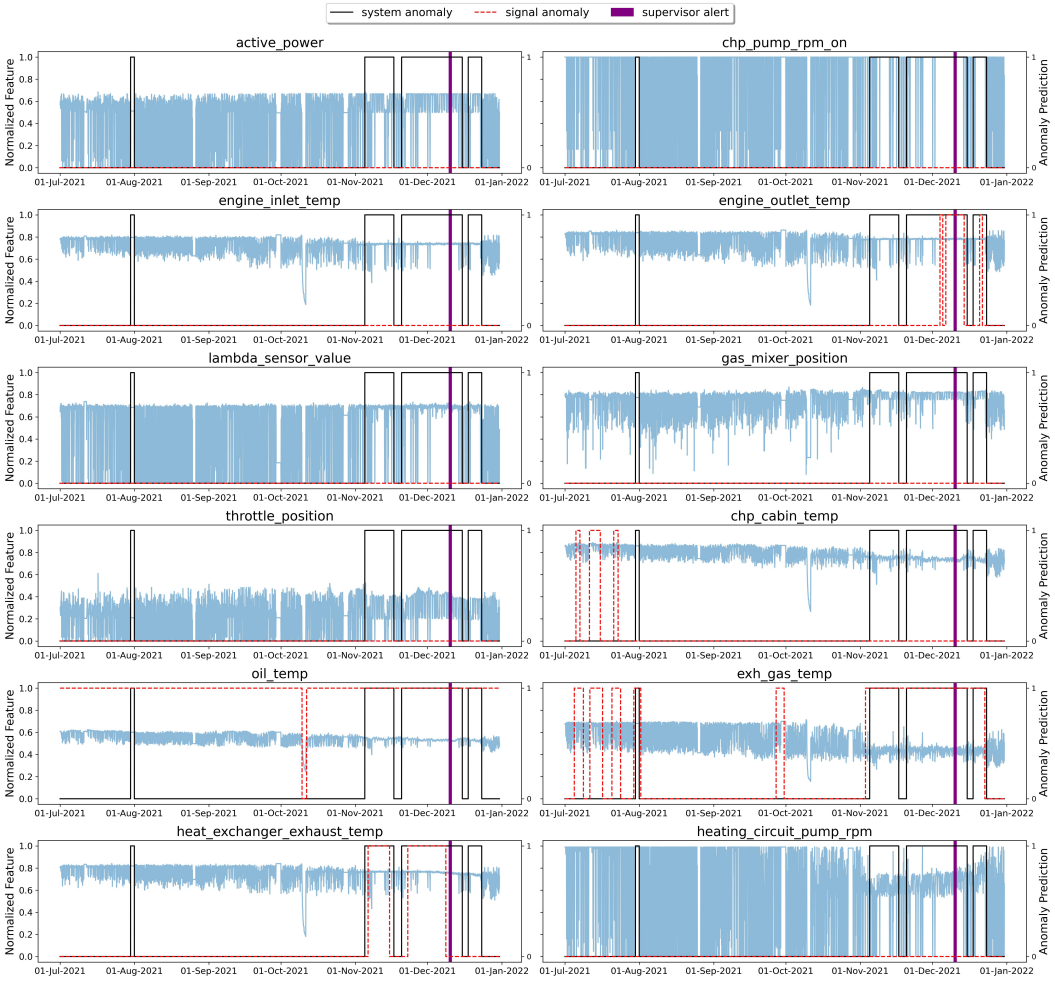

Figure 6.12: Time-series of the normalized features used to test the ConvAE in the presence of anomalies.

to be implicated in the malfunction. A global anomaly is predicted when certain correlated signals simultaneously exhibit anomalies, thereby making a fundamental contribution to the global anomaly signal itself.
Another notable scenario occurs when the global anomaly is absent, but certain signals exhibit anomalies, as observed, for example, in the September 2021 period for the oil temperature. This situation is of considerable interest to the machine service manager. Here, it becomes possible to identify a specific subcomponent anomaly or a malfunction in the measurement sensor. Moreover, this condition could pose a health hazard for the component itself or serve as an early symptom of failure in other components. In other words, it can act as a precursor to a global and more severe anomaly. An instance of this can be seen in the monitoring of oil temperatures. Although the presence of this irregularity in the data was not initially known, it was identified by the ConvAE and subsequently verified by the YANMAR maintenance department. It was determined that the oil temperature was consistently falling within a lower range due to abnormal operation of the engine cooling system. The alarm raised by ConvAE enabled the maintenance engineers to uncover this underlying malfunction in the CHP system and make necessary improvements to enhance its efficiency. Interestingly, in this case, the model was able to distinguish individual signal malfunction and not transfer it to a global anomaly. The failure of this signal alone was not sufficient to trigger the system anomaly.
Another significant case, considering the primary anomaly identified in the heat exchanger back in December 2021, holds importance in the early detection of irregularities in the exhaust gas temperature signal. This specific measurement can be seen as an indicator of the heat exchanger's condition, as illustrated in Figure 4.2, which depicts the cooling circuit diagram. The heat extracted from the flue gas indirectly affects the heat transfer to the heat exchanger by influencing the coolant temperature. Describing the physical correlation or sequence of phenomena involved in heat transfer is neither intuitive nor simple. However, the model detects local anomaly patterns in the exhaust gas signal as early as July-August 2021 and later in October 2021. These patterns are impossible to recognize through standard data analysis techniques or by relying on classical thresholds for the measured value. Hence, the significance of the model becomes evident as it identifies a local anomaly in the exhaust gas system, serving as an early indication of a more critical anomaly in the heat exchanger. Detecting these warnings

several months in advance allows proactive measures to be taken. Other precursors of the anomaly, identified by the anomaly detection routine alongside the exhaust gas temperature, include the CHP cabin temperature (ConvAE detects a first anomaly in July 5^{th}, 2021) and the heat exchanger exhaust temperature (ConvAE detects a first anomaly in date November 11^{th}, 2021). Similar observations can be made for these signals.
In the heatmap displayed in Figure 6.13, a comprehensive overview reveals the signals co-participating in the global anomaly. Evidently, the block heat exchanger temperature and exhaust gas temperature, as expected, exhibit simultaneous anomalies during the anomaly presence period. Anomalies in engine outlet temperature emerge later in time (mid-December 2021), just before detection in the field. When this measurement displays non-normal patterns, the anomaly has already significantly progressed. Interestingly, the model recognizes the anomaly long before this specific signal also contributes directly. The presence of precursors or early detection of anomalies (in com-

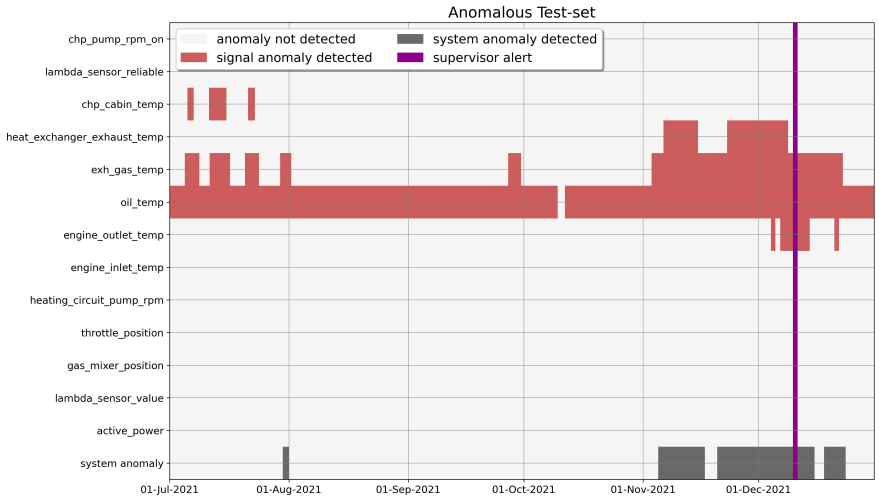

Figure 6.13: Heatmap of anomalous test-set, ConvAE.

parison to the field supervisor's alert) raises several important issues that need to be addressed and discussed with YANMAR's service department. Specifically, questions arise, such as when to intervene on the machine in the event of an anomaly and how to determine whether an anomaly serves

as a precursor. These answers are not straightforward but suggest further enhancements to the already promising model. Introducing a degree of confidence, as subsequently discussed in Chapter 7, is crucial to distinguish between critical faults and insignificant deviations from the learned behavior. In fact, to develop a CBM strategy, the anomaly detection routine should provide a level of confidence that highlights the impact of abnormal behavior on the system.
At the beginning of July 2022, global maintenance was performed and the heat exchanger has been substituted. Figure 6.14 shows the *Healthy Test-set* where the proposed algorithm detects neither system nor signal anomalies. Both CHP's on-board system and YANMAR engineers do not report any abnormal behavior in accordance with ConvAE predictions.
For direct comparison with the anomalous case (Figure 6.13) the heatmap of the healthy case is shown in Figure 6.15 from which to have a global and quick look at the prediction of the model. This second test-set, where no anomaly is present, confirms the well functioning of the FFT filtering procedure described in Section 6.2.3. This post-process procedure properly limits the most underlined flaw of semi-supervised systems in literature: the false-positive detection. From Figure 6.16 it can be seen that in fact, the model without filtering block predicts a number of false positives, which, as discussed above are disturbances of the prediction and not real anomalies. The robustness of the system is ensured precisely by the filtering block. By managing the filtering threshold, it is possible for the user to adjust the sensitivity of the service.

6.4.1 Benchmark models

The first benchmark model, the vanilla AE, performs excellently in detecting anomalies in the cooling oil circuit but proves to be less sensitive in detecting fouling anomalies on the heat exchanger (Figure 6.17). A system anomaly is briefly identified only well before the target anomaly, while it is not signaled in proximity to the fouling of the heat exchanger. A precursor is identified at the appropriate time in the CHP pump speed, but it alone may be too ambiguous to uniquely identify the fouling anomaly. Conversely, during the normal period, no anomalies are detected as desired (Figure 6.18). Regarding the second benchmark model, the OCSVM, it has an intrinsic disadvantage: it cannot perform RCA to determine which feature led to an anomaly. However, its performance in detecting a system anomaly can still

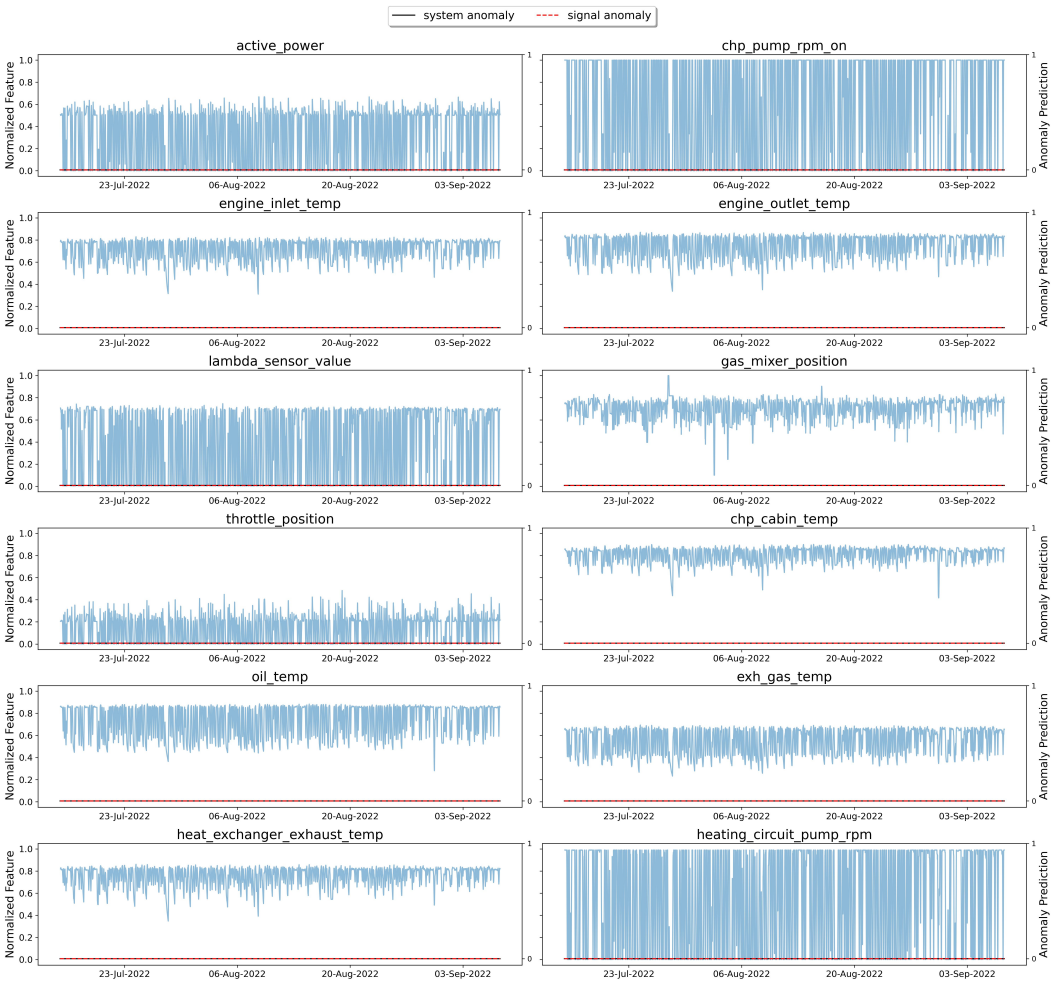

Figure 6.14: Time-series of the normalized features used to test the ConvAE in the absence of anomalies.

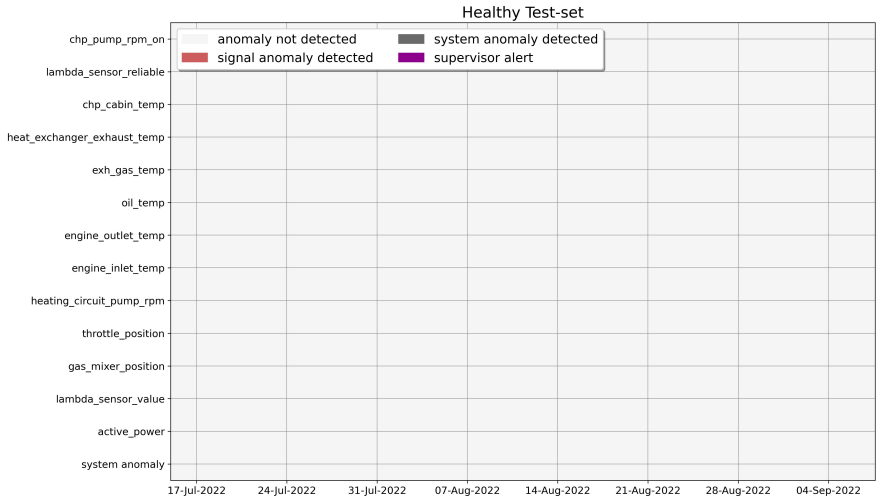

Figure 6.15: Heatmap of healthy test-set, ConvAE.

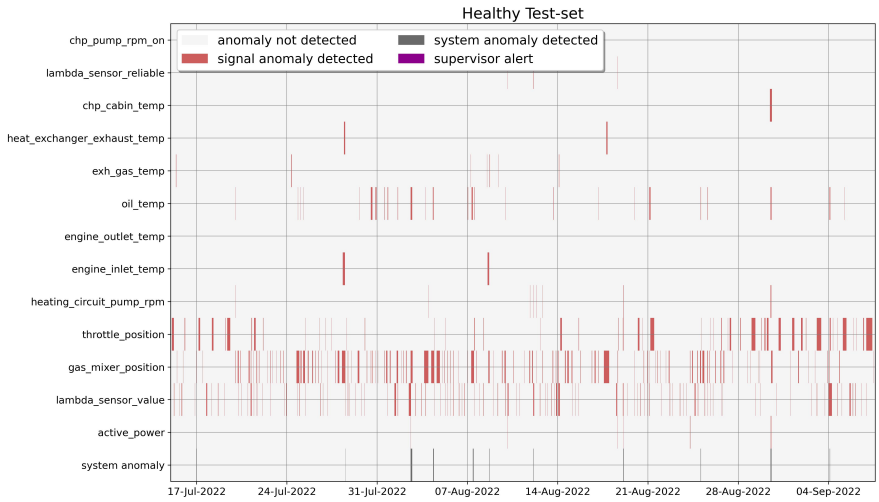

Figure 6.16: Predictions before the filtering block is applied to ConvAE output.

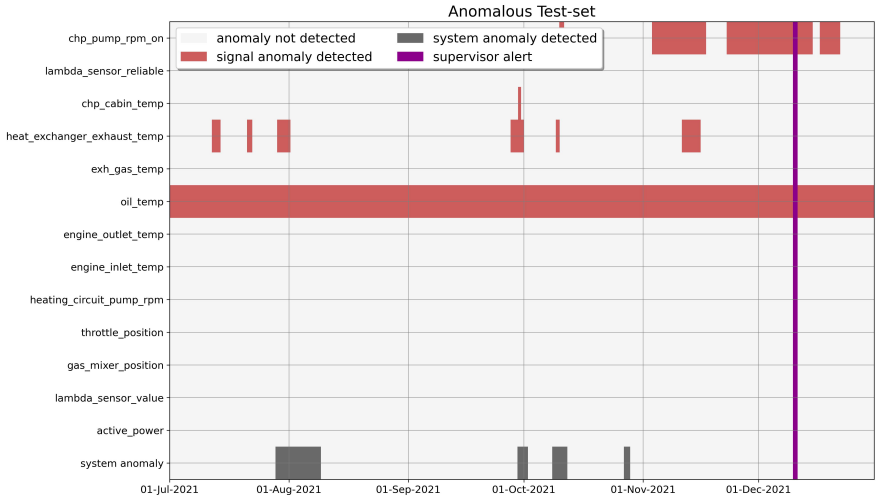

Figure 6.17: Heatmap of anomalous test-set, Vanilla AE.

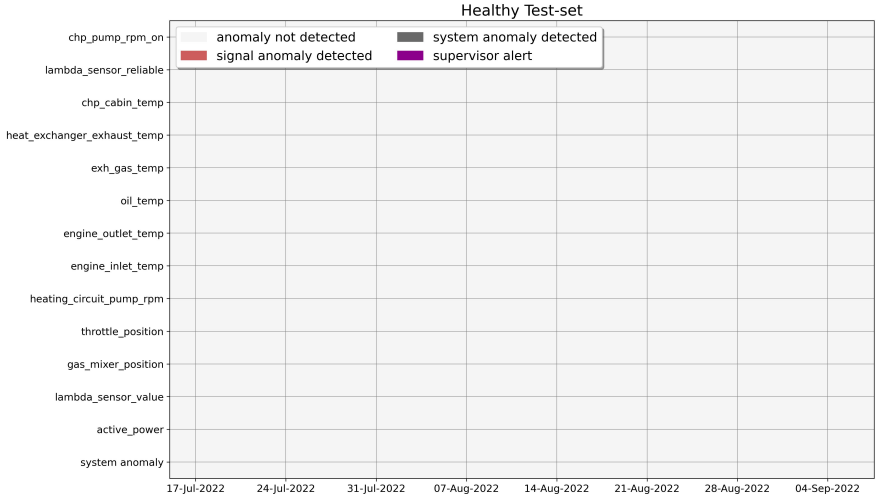

Figure 6.18: Heatmap of healthy test-set, Vanilla AE.

be evaluated.
Unfortunately, despite the OCSVM, like the ConvAE and AE, being coupled with low-frequency anomaly filtering, it still fails to demonstrate robustness against false positives. In fact, while the continuous detection of a system anomaly can be accepted during the first test period (Figure 6.19) due to the presence of two issues (one related to oil temperature and the other due to fouling), the same cannot be said for the second test period (Figure 6.20), where maintenance has just been carried out and no faults are present.

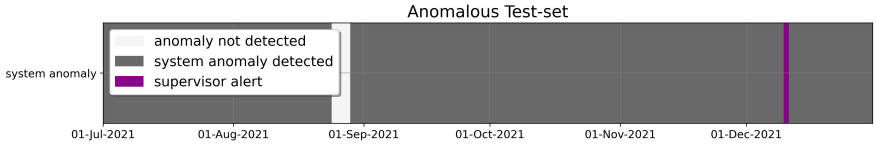

Figure 6.19: Heatmap of anomalous test-set, OCSVM.

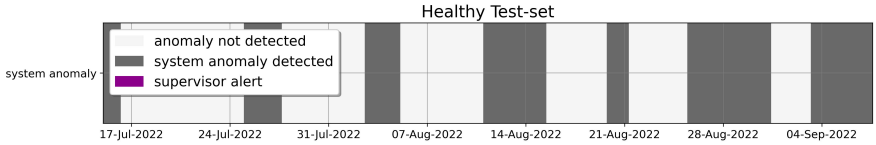

Figure 6.20: Heatmap of healthy test-set, OCSVM.

6.5 Chapter Conclusions

In this chapter, the object of the research is a 20 kW_e YANMAR CHP used to supply a school facility located in Germany. An ad-hoc structure of autoencoder with convolutional layers is applied to a problem of anomaly detection in a time-series framework. Reduction of false positives is addressed by employing a frequency-based technique and, in addition, a failure RCA is performed to understand what features carry the content of abnormality. The deep learning model has been trained on 6 weeks of normality and then tested over two frames of data: in the first frame, some abnormal behavior

was present while in the second frame, the CHP was serviced and the functionalities were restored.
The proposed methodology yielded successful results predicting the anomaly more than one month before and giving useful indications about the failing component. Furthermore, the methodology proved to be robust not generating any false positives in the set of data where the CHP was running in normal conditions. To validate the results, the outcomes of the proposed methodology were compared with two benchmark algorithms, demonstrating the superiority of the proposed ConvAE structure. It is important to underline that the proposed routine needs to be tailored to customers' needs and on the specific case of application. A number of next developments have already been identified, marking significant progress in the proposed approach. The first crucial step involves the implementation of an online system, where the anomaly detection system will operate in real-time through a cloud-based service, ensuring continuous monitoring and prompt detection of anomalies as they occur. This real-time capability is fundamental for efficient monitoring, enabling the quick identification of deviations from the norm. As discussed in the subsequent chapters, the introduction of a degree of confidence is essential. It serves as a critical factor, allowing to distinguish between potentially dangerous faults and irrelevant deviations, thereby supporting targeted maintenance operations on the actual product. Additionally, this thesis will focus on aspects such as domain adaptation, enabling the model to adjust to non-anomalous changes in the machine over time. In addition, transfer learning will be explored, investigating the feasibility of transferring a pre-trained autoencoder to different energy generators. These investigations have already begun with some preliminary tests and will continue in the short term, albeit outside the scope of the current thesis.

Chapter 7

A Bayesian Approach to Improve Detection Robustness

This chapter proposes a deep learning methodology to detect anomalies on a real CHP unit supplying a school in Germany. The core of the work is a convolutional autoencoder trained on the normal behavior of the energy generator. The autoencoder is enhanced with a Bayesian technique, the Monte Carlo dropout, used to add a stochastic component to the model to quantify the uncertainty degree of the detection. This information is crucial to determine if or when an action is actually needed, optimizing the service and maintenance strategy. The proposed approach was applied to a real case study and was found to be effective; heat exchanger fouling was detected 5 weeks before the standard detection system. The algorithm returns high confidence in system anomalies and low detection confidence for minor alterations in behavior, less risky for the machine. Section 7.1 provides comprehensive insights into the available dataset. Section 7.2 describes the model architecture and addresses the proposed methodology. The obtained results from two distinct test-sets will be examined in Section 7.3. Additionally, conclusions are drawn in Section 7.4. [1]

[1]Part of the contents of this chapter has been published as "Applied Anomaly Detection: a Bayesian Approach to Improve Robustness" in *Proc. of IEEE International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2023*, [23]

7.1 Dataset

The YANMAR micro-CHP under analysis provides electricity and thermal energy to a school facility located in Germany and is the same as described in Section 6.1.
To train the model, a 6-week training period is required (as demonstrated in Chapter 5) where the CHP behaves normally. Two test-sets are also required (as detailed in Table 6.4): the first one where a certain quantity of abnormalities happens to check that the model correctly detects the presence of anomalies and the second one where the CHP does not present a particular deviation from the learned behavior to verify that the model does not present false positives.

7.2 Proposed Model

The same problem has been addressed in the previous Chapter using an autoencoder with convolutional layers (ConvAE) that considers temporal information. In this Chapter, a new feature is introduced by using a Bayesian approach to estimate the confidence with which the autoencoder detects an anomaly.
As delineated in Section 2.6, compared to VAE, the MCD approach facilitates the introduction of stochasticity in the model without necessitating any modification in the original architecture of the deep learning model, if a dropout layer has already been incorporated.
In the reference architecture, dropout layers are not included, yet it has produced excellent results (hyper-parameters are reported in Table 6.1), and therefore, keeping it similar is desirable. To compensate, changes were made to the original ConvAE model by incorporating MCD to maintain the same level of performance when it comes to reconstruction errors. In fact, since MCD functions as a regularization technique during the training process, the original architecture required additional complexity to maintain the same level of information retention during learning.
Hyperparameters employed for both models are presented in Table 7.1, which highlights the augmentation of convolutional filter numbers, from 10 to 18, and an expansion of the latent space from 3 to 5 dense neurons. Additionally, an MCD of 5% was incorporated, and the inference was conducted 50 times. The number of inferences was chosen to be 50 to strike a balance between

the model's variability and mitigating the computational time required for making multiple predictions. The architecture of the Convolutional autoen-

hyper-parameter	**ConvAE**	**ConvAE MCD**
sliding window	10	10
input dimension	(10,13)	(10,13)
latent dimension	3	5
output dimension	(10,13)	(10,13)
activation function	*relu*	*relu*
batch size	32	32
learning rate	10^{-4}	10^{-4}
l1 regularization	0	0
dropout	0	-
padding	same	same
strides	1	1
filters number	10	18
kernel size	5	5
optimizer	Adam	Adam
loss function	MSE	MSE
validation split	0.1	0.1
Montecarlo dropout	-	0.05
Montecarlo samples	-	50

Table 7.1: Hyperparameters of Configurations for the Two Models.

coder with Monte Carlo Dropout (ConvAE MCD) is reported in Fig. 7.1. The ConvAE MCD exhibits an asymmetrical arrangement, where the encoder is marginally bigger than the decoder. The input data comprising 13 features are initially segmented into 10-sample moving windows and then processed using a 1D-convolutional layer composed of 18 filters. The output of this layer then proceeds to the bottleneck of the autoencoder through a fully connected layer. On the decoding side, the architecture is identical, except for the convolutional layer, which employs 13 filters to align with the number of feature reconstructions.
During the training stage, the model strives to minimize the reconstruction error of the inputs. In accordance with the reference model, a threshold is set using the 99^{th} percentile of the training error to discern between normal and abnormal samples during the testing phase. As in the previous

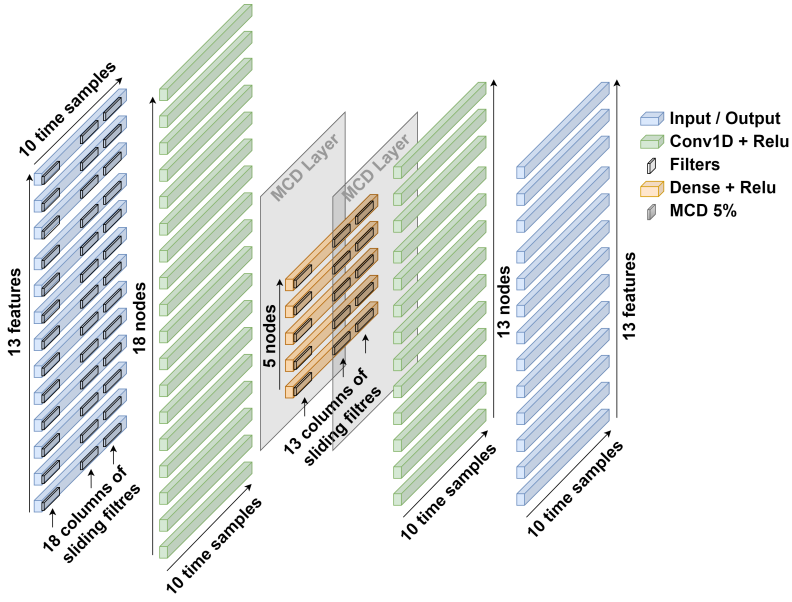

Figure 7.1: Architecture of the Convolutional Autoencoder with MCD.

Chapter, two types of anomalies are generated. The first type is generated for the entire system if the average error over the 13 features exceeds the 99^{th} percentile threshold calculated on the train-set reconstructed inputs. The second type of anomaly is generated for each signal: a specific reconstruction error is calculated and compared with a 99^{th} percentile threshold obtained by considering the error committed during the training phase inference of the particular signal itself.
As discussed in Section 6.2.3, the autoencoder generates anomaly signals that may be susceptible to yielding false positives. To mitigate this challenge, a post-processing low pass filter has been devised to disregard any alarms that persist for less than 12 hours. The methodology proposed in the Chapter is illustrated in Fig. 7.2, which presents a flowchart outlining the steps followed in the study.

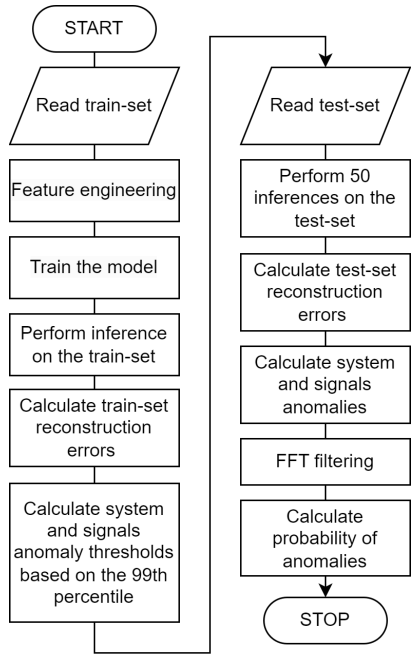

Figure 7.2: Flowchart of the proposed methodology.

7.3 Results

As mentioned in Section 6.2, the autoencoder model has been designed to reconstruct 13 features but, as a matter of simplicity, in the following, only the most interesting signals recorded during the *Anomalous Test-set* and the *Healthy Test-set* are described.
During periods where anomalies are present the heat exchanger becomes dirty, and the efficiency of the CHP has a negative trend, which leads to a significant increase in costs. Fig. 6.12 shows how the anomaly was discovered only on December 10^{th}, 2021 (vertical purple line). However, the proposed algorithm detects the first system anomaly (black line) on October 10^{th}, due to an unexpected shutdown. Subsequently, a second system anomaly is detected on November 7^{th}. In this case, the anomaly depends on the fouling of the heat exchanger, since signals related to the heat exchanger produce specific alarms (red lines): in particular, it can be seen how the exhaust gas temperature decreases significantly on average.
The discussed alarms are considered extremely reliable by the model as the percentage of Monte Carlo experiments generating an anomaly is 100%. However, the behavior of the alarm linked to the heating circuit pump is different. The pump starts operating at lower speeds, initially generating an anomaly confidence level of around 60%, which then decreases ranging between 20% and 10% when the pump speeds up. In contrast, the alarm signal of the CHP cabin temperature shows a confidence level of around 40%, which then increases to 100%.
Ultimately, the algorithm detected the heat exchanger fouling 5 weeks before the actual detection made by the YANMAR supervisor. On the other hand, if the heat exchanger anomaly was a late detection by the engineer supervising the plant, also in this variant, the algorithm revealed another problem of the cogeneration unit that went totally untracked but then acknowledged by the YANMAR maintenance department: the engine cooling system exhibited abnormal functionality, resulting in a lower range of oil temperature. As a result, the algorithm yielded a probability of anomaly with complete certainty, i.e., 100%.
It is noteworthy that the model has changed compared to the one used in the previous chapter, thus leading to slight variations in the detections. In both approaches, anomalies in oil temperature are detected, and both identify irregularities in certain signals related to the heat exchanger. While in Chapter 6, the signals susceptible to anomalies are the heat exchanger

exhaust temperature, engine outlet temperature, and exhaust gas temperature, in the approach presented in this chapter that utilizes MCD, the signals of interest include not only the exhaust gas temperature but also the CHP cabin temperature and the heating circuit pump rate. To establish not only the algorithm's sensitivity to anomalies but also its robustness against false positives, it's crucial to utilize a dataset where the CHP's proper functioning without any malfunctions is certain. This aspect has already been addressed in the previous Chapter, where a dataset was collected before a planned global maintenance intervention in early July 2022, during which the heat exchanger was replaced.
Fig. 7.4 shows how all signals recovered and no anomalous trend is present: indeed, the algorithm proved to be well-performing also when the CHP is healthy and to be not prone to false positives. The presented data demonstrate that the signals, which are directly related to the heat exchanger, have regained normal behavior, without any discontinuities. Fig. 6.12 illustrates that all temperature signals (cabin, oil, and gas exhaust) exhibited a negative temporal trend, which has now vanished in Fig. 7.4. Furthermore, the pump of the heating circuit has resumed its healthy pattern of modulation, operating within the range of 0% to 100%.
By comparing the plots of these signals before and after the maintenance intervention, can be concluded that the proposed routine exhibits sensitivity towards the anomaly presented and effectively distinguishes when a healthy condition has been restored.

7.4 Chapter Conclusions

Major companies are striving to migrate their business from the production and sale of products to the provision of services. In this context, data-driven approaches can flourish and provide significant added value. Specifically, with growing attention to climate change and energy and economic savings, the transition from time-based maintenance techniques to CBM plays a primary role.
This Chapter proposes a deep learning technique that, when trained on 6 months of data under normal conditions of a YANMAR micro-cogenerator, can provide both qualitative and quantitative indications of the CHP's behavior. As in Chapter 6, two test-sets were chosen: the first where deterioration of the heat exchanger was known a-priori, and the second where

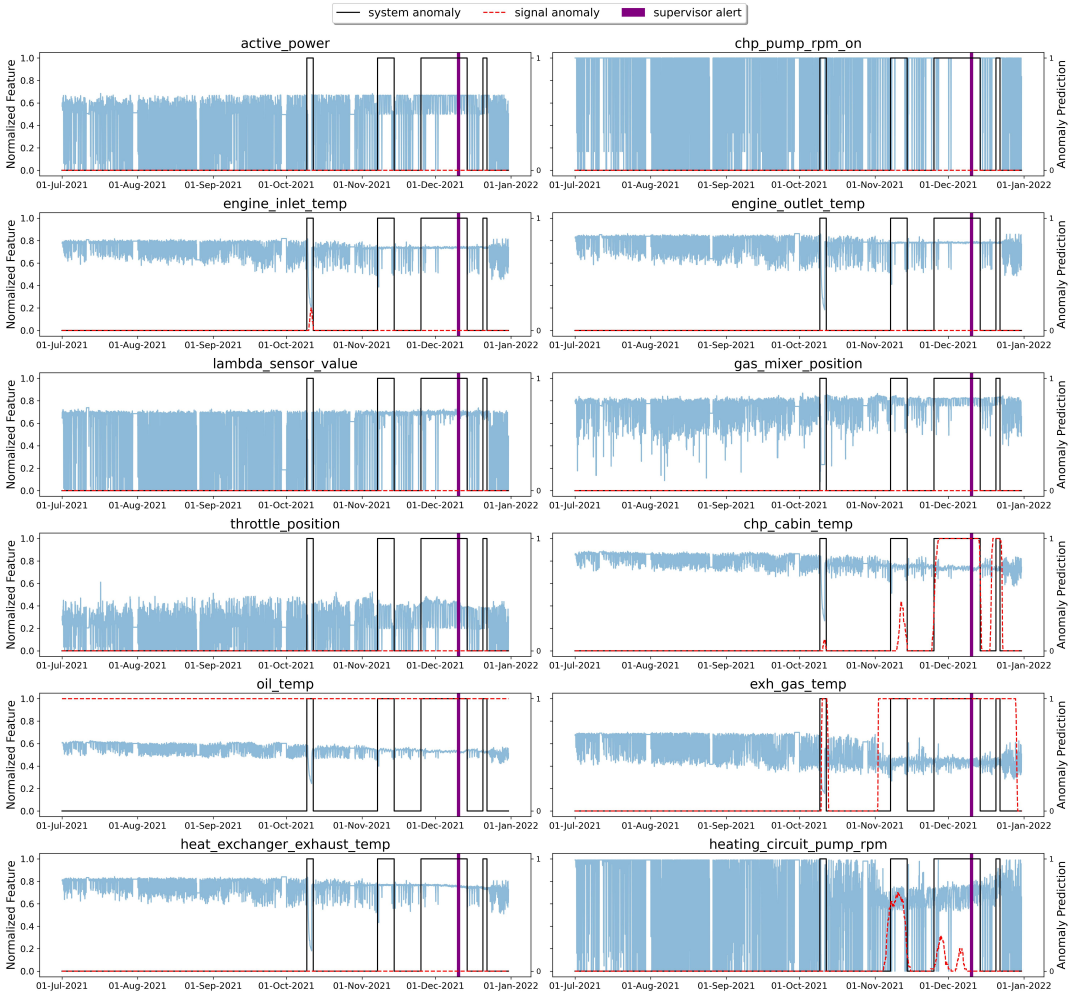

Figure 7.3: Time-series of the normalized features used to test the ConvAE MCD in the presence of heat exchanger fouling.

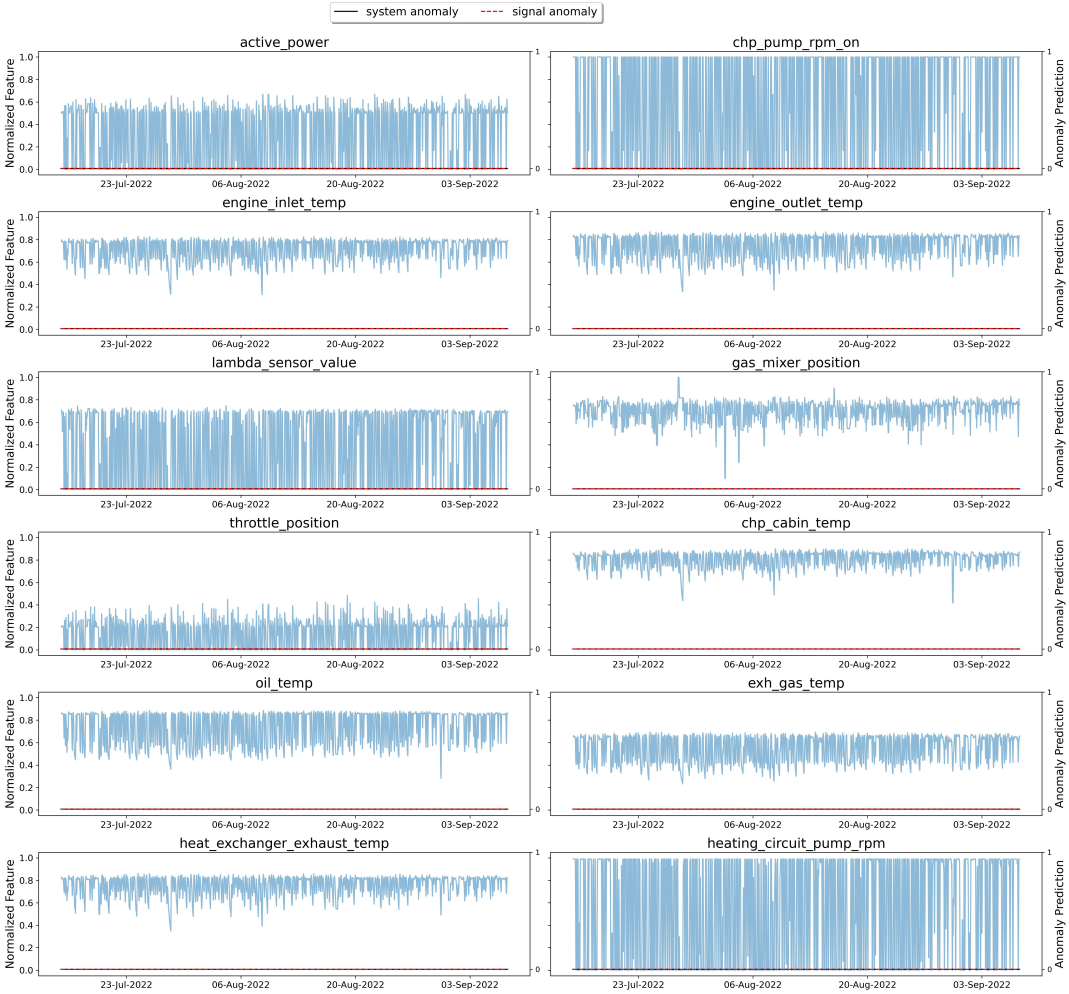

Figure 7.4: Time-series of the normalized features used to test the ConvAE MCD in the absence of anomalies.

maintenance and replacement of faulty parts had just occurred. The deep learning model consists of an asymmetric autoencoder embedded with 1D-convolutional layers, MCD layers, and a fully connected bottleneck. At the end of the detection pipeline, a frequency-based filter is used to reduce false positive alarms and increase robustness. The proposed algorithm demonstrated predictive capabilities by detecting heat exchanger fouling five weeks before the plant supervisor noticed and by revealing an anomaly in the oil temperature measure that had gone disregarded. Furthermore, the algorithm demonstrated not to be prone to false positives by not detecting any anomalies during periods of normality.
The MCD layers allow the introduction of stochasticity in the diagnosis process and by performing a certain amount of inferences it is possible to get a quantification of the uncertainty in detecting the anomalies adding a piece of important information to decide if the maintenance intervention is urgent or must be planned in the short future.
A lesson learned from this application is that if there is a future intention to extract a confidence level from the algorithm's predictions, then when using neural networks, it is advisable to size the model with a dropout layer from the outset. This allows the implementation of MCD with minor modifications to the code at a very limited time cost.

Chapter 8

Autoencoder Fine Tuning for Domain Adaptation

This chapter presents a fine-tuning application for anomaly detection in a YANMAR CHP system. The proposed method aims to address the challenge of detecting anomalies, specifically focusing on gas pressure fluctuations that could potentially lead to CHP breakdown. Initially, the autoencoder was trained on 6 weeks of normal operational data to learn the underlying patterns of system behavior. However, the original model's performance led to undesired false positive detections, mainly when applied to new data collected during a different season, approximately three years after the initial training conducted during a summer month. To tackle the issue, the existing model was updated using 1 week of newly sampled data to adapt the model to the system's current behavior. This process allowed the autoencoder to better capture the system's dynamics during the different seasons, reducing false alarms without creating false negatives. Section 8.1 describes the dataset, Section 8.2 explains the fine-tuning process, Section 8.3 reports results, and Section 8.4 draws the conclusions.
[1]

[1]Part of the contents of this chapter has been published as "Anomaly Detection in a Micro-CHP Using Convolutional Autoencoders and Fine Tuning for Domain Adaptation" in *Proc. of IEEE International Conference on System Reliability and Safety (ICSRS), 2023*, [22]

8.1 Dataset

The energy generator studied in this chapter has already been described in Section 6.1. It is a YANMAR micro-cogenerator that is currently installed in a school in Germany to deliver electrical and thermal energy.

8.2 Fine Tuning

This chapter inherited the same anomaly detection routine employed in Chapter 6 and in particular in Section 6.2 where the main character is an autoencoder with convolutional layers (ConvAE) that retain temporal trends. The autoencoder exploited was trained using 6 weeks of data collected in July 2019. Previous work conducted in Chapter 5 had demonstrated that for a 20 kW_e YANMAR CHP, an appropriate amount of data for training is 6 weeks of normal operational conditions.
However, obtaining 6 consecutive weeks of normal data for the specific CHP under consideration posed a challenge. Therefore, it was necessary to retrieve historical data from earlier periods. This introduces two potential issues: firstly, as the years pass, the CHP may undergo natural wear and tear, leading to changes in performance and internal data patterns. Secondly, training the model on data from a particularly hot month may result in a slightly different statistical distribution compared to the data on which inference is performed, especially if the data comes from colder months of the year.
With the assistance of specialized YANMAR technicians, the exhaust gas temperature was identified as a sentinel signal for the phenomenon of temperature reduction during colder months. It was observed that this temperature is significantly influenced by external temperature variations, leading the autoencoder to generate numerous false positives during the colder months.
To address this issue, a specific week of data was selected during which the CHP operates normally, but the autoencoder detects anomalies in signals influenced by external temperature variations. This data week was used to refine the model's weights, aiming to alleviate the problem of false positive detection and improve the accuracy of anomaly identification.
Fig. 8.1 illustrates how the anomaly detection routine predicts signal false positives during the first week of November 2022. In the image, two signals have been chosen for simplicity: one signal influenced by seasonal climate

variations (the exhaust gas temperature) and another signal that remains unaffected by these fluctuations (gas mixer valve position).

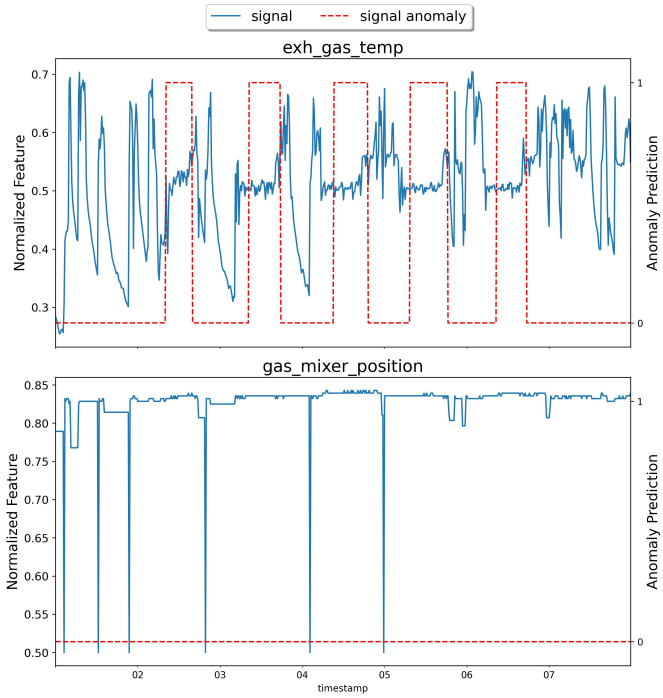

Figure 8.1: Gas mixer valve position and exhaust gas temperature behavior in the week selected for fine-tuning (November 2022).

Since it is not feasible to retrain the model entirely using only one week of data, it was decided to fine-tune the weights, restarting the training process from the optimized weights obtained during the 6-week training in July 2019. The data from the first week of November 2022 was utilized only for this fine-tuning, aiming to address the issue of false positives caused by temperature variations.
Fig. 8.2 presents the evolution of the learning curves throughout the fine-tuning process, showcasing the gradual refinement of weight adjustments accompanied by a gradual reduction in the loss function. The training loss and validation loss display a paired pattern, indicating the absence of both overfitting and underfitting.

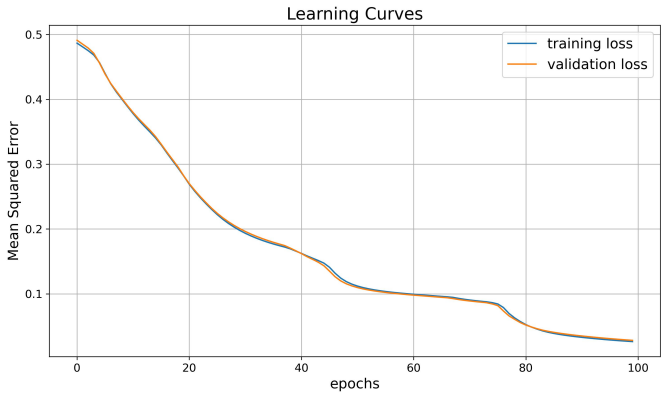

Figure 8.2: Learning curves during ConvAE fine-tuning.

Fig. 8.3 presents the obtained reconstruction errors, demonstrating that the newly calculated anomaly/normal discrimination threshold keeps the 99% of the recent acquisitions below.

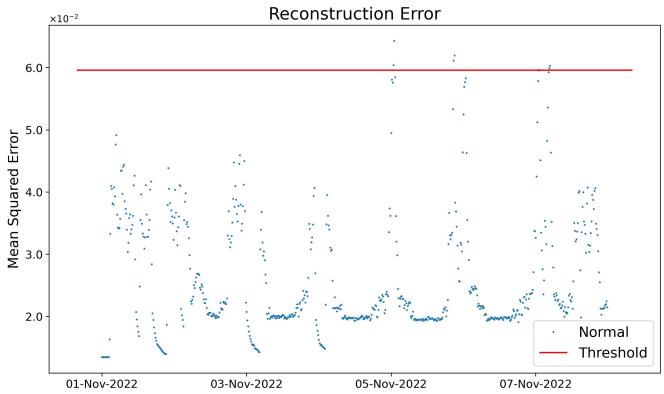

Figure 8.3: Reconstruction errors in the week selected for fine-tuning.

8.3 Results

To demonstrate the effectiveness of the fine-tuning approach, both the original algorithm and the retrained one were tested over a specific time period during which an anomaly occurred. The goal is to show that the numerous false positives observed in signals affected by external conditions can be attenuated without generating false negatives for actual existing anomalies. The period selected for testing spans from the 10^{th} of November 2022 (shortly after the retraining) to the end of April 2023, as detailed in Table 8.1.

	From	**To**	**CHP status**
Train-set	1 Jul 2019	12 Aug 2019	CHP Healthy
Retrain-set	1 Nov 2022	7 Nov 2022	CHP Healthy
Anomalous Test-set	10 Nov 2022	30 Apr 2023	CHP Faulty

Table 8.1: Datasets details: dates and CHP status.

Within this timeframe, two dangerous behaviors of the CHP unit occurred. The first one occurred on the 17^{th} of November 2022 and lasted for 6 days; it went totally unnoticed by both the machine's onboard alarm system and YANMAR technicians. The second one, of a bigger entity, occurred on the 24^{th} January 2023 and was detected by the onboard alarm system only after the CHP shutdown and the consequent partial load operation. These two failures were caused by an abnormal decrease in the plant's gas pressure, affecting only one signal of the CHP, namely, the gas mixer valve position. As a consequence, the CHP, lacking sufficient fuel power, attempted to open the gas valve in an unconventional manner, with the abnormal opening lasting for about 2 to 5 minutes. This anomaly is not visible to the naked eye due to the fact that the sampling time of this CHP is set at 15 minutes.
Fig. 8.4 shows the original algorithm's performance. It can be observed that both the exhaust gas temperature and the rotation speed of the heating circuit pump show many false positives during the periods when no real anomaly is present. In addition, some incorrect system anomalies are predicted after the system reset. On the one hand, one positive aspect is that the original algorithm detects several anomalies related to the opening of the gas mixer valve, which could have alerted YANMAR maintainers and prevented the two dangerous events. However, the presence of numerous false positives on signals that are not related to the valve anomaly (e.g., exhaust gas temperature) may have misled the monitoring personnel. In Fig. 8.5,

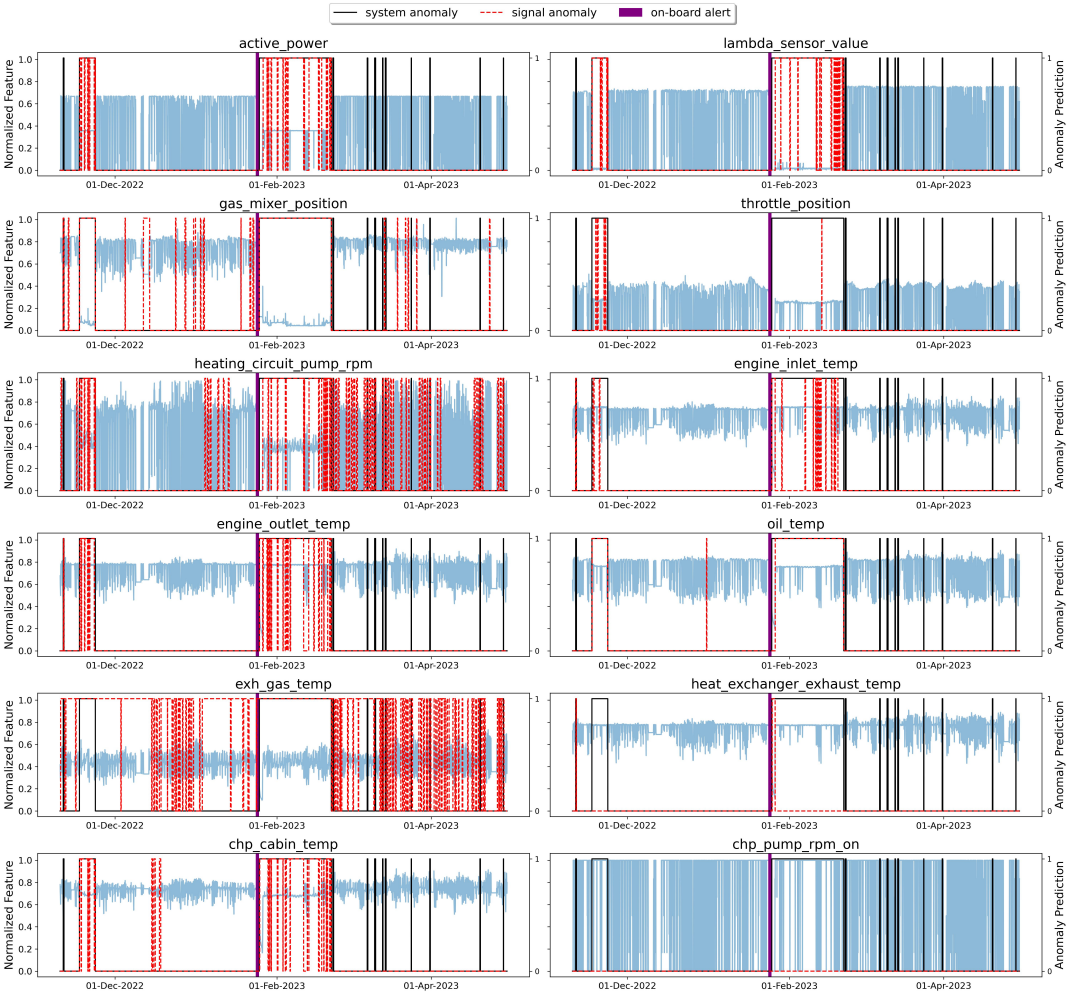

Figure 8.4: Time-series of the normalized features used to test the ConvAE in the presence of plant gas pressure anomaly - Before fine-tuning.

it is evident that the tuned algorithm no longer reports any type of false positive. It successfully generates real alarms for the gas mixer valve, both before the first event, which went unnoticed, and before the second event, which caused a prolonged malfunction.
Indeed, it successfully generates actual alarms only for the gas mixer valve, both before the first event (November 17^{th}, 2022), which went unnoticed, and before the second event (January 24^{th}, 2023), which caused a prolonged malfunction. In summary, both the original and the improved algorithm predicted a failure (signal anomaly) of the cogenerator gas mixing valve in the early morning hours of November 11^{th}, 6 days before the first unobserved failure. The second failure was predicted seventy-four days earlier than the onboard system. In addition, during the incorrect operation of the cogenerator, both algorithms detected a system failure(system anomaly): a symptom of a highly impactful misoperation.
While both algorithms show sensitivity to real anomalies, the fine-tuned one exhibits not only sensitivity but also robustness, enabling it to adapt effectively to domain shifts without compromising detection performance.

8.4 Chapter Conclusions

In conclusion, this work presents a fine-tuning application for anomaly detection in a YANMAR micro-cogeneration system using an autoencoder with 1D-convolutional layers to capture temporal dynamics. The proposed method addresses the challenge of detecting gas pressure fluctuations that could potentially lead to CHP breakdown. The study initially used an autoencoder pre-trained on 6 weeks of normal operational data, but it led to undesired false positive detections when applied to new data collected during a different season, approximately three years after the initial training.
To overcome the issue of false positives, a fine-tuning process was introduced. The existing autoencoder model was fine-tuned using one week of newly sampled data, adapting the model to the system's current behavior. This fine-tuning process successfully enabled the autoencoder to better capture the system's dynamics during different seasons and effectively reduced false alarms.
The results demonstrate the efficacy of the applied method: the fine-tuned autoencoder successfully detected anomalies in gas pressure, potentially leading to CHP malfunction, almost three months in advance with respect to the

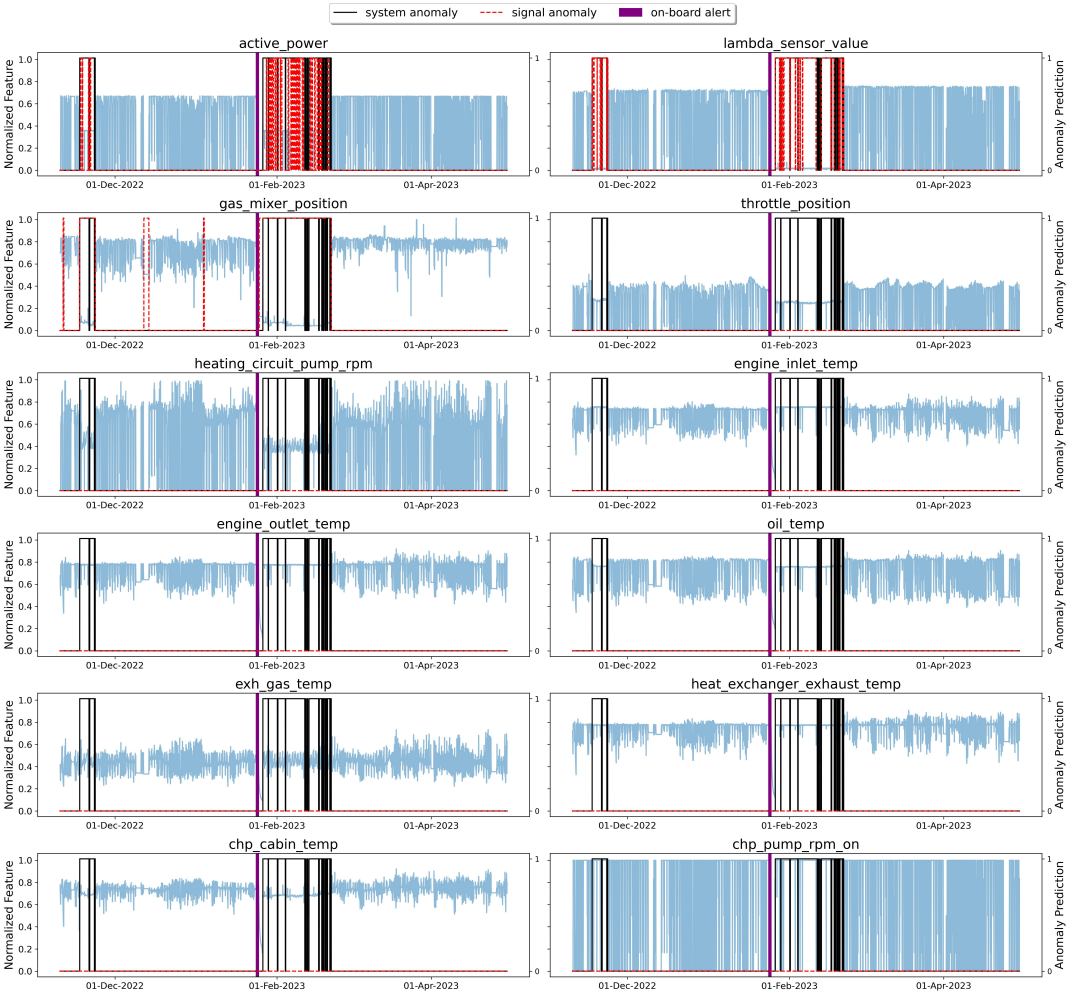

Figure 8.5: Time-series of the normalized features used to test the ConvAE in the presence of plant gas pressure anomaly - After fine-tuning.

currently used system. Additionally, false anomaly detections related to the whole system and irrelevant signals were effectively filtered out by the fine-tuned model.
The testing results comparing the performance of the original algorithm and the retrained one show that the fine-tuned algorithm significantly attenuates false positives without generating false negatives for actual anomalies. Notably, the tuned algorithm successfully identified anomalies in the gas mixer valve position, which could have alerted YANMAR maintainers and prevented a dangerous shutdown.
The study's key insight is the importance of fine-tuning based on recent data to enable the model to adapt to changing conditions and mitigate the impact of seasonality on anomaly detection accuracy. This research has significant implications for enhancing the reliability and maintenance of micro-CHP systems in real-world applications. By improving anomaly detection and reducing false positives, the proposed approach can contribute to better operational efficiency and increased safety in cogeneration systems.
The results achieved in this study lay the groundwork for the development of an anomaly detection service that incorporates continuous retraining to adapt the model to the current engine conditions. Additionally, it ensures the capability to update a model using data from a few days of normal operation rather than an extended period of 6 weeks, which, depending on the applications, can be resource-intensive to collect and analyze. Furthermore, an important next step would involve attempting to apply fine-tuning to different CHP systems. This would enable the initiation of data-driven maintenance strategies, starting from models pre-trained on already installed CHP systems and subsequently refining the model on newly commissioned CHP systems.

Chapter 9

Conclusion

9.1 Summary of contribution

In recent years, significant efforts have been devoted to developing anomaly detection techniques aimed at enhancing maintenance strategies in the energy generation sector. Whereas the focus was traditionally on large-scale generators, the escalating costs of energy and heightened concerns regarding environmental impact and resource conservation have redirected research attention toward moderate-sized machines with relatively affordable acquisition costs.
The development of advanced maintenance techniques offers substantial benefits to manufacturing companies. Firstly, it enables these companies to transition from conventional product-centric entities to service-oriented businesses. This shift allows not only to sell the machinery but also to provide advanced maintenance services. Secondly, it fosters customer loyalty towards a brand that demonstrates reliability, durability, and efficiency in energy production. These advancements mark a significant step toward sustainable, customer-focused energy solutions.
This thesis focuses on the study of a 20 kW_e YANMAR Combined Heat and Power (CHP) system, which is the best-selling model in Europe by YANMAR and is particularly attractive due to its ability to modulate power output. Additionally, its interest is enhanced by the fact that this model is installed at YANMAR's headquarters in Germany, making it even more accessible for internal analyses. However, all methodologies and considerations obtained are extendable and applicable to other sizes and products.

Initially, a study was conducted to develop a generalizable methodology to understand the optimal amount of data required to train an autoencoder to learn the normal behavior of a generator under analysis. To achieve this, data from a CHP unit about which YANMAR had comprehensive and reliable information needed to be obtained. Therefore, the unit satisfying the energetic needs of the facility where YANMAR CHPs are produced was chosen. One year of data, during which the CHP was confirmed to be operating under normal conditions, was acquired. A methodology was developed and applied to these data and later extended to two additional case studies. This approach allowed establishing that, unless some specific operational scenarios happen, a dataset spanning 6 weeks maximizes the performance of the model in anomaly detection tasks for a 20 kW_e CHP. This outcome holds significant advantages, as it allows offering anomaly detection services to customers after just 1 month and a half of collecting normal operation data when installing a new CHP system. Building upon the obtained result, a search was conducted for a 20 kW_e CHP system for which YANMAR could confirm the presence of at least 6 weeks of normal operation data, along with awareness of a past anomaly that had occurred and was subsequently resolved. This specific CHP system was identified in a unit satisfying the energy needs of a school in Germany. A convolutional autoencoder was trained on 6 weeks of normal data from July 2019 and then tested on two separate periods: one during which an anomaly occurred in the heat exchanger and another immediately following a maintenance intervention that restored normal conditions.

The algorithm demonstrated sensitivity to the existing anomaly, detecting it over a month in advance and providing crucial insights for identifying the component responsible for the issue. Furthermore, it exhibited robustness against false positives, as it did not produce any detections in the post-maintenance period. This outcome underscores the algorithm's efficacy in early anomaly detection and its resilience against false alerts, marking a significant step forward in anomaly detection strategies for micro-CHP systems.

In addition to providing alerts related to the detection of a failure, there was an evaluation of incorporating an uncertainty level into the algorithm's output using Bayesian techniques. In this context, a Monte Carlo dropout layer was introduced into the previously used autoencoder, and the architecture was revised to ensure coherent reconstruction results. Subsequently, the autoencoder was tested on the same periods described earlier, demonstrating

that, besides predicting anomalies, the percentage level of anomaly was consistent with observed behavior.
This result paves the way for the development of maintenance routines that minimize unnecessary interventions. Machine downtime or maintenance could be scheduled only when the algorithm produces high certainty levels, optimizing operational efficiency by ensuring actions are taken based on reliable and precise anomaly predictions. Finally, it was observed that over time, the algorithm's robustness degraded due to being trained on data from over two years before, collected during the warmest month of the year. This led to false positives, originating from both a temporal shift causing a natural degradation in performance and lower temperatures in winter months. To address this issue, only 1 week of normal data, including false positives, was used to retrain the autoencoder. During subsequent testing, where a real anomaly occurred in the gas mixer valve, all false positives were filtered out, and the actual anomaly was detected nearly 3 months in advance. The significant advantage of this refinement lies in the fact that adapting a pretrained model now requires only a reduced amount of normal data, easily verifiable by YANMAR personnel, making the retraining process more efficient and adaptable to changing operational conditions.
In the context of this scientific thesis, it is worth noting that a significant challenge in its development, apart from selecting and employing deep learning technologies suitable for solving the described problems, has been dealing with real-world issues and interacting with the various stakeholders involved in managing the problem. Working with real data, as opposed to data originating from well-known benchmark datasets within the academic research community, has added complexity to evaluating methodologies, given the lack of precise knowledge regarding the presence or absence of anomalies. The findings of the research were presented with an emphasis on early anomaly anticipation compared to real field detection rather than relying on comparative tables built upon well-defined metrics prevalent in the existing literature. This approach enabled a detailed study of the topic in real-life situations.
In conclusion, the significant advantage of implementing automatic anomaly detection routines lies not only in the ability to detect hidden anomalies, such as those related to the gas mixer valve but also in identifying issues visible to a maintainer, like those concerning the heat exchanger. Manufacturing companies, such as YANMAR, often employ small maintenance teams

unable to monitor every machine in real-time. Typically, interventions occur only in case of shutdowns, breakdowns, or customer-reported malfunctions. Having reliable and robust alarms can prevent many breakdowns and efficiency losses, saving customers costs while enhancing the manufacturer's reputation. Additionally, it enables the sale of advanced maintenance services, fostering brand trust and loyalty.

9.2 Directions for future work

The anomaly detection routine described in this thesis has been containerized and deployed on local servers, operating in real-time for three 20 kW_e CHP systems. Shortly, it will be extended to cover multiple CHP units, and the routines will be migrated to the cloud environment. Furthermore, the studies conducted will be expanded to machines of different sizes (e.g., 11, 16, and 50 kW_e) and diverse models (e.g., YANMAR Gas Heat Pump).
Significant interest has been generated in the application of domain adaptation techniques and algorithm refinement. This interest derives from the possibility of aligning the algorithm with real-time machine conditions with minimal effort. This concept opens roads for the implementation of transfer learning, allowing the knowledge acquired from one context to be effectively applied to another. Indeed, an aspect currently under evaluation is the feasibility of using a pre-trained model from a particular CHP system for other CHPs of the same size, especially when 6 weeks of normal data are not available. This approach involves refining the model using just one week of data from the new CHP and could significantly simplify the training process, leveraging knowledge from similar systems and adapting it rapidly to specific machinery. By incorporating this methodology, the industry can enhance the efficiency of anomaly detection models and accelerate the implementation of reliable and accurate predictive maintenance solutions, benefiting both manufacturers and end-users.

Appendix A

Publications

This research activity has led to several publications in international journals and conferences. These are summarized below.[1]

International Journals

1. **Piero Danti***, A. Innocenti, S. Sandomier. "Anomaly Detection and Root Cause Analysis using Convolutional Autoencoders: a Real Case Study", *Energy*, under review, 2023. [DOI: 10.2139/ssrn.4574041]
2. **Piero Danti***, A. Innocenti. "A methodology to determine the optimal train-set size for autoencoders applied to energy systems", *Advanced Engineering Informatics*, vol. 58, p. 102139, 2023. [DOI: 10.1016/j.aei.2023.102139]

International Conferences and Workshops

1. **Piero Danti***, A. Innocenti. "Anomaly Detection in a Micro-CHP Using Convolutional Autoencoders and Fine Tuning for Domain Adaptation", in *Proceedings of the 7th IEEE International Conference on System Reliability and Safety (ICSRS 2023)*, Bologna (Italy), 2023.
2. **Piero Danti***, A. Innocenti. "Applied Anomaly Detection: a Bayesian Approach to Improve Robustness", in *Proceedings of the 3rd IEEE International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME 2023)*, Tenerife (Spain), 2023.
3. M. Latinov*, N. Fiorini, G. Vichi, A. Innocenti, **Piero Danti**. "A Machine Learning Approach for Engine Model-Based Control on NOx Emissions",

[1] The author's bibliometric indices are the following: H-index = 5, total number of citations = 59 (source: Google Scholar on Month 01, 2024).

in *Proceedings of Powertrains, Energy and Lubricants International Meeting (JSAE/SAE 2023)*, Kyoto (Japan), 2023. **(Best paper award)**

4. **Piero Danti***, G. Vichi, R. Minamino. "Wrong Injection Detection in a Small Diesel Engine, a Machine Learning Approach", in *Proceedings of the 7th European Conference of the Prognostics and Health Management Society (PHM 2022)*, Turin (Italy), 2022.
5. C. Giola*, **Piero Danti***, S. Magnani. "Learning Curves: A Novel Approach for Robustness Improvement of Load Forecasting", in *Proceedings of The 7th International Conference on Time Series and Forecasting (ITISE 2021)*, Gran Canaria (Spain), 2021.

Bibliography

[1] M. Abdar *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, Dec. 2021.

[2] R. Ahmad and S. Kamaruddin, "An overview of time-based and condition-based maintenance in industrial application," *Computers & Industrial Engineering*, pp. 135 – 149, 2012.

[3] A. L. Alfeo, M. G. Cimino, G. Manco, E. Ritacco, and G. Vaglini, "Using an autoencoder in the design of an anomaly detector for smart manufacturing," *Pattern Recognition Letters*, vol. 136, pp. 272–278, Aug. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167865520302269

[4] E. Anderlini, G. Salavasidis, C. A. Harris, P. Wu, A. Lorenzo, A. B. Phillips, and G. Thomas, "A remote anomaly detection system for Slocum underwater gliders," *Ocean Engineering*, vol. 236, p. 109531, Sep. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801821009240

[5] D. B. Araya, K. Grolinger, H. F. ElYamany, M. A. Capretz, and G. Bitsuamlak, "An ensemble learning framework for anomaly detection in building energy consumption," *Energy and Buildings*, vol. 144, pp. 191–206, Jun. 2017.

[6] H. Bahlawan, M. Morini, M. Pinelli, P. Spina, and M. Venturini, "Development of reliable narx models of gas turbine cold, warm and hot start-up," *Proceedings of the ASME Turbo Expo*, vol. 9, 2017.

[7] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018.

[8] A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, M. Rampazzo, and F. Simmini, "Data-driven fault detection and diagnosis for hvac water chillers," *Control Engineering Practice*, vol. 53, pp. 79–91, Aug. 2016.

[9] I. Bellanco, E. Fuentes, M. Vallès, and J. Salom, "A review of the fault behavior of heat pumps and measurements, detection and diagnosis meth-

ods including virtual sensors," *Journal of Building Engineering*, vol. 39, p. 102254, 2021.

[10] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," Sep. 2012.

[11] A. Borghesi, A. Bartolini, M. Lombardi, M. Milano, and L. Benini, "A semisupervised autoencoder-based approach for anomaly detection in high performance computing systems," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 634–644, Oct. 2019.

[12] A. Borghesi, M. Molan, M. Milano, and A. Bartolini, "Anomaly Detection and Anticipation in High Performance Computing Systems," *IEEE Trans. Parallel Distrib. Syst.*, pp. 1–1, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9439169/

[13] A. Capozzoli, M. S. Piscitelli, S. Brandi, D. Grassi, and G. Chicco, "Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings," *Energy*, vol. 157, pp. 336–352, Aug. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544218309617

[14] M. Carletti, C. Masiero, A. Beghi, and G. A. Susto, "A deep learning approach for anomaly detection with industrial time series data: a refrigerators manufacturing case study," *Procedia Manufacturing*, vol. 38, pp. 233–240, 2019.

[15] M. Castangia, R. Sappa, A. A. Girmay, C. Camarda, E. Macii, and E. Patti, "Anomaly detection on household appliances based on variational autoencoders," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100823, Dec. 2022.

[16] V. Cerqueira, L. Torgo, and I. Mozetic, "Evaluating time series forecasting models: An empirical study on performance estimation methods," *Mach Learn*, vol. 109, pp. 1997–2028, 2020.

[17] R.-Q. Chen, G.-H. Shi, W.-L. Zhao, and C.-H. Liang, "A Joint Model for IT Operation Series Prediction and Anomaly Detection," *Neurocomputing*, vol. 448, pp. 130–139, Aug. 2021, arXiv: 1910.03818. [Online]. Available: http://arxiv.org/abs/1910.03818

[18] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of Computation*, pp. 297–301, 1965.

[19] A. Copiaco, Y. Himeur, A. Amira, W. Mansoor, F. Fadli, S. Atalla, and S. S. Sohail, "An innovative deep anomaly detection of building energy consumption using energy time-series images," *Engineering Applications of Artificial Intelligence*, vol. 119, p. 105775, Mar. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197622007655

[20] A. Cossu, A. Carta, V. Lomonaco, and D. Bacciu, "Continual learning for recurrent neural networks: An empirical evaluation," *Neural Networks*, vol. 143, pp. 607–627, Nov. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608021002847

[21] M. Cozzatti, F. Simonetta, and S. Ntalampiras, "Variational autoencoders for anomaly detection in respiratory sounds," *arXiv*, 2022, accessed: Mar. 30, 2023. [Online]. Available: http://arxiv.org/abs/2208.03326.

[22] P. Danti and A. Innocenti, "Anomaly detection in a micro-chp using convolutional autoencoders and fine tuning for domain adaptation," in *The 7th International Conference on System Reliability and Safety*. IEEE, 2023, iN PRESS.

[23] ——, "Applied Anomaly Detection: a Bayesian Approach to Improve Robustness," in *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. Tenerife, Canary Islands, Spain: IEEE, Jul. 2023, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/10252975/

[24] ——, "A methodology to determine the optimal train-set size for autoencoders applied to energy systems," *Advanced Engineering Informatics*, vol. 58, p. 102139, Oct. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474034623002677

[25] P. Danti, A. Innocenti, and S. Sandomier, "Anomaly Detection and Root Cause Analysis Using Convolutional Autoencoders: A Real Case Study," SSRN, preprint, 2023. [Online]. Available: https://www.ssrn.com/abstract=4574041

[26] H. L. Dawson, O. Dubrule, and C. M. John, "Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification," *Computers & Geosciences*, vol. 171, p. 105284, Feb. 2023.

[27] Z. Deng, Y. Li, H. Zhu, K. Huang, Z. Tang, and Z. Wang, "Sparse stacked autoencoder network for complex system monitoring with industrial applications," *Chaos, Solitons & Fractals*, vol. 137, p. 109838, Aug. 2020.

[28] C. Ding, Z. Wang, Q. Ding, and Z. Yuan, "Convolutional neural network based on fast fourier transform and gramian angle field for fault identification of hvdc transmission line," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100888, Dec. 2022.

[29] X. Ding, Y. Li, A. Belatreche, and L. P. Maguire, "An experimental evaluation of novelty detection methods," *Neurocomputing*, vol. 135, pp. 313–327, Jul. 2014.

[30] I. M. Dupont, P. C. M. Carvalho, S. C. S. Jucá, and J. S. P. Neto, "Novel methodology for detecting non-ideal operating conditions for

grid-connected photovoltaic plants using Internet of Things architecture," *Energy Conversion and Management*, vol. 200, p. 112078, Nov. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890419310842

[31] A. El-Nouby, G. Izacard, H. Touvron, I. Laptev, H. Jegou, and E. Grave, "Are large-scale datasets necessary for self-supervised pre-training?" Dec. 2021, arXiv:2112.10740 [cs].

[32] C. Fan, F. Xiao, Y. Zhao, and J. Wang, "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data," *Applied Energy*, vol. 211, pp. 1123–1135, Feb. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261917317166

[33] C. Feng, C. Liu, and D. Jiang, "Unsupervised anomaly detection using graph neural networks integrated with physical-statistical feature fusion and local-global learning," *Renewable Energy*, vol. 206, pp. 309–323, Apr. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148123001970

[34] X. Fu and P. A. Bates, "Application of deep learning methods: From molecular modelling to patient classification," *Experimental Cell Research*, vol. 418, p. 113278, 2022.

[35] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," *arXiv*, Oct. 2016, accessed: Mar. 30, 2023. [Online]. Available: http://arxiv.org/abs/1506.02142.

[36] Z. Ghafoori, S. M. Erfani, S. Rajasegarar, J. C. Bezdek, S. Karunasekera, and C. Leckie, "Efficient Unsupervised Parameter Estimation for One-Class Support Vector Machines," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 10, pp. 5057–5070, Oct. 2018.

[37] C. Giola, P. Danti, and S. Magnani, "Learning curves: A novel approach for robustness improvement of load forecasting," in *The 7th International conference on Time Series and Forecasting*. MDPI, 2021, p. 38.

[38] A. Glowacz and Z. Glowacz, "Diagnosis of stator faults of the single-phase induction motor using acoustic signals," *Applied Acoustics*, vol. 117, pp. 20–27, Feb. 2017.

[39] M. Gokhale, S. K. Mohanty, and A. Ojha, "A stacked autoencoder based gene selection and cancer classification framework," *Biomedical Signal Processing and Control*, vol. 78, p. 103999, 2022.

[40] A. González-Muñiz, I. Díaz, A. A. Cuadrado, D. García-Pérez, and D. Pérez, "Two-step residual-error based approach for anomaly detection in engineer-

ing systems using variational autoencoders," *Computers and Electrical Engineering*, vol. 101, p. 108065, Jul. 2022.

[41] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[42] V. Gori, G. Veneri, and V. Ballarini, "Continual Learning for anomaly detection on turbomachinery prototypes - A real application," in *2022 IEEE Congress on Evolutionary Computation (CEC)*. Padua, Italy: IEEE, Jul. 2022, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/9870234/

[43] F. F. Gulamali, A. S. Sawant, P. Kovatch, B. Glicksberg, A. Charney, G. N. Nadkarni, and E. Oermann, "Autoencoders for sample size estimation for fully connected neural network classifiers," *npj Digital Medicine*, vol. 5, no. 1, pp. 1–8, Dec. 2022.

[44] J. Gütter, A. Kruspe, X. Zhu, and J. Niebling, "Impact of training set size on the ability of deep neural networks to deal with omission noise," *Frontiers in Remote Sensing*, vol. 3, Jul. 2022.

[45] R. Hadsell, D. Rao, A. A. Rusu, and R. Pascanu, "Embracing Change: Continual Learning in Deep Neural Networks," *Trends in Cognitive Sciences*, vol. 24, no. 12, pp. 1028–1040, Dec. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364661320302199

[46] H. Han, X. Cui, Y. Fan, and H. Qing, "Least squares support vector machine (ls-svm)-based chiller fault diagnosis using fault indicative features," *Applied Thermal Engineering*, vol. 154, pp. 540–547, May 2019.

[47] F. Harrou, A. Dairi, B. Taghezouit, and Y. Sun, "An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine," *Solar Energy*, vol. 179, pp. 48–58, Feb. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X18312209

[48] Z. He, S. Deng, and X. Xu, "An optimization model for outlier detection in categorical data," in *Advances in Intelligent Computing*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, pp. 400–409.

[49] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, "Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives," *Applied Energy*, vol. 287, p. 116601, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261921001409

[50] Y. Jiang and C. Zhao, "Attention classification-and-segmentation network for micro-crack anomaly detection of photovoltaic module cells," *Solar*

Energy, vol. 238, pp. 291–304, May 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X22002602

[51] E. Keogh, J. Lin, S.-H. Lee, and H. V. Herle, "Finding the most unusual time series subsequence: algorithms and applications," *Knowledge and Information Systems*, vol. 11, no. 1, pp. 1–27, Jan. 2007.

[52] T. Kieu, B. Yang, C. Guo, and C. S. Jensen, "Outlier detection for time series with recurrent autoencoder ensembles," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 7 2019, pp. 2725–2732. [Online]. Available: https://doi.org/10.24963/ijcai.2019/378

[53] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv:1312.6114 [cs, stat]*, May 2014, accessed: Mar. 17, 2022. [Online]. Available: http://arxiv.org/abs/1312.6114.

[54] ——, "An introduction to variational autoencoders," *FNT in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.

[55] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, 2021.

[56] A. Korovin, A. Vasilev, F. Egorov, D. Saykin, E. Terukov, I. Shakhray, L. Zhukov, and S. Budennyy, "Anomaly detection in electroluminescence images of heterojunction solar cells," *Solar Energy*, vol. 259, pp. 130–136, Jul. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X23003043

[57] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. IEEE, 2010, pp. 253–256.

[58] D. Lee, M.-H. Chen, and G.-W. Lai, "Achieving energy savings through artificial-intelligence-assisted fault detection and diagnosis: Case study on refrigeration systems," *Case Studies in Thermal Engineering*, vol. 40, p. 102499, 2022.

[59] A. Legrand, H. Trannois, and A. Cournier, "Use of uncertainty with autoencoder neural networks for anomaly detection," in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, Sardinia, Italy, 2019, pp. 32–35.

[60] J. Li, H. Izakian, W. Pedrycz, and I. Jamal, "Clustering-based anomaly detection in multivariate time series data," *Applied Soft Computing*, vol. 100, p. 106919, Mar. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1568494620308577

[61] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, and L. Li, "Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis," *Measurement*, vol. 46, no. 1, pp. 259–271, Jan. 2013.

[62] Y. Lin and J. Wang, "Probabilistic Deep Autoencoder for Power System Measurement Outlier Detection and Reconstruction," *IEEE Trans. Smart Grid*, vol. 11, pp. 1796–1798, Mar. 2020.

[63] F. Liu, C. Zeng, L. Zhang, Y. Zhou, Q. Mu, Y. Zhang, L. Zhang, and C. Zhu, "Fedtadbench: Federated time-series anomaly detection benchmark," Dec. 2022.

[64] Y. Liu, Y. Zhou, K. Yang, and X. Wang, "Unsupervised deep learning for iot time series," *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[65] W. H. Lopez Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, "Autoencoders," in *Machine Learning*. Elsevier, 2020, pp. 193–208.

[66] V. Losing, B. Hammer, and H. Wersing, "Incremental on-line learning: A review and comparison of state of the art algorithms," *Neurocomputing*, vol. 275, pp. 1261–1274, Jan. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231217315928

[67] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377–388, Jan. 2017.

[68] Q. Luo, J. Chen, Y. Zi, Y. Chang, and Y. Feng, "Multi-mode non-Gaussian variational autoencoder network with missing sources for anomaly detection of complex electromechanical equipment," *ISA Transactions*, p. S0019057822004669, 2022.

[69] S. Maleki, S. Maleki, and N. R. Jennings, "Unsupervised anomaly detection with lstm autoencoders using statistical data-filtering," *Applied Soft Computing*, vol. 108, p. 107443, Sep. 2021.

[70] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Lstm-based encoder-decoder for multi-sensor anomaly detection," *arXiv:1607.00148 [cs, stat]*, 2016.

[71] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," Apr. 2018.

[72] T. Matsui, K. Yamamoto, and J. Ogata, "Anomaly detection for wind turbine damaged due to lightning strike," *Electric Power Systems Research*, vol. 209, p. 107918, Aug. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779622001481

[73] D. Miljkovic, "Review of novelty detection methods," in *The 33rd International Convention MIPRO*, 2010, pp. 593–598.

[74] K. Müller, J. Ooms, D. James, S. DebRoy, H. Wickham, and J. Horner, *RMariaDB: Database Interface and MariaDB Driver*, 2022, https://rmariadb.r-dbi.org, https://github.com/r-dbi/RMariaDB, https://downloads.mariadb.org/connector-c/.

[75] A. Netti, Z. Kiziltan, O. Babaoglu, A. Sîrbu, A. Bartolini, and A. Borghesi, "A machine learning approach to online fault classification in HPC systems," *Future Generation Computer Systems*, vol. 110, pp. 1009–1022, Sep. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167739X1932045X

[76] A. Ng *et al.*, "Sparse autoencoder," *CS294A Lecture notes*, vol. 72, no. 2011, pp. 1–19, 2011.

[77] H. Nguyen, K. Tran, S. Thomassey, and M. Hamad, "Forecasting and anomaly detection approaches using lstm and lstm autoencoder techniques with the applications in supply chain management," *International Journal of Information Management*, vol. 57, p. 102282, Apr. 2021.

[78] Z. Noumir, P. Honeine, and C. Richard, *On simple one-class classification methods*. IEEE, Jul. 2012.

[79] A. Oluwasegun and J.-C. Jung, "A multivariate gaussian mixture model for anomaly detection in transient current signature of control element drive mechanism," *Nuclear Engineering and Design*, vol. 402, p. 112098, Feb. 2023.

[80] U. Otamendi, I. Martinez, M. Quartulli, I. G. Olaizola, E. Viles, and W. Cambarau, "Segmentation of cell-level anomalies in electroluminescence images of photovoltaic modules," *Solar Energy*, vol. 220, pp. 914–926, May 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X21002462

[81] X. Ou, G. Wen, X. Huang, Y. Su, X. Chen, and H. Lin, "A deep sequence multi-distribution adversarial model for bearing abnormal condition detection," *Measurement*, vol. 182, p. 109529, Sep. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0263224121005091

[82] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural Networks*, vol. 113, pp. 54–71, May 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608019300231

[83] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, Jun. 2014.

[84] J. Qian, Z. Song, Y. Yao, Z. Zhu, and X. Zhang, "A review on autoencoder based representation learning for fault detection and diagnosis in industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 231, p. 104711, Dec. 2022.

[85] C. Qu, Z. Zhou, Z. Liu, and S. Jia, "Predictive anomaly detection for marine diesel engine based on echo state network and autoencoder," *Energy Reports*, vol. 8, pp. 998–1003, 2022.

[86] Z. Qu, H. Liu, Z. Wang, J. Xu, P. Zhang, and H. Zeng, "A combined genetic optimization with AdaBoost ensemble model for anomaly detection in buildings electricity consumption," *Energy and Buildings*, vol. 248, p. 111193, Oct. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778821004771

[87] P. M. Radiuk, "Impact of training set batch size on the performance of convolutional neural networks for diverse datasets," *Information Technology and Management Science*, vol. 20, no. 1, Jan. 2017.

[88] J. Saari, D. Strömbergsson, J. Lundberg, and A. Thomson, "Detection and identification of windmill bearing faults using a one-class support vector machine (svm)," *Measurement*, vol. 137, pp. 287–301, Apr. 2019.

[89] M. A. M. Sadr, Y. Zhu, and P. Hu, "An anomaly detection method for satellites using monte carlo dropout," *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1–9, 2022.

[90] M. S. Safizadeh and S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Information Fusion*, vol. 18, pp. 1–8, Jul. 2014.

[91] P. Sarajcev, A. Kunac, G. Petrovic, and M. Despalatovic, "Power System Transient Stability Assessment Using Stacked Autoencoder and Voting Ensemble," *Energies*, vol. 14, p. 3148, Jan. 2021.

[92] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support Vector Method for Novelty Detection," in *Advances in Neural Information Processing Systems*, vol. 12. MIT Press, 1999. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/1999/hash/8725fb777f25776ffa9076e44fcfd776-Abstract.html

[93] X. Shu, T. Bao, R. Xu, Y. Li, and K. Zhang, "Dam anomaly assessment based on sequential variational autoencoder and evidence theory," *Applied Mathematical Modelling*, vol. 98, pp. 576–594, Oct. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0307904X21002778

[94] S. Skansi, *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*, ser. Undergraduate Topics in Computer Science. Springer International Publishing, 2018.

[95] M. Sohaib, C.-H. Kim, and J.-M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, no. 12, p. 2876, Dec. 2017.

[96] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, Jun. 2014.

[97] C. Sun, Z. He, H. Lin, L. Cai, H. Cai, and M. Gao, "Anomaly detection of power battery pack using gated recurrent units based variational autoencoder," *Applied Soft Computing*, vol. 132, p. 109903, Jan. 2023.

[98] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 185–195, 2018.

[99] O. Surucu, S. A. Gadsden, and J. Yawney, "Condition monitoring using machine learning: A review of theory, applications, and recent advances," *Expert Systems with Applications*, vol. 221, p. 119738, Jul. 2023.

[100] Y. Tan, C. Niu, H. Tian, L. Hou, and J. Zhang, "A one-class SVM based approach for condition-based maintenance of a naval propulsion plant with limited labeled data," *Ocean Engineering*, vol. 193, p. 106592, Dec. 2019.

[101] M. Thill, W. Konen, H. Wang, and T. Bäck, "Temporal convolutional autoencoder for unsupervised anomaly detection in time series," *Applied Soft Computing*, vol. 112, p. 107751, 2021.

[102] J. Urmeneta, J. Izquierdo, and U. Leturiondo, "A methodology for performance assessment at system level—Identification of operating regimes and anomaly detection in wind turbines," *Renewable Energy*, vol. 205, pp. 281–292, Mar. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148123000423

[103] J. von Schleinitz, M. Graf, W. Trutschnig, and A. Schröder, "VASP: An autoencoder-based approach for multivariate anomaly detection and robust time series prediction with application in motorsport," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104354, Sep. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0952197621002025

[104] S. N. Wanasundara, A. Wickramasinghe, M. Schaubroeck, and S. Muthukumarana, "Detecting thermal anomalies in buildings using frequency and temporal domains analysis," *Journal of Building Engineering*, vol. 75, p. 106923, Sep. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352710223011026

[105] W. Xiaolan, M. Manjur Ahmed, M. Nizam Husen, Z. Qian, and S. B. Belhaouari, "Evolving anomaly detection for network streaming data,"

Information Sciences, vol. 608, pp. 757–777, Aug. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025522006582

[106] F. Xu and L. Wang, "Constructing a health indicator for bearing degradation assessment via an unsupervised and enhanced stacked autoencoder," *Advanced Engineering Informatics*, vol. 53, p. 101708, 2022.

[107] S. Yan, H. Shao, Y. Xiao, B. Liu, and J. Wan, "Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises," *Robotics and Computer-Integrated Manufacturing*, vol. 79, p. 102441, 2023.

[108] *neoTower® 20.0 datasheet*, YANMAR Energy System Europe, 2022, https://www.rmbenergie.com/en/neotower/#nt20.0en.

[109] B. Yao, P. Zhen, L. Wu, and Y. Guan, "Rolling Element Bearing Fault Diagnosis Using Improved Manifold Learning," *IEEE Access*, vol. 5, pp. 6027–6035, 2017.

[110] D. Yu, J. Yu, F. Sun, Y. Deng, Q. Wu, and G. Cong, "Research on the pca-based intelligent fault detection methodology for sewage source heat pump system," *Procedia Engineering*, vol. 205, pp. 1064–1071, 2017.

[111] W. Yu, I. Y. Kim, and C. Mechefske, "Analysis of different RNN autoencoder variants for time series classification and machine prognostics," *Mechanical Systems and Signal Processing*, vol. 149, p. 107322, Feb. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0888327020307081

[112] C. Zhang, J. Lu, and Y. Zhao, "Generative pre-trained transformers (GPT)-based automated data mining for building energy management: Advantages, limitations and the future," *Energy and Built Environment*, Jun. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666123323000521

[113] C. Zhang, D. Hu, and T. Yang, "Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost," *Reliability Engineering & System Safety*, vol. 222, p. 108445, 2022.

[114] S.-s. Zhang, J.-w. Liu, and X. Zuo, "Adaptive online incremental learning for evolving data streams," *Applied Soft Computing*, vol. 105, p. 107255, Jul. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1568494621001782

[115] J. Zheng, J. Du, Y. Liang, Q. Liao, Z. Li, H. Zhang, and Y. Wu, "Deeppipe: A semi-supervised learning for operating condition recognition of multi-product pipelines," *Process Safety and Environmental Protection*, vol. 150, pp. 510–521, Jun. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957582021002172

[116] F. Zhou, Y. Gao, and C. Wen, "A Novel Multimode Fault Classification Method Based on Deep Learning," *Journal of Control Science and Engineering*, vol. 2017, p. e3583610, Mar. 2017.

[117] H. Zhou, K. Yu, X. Zhang, G. Wu, and A. Yazidi, "Contrastive autoencoder for anomaly detection in multivariate time series," *Information Sciences*, vol. 610, pp. 266–280, 2022.

[118] Y. Zhou, X. Song, Y. Zhang, F. Liu, C. Zhu, and L. Liu, "Feature encoding with autoencoders for weakly-supervised anomaly detection," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 33, pp. 2454–2465, Jun. 2022.