Vaasan yliopisto
UNIVERSITY OF VAASA

Hasibur Rahman

Architecture-Level Fusion of SCADA and ERA5 Reanalysis Data for Enhanced Winter Fault Detection in Offshore Wind Turbines

Master Thesis

School of Technology and Innovation
Master of Science in Technology
Master's Program in Sustainable and Autonomous Systems

Vaasa 2026

UNIVERSITY OF VAASA
School of Technology and Innovation
Author: Hasibur Rahman
Title of the thesis: Architecture-Level Fusion of SCADA and ERA5 Reanalysis Data for Enhanced Winter Fault Detection in Offshore Wind Turbines Master Thesis
Degree: Master of Science in Technology
Degree Programme: Master’s Program in Sustainable and Autonomous Systems
Supervisor: Mazaher Karimi
Instructor: Hossein Naderian
Year: 2026 **Pages:** 108

ABSTRACT:

Wind energy is a crucial part of the global shift towards low-carbon energy systems; however, the economic uncertainties of offshore wind energy remain driven by high offshore operation and maintenance (O&M) costs in harsh environments. Predictive maintenance is vital in offshore operations to minimize downtime and avoid catastrophic failures. But current techniques are mainly based on supervisory control and data acquisition (SCADA) data, which is susceptible to environmental effects and can become unreliable in winter.

This thesis assesses the benefit of incorporating weather data for anomaly detection in offshore wind turbines during winter conditions. Using the CARE to Compare benchmark dataset, a three-step approach is proposed. First, a SCADA-based one-dimensional neural network (1D-CNN) is reproduced to serve as a baseline. A data-level fusion model is then used to increase SCADA inputs using meteorological information produced by ERA5. Lastly, an architecture-level fusion model is developed, which relies on two parallel branches with a global gating mechanism to control the weight of environmental information.

The performance of the models is evaluated on the full test set and the winter subset (December-February). On the winter subset, the SCADA-only baseline has a high recall (0.9500) and low precision (0.6472), which produces a significant number of false alarms. Under the same winter evaluation, data-level fusion also enhances precision (0.8126) and reduces recall (0.8249), so fusion is more conservative in its detection behaviour. The proposed architecture level fusion model further achieves optimum performance with a precision of 0.7995, a recall rate of 0.9234 and the highest F1-score of 0.8570 and reduces false positives by over 55 percent as compared to the baseline. These results show how weather context can improve predictive maintenance in winter and can be integrated in effective ways to achieve better predictive maintenance performance. In terms of operational implications, lowering false alarms facilitates maintenance scheduling by reducing wastage of resources on false alarms. Finally, the research exports the model's detection of winter abnormalities as data inputs to an Operations Research optimization model, bridging the gap between predictive maintenance and maintenance operations in offshore wind farms.

KEYWORDS: Anomaly Detection, Deep Learning, Data Fusion, ERA5 Reanalysis, Offshore Wind Turbines, Predictive Maintenance, SCADA Data, Winter Conditions

Contents

Figures

Tables

Abbreviations

1D-CNN	One-Dimensional Convolutional Neural Network
AI	Artificial Intelligence
ANN	Artificial Neural Network
CM	Confusion Matrix
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
DL	Deep Learning
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	Fifth-generation ECMWF Atmospheric Reanalysis
EWMA	Exponentially Weighted Moving Average
FN	False Negative
FP	False Positive
LCOE	Levelised Cost of Energy
LSTM	Long Short-Term Memory
MAD	Median Absolute Deviation
MILP	Mixed-Integer Linear Programming
ML	Machine Learning
MonVal	Monitoring Validation
NBM	Normal Behaviour Model
O&M	Operation and Maintenance

PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SCADA	Supervisory Control and Data Acquisition
TCN	Temporal Convolutional Network
TN	True Negative
TP	True Positive
UTC	Coordinated Universal Time

1 Introduction

Offshore wind energy is the most crucial as the world moves towards the use of renewable energy sources which are scalable and sustainable in the face of the increasing energy demands as the world tackles the issue of global warming. Offshore turbines have an advantage over onshore wind farms because the wind speeds are higher and more consistent and they have a greater potential to generate energy (Bilgili & Alphan, 2022). Nevertheless, these benefits are offset by much higher operation and maintenance (O&M) expenses that still are a major challenge to the large-scale deployment of offshore wind energy. Complexity of maintaining turbines in the harshness of marine environments coupled with difficulties in access and seasonal uncertainty of weather conditions are one of the main causes of high offshore O&M costs (Fox et al., 2022; Wrzask et al., 2025).

Offshore wind turbines are located in areas that cannot be reached easily, unlike onshore turbine. The maintenance activities need special vessels and trained professionals to carry out maintenance tasks (Hassan, 2025). In addition, most maintenance tasks in these locations would always be delayed due to high wind speeds, large wave heights, low temperatures, as well as short daylight during the winter months (Dowell et al., 2013). Such limitations result in long downtimes and huge losses in production unless the faults are identified and dealt with in a timely manner. Efficient fault detection and predictive maintenance are, therefore, not only necessary to reduce operational interruptions but also to ensure that the offshore wind farms remain economically viable, especially during the winter, when there is limited time to carry out maintenance.

1.1 Background

The shift to renewable energy, especially offshore wind power, is key to climate mitigation and energy security goals. However, the technical and logistical issues with offshore turbines bring serious technical and logistical challenges. The constant exposure to a harsh environment in the sea causes corrosion, extreme weather effects, and greater degradation in the systems (Wang et al., 2022; Wrzask et al., 2025). **Figure 1** shows the usual infrastructure and transmission arrangement of an offshore wind farm.

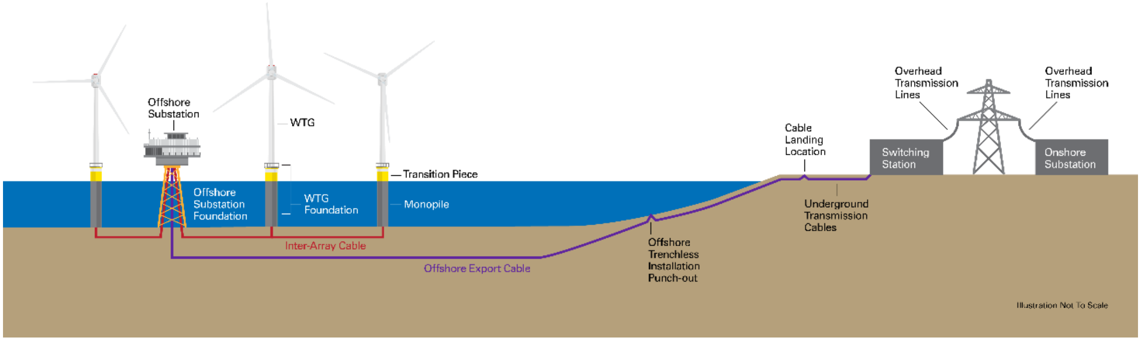

Figure 1. Offshore wind energy infrastructure, Source: [Offshore Wind Power]

Offshore maintenance is always complicated in nature. In many cases, the turbines are located many miles out in the water, which necessitates well-coordinated logistics of vessels, personnel, and favourable weather conditions to transport. These limitations augment the risk in operations as well as expenses of maintenance. Disruption that are associated with weather especially in winter makes it even more difficult to detect and fix the fault in time. This has made predictive maintenance a vital approach of enhancing the reliability of turbines by detecting possible failures before they turn into costly failures (Fox et al., 2022). The data-oriented methods have been specifically promising in minimization of downtime and improvement of system performance.

1.2 Research Motivation

Although the predictive maintenance methods have advanced, offshore wind operation processes are still very sensitive to environmental changes. Specifically, the winter conditions impose more uncertainty on the behaviour of turbines and the decisions regarding maintenance. In these circumstances, it becomes more challenging to draw a line separating the anomalies that are caused by fault and those that are environmentally induced (Tautz-Weinert & Watson, 2017; Wrzask et al., 2025).

From an operational perspective, inaccurate fault detection has direct economic implications. False alarms can cause unnecessary planning of its maintenance and resource allocation, whereas missed detects may cause severe component failure and intensive downtime (Fox et al., 2022). Since offshore interventions are expensive, even modest advances in detectability can place the process on a more efficient trajectory, thus lowering overall O&M costs (Turnbull & Carroll, 2021).

New possibilities to mitigate these challenges are offered by recent progress in machine learning and data-driven modeling. Specifically, the combination of environmental data and turbine-functional data presents the opportunity to enhance the contextual interpretation of turbine behavior. This encourages an organized inquiry on whether meteorological data can facilitate predictive maintenance operations, particularly in seasonal extreme weather conditions.

1.3 Predictive Maintenance in Offshore Wind

Predictive maintenance has proven to be one of the main solutions to reduce downtime in the operational process in offshore wind turbines and achieve efficiency in operations. This method utilizes the power of modern data analytics (such as machine learning and artificial intelligence) to identify which faults might occur before they happen. Predictive maintenance helps operators to plan interventions in advance and synchronize the work done on maintenance with good weather conditions (Maron et al., 2022).

The majority of predictive maintenance models are based on the data obtained using Supervisory Control and Data Acquisition (SCADA) systems, which monitor operational parameters of a turbine, including power production, rotor speed, temperatures, and control signals (Tautz-Weinert & Watson, 2017). These high-frequency measurements are the foundations of the data-driven anomalies detection. The deep learning models, especially Convolutional Neural Networks (CNNs), have shown high potential at predicting temporal dependencies in SCADA data and detecting early fault indicators (Ulmer et al., 2020; H. Zhao et al., 2018). The general flow from turbine sensing and data storage to fault diagnosis and maintenance planning is illustrated in **Figure 2**.

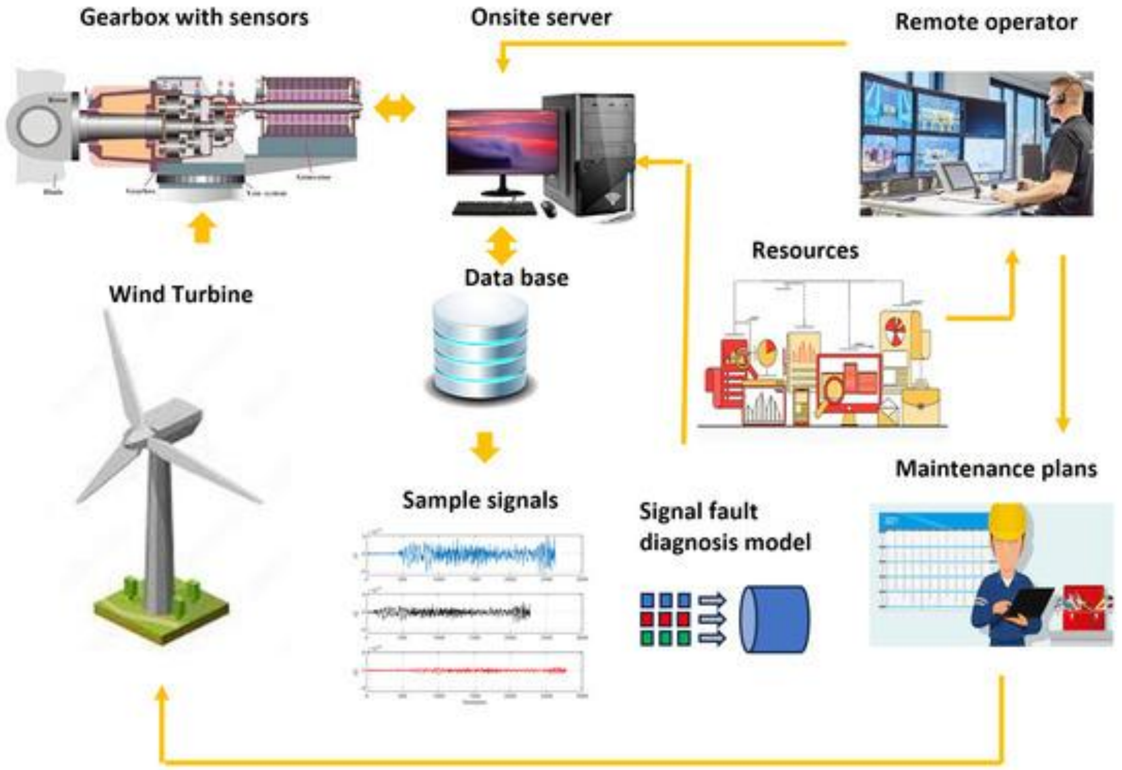

Figure 2. Conceptual framework for intelligent fault detection and predictive maintenance in wind turbine (Santiago et al., 2024).

Although these techniques have displayed some encouraging performance, their performance is often questioned in the offshore locations, especially when there are winter months.

1.4 Research Problem

Predictive maintenance models built on SCADA offer useful insights into the turbine functionality; nevertheless, they are also susceptible to external environmental factors. Offshore wind turbines work in constantly changing atmospheric and marine conditions, directly affecting SCADA measurements. Consequently, variability caused by wind, temperature, or turbulence might be incorrectly interpreted as an indicator of errors and not the expected behavior provided by the environment (Tautz-Weinert & Watson, 2017; Wrzask et al., 2025).

This issue is more severe in the winter where seasonal influences on turbine performance, including icing, cold-weather derating and greater turbulence, tend to cause significant changes in operating patterns. In this case, distinguishing a steady state of normal behavior becomes even more challenging, causing predictive models to produce even more false positives. In contrast, models tuned to minimize false alarms can be overly conservative, leading to failure to detect faults.

This is a trade-off between false positives versus false negatives, which identifies a core challenge of SCADA-based fault-detection in offshore environments where maintenance decisions have tremendous operational and economic implications. To address this challenge, we need to merge environmental variables in an organized way which allow us to distinguish between real fault events and environmental influence.

1.5 The CARE Benchmark and Reproducible Modeling

One of the main difficulties in creating effective predictive maintenance models is the unavailability of standardized datasets and assessment frameworks. The CARE to Compare project can help resolve this problem by offering a large-scale benchmark dataset and a uniform preprocessing and evaluation procedure (Gück et al., 2024). The framework enhances reproducibility and provides a fair comparison among various modeling methods.

Based on the CARE benchmark, (Hassan, 2025) has shown that deep learning models especially 1D-CNN architectures have the potential at detecting early signs of fault in SCADA data effectively through a structured preprocessing pipeline. The research however also pointed out that model performance can be affected by the environmental variability, leading to the speculation that more contextual information can be required to enhance robustness. In this thesis, the CARE framework will be used systematically as a guideline for exploring the possibility of environmental data contribution to fault detection.

1.6 Role of Meteorological Data in Offshore Maintenance

The weather conditions are very important in the operation and maintenance planning of turbines. Environmental factors like wind speed and wave height not only affect the performance of turbines but also determine the viability of maintenance operations (R. K. Pandit et al., 2020). The inclusion of weather data in the maintenance planning process has been observed to minimize waiting time and enhance coordination of resources (Irawan et al., 2017; Schrotenboer et al., 2018).

In this regard, including meteorological information in predictive maintenance models provides a fruitful avenue to enhance detection of anomalies. In this thesis, environmental variables derived from the ERA5 reanalysis dataset (Hersbach et al., 2020), including wind and turbulence are combined with SCADA data. This integration offers better representation of operating condition of turbines, especially during winter when environmental effects are much more pronounced.

1.7 Data Fusion Strategies

The process of combining different data sources is often known as data fusion in machine learning. Fusion can be applied at various levels in predictive maintenance depending on how the data are combined in the modeling process.

Data-level fusion refers to the process of integrating SCADA and meteorological variables and then combining them into a single feature space before model training. This will enable the model to directly learn cross-modal relations. As it was shown by (Attouri et al., 2026), early fusion techniques can be used to extract intermodal dependencies;

however, these techniques can also be discussed as introducing noise and obscuring modality-specific patterns.

Conversely architecture level fusion works with different data modalities by differing model parts and then combines them. This allows learning of specialized representations of each modality whilst maintaining their own unique properties. This thesis proposes a dual-branch CNN architecture, in which SCADA and meteorological data streams are processed independently and combined with a learned scalar gating mechanism. This strategy enables moderate acquisition of environmental context in preserving the power of turbine-functional signal learning.

1.8 Research Aim and Questions

This thesis explores the impact of various data fusion strategies on the performance of anomaly detection in offshore wind turbines. Special attention is paid to winter operating conditions, where the most significant environmental impacts are observed and where the contingency of faults is the most important consideration when making maintenance decisions.

To do this, the study uses a comparative modeling approach which builds on a SCADA-only baseline model and then expands respectively to data-level fusion and architecture-level fusion approaches. The design allows us to systematically assess whether the addition of meteorological information increases fault detection, and whether the integration method makes a difference towards better fault sensitivity or alert reliability.

The subsequent research questions guide the investigation:

1. What is the performance of the replicated SCADA-only 1D-CNN baseline in the detection of winter anomalies in offshore wind turbines?
2. Does data-level fusion of SCADA and ERA5 meteorological variables improve the performance of anomaly detection in comparison to the baseline of SCADA-only?

3. Does architecture-level fusion offer any extra advantages over data-level fusion in terms of robustness and fault detection performance in winter seasons?

1.9 Overview of thesis structure

The overall structure of the thesis is divided into six chapters for systematic presentation:

Chapter 1: Introduction

This chapter presents a background of the offshore wind energy and the problems involved in predictive maintenance in a harsh environment. It reports the research problem, outlines the study objectives, and develops the research questions that conduct the study.

Chapter 2: Literature Review

In this chapter the literature on predictive maintenance, condition-based maintenance, and SCADA based fault detection in wind turbines is reviewed. It also analyzes the importance of environmental data, data fusion strategy, and current limitations, as well as research gaps that are filled in this thesis.

Chapter 3: Methodology

This chapter outlines the general research design, data sources, preprocessing plan and model development plan. It describes how the SCADA-only baseline, data-level fusion, and architecture-level fusion pipelines are implemented as one experimental design.

Chapter 4: Model Design and Analytical Assessment

The chapter shows the process of development of the three predictive models, which explains the design options and the reasoning behind the choice of each model. It also explores the representation and use of SCADA and environmental data in the various model architectures.

Chapter 5: Results and Discussion

In this chapter, the experiments results are reported and the performance of the proposed models compared using general and specific (winter) test sets. It gives interpretation of the findings in terms of precision- recall tradeoff and discusses their implications on offshore maintenance operations.

Chapter 6: Conclusion and Future Scope

This chapter sums up the key findings of the study and points out the theoretical and practical contributions. It also identifies some of the possible future research directions especially in enhancing model generalization and ways of incorporating the predictive outputs into maintenance optimization models.

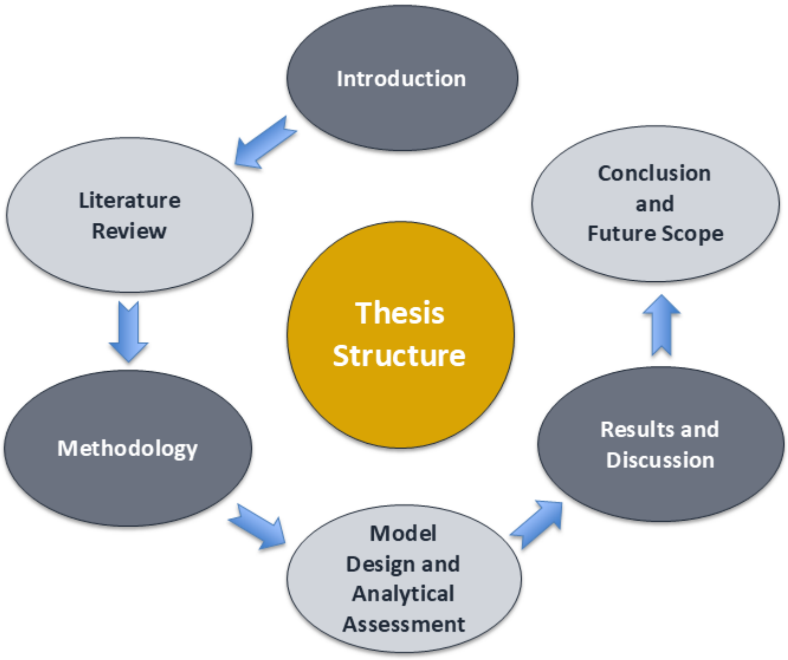

Figure 3. Thesis Structure

2 Literature Review

This chapter summarizes some of the literature associated with offshore Wind Turbine predictive maintenance, anomaly detection using SCADA, integration of environmental context and data fusion approaches. First, the review focuses on operational and maintenance issues of offshore wind systems, especially during the winter season, and then concentrates on fault detection approaches based on SCADA data using statistical, machine learning and deep learning techniques.

Next, the chapter addresses the use of meteorological information in offshore wind operation and surveys the current state of multimodal fusion for the integration of both operational and environmental information. Finally, the key research gaps found in the literature are summarized to provide motivation and methodological foundation for the proposed models developed in this thesis.

2.1 Offshore Wind Operations and Maintenance: Challenges and Economic Context

Offshore wind farms have grown significantly in the last 20 years and play an increasingly key role in large-scale renewable energy development strategies (Bilgili & Alphan, 2022). Operational and maintenance (O&M) challenges have been commensurate with turbine size and increased with distance from the coast and the harsher environment. To understand the significance of predictive maintenance and in particular, early anomaly detection for offshore O&M, it is important to understand the economic structure of offshore O&M.

2.1.1 Economic Structure of Offshore Wind O&M

O&M expenses are a disproportionately higher proportion of total lifetime expenditure on offshore wind, compared to onshore installations. Offshore O&M is always estimated between 20 and 35 percent of the levelised cost of energy (LCOE), as compared to 10 to

15 percent in the case of onshore wind (Ren et al., 2021). This difference evidences the compounding effect of the severity of marine environments, accessibility, vessel requirements, and extended mobilization times. Unplanned maintenance events turn out to be more costly in terms of expenses and lost time than proactive strategies, with operators required to coordinate vessel fleets, manage spares inventory and deal with the uncertainty of weather conditions to safely complete their transfers (Fox et al., 2022). (Turnbull & Carroll, 2021) demonstrate through cost-benefit analysis that the financial case for advanced monitoring and predictive maintenance strategies is sound even when the costs of implementation are factored in, even in major component drives of the moving machines.

The specialty of offshore O&M economics is the connection between the severity of failure and the intervention cost, which is non-linear. The major component failures require complicated logistics solutions and the operation and maintenance costs of ships and helicopters are high which makes these events financially costly (Kou et al., 2022). (Ren et al., 2021) emphasize this fact, demonstrating that the failure of coordination between planned and unplanned maintenance is a critical variable of operations: unforeseen failures disrupting the schedule, putting a strain on expensive equipment resources (the use of special vessels), and leading to significant loss of revenue due to prolonged outage. The economic rationale is then clear: realizing developing faults early on to turn unplanned interventions into planned ones is among the most valuable activities, which can be implemented amongst the operators of offshore wind.

2.1.2 Operational Constraints in Offshore Environments

Offshore wind has big physical and logistical limitations in addition to costs. The marine environment presents additional challenges not found in the onshore environment: corrosion is accelerated by salt spray, vibrations caused by waves increase fatigue loads and biofouling makes inspection of the substructures difficult (Wrzask et al., 2025). Offshore turbines can only be accessed during favorable weather conditions with the significant wave height, wind speed and visibility conditions being thresholds for crew transfer vessels to safely approach platforms (Ren et al., 2021). (de Boer & Xydis, 2024) state that during the winter the weather windows for maintenance can be a few days per month, which causes scheduling problems. The logistics of offshore maintenance – vessel scheduling, availability of spare parts and personnel – need to be coordinated with weather forecasts to minimise waiting time and maximise the use of the available access windows (Koltsidopoulos Papatzimos, 2020).

Such limitations mean that offshore operators can't conveniently respond to failures as they happen. Unplanned downtime from component failure may be compounded by environmental limitations since mobilization and maintenance activities may be delayed by adverse weather conditions (Koltsidopoulos Papatzimos, 2020). (Rinaldi et al., 2021) point out that the switch from reactive to predictive maintenance is key to cost reduction, but it gets more technically challenging as wind farms expand into deeper waters.

2.1.3 Failure Modes and Component Reliability

It is critically important to understand what components are likely to fail and why they are failing in order to target the predictive maintenance efforts. The most critical ones are gearboxes, generators, power converters and pitch systems due to their frequency of failure and downtime (Ren et al., 2021). Based on large-scale analyses of maintenance and failure data, it is observed that failures in the electrical and control systems are common but are generally easy and quick to fix, whereas mechanical failures in the drivetrain are less common but much slower to repair and more expensive (Carroll et al., 2016; Faulstich et al., 2011).

Importantly, many failure modes evolve over time and may be noticed over a time span of weeks to months. This aspect is the cornerstone of the potential benefits of predictive maintenance: by detecting signs of potential failure early on, data-driven monitoring can enable operators to schedule maintenance actions during optimal weather conditions, order spare parts in advance, and prevent the costly emergency mobilization process.

2.2 Winter-Specific Challenges in Offshore Wind Operations

Offshore wind is difficult to operate all year round, but the winter season poses new operational and technical challenges, which will increase O&M costs and make maintenance planning more complex. These winter-specific challenges lay the foundations for understanding the need for seasonal anomaly detection capabilities.

2.2.1 Seasonal Variability in Offshore Conditions

There is significant seasonal variation in offshore wind resources. Offshore wind power density is highly seasonal and varies locally according to the global evaluation based on ERA5 (Soares et al., 2020). Long-term modelling also shows that the winter months make a disproportionate energy contribution to the annual energy production.

During the winter months with the increased wind speeds, the conditions are more extreme. Offshore maintenance windows during the winter are much shorter and more variable than summer windows, and wave heights and wind speeds are often above safe access limits (Martini et al., 2017). This creates the paradox that winter is the highest revenue season for turbine but is also the hardest time of year to access a turbine and the season with the highest risk of turbine failure.

2.2.2 Winter-Specific Failure Mechanisms

Conditions in the winter season induce failure mechanisms that are not present or not as severe in other seasons. (Y. Zhang et al., 2024) emphasize that icing as a winter-specific hazard results from rotational imbalance due to ice accretion on the blades, causing damage to turbine structures, power loss and revenue loss. The occurrence of icing events can be detected by the power output, rotor speed and temperature signals by distinctive patterns prior to total ice accumulation.

In addition to icing, winter conditions may exacerbate mechanical stresses in low temperatures, thermal cycling, and salt freezing exposure. Failure rates and failure modes can vary greatly between seasons, but most anomaly detection models are trained with data from all seasons without taking the seasonality into account.

2.2.3 Maintenance Access and Planning in Winter

Winter maintenance planning is more difficult due to the increased risk of failure and the decreased access windows. (Tapoglou et al., 2021) use machine learning in sea-state prediction, highlighting the importance of short-term accurate wave condition forecasting for optimization of maintenance vessel scheduling. But despite better forecasting, meteorological conditions that can limit windows to operate in winter remain beyond operators' control.

This constraint has direct economic consequences. If a turbine fails in the wintertime, operators will have to choose between an immediate high-risk, high-cost callout (depending on weather conditions) or wait for a better weather window by accepting longer downtime during the peak revenue season. It's particularly valuable in winter, when early detection of faults gives time to plan the intervention, rather than acting on failures as they happen.

2.2.4 Existing Research on Winter Offshore Operations

Winter issues are well known in the offshore wind literature, but there are relatively few papers focused on winter anomaly detection. (Y. Zhang et al., 2024) propose a temporal convolutional network (TCN) framework for icing event prediction up to 48 hours ahead, with an average accuracy of 77.6%. Their focus is, however, on predicting icing events, not anomaly detection generally during the winter.

2.3 SCADA-Based Condition Monitoring and Anomaly Detection

The main source of data for condition monitoring of wind turbines is a Supervisory Control and Data Acquisition (SCADA) system. The recent advancements in anomaly detection using SCADA are vital to understanding the wider research context of this thesis, from traditional statistical approaches to the recent advances based on deep learning.

A standard SCADA-based anomaly detection workflow generally progresses from data collection and preprocessing to model training, threshold selection, and anomaly detection, as illustrated in **figure 4**.

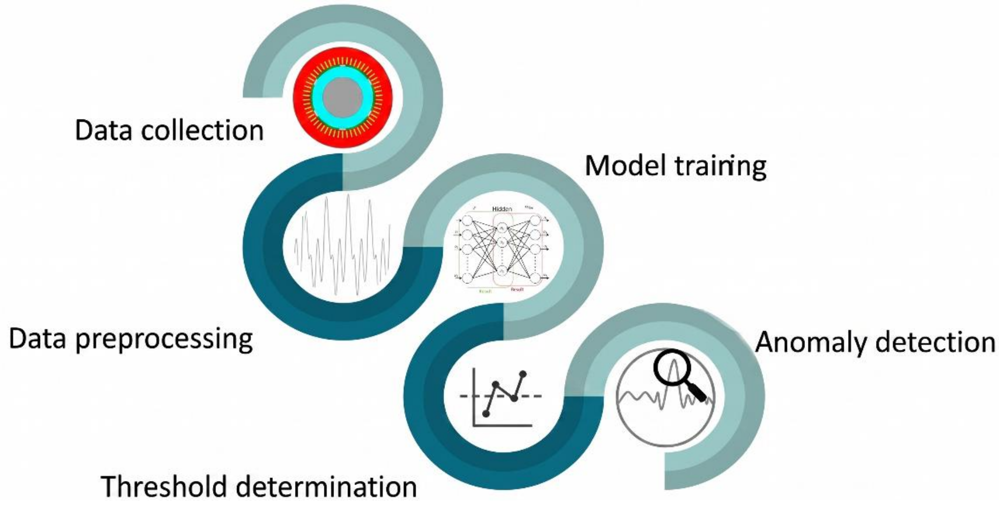

Figure 4. General Workflow of Data-Driven Anomaly Detection for Wind Turbine Condition Monitoring (Dibaj et al., 2024)

2.3.1 SCADA Data in Wind Turbine Monitoring

Operational parameters such as power output, rotor speed, temperature of the nacelle and other sensor channels are continuously recorded at 10-minute intervals on SCADA (Tautz-Weinert & Watson, 2017). SCADA data are also readily available at large scales, and offer operationally cheap data, but suffer from coarse temporal resolution (usually

10 minutes), limiting their ability to capture high-frequency phenomena (Tautz-Weinert & Watson, 2017).

With this point of view, (Maldonado-Correa et al., 2020) make the point that SCADA has value not as a substitute for high frequency vibration monitoring, but as a complement to it. SCADA offers much needed information regarding the system level context and can monitor components which are not instrumented with dedicated monitoring sensors. But a fundamental problem with SCADA-based anomaly detection is that there is some variation in normal operation due to environmental changes, so it can be difficult to tell if these variations are caused by degradation of the components or just due to normal operation.

2.3.2 Traditional Statistical Approaches to Fault Detection

Early fault detection using SCADA systems was based on normal behavior models (NBMs), which use operational and environmental variables to estimate expected sensor values; a deviation from these estimated sensor values is considered to be an anomaly. The classical NBM approaches include polynomial regression, Gaussian process regression and support vector regression, and the residuals are generally monitored using statistical control methods like exponentially weighted moving averages (EWMA) (R. Pandit & Wang, 2024).

The fundamental principle of an NBMs is that if a turbine is operating in a healthy condition, there are well-defined relationships that are predictable from the physical behaviour and the turbine's control-system response. If the model is trained on known healthy data, the difference between the expected and observed sensor value is a residual, and if the residual is abnormally high, the sensor may be operating abnormally.

Despite the growing popularity of modern deep learning methods, traditional statistical NBMs remain useful due to their high computational efficiency, interpretability, and the relative lack of required training data. Their shortcomings, however, become evident with respect to temporally dependent and high-dimensional nonlinear behavior of the turbine. Traditional methods are particularly weak at addressing the complex temporal relationships of continuous dynamic processes where the behavior of the turbine depends on the recent operating characteristics (Lei et al., 2019). The limitation drives the need for more flexible or sequence-aware learning approaches for anomaly detection in SCADA-based environment.

2.3.3 Machine Learning for SCADA-Based Fault Detection

The shift from traditional statistical methods to machine learning (ML) added flexibility to the modelling of nonlinear relationships and high dimensional SCADA feature spaces. Supervised ML methods, e.g., random forests and gradient-boosted trees, have been applied extensively in wind turbine fault detection, as they can identify complicated decision boundaries without the need of making fixed functional assumptions (Tang et al., 2020; D. Zhang et al., 2018).

In general, machine learning approaches can be classified as supervised, unsupervised, or reinforcement learning, as shown in **Figure 5** with classification and regression being two of the most important supervised learning tasks for fault detection and predictive maintenance (Rinaldi et al., 2021).

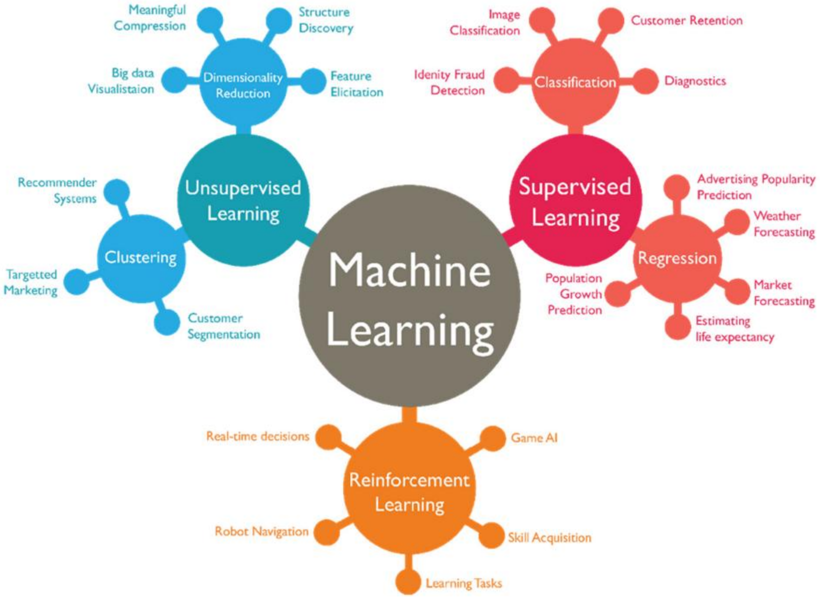

Figure 5. Main Machine Learning Categories, Techniques, and Typical Applications Relevant to Predictive Maintenance and Fault Detection (Rinaldi et al., 2021)

While ML methods could outperform previous statistical methods, they still rely heavily on feature selection and data representation (Jankauskas et al., 2023). Models obtained from one turbine might not be as well-transferred to another turbine without retraining due to differences in turbine aging, maintenance, environmental exposure, or sensor calibration. This transferability concern is particularly relevant for offshore wind farms as the ideal monitoring system should be viable among several turbines, not just one (Masoumi, 2023).

Another constraint is that feature engineering is still required. It is common for ML-based condition monitoring to demand domain knowledge regarding which SCADA channels to consider, how to handle missing data, what derived features might be of value, and how to normalize variables prior to training. Both these preprocessing decisions can have a significant impact on the performance of the model, further strengthening the case for learning architectures that are more able to learn from structured time-series inputs (Jankauskas et al., 2023; Udo & Muhammad, 2021).

Despite the development of advanced ML techniques, fault detection is still limited by the deceptiveness and weakness of fault precursors. Some failures are slow and result in recognizable patterns, others happen in an instant or are obscured by the normal variations in operation (Jankauskas et al., 2023). Hence, the selection of the algorithm is not the only important factor in the use of machine learning for fault detection in SCADA; the pre-processing, the temporal representation, and the capability of the model in distinguishing the actual degradation from normal operating changes are also significant (Jankauskas et al., 2023; Udo & Muhammad, 2021).

2.3.4 Deep Learning for Anomaly Detection

Deep learning (DL) techniques, especially multi-layered neural networks have proven to be an efficient way for detecting anomaly through SCADA. (Helbing & Ritter, 2018) review the applications of deep learning for wind turbine fault detection. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are able to learn representations of hierarchical features automatically from raw or minimally processed data, eliminating the need of manual feature engineering (R. Zhao et al., 2019).

Bearing failure detection is carried out using a convolutional Autoencoder based on SCADA data, where the false alarm rate is reduced and the bearing failures can be detected months prior to failure (Tutivén et al., 2023). A normal operating behaviour is learned from autoencoders and inputs that cannot be reconstructed accurately are identified as potential anomalies (Tutivén et al., 2023). Similarly, (S. Zhang et al., 2023) use LSTM networks for predictive fault diagnostics, where the LSTM architecture also has the ability to store time-lagged information and to model long-term temporal dependencies in sequential data.

Transfer learning is investigated for fault detection, showing that CNNs that have been trained on one turbine can be adapted to new turbines with limited data, thereby increasing the scalability (Zgraggen et al., 2021). (Schröder et al., 2022) then build on this idea with physics-informed transfer learning, in which also artificial neural networks

(ANNs) and autoencoders are pre-trained using simulation data and subsequently fine-tuned using SCADA data, achieving greater sample efficiency by exploiting the limited amount of measurement data.

The common general Deep Learning pipeline for anomaly detection consists of feature extraction and subsequent decision making in a learned representation space, as shown in **figure 6**.

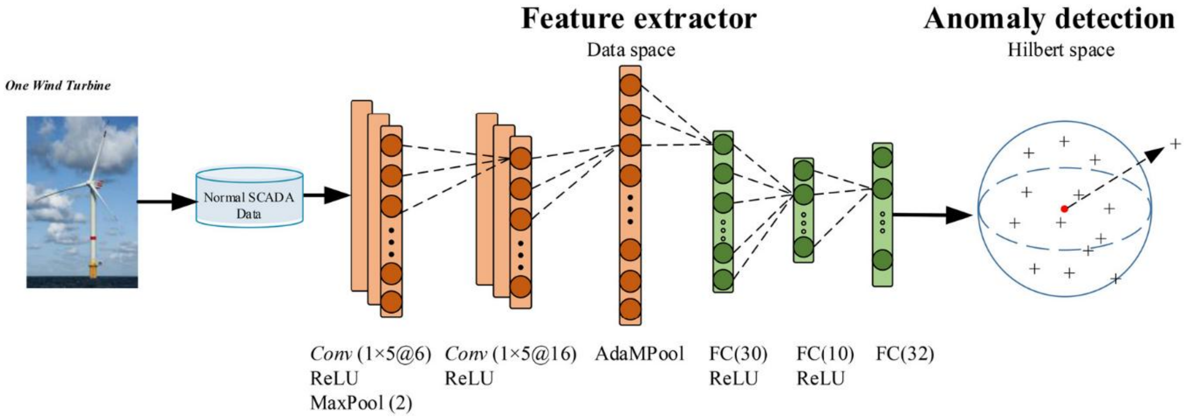

Figure 6. CNN-Based Feature Extraction and Anomaly Detection Architecture for SCADA-Based Fault Detection (Peng et al., 2023)

Therefore, deep learning models especially CNN based models would be suitable to capture temporal dependency in SCADA data and thus they are used as the base modelling approach in this thesis.

2.4 The CARE to Compare Benchmark: Toward Reproducible Evaluation

One of the common difficulties in the research on wind turbine fault detection is the absence of standardization of the datasets and evaluation procedures, which makes it difficult to compare methods from different studies. The CARE to Compare benchmark fills this gap and establishes a framework of reproducible research.

2.4.1 The Reproducibility Challenge in Wind Turbine Fault Detection

One major challenge in developing wind turbine fault detection research is the variety of data usage and evaluation methodologies across the different studies, making it difficult to compare different approaches. (Helbing & Ritter, 2018) note that many studies on fault detection use their own proprietary datasets, and the preprocessing pipeline, the way in which the features are selected, as well as the way the results are evaluated vary from study to study, making it difficult to make comparisons between them. Moreover, the notion of "anomaly" varies from study to study: some define abnormal signals as any signal that is not normal; others define abnormal signals as a persistent deviation from normal behavior, or a particular failure mode, which can further confuse the definition of "normal" vs "abnormal" signals (Helbing & Ritter, 2018). There is also different evaluation metrics used in different studies such as accuracy, precision-recall metrics, or domain specific metrics like detection lead time or false alarm rate (Stetco et al., 2019). In the absence of common standards and publicly available datasets, it is hard to judge if performance is better in some cases than in others because of true methodological improvements or due to the different nature of the datasets or the design of their evaluation (Stetco et al., 2019).

(Helbing & Ritter, 2018) point out that creating labeled datasets in the public domain has been a major challenge to the development of supervised learning methods. In most cases, the wind turbine SCADA data remain proprietary and are controlled by the turbine's manufacturer and/or wind farm owner, who is hesitant to share it for commercial reasons. Available data are very hard to find, and even if available, may only cover a small number of turbines or a limited time period, without detailed failure labeling. Data scarcity means that researchers are forced to devise methods on small idiosyncratic datasets, which means that generalisability is limited and it is difficult to determine whether the methods will work on new turbines or wind farms (Helbing & Ritter, 2018).

(Badarinath et al., 2024) explain that SCADA issues can be an actual mechanical problem or a false indication of a damaged, dirty, misaligned or communication sensor. The test result of an anomaly detection model can be ambiguous when there are no actual ground truth maintenance logs to differentiate actual component failures and sensor failures.

2.4.2 CARE to Compare Framework Overview

(Gück et al., 2024) introduced the "CARE to Compare" dataset, a publicly available and openly licensed high-frequency dataset from multiple wind turbines with known failure events, which serves as an essential testbed for wind turbine anomaly detection research. With this resource, (Hassan, 2025) created a standardised evaluation protocol that outlines strict data partitioning (e.g., temporal splits to prevent leakage of information), evaluation metrics (precision, recall, F1-score, and early warning lead time) and baseline models for comparison.

This sort of standardisation has to be a prerequisite to the growth of the field. The dataset "CARE to Compare" provides a transparent and common benchmark to rigorously compare new approaches to the existing ones and find reliable advances and prevent designing a method that is too dependent on the particular features of a proprietary dataset (Hassan, 2025).

2.4.3 Baseline CNN Architecture

(Hassan, 2025) develops a base CNN architecture in the CARE to Compare framework. The proposed 1D-CNN comprises of two convolutional blocks with batch normalisation, ReLU activation, max-pooling, dropout, and dense layers that are trained on fixed-length multivariate sequences (20 time steps x 100 components). The model shows that it has a high anomaly detection capability in the benchmark environment and the model also has the capability of early warning before failure events can occur as predicted.

Notably, the research proves that the use of sequence-aware neural networks significantly outperforms the use of snapshot-based methods, which indicates the relevance of time-related context in the detection of faults in SCADA. The CNN baseline then serves as a stringent and replicable benchmark to assess additional architectural extensions and environmental data integration plans in this thesis.

2.4.4 Identified Limitations and Extension Opportunities

The CARE to Compare baseline sets a high standard for performance but (Hassan, 2025) identifies several limitations and extension opportunities. The study centres on only four representative channels of SCADA data; other sensors, e.g. accelerometers at the blade-roots or stray-current monitors, could have other predictive value. Moreover, the model includes local environmental parameters like ambient temperature but mainly depends on the SCADA feeds. Incorporating external meteorological context, such as satellite-derived sea-surface temperature and ice formation indicators, is recognized as a key next step to enhance the environmental robustness and understanding of how the large-scale weather and environmental conditions affect turbine performance in various operating environments.

2.5 Meteorological Data and Environmental Context in Wind Energy

Weather is a key factor for wind turbine operations. To contextualize meteorological data, such as ERA5 reanalysis, in anomaly detection models, it is important to understand how these data have been used in wind energy research.

2.5.1 Weather Dependence of Wind Turbine Operations

The environmental envelope of wind turbines is constantly changing, and consists of a number of meteorological parameters such as wind speed, wind direction, air temperature, turbulence intensity, air density and more. These environmental conditions influence power output, rotor speed, blade pitch angle and component temperatures in a dynamic manner. This weather dependency makes anomalies inherently difficult to detect, since what is considered an abnormal value for one sensor can be a normal value for the turbine under certain weather conditions (Tutivén et al., 2023).

Turbine operating behavior is affected by environmental conditions in nonlinear and complex ways. Power output and component temperatures could fluctuate significantly with wind speed, turbulence, temperature, and atmospheric conditions, and these variations in operation may appear to be fault signatures. Consequently, separating the real component degradation from variation caused by the environment is a significant issue in environment-induced operational variation in SCADA-based condition monitoring (Stetco et al., 2019; Zgraggen et al., 2021).

(R. Pandit & Wang, 2024) stress that environment and operation variability is one of the most important sources of false alarms in anomaly detection using SCADA. Any changes to SCADA measurements could be either due to developing faults or normal environmental influences, and it's not easy to determine whether the change is caused by a true degradation or due to the weather driving operational changes. In addition, even though some environmental parameters are recorded in SCADA systems, sensors directly installed on the turbine may be affected by the behavior of the turbine, such as yaw error of the nacelle, and may not necessarily capture the true surrounding wind conditions.

During high meteorological variability, this is an environmental confounding problem that is especially problematic. In conditions of relatively stable weather, the behavior of the turbine is more predictable and any deviation from the operating behavior is more likely to be a true fault. But, in rapidly shifting weather situations, the turbine behavior can vary highly depending on the external environmental forcing, which makes the identification of the anomalies caused by faults less reliable. This challenge is amplified during the winter season with more extreme conditions and weather variation in the offshore environment.

2.5.2 ERA5 Reanalysis Data for Wind Energy Applications

The fifth-generation atmospheric reanalysis (ERA5) by the European Centre for Medium-Range Weather Forecasts (ECMWF) offers high spatial resolution (around 31 km) meteorological data every hour from 1950 onward on a global scale (Hersbach et al., 2020). (Hersbach et al., 2020) detail the data assimilation system of ERA5, which integrates data from satellites, weather stations and other observation sources with the numerical weather prediction models to generate a physically coherent and complete picture of the atmosphere.

ERA5 is well validated for wind energy applications. ERA5 wind speed and marine estimates were shown in independent studies to be well correlated with in-situ measurements, as summarised by (Hersbach et al., 2020). Based on this accuracy, (Soares et al., 2020) have used ERA5 to characterise global offshore wind resources, with good capture of seasonal and interannual variability. Additionally, using ERA5 data and simulating long-term power generation, (Hayes et al., 2021) demonstrate that ERA5-based simulations are able to reproduce observed capacity factors and seasonal patterns. These validation studies make ERA5 a trusted resource for providing environmental context for offshore wind applications.

2.5.3 Meteorological Information in Predictive Maintenance

ERA5 has been widely applied to the field of resource assessment and power forecasting, but not so many other use-cases have been applied to the field of predictive maintenance and anomaly detection. O&M tools are now available which are based on data and include weather forecasting to allow for better planning of shipping for maintenance tasks (Koltsidopoulos Papatzimos, 2020). The integration, however, is limited to planning logistics from the weather forecast, and not to including meteorological variables into fault detection models themselves.

(Y. Zhang et al., 2024) form a first step towards the integration of meteorological data in fault detection and use external variables from the weather database in addition to SCADA data to predict icing events. They are, however, specifically targeting this winter hazard (icing) and not general anomaly detection throughout all seasons and failure modes.

2.5.4 Limited Integration of ERA5 in Anomaly Detection Models

Limited exploration of systematic integration of ERA5 meteorological data in anomaly detection model within the reviewed literature on the CARE to Compare framework could be found, despite the proven usefulness of ERA5 for wind energy applications and the identified problem of environmental confounding in anomaly detection based on SCADA data. A remarkable gap is the potential to use the environmental context to better differentiate between operationally induced, weather-related variability and true component degradation in order to reduce false positive alarms, increase detection sensitivity, and strengthen the models during seasonality, especially during the winter time when environmental variability is highest and the operational impact of missed faults and unnecessary maintenance dispatches is most critical.

2.6 Data Fusion Strategies in Condition Monitoring

Data fusion, which combines data from various data sources, is a well-known approach in condition monitoring. To design multi-input anomaly detection models, it is important to know about the different fusion architectures and the associated pros and cons.

2.6.1 Fundamentals of Multi-Modal Data Fusion

Data fusion (also called data integration) is the task of combining data from several different, heterogenous data sources to create a more complete and more correct picture of the situation than what could be derived from any one source alone. In offshore wind turbine monitoring and often other applications, it means combining operational SCADA data with data from outside the turbine systems, such as wind or weather data, to complement the understanding of wind turbine behaviour under different operational circumstances (Y. Zhang et al., 2024). The complementary nature of the data from different sources can assist in reducing ambiguity in anomaly interpretation, enhancing robustness against system operation variability, and contributing to richer contextual information for fault detection and maintenance decision making (R. Pandit & Wang, 2024).

In today's machine learning systems, fusion techniques are usually classified based on the level of fusion, which is either data-level fusion (early fusion), feature-level fusion (intermediate fusion), or decision-level fusion (late fusion) (Brena et al., 2020; T. Lin et al., 2025). The fusion levels are conceptually shown in the **Figure 7**, in which various information streams are fused at various levels of the learning architecture. Each approach comes with its own set of compromises with regard to representation learning, computational complexity, interpretability and robustness of conditions monitoring applications (T. Lin et al., 2025).

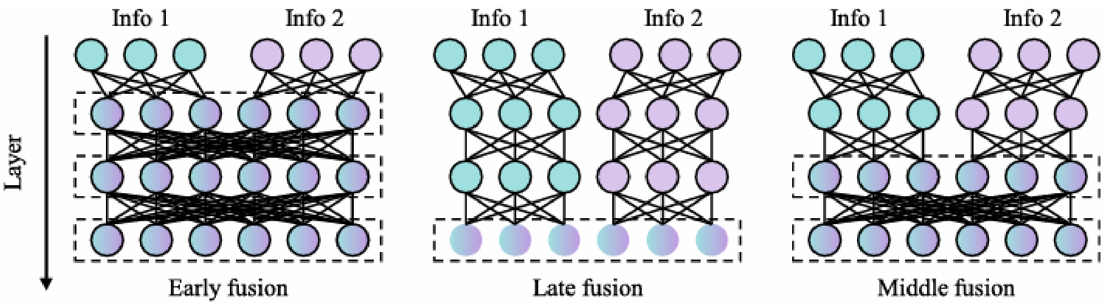

Figure 7. Illustration of Early, Late, and Middle Fusion Strategies in Neural Network Architecture (T. Lin et al., 2025)

2.6.2 Data-Level Fusion

Data-level fusion is a method that aggregates raw or preprocessed data from multiple data sources and fuses them into a single data representation as an input for model training. In the context of the interaction between SCADA and ERA5, this would be a feature vector (or tensor) that combines SCADA sensor channels with ERA5 variables, and then train a unique model with this combined input (Attouri et al., 2026).

The main advantage of the data-level fusion is that it can learn the interaction between the SCADA and meteorological variables directly from the data. For instance, an apparently unusual temperature under one environmental condition may actually be a normal turbine reading under a different environmental condition (R. Pandit & Wang, 2024). The model is able to capture cross-modal dependencies because these relationships are learnt jointly. Furthermore, data-level fusion is conceptually trivial and can be easily performed in a computational-efficient manner, since all variables go through the same learning pipeline (T. Lin et al., 2025).

But in data-level fusion, it is assumed that the information coming from different data sources can be related at compatible temporal and spatial scales. The measurements in SCADA mainly occur on turbine level, with a high temporal resolution, while ERA5 variables are meteorological estimates that are spatially averaged with low temporal resolution. Therefore, temporal synchronization and spatial alignment should be performed prior to the training of the model (Xu et al., 2026).

A crucial thing to remember is that features need to be scaled and the preprocessing methods need to be consistent. SCADA and meteorological variables have different physical ranges and units, and if they are not normalized properly, some variables may be given greater weight in the learning process. Standard practice thus entails implementing normalization procedures with only training data parameters, which are to be excluded from the information in order to avoid information leakage and ensure the same evaluation conditions (Hassan, 2025; Xu et al., 2026).

2.6.3 Architecture-Level Fusion

Architecture-level fusion receives multiple modalities of data from the various branches of the architecture and then it fuses the learning representations for the final prediction.

In the context of offshore wind turbine monitoring, this means that one branch is used for processing operational turbine data (e.g. SCADA data) and another branch is used for processing contextual environmental data for predictive maintenance and anomaly detection.

The primary benefit of architecture-level fusion is that it can be used to learn modality specific patterns separately in each branch. A SCADA branch can record operational and temporal turbine behaviour and a meteorological branch can model environmental variability and seasonality. The separation of the learning process prior to integration allows better adaption to the differing statistical properties and characteristics of heterogeneous data sources while maintaining modality-specific feature representations (Xu et al., 2026).

Architecture-level fusion has more flexibility in learning the contexts between operational and environmental information than direct data-level concatenation. Moreover, since each branch processes its own input separately then fuses them together, the fusion process can be better suited for dealing with different characteristics of the data, and better representation learning for multimodal inputs (T. Lin et al., 2025). Architecture-level fusion can then be more effective in detecting anomalies in context for highly heterogeneous data sources like SCADA data and ERA5 variables (Liu & Hui, 2024).

2.6.4 Fusion in Wind Turbine Monitoring: Current State

(W.-H. Lin et al., 2021) serve as an important precursor in the field of wind energy for operational and meteorological data fusion. They build a Temporal Convolutional Network (TCN) architecture, combining SCADA power generation outputs with historical meteorological data for wind power forecasting and obtain good prediction results. This work is a proof of concept for which external weather data can be successfully integrated with the SCADA data in the deep learning architectures.

The study is aimed at power forecasting, and not at anomaly detection (W.-H. Lin et al., 2021). Forecasting and anomaly detection are drastically different tasks, with different goals, evaluation measures and modeling needs. Forecasting involves predicting future power output based on past and environmental conditions, whereas anomaly detection attempts to provide information on variation of operational behavior from normal conditions, which can be seen as signs of emerging faults. Therefore, while the good performance of the fusion of SCADA–meteorological data in forecasting could give hope in using the fusion for anomaly detection in the predictive maintenance context, it cannot be automatically assumed that this would be the case.

Moreover, there are few studies that have explored SCADA-ERA5 fusion in standard anomaly detection benchmark schemes like CARE to Compare. This gap presents the need to explore whether the robustness and reliability of offshore wind turbine anomaly detection models can be enhanced by incorporating environmental context, especially in the winter operating regime.

2.7 Synthesis of Literature Gaps

Although the study of SCADA-based predictive maintenance and anomaly detection exists for offshore wind turbines, the literature has some key limitations yet to be addressed. Such gaps are especially noticeable in terms of considering the environmental trends, assessing winter conditions and using multimodal fusion strategies as part of benchmarking frameworks.

One of the key drawbacks of existing baseline models used for CARE to Compare is that they are only based on the data provided by the SCADA systems on the turbine, and do not include external meteorological data (Hassan, 2025). This omission is important as the changes in environmental conditions affect the measurements taken by SCADA that can result in variations in operation which are interpreted as faulty behavior. Therefore, models that do not include free-stream environmental input could yield false positives due to normal turbine response to varying conditions offshore before the actual component degradation occurs.

A second important gap relates to the minimal amount of investigation of anomaly detection performance in winter. While the operation difficulties and weather-induced failures like icing are well recognized in the research of offshore wind (de Boer & Xydis, 2024; Y. Zhang et al., 2024), a systematic analysis of the performance of anomaly detection models trained with multi-seasonal data in winter conditions has not been conducted so far. Specifically, there has been little research into the possibility of improving the reliability of fault detection under circumstances of high meteorological fluctuation and limited maintenance access.

Furthermore, fusion of operational and meteorological data has proved successful in related wind energy applications but is not widely used for the detection of anomalies in the standardized offshore wind benchmarking frameworks. (W.-H. Lin et al., 2021) show that a multimodal fusion strategy using the outputs of the SCADA and historical meteorological data can enhance the performance of wind power forecasting; however, this fusion strategy has not been systematically investigated in the framework of CARE to Compare for anomaly detection. This is particularly significant as the CARE benchmark allows a replicated evaluation scene, providing a testing ground to determine if using meteorological data is truly beneficial in detecting offshore wind turbine anomalies.

3 Methodology

This chapter outlines the methodological approach that would be used to develop, train, and evaluate the predictive maintenance models proposed in this paper. The methodology plan was to respond to two objectives that are correlated. The initial goal was to construct a technically sound and reproducible machine learning pipeline to detect faults in wind turbines using supervisory control and data acquisition (SCADA) data. The second aim was to determine whether the use of meteorological data based on ERA5 reanalysis could lead to a better fault detection performance, especially when the operating conditions are winter offshore, where the environmental severity can be incredibly crucial to both turbine operations and maintenance decisions. To meet these goals, three models were designed and evaluated within a common set of criteria: a SCADA-only CNN (i.e. convolutional neural network), a data-level fusion CNN where all variables in SCADA and ERA5 were fused at the features level and feed into sequence learning, and an architecture-level fusion CNN where both SCADA and ERA5 variables were run through independent branches and combined by adopting a learnable scalar gating mechanism. The general process of the thesis, as depicted in **Figure 8**, shifted to the selection of the turbine and SCADA preprocessing to model creation, season assessment, and ultimate winter caution export to optimize downstream.

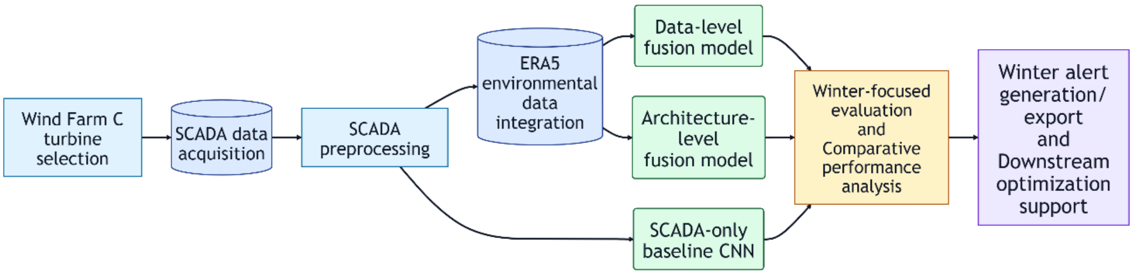

Figure 8. Overall Methodological Approach for the Proposed Offshore Wind Predictive Maintenance Framework

The methodological design gives special consideration to four principles. To begin with, any preprocessing and transformation made was designed in a way that reduced leakage.

Second, windowing logic and train-only preprocessing and shared evaluation metrics across pipelines permitted the comparison of models to be fair. Third, the research diligently segregated the total test assessment and winter evaluation only to bring the research design to meet the main seasonal argument of the thesis. Fourth, the final architecture-level fusion model was later expanded beyond the area of classification by exporting winter alerts in a structured form that could be beneficial to an optimization step in the maintenance optimization problem based on mixed-integer linear programming.

3.1 Research Design

This research employed a comparative experimental approach, where three models for predictive inference were constructed and tested using the same data processing and testing procedures. The design was progressive in nature. Rather than providing meteorological information in its most advanced form, the study started with a baseline model, then elaborated towards structured fusion strategies. This allowed a better understanding of the contributions of environmental information to the detection process.

The initial step was a replicated baseline 1D-CNN model (Hassan, 2025), trained solely on a cleaned and PCA-compressed SCADA dataframe. This formed the baseline for assessing the effect of fusion strategies. The second model presented the data-level fusion whereby the ERA5 weather variables were added to the cleaned turbine data frame and then train-only scaling, dimensionality reduction, and CNN classification were performed. The third model was architecture-level fusion, which involved separate processing of SCADA and ERA5 data and then fused at the learned embedding phase by a trainable scalar gating mechanism.

A winter evaluation strategy was also investigated in this study. Instead of a dedicated winter model, all models were trained on the entire season range to ensure robustness. Winter was then used to provide a dedicated subset of the held-out test data. This enabled the thesis to rigorously address its main point: if the fusion models showed the greatest relative benefits in winter, this would suggest that weather context is particularly useful in adverse seasonal operating conditions.

3.2 Data Sources

3.2.1 SCADA Data Source

The main data set used in this study is the "CARE to Compare" wind turbine anomaly data set (Gück et al., 2024). This includes real SCADA data recordings from three different wind farms: Farm A (onshore, Portugal), and Farms B and C (offshore, Germany). The CARE wind turbine data set was the main source of data for this work, specifically Wind Farm C files. Our working directory initially included 116 CSV files (turbine records). But it contained duplicate files for some turbines, usually in different delimiter formats (e.g. with normal names and names prefixed by comma_). To resolve this, deduplication was performed. A single file associated with each turbine identifier was kept, preferring the non-comma_ file when both were present. This brought the list down to 58 turbines.

3.2.2 ERA5 Meteorological Data Source

In order to introduce environmental data, this study downloaded ERA5 reanalysis data from the Copernicus Climate Data Store. ERA5 was selected since this dataset provides physically meaningful atmospheric and surface variables on a regular time basis and is suitable for estimating environmental conditions at offshore sites when climate conditions from the sites are not completely available or well integrated with the turbine-level data set.

ERA5 data were downloaded based on the UTC time range needed by the turbines those were selected. Data was downloaded monthly from January 2022 to January 2024 for a selected bounding box encompassing the German North Sea. The following 6 ERA5 variables were selected as environmental information relevant for offshore turbines: air temperatre (t2m), dewpoint temperature (d2m), wind component (u10), wind component (v10), total cloud cover (tcc) and sea surface temperature (sst). These raw fields were used to calculate five fields: air temperature in degrees Celsius, dewpoint temperature in degrees Celsius, resultant wind speed $\left(\sqrt{u_{10}^2 + v_{10}^2}\right)$, cloud cover fraction, sea surface temperature in degrees Celsius.

ERA5 should be considered an environmental context rather than a turbine sensor data source. Its value in this study is not to substitute SCADA data, but rather it should be used to give a broader atmospheric and marine context which can explain anomalous SCADA behavior, in particular, offshore operation in winter.

3.3 Turbine Selection Logic

Raw turbine files included operational variables and status labels as a timestamp. A turbine selection strategy was used on the basis of the percentage of anomalous records on the prediction section of each file to support focused but analytically significant development of the models. The binary target variable in this thesis was developed using the field status_type_id. When the target label had been converted to numeric form, its definition was as follows:

$$y_t = \begin{cases} 1, \text{if } status\_type\_id > 0 \\ 0, \text{if } status\_type\_id = 0 \end{cases} \qquad (1)$$

where $y_t = 1$ represents an anomalous or fault related state and $y_t = 0$ represents normal state.

There was consistency in the application of this binary definition within the three models. The last chosen subset was the three most anomalous turbines and the three most normal-like turbines, which made a six-turbine experimental sample that represented a broad range of fault prevalence.

This methodologically significant choice strategy was based on two reasons. To begin with, it established a training and evaluation environment, which involved a heavy-fault and predominantly healthy operational regime. Second, it minimized the risk of constructing a model that was only able to gain knowledge by observing one end of the fault distribution. The study maintained meaningful variability in turbine condition profiles by incorporating turbines with a range of anomaly fractions (between about 0.0058 and 0.8069).

3.4 SCADA Data Cleaning and Preprocessing

The data acquired by the SCADA implemented in wind turbines is usually dirty, inaccurate, and predisposed towards sensor instability. This means that raw turbine data cannot be used directly to process sequence learning without first being processed. Preprocessing was considered as an important methodological aspect in this study and not a small implementation issue. Its goal was to maintain significant temporal dynamics without deterring the impact of missingness, sudden impulsions, and sensor effects that otherwise can alter model learning.

Our thesis-style pipeline was made up of delimiter detection and timestamp validation, chronological sorting, handling missing-value with gap-sensitive rules, roll-window outlier correction, filtering and missingness at row level, feature-train only filtering, and model safety final fill.

Automatic delimiters detection was used to first read each turbine file since files were separated by semicolons or commas. The time stamp column was changed to UTC-aware Datetime. The invalid timestamps in records were deleted, and the rest of the rows became arranged by the ascending chronological order. This time sequence was necessary since all the following processes of the study hinged on maintaining the true operational sequence of each turbine.

Missing-value treatment has been developed so that short local gaps, which can be reasonably imputed, and long missing runs, which should not be aggressively reconstructed are separated. Each numeric feature had any runs of five consecutive missing values or fewer replaced with short gaps and subsequently filled in with forward fill and then backward fill. Any longer gaps were considered missing, but the linear interpolation was done with only a restricted range of three observations in both directions. This discouraged the establishment of artificial long trajectories which were literally unobserved.

Following missing-value management, each feature was subjected to a two-stage rolling outlier treatment consisting of rolling z-score clipping and rolling MAD replacement. In rolling z-score clipping, the centred twelve-observations window was employed and three standard deviations were used as the clipping threshold. The z-score at time step t was calculated as:

$$z_t = \frac{x_t - \mu_t}{\sigma_t} \qquad (2)$$

Where t is the time step, x_t is the observed value of the feature at that time step, μ_t indicates the rolling mean around time step t, σ_t is the rolling standard deviation around time step t and z_t is the standardized value (z-score) which is the distance between x_t and the local mean itself in terms of standard deviations.

When the absolute v-score was greater than three, it was trimmed down to the local boundary. Rolling median absolute deviation replacement was used in the second stage. With a rolling median and a threshold of five, the values which had an unusually large absolute deviation around the local median were substituted with the median itself. The two phases collectively mitigated implausible spikes while preserving the typical temporal variance of turbine signals.

Following the management of missing values and the outlier treatment, a tight row-level missingness filter was implemented. Any row that had over twenty percent of missing features values were eliminated. Then at the stage of train, validation and test split, the feature level filter was performed with only the training split. Any attribute whose percentage of missingness was above 40% in the training data was dropped, and the identical set of attributes imposed on validation and test data.

Lastly, missing values were filled after filtering with forward fill and backward fill to create complete number matrices that can be fed into CNN. This last fill process was used as a polish after rows of poor quality and too sparse features had been eliminated.

3.5 Data Preparation: Temporal Splitting, Train-Only Preprocessing, and Sequence Construction

In all 3 pipelines, cleaned turbine dataframe was chronologically divided into 70% of training, 15% of validation, and 15% of testing. This time period division was done separately on each turbine prior to the construction of the windows. A time-based split is better than a random row-based split since it is more realistic to predictive environment where a model is trained with past turbine behavior, and then used in unseen periods in the future.

The original SCADA feature space had around 952 features at each time, which is high-dimensional to a concise CNN. In order to lessen repetition and normalize learning, a train-only preprocessing stack was used: RobustScaler, StandardScaler and principal component analysis. Only the training split was fitted with PCA, which was used to retain 100 principal components. The fitted training scalers and PCA model were used to transform validation and test data.

This step was performed on a file-to-file basis and sequence concatenation of the turbines followed. The study ensured that preprocessing fit had a clean experimental boundary between observed historical data and the unseen future assessment data, and it was confirmed that preprocessing fit was restricted to the training data.

The three models used fixed-length temporal windows of twenty timesteps, each of which corresponded to ten minutes. Here thus two hundred minutes of the recent turbine history were archived in each window. The sequence label was determined by the last timestep in the window and the sequence timestamp was also derived by the last timestep. The winter subset was selected based on the month of the window-end timestamp, so this design was particularly important to the winter evaluation.

The construction of windows was done following train-only preprocessing. In the case of the baseline and data-level fusion pipelines, the shape of every window was twenty and one hundred and then transposed to channel-first format. In the case of the architecture-level fusion pipeline, each sample had two matching windows; a SCADA window with shape twenty by one hundred (20, 100) and an ERA5 window with shape twenty by five (20, 5).

3.6 ERA5 Alignment and Fusion Preparation

3.6.1 ERA5 Temporal Alignment

ERA5 hourly data were initially resampled to a ten-minute time lag using a forward fill. A clean merge table including a timezone-aware UTC timestamp column was then created. In order to bring ERA5 and SCADA in a certain order, the SCADA calendar was capped to the closest ten minutes, and merge asof was run with direction set to backwards and an exact match maximized. This backward-only alignment method has been selected to prevent leakage.

3.6.2 Data-Level Fusion Construction

The ERA5 variables were added directly to the cleaned turbine dataframe in the data-level fusion pipeline and then train-only scaling and PCA steps were performed. This generated one overall space of fusion where SCADA and weather variables were considered together as a feature matrix. The advantage of this strategy is due to its simplicity but also presupposes that the combination of environmental and turbine-operational variables must occur before the representation learning.

3.6.3 Architecture-Level Fusion Construction

To implement fusion at the architecture level, SCADA and ERA5 variables were already maintained as independent features. As previously described, the SCADA branch was subjected to the robust scaling, standard scaling, and PCA pipeline. A train-only StandardScaler was used to standardize the ERA5 branch, with no PCA applied, as the ERA5 branch had only five variables that were interpretable enough on their own. This branch preprocessing was able to save the statistical difference between the modalities on a per-branch basis.

3.7 Model Architecture Specification

To build the predictive maintenance models in this study, convolutional neural network architectures for multivariate temporal sequence classification were used. CNN-based models are highly appropriate for extracting hierarchical temporal features from structured turbine-operational sequences, while keeping the computational efficiency. **Figure 9** shows a generalized CNN-based sequence learning structure.

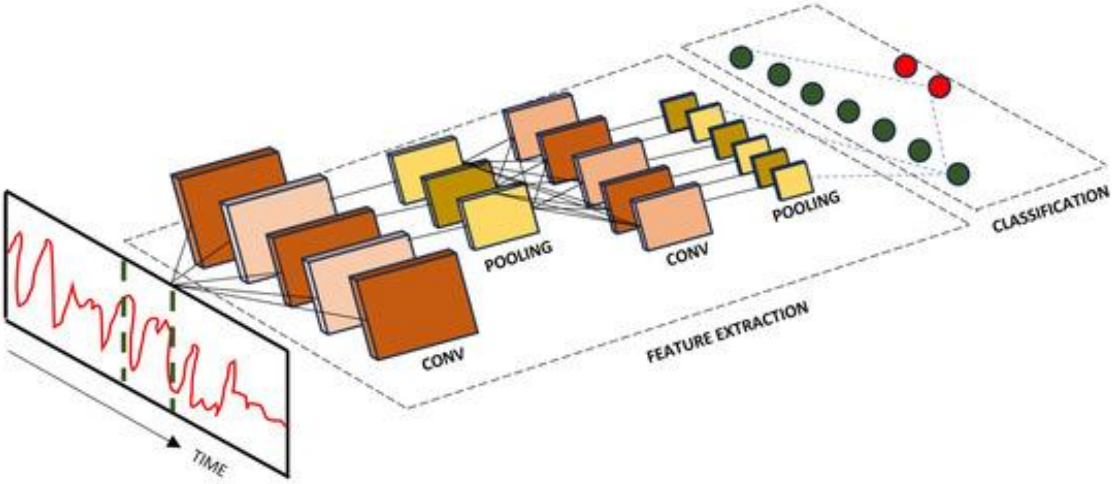

Figure 9. General CNN-Based Temporal Feature Extraction and Classification Architecture (Santiago et al., 2024)

3.7.1 Baseline SCADA-Only CNN

The baseline model consisted of a one-dimensional convolutional neural network that included two convolutional blocks and a fully connected classifier. The SCADA input was in the form of PCA compressed windows of shape one hundred and twenty. Its architecture consisted of two convolutional blocks with batch normalization, ReLU activations, max pooling, dropout, and a fully classification head consisting of two hidden dense layers and a final output layer which produced a single logit. The overall preprocessing, baseline modelling, and winter evaluation workflow of the SCADA-only pipeline is shown in **Figure 10**.

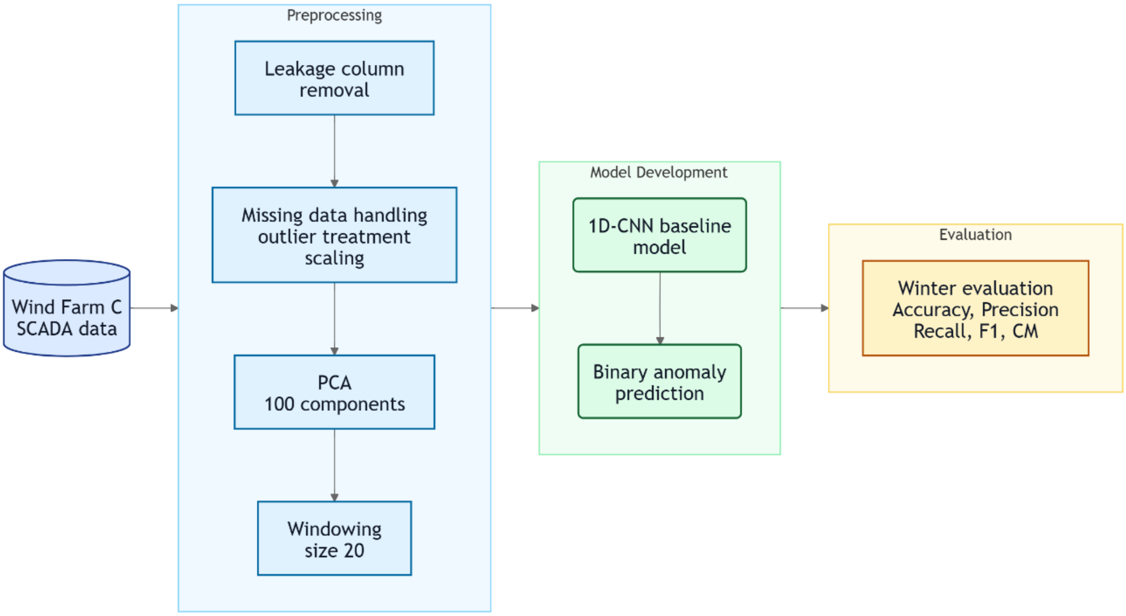

Figure 10. Workflow of the Baseline SCADA-only Anomaly Detection Pipeline

The baseline architecture was carefully built to be a small sequence-learning network, not a very deep network. This design decision contributed to maintaining the interpretability of the model and preventing overfitting while learning from temporally structured SCADA sequences. The fully connected classifier was applied to convert the learned sequence representation to a binary anomaly prediction output, while the convolutional layers were able to capture local temporal patterns and operational transitions in the PCA-compressed turbine signals.

3.7.2 Data-Level Fusion CNN

The data-level fusion model employed the identical CNN structure as the baseline model. The only distinction was that now the model input was a combination of a fused feature matrix with SCADA-derived and ERA5-derived data. Holding the classifier fixed was done to isolate the effect of feature level fusion itself. The complete workflow of the data-level fusion pipeline, including temporal alignment, feature-level integration, preprocessing, and CNN-based anomaly classification, is illustrated in **Figure 11**.

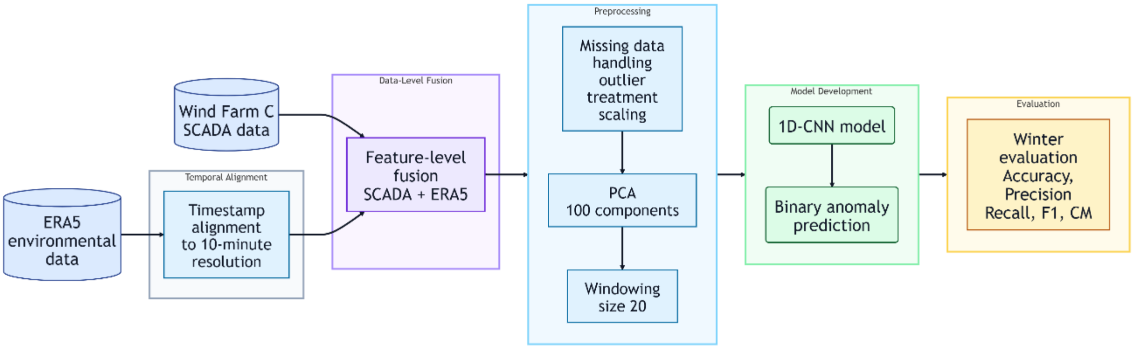

Figure 11. Workflow of the Data-Level Fusion Pipeline Integrating SCADA and ERA5 Variables Before CNN-Based Anomaly Detection

In this pipeline, the meteorological variables were directly fed into the same feature space as the SCADA variables before sequence learning. This way, the CNN was able to learn both operational and environmental information together from the unified representation. The experiment kept the same downstream architecture as the baseline model, eliminating the effect of increasing network complexity or changing the structure of the classifier on the proportion of improvements in anomaly prediction.

3.7.3 Architecture-Level Fusion CNN

The architecture-based model used a dual-input CNN. The SCADA sequence was processed by the first branch and the ERA5 sequence was processed by the second branch. A sixty-four (64) dimensional hidden embedding was achieved in the SCADA branch. ERA5 branch was trained on a thirty-two (32) dimensional pooled hidden embedding. One important methodological input of the study is the fact that the environmental embedding was not merely concatenated with the SCADA embedding. Rather, before fusion a learnable global scalar gate, α, was applied to the ERA5 embedding.

$$h'_{\text{env}} = \alpha \cdot h_{\text{env}} \tag{3}$$

where $\alpha \in \mathbb{R}$ is a trainable scalar parameter, starting with a value of 1.0. The resultant fused representation is:

$$h_{\text{fused}} = \left[h_{\text{scada}}, h_{\text{env}}^{'}\right] \qquad (4)$$

This fused vector is then passed through the classifier:

$$\hat{y} = f_{\text{classifier}}(h_{\text{fused}}) \qquad (5)$$

This mechanism is a learnable global volume control, which modulates the overall input of the environmental branch to the ultimate prediction. The entire architecture-level fusion workflow, including the dual-branch representation learning and learnable scalar gating mechanism, is presented in **Figure 12**.

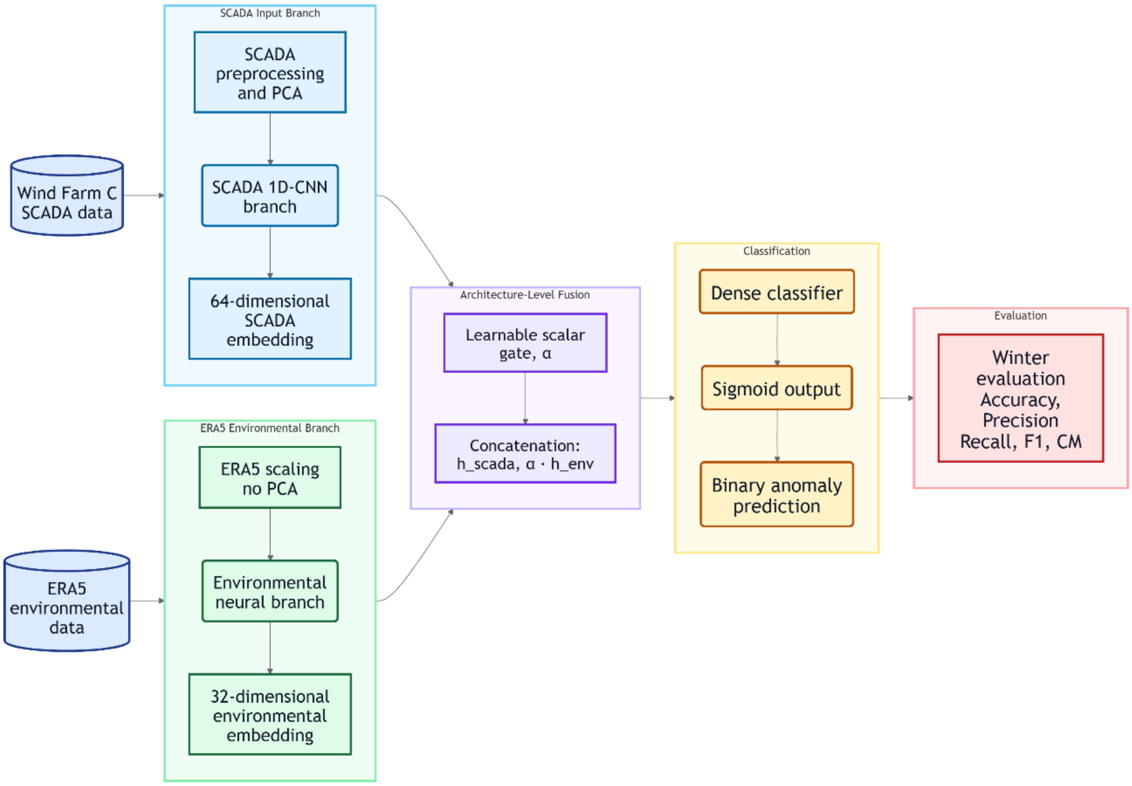

Figure 12. Dual-Branch Architecture-Level Fusion Model with Learnable Scalar-Gated Environmental Integration for Anomaly Detection

3.8 Class Imbalance Handling and Data Augmentation

The wind turbine fault detection is an imbalanced classification problem. In the baseline model, the loss function BCEWithLogitsLoss took a positive-class weighting factor of the training distribution. In the case of the fusion models, a target positive class ratio of 0.20 at the time of batch construction was approximated through the use of a WeightedRandomSampler. Besides this, dynamic Gaussian noise injection was applied with a standard deviation of 0.02 and only on the training stage as well as on positive samples. In the architecture-level fusion model, this augmentation was only done to the SCADA branch but not on the ERA5 branch. Without augmentation, validation and test datasets were instantiated.

3.9 Model Training Procedure

Each model was implemented in PyTorch and trained with mini-batch gradient descent and Adam optimizer and an initial learning rate of 10^{-3}. Adam was selected due to its adaptive learning ability being suitable in high-dimensional time-series data and enabling stable convergence in varying model configurations. To achieve effective training and stability of the gradient, a batch size within the capacity of available computational resources was employed.

A ReduceLROnPlateau scheduler was used to enhance convergence behaviour. This scheduler kept track of the validation loss and decreased the learning rate by half when it did not improve after three consecutive epochs. This dynamic adaptation enabled the models to shift between coarse and fine-grained learning especially during late training phases.

To stabilize the parameter updates, particularly at much deeper convolutional layers, gradient clipping with a maximum norm of 1.0 was used, as this helps reduce the likelihood of gradient explosion. Each model was trained up to 50 epochs. Nevertheless, they used early stopping to avoid overfitting with a patience of 10 epochs and a minimum improvement of 10^{-4}. Early stopping was monitored using loss of the pooled monitoring validation (MonVal) set.

In training, model checkpoints were stored whenever the validation loss decreased below the set threshold. Training was ended in case no further improvement was noticed during the patience period, and the model parameters that made the least loss on validation saved. This guaranteed that the end model performance represented the optimal generalization point as opposed to the last training epoch. All three model configurations underwent the same training process to facilitate a fair and controlled comparison of the baseline, data-level fusion, and architecture-level fusion methods.

3.10 Evaluation Strategy

3.10.1 Overall Evaluation

The initial testing phase determined the capacity of all three models on the entire combined test set. This step was used in order to determine the general model validity. Fusion that only works in a seasonal subset but not in the wider operating distribution would not be convincing enough as a robust predictive maintenance solution.

3.10.2 Winter Subset Evaluation

A second set of evaluation was performed on a subset of the held-out test set during the winter. The month of the window-end timestamp was used to define winter.

$$\text{Winter} = \{\, t\text{:month}(t) \in \{12,1,2\}\} \tag{6}$$

The winter subset encompassed all the test windows that concluded with December, January or February. Notably, this subset was outlined once the models were trained and after building the entire held-out test set. The models were not re-trained using winter data. This maintained methodological rigour and directly tested the hypotheses that environmental fusion should prove maximum advantages in severe offshore winter circumstances.

3.10.3 Evaluation Metrics

This study employed five complementary measures of classification to assess predictive performance of proposed models: accuracy, precision, recall, F1-score, and the confusion matrix. The metrics were calculated on the held-out test data and interpreted on the operational goals of offshore wind predictive maintenance. Accuracy, precision, recall, and F1-score have continued to be relevant in classification research as they are understandable and results can be directly related to real-world decisions, although applying these metrics may demand various focuses depending on the context of the evaluation (Yacouby & Axman, 2020).

For the purposes of this thesis, the positive class is related to a state of anomaly or fault, and the negative class is related to a normal state. Where TP, TN, FP and FN represent the numbers of true positives, true negatives, false positives and false negatives respectively (Hassan, 2025). The confusion matrix gives the fundamental framework upon which all four scalar metrics can be obtained and is particularly effective since it indicates a model is erroneous not simply that it is correct. This is especially pertinent in the case of offshore maintenance, when a false alarm and a missed fault are not equally significant in terms of operational impact. False positive can lead to undesirable planning work or dispatched pressure, whereas false negative can lead to a real turbine fault left to run wild. This generalized cost asymmetry also aligns with the way Hassan formulates offshore fault prediction, which measures the precision and recall in the context of the practical implications of false alarms and false negative outcomes (Hassan, 2025).

Accuracy is the total ratio of the correct classification of observations, and it is defined to be:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (7)$$

Here, TP = the number of anomalous windows labeled anomalous, TN = the number of normal windows labeled normal, FP = the number of normal windows labeled anomalous, and FN = the number of anomalous windows labeled normal. The usefulness of accuracy is that it can be calculated as a single general summary measure but it cannot be interpreted individually when the dataset is imbalanced, as it might provide an overly optimistic view of model performance.

Precision is used to measure reliability of positive predictions and it is defined as:

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (8)$$

where TP = the value of correctly identified faults and FP = the value of the number of false positives. Precision thus represents the frequency of an elevated alarm to an actual fault. This measure is especially crucial in the offshore maintenance context since false alarms can result in unnecessary planning, avoidable vessel mobilization or unnecessary operator intervention. (Hassan, 2025) also explains precision as the extent of consistent warnings in comparison to actual pre-fault events and attributes this directly to reducing unnecessary maintenance measures.

Recall, also known as sensitivity, is the ratio of actual positive cases which the model identifies correctly:

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (9)$$

where TP once again represents the quantity of anomalies found accurately and FN represents the quantity of missed anomalies. The question answered by recall is: of all real anomalous cases, what fraction of those were captured by the model? High recall is essential in predictive maintenance as false alarms can cause the true defects to go unremedied/untreated until they turn into more serious failures. Traditional arguments on precision and recall point out that the two are indicators of varying dimensions of retrieval performance and that a trade-off between them is often present in real-world systems (Buckland & Gey, 1994). If one gets better, it influences the other and that is why they should be reported together and not separately.

Precision and recall tend to vary in tension and therefore F1-score is used to balance both indicators, thus the F1-score helps to reveal only one number that represents the trustworthiness and completeness of the faults (Leahy et al., 2018; Stetco et al., 2019). It is the harmonic average of recall and precision:

$$F_1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \qquad (10)$$

where Precision is the proportion of true positive to the total number of predicted positive, and Recall is the proportion of true positive to the total number of the actual positive. F1-score was specifically helpful in this work since the fundamental goal was not only to maximize the detection of anomalies, but also to identify a practically acceptable compromise between reliability of alerts and fault detection. (Hassan, 2025) used the same interpretive logic in introducing F1-score as a single indicator of reliability as well as completeness of fault detection in predictive maintenance through class-imbalanced conditions.

Such assessment metrics give a more comprehensive view of model performance than individual indicators. A broad measure of overall correctness is the accuracy, alert reliability is quantified by precision, fault sensitivity is quantified by recall, and their combination is summarized by the F1-score. (Yacouby & Axman, 2020) also observe that judgment decisions must be linked to the practical conditions of the classification task, which exist in the real world and this observation correlates with the current research on the needs of anomaly detection of winter under offshore maintenance conditions.

3.11 Export of Winter Alerts for Optimization

After establishing the architecture level fusion model as the most powerful winter performer, its winter test predictions were brought out in structured format to be later integrated with an operations research solver. The last alert file had 7,995 winter alerts and consisted of the following fields, Task_ID, Timestamp, Turbine_ID, True-Label, Predicted_Probability, ERA5_Wind_Speed, ERA5_Temperature, and ERA5_Sea_Temperature.

3.12 Chapter Summary

The entire methodological framework of the study has been unveiled in this chapter. It explained the CARE SCADA data and ERA5 environmental data, the six-turbine selection rationale, data cleaning and preprocessing regulations, the train- only transformation pipeline, and the sequence building process. It also outlined the 3 modelling strategies, the class-imbalance treatment, the training process and the evaluation strategy. Above all, the chapter institutionalized the two-layered assessment plan of the thesis: general test assessment to identify the broad level of robustness and winter subset assessment to verify the main hypothesis on the seasonality.

4 Model Design and Analytical Assessment

This chapter describes the development of the three predictive models based on the cleaned turbine and environmental data, and the development of the logic used in the analysis based on the baseline SCADA-only model to the two fusion-based options. Although Chapter 3 defined the methodological framework in formal terms, the current chapter is more about the process of model development: how the characteristics of the data influenced the design decisions, how the strategies of data fusion varied in their analytic nature, and what the process of the pipelines tell us about the problem of classification.

4.1 Development of the Baseline SCADA-Only Model

The baseline model was first constructed since any improvement based on credible fusion must be evaluated against a healthy SCADA-only reference. The six selected files were then put through the data cleaning pipeline outlined in Chapter 3. The size of the prediction-segment samples after cleaning differed significantly among turbines, as turbine operations and prevalence of anomalies are heterogeneous. Turbine 16, as an example, had 2,304 observations left after cleaning with a positive rate of 0.8069 whereas turbine 62 only had 1,032 observations with a positive rate of just 0.0058.

Train only scaling followed by PCA resulted in the pooled baseline sequence corpus consisting of 228,027 training windows, 48,774 monitoring-validation windows and 48,777 test windows. These numbers indicate that the model was trained on a big sequence-level dataset as opposed to a small and hand-selected set of events. The baseline thus offers a good empirical perspective of comparison.

4.2 Analytical Interpretation of the Baseline Representation

The default model was a 1D-CNN on PCA-compressed SCADA sequences. The hypothesis behind this design is that recent local temporal dynamics in turbine behavior hold enough information to differentiate anomalous and normal conditions. The PCA step compressed the original size of the feature matrix of about 952 numerical variables in a single timestep down to 100 principal components, enabling the network to learn on dense latent patterns instead of a very expansive raw sensor space.

The training log illustrated that the baseline was learned successfully. Training loss was decreasing at a gradually decreasing rate, with 0.2511 at the first epoch through 0.0709 at the last epoch, and validation loss was increasing steeply initially and maintaining the trend. The optimal validation checkpoint was regained through early stopping. This takes into consideration that the model was large enough to learn the sequence task but it also required regularization and early stopping against overfitting.

4.3 Motivation for Extending Beyond SCADA Alone

Even though the base was very strong, its emerging nature revealed a general weakness of only SCADA learning. SCADA signals indicate internal or proximate operation of turbines, without directly indicating why the turbine might be operating abnormally. Abnormal SCADA trends in offshore wind can be an indication of true internal faults; however, they can also show forced operating stress, particularly during winter conditions.

This issue is particularly important in case the target of operation goes further than an anomaly capture on a raw level to the actual maintenance planning. A model minimizing the number of weather-related yet non-faulty deviations as anomalies can have good recall but will also experience false alarms as well. The following two pipelines were then constructed to determine whether the environmental context could assist the model to distinguish between operational irregularity and the actual fault-like behavior.

4.4 Development of the Data-Level Fusion Model

The data-level fusion model maintained the original CNN structure but changed the input construction phase. ERA5 weather variables were downloaded, converted into five engineered environmental variables to match the cleaned turbine data temporally, a merge strategy (backward-only) was used. These are then directly added to the SCADA dataframe followed by train-only scaling and PCA.

Following the preprocessing, the fused cleaned dataframes were significantly bigger than the prediction-only files at the baseline model as the dataframes retained training and prediction rows prior to the 70/15/15 chronological split was applied. These cleaned fused data frames which were cleaned per file had an approximate size of around 53,000 to 56,000 rows and positive rates of between about 0.0487 to 0.1558. The positive ratio of the pooled fused training set was 0.0764 and the test set increased to 0.1662 which was a temporal change in the prevalence of anomalies.

4.5 Analytical Interpretation of Data-Level Fusion

The primary analytical characteristic of data-level fusion is early entanglement. The model used ERA5 variables directly within the input matrix, and thus, it used environmental context in the same feature space as SCADA until the commencement of representation learning. This is a practical form of simplicity; along with it, a strong assumption must be made that information about weather and turbines in operation is collectively compressed prior to the network learning the temporal attributes.

Since the SCADA block is very high-dimensional and contains a lot of signal structure specific to turbines, the weather information can be diluted or merge into latent directions that contains SCADA variance. The subsequent test outcomes demonstrate that the model turned out to be more selective when making positive predictions that led to fewer false alarms and a greater number of missed true anomalies. This fits with a classifier that has been made conservative with the addition of environmental context but does not have a formal mechanism to determine how the environmental context ought to impact the final verdict.

4.6 Development of the Architecture-Level Fusion Model

Architecture level fusion model came in an attempt to overcome the representational drawbacks of the data-level fusion. This model did not try to force SCADA and ERA5 variables into a common feature space at the outset, but instead, considered them as individual, yet complementary modalities. The stronger representational pipeline was adopted in the SCADA branch due to the high-dimensional SCADA data and centralized operation. ERA5 branch was specifically light, taking into consideration the lesser and more contextual role of environmental variables.

Following preprocessing, each sample was made up of two synchronized temporal windows: a SCADA sequence of shape twenty by a hundred (20, 100) and an ERA5 sequence of shape twenty by five (20, 5). These were convoluted separately through branches. The SCADA branch generated a hidden embedding of dimension sixty-four; post-temporal-pooling, and the ERA5 generated a hidden embedding of dimension thirty two.

The addition of the global scalar gate to the environmental branch was the most crucial step of the development. This enabled the network to learn the importance with which the ERA5 embedding should provide a contribution with respect to SCADA embedding. This architecture is particularly well-suited to the issue tackled in this thesis weather is a factor but not projected to control turbine-internal failure signatures. This is precisely the intuition that is captured in section 3.7.3 of the global gate in learnable mathematical form.

4.7 Interpretation of the Learned Environmental Gate

One of the most informative indicators of the balance between the SCADA and meteorological information in the architecture-level fusion model is the evolution of the environmental gate during the training. The gate was set to 1.0, which meant that, at the onset of the training, the ERA5-computed embedding did not receive any attenuation and was equally influential as the SCADA representation.

Nevertheless, we can find a persuasive and systematic fine-tuning training path. The gate dropped to about 0.791, as it was in the initial generation, demonstrating that there was an early decrease in environmental impact. It depreciated even more to approximately 0.470 by epoch 4, as the contribution of ERA5 continued to be reweighted. With further training, the gate tended to stabilize around a smaller range, changing between some 0.22 to 0.26 and finally, reaching towards 23.8 per cent at the most successful validation point.

This is an analytically important development. When the gate went down to zero it would indicate a lack of useful information in the environmental variables but a value near one would be an indication that meteorological context was dominating. Rather, the fact that the percentage seems to converge to about 23-25 percent of the original scale, indicates that the model still had a stable yet scaled down reliance on environmental features.

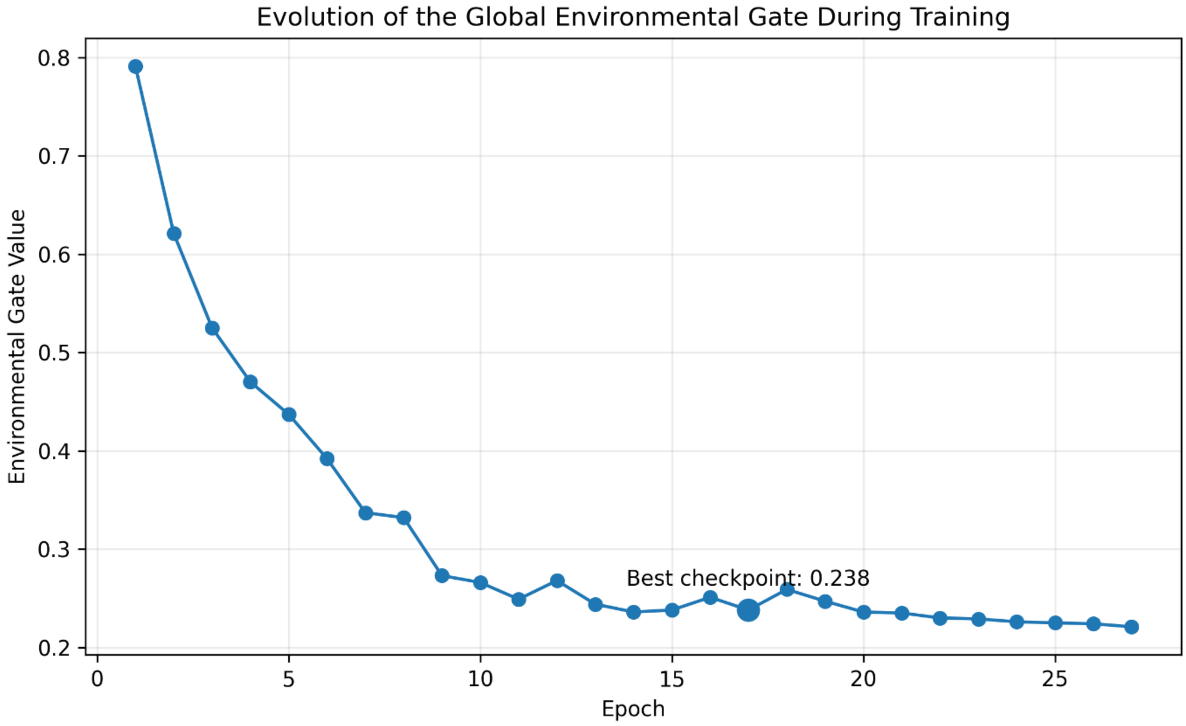

Figure 13. Evolution of the Global Environmental Gate During Training

This leads to the interpretation that ERA5 variables were complements to SCADA signals and not substitutes. The gate substantially was trained to down-weight big-environmental information but maintain the performance of figuring out unclear operational pattern. This implies that meteorological information is not the main driver in classification of anomalies but more of a contextual adjustment mechanism.

4.8 Class Imbalance and Augmentation Analysis

The data distributions that were experienced in the development of models justify the need to handle the issue of class imbalance. There was no even distribution of fault windows and relative prevalence of positive labels varied among turbines and over time. The paper has tackled this by using weighted loss or weighted sampling as well as by using dynamic Gaussian augmentation of positive training sequences.

The fusion model at the architecture level used this augmentation on the SCADA branch only. This is analytically justified since SCADA data is high-dimensional, and it is subjected to small perturbation-based regularization, whereas ERA5 variables are low-dimensional contextual features that must not be physically un-realistic but should be instead binary and physically coherent features.

4.9 Model Configuration Decisions and Empirical Rationale

The ultimate framework of the three model pipelines was developed using controlled and comparison-based design. The primary objective of the experiment was to estimate the impact of meteorological fusion in a fair and reproducible configuration, rather than to optimize the performance of any of the models by extensive application of tuning. Due to this very reason, a number of core settings were maintained constant throughout the experiments other than a certain alteration was required in order to assist the fusion mechanism.

The input sequence length was established at 20 timesteps, equating to 200 minutes of turbine history, based on the normal 10-minute SCADA reporting period. This option was an indicator of a compromise between time coverage and redundancy. Longer windows failed to reduce the computational cost and slowed the local fault signature and shorter windows lacked enough operational context prior to predict labels of anomalies. Train-only preprocessing was used to reduce the SCADA representation to 100 principal components to maintain the high-dimensional multivariate structure of the original sensor space and hold the inputs constant and computationally feasible.

The similar 1D-CNN backbone was employed in the case of the baseline SCADA-only model and the data-level fusion model to ensure the comparison. In this backbone, two convolutional steps 32 and 64 channel dimensions, 3 kernel size, 1 padding, and 0.2 dropout probability were adopted everywhere. The classifier was maintained to be fully connected, and the size of the network was 320→128→32→1, such that the network was expressive, yet not too deep or too parameter heavy. This same architecture was intentionally maintained in the data-level fusion model, to ensure that any changes could be explained by feature-level fusion and no longer by structural capacity expansion of the model.

The architecture-level fusion model represented with a fine-tuning instead of a complete redesign. Its SCADA division shared the identical 32 → 64 convolutional progressions as the foundation with the environmental division being intentionally much lighter with 16 → 32 channels per five ERA5 inputs. This asymmetrical architecture was based on the varying roles of the two modalities, with SCADA as the primary high-dimensional operational signal, and ERA5 as low-dimensional contextual input.

The handling of class imbalance was also done in a conservative way. The training process in the fusion pipelines aimed at achieving the positive sampling ratio of 0.20; augmentation of minority-class using Gaussian noise was injected at 0.02. These settings were chosen to increase anomalous sensitivity while without overwhelming the original data distribution or causing unrealistic synthetic variation. In general, the chosen design focused on the stability, interpretability, and the comparison fairness of the three model pipelines.

4.10 Why the Winter Subset Matters for Model Development

The winter subset is the key point to the development of the models as they were developed. The idea behind the thesis is offshore winter maintenance, in which severe environmental factors influence the behaviour of turbines, as well as the logistics of turbine maintenance. Anywhere that the fusion adds value, it is expected to show up most in winter test data.

4.11 Exporting Alerts as a Developmental Extension

A unique aspect of this research is that the ability to develop models was not confined to the classification measures. Once the architecture-level fusion model was chosen as the best winter performer, the winter alerts in this model were exported to a structured, downstream-optimization format. The last export involved 7,995 positive winter alerts. These are not model outputs in the statistical sense; these are potential maintenance tasks.

4.12 Challenges Encountered During Pipeline Construction

Several iterative corrections, stages of debugging and methodological refinements were associated with the construction of the three model pipelines. These issues were especially significant since they directly affected the validity and reliability of the final experimental outcomes. Unintended data leakage was considered one of the first problems during the development of the baseline. First, label and metadata fields were erroneously added in the input feature space which artificially enhanced the model performance. Moreover, there was the stacking of several turbine files prior to temporal splitting and window building, which led to the fact that sliding windows cut through file boundaries. For this, the independence of turbine-level sequences was broken, and it injected a latent temporal leakage. These were subsequently addressed by keeping feature and label spaces strictly separated and performing per-file preprocessing before window generation so that the pipeline maintains consistency with the desired method.

The second problem was a result of inappropriate division of data and the gap between preprocessing implementations. The approach to missing-value and rolling treating outliers as described was not enforced in early implementations of the pipeline. Moreover, the logic of splitting was not within the necessary time frame. The solutions to these discrepancies were to ensure that the methodology was closely adhered to so that all the transformations were executed in the right sequence and only on training data to prevent any indirect information leakage.

A third feasibility challenge pertained to the dataset construction plan. A first attempt to utilize the complete range of time of Wind Farm C was shown to be computationally inefficient and incompatible with the thesis replication design. This was subsequently superseded by the six-file selection (three anomalous and three normal turbines) after much debugging to give a more well-balanced and reproducible experimental setup. The incorporation of ERA5 meteorological information added even more complexity. Difficulties associated with timestamp alignment, rounding to 10 minutes and forward-filling strategies had to be experimented on several times until a consolidated merged dataset was reached.

Class imbalance handling was another major problem. Although normal random oversampling is a common method of enhancing minority-class learning, it is also known to elevate the chances of model overfitting, as it generates same representations of the existing ones (Mohammed et al., 2020). To address this identified shortcoming, our original model was improved with Gaussian noise augmentation and class weights. Once this dynamic noise-injection scheme was in place, performance became in line with expectations.

Lastly, in the extension to fusion models, creating augmented data of multiple modalities posed extreme computational limitations. Large and complex data processing inherently conflict with the limits of computational feasibility and memory bandwidth as a general hardware and scaling problem in contemporary deep learning (Thompson et al., 2022). To address these computational bottlenecks and avoid session crashes without compromising the methodological consistency, we did not use our static dataset-based augmentation and instead directly applied the noise injection to our Dataset PyTorch class, which subsequently did not crash during our experiments.

In general, these obstacles emphasize that the final pipeline design was created by the process of correction and validation, instead of an immediate implementation, enhancing the validity of the experimental structure.

4.13 Chapter Summary

This chapter has demonstrated how the three models were implemented in a three-stage maturation of a robust SCADA-only framework to progressively structured fusion frameworks. It has claimed that the baseline model demonstrated the feasibility of CNN-based fault detection of cleaned SCADA sequences, and also revealed the constraints of environmentally challenging SCADA-only interpretation. It has revealed that data-level fusion added beneficial environmental context, however, it used a strategy of a first-mixing that had constraints of representational issues. Lastly, it has shown that architecture level fusion provided a more suitable developmental solution by decoupling modalities, maintaining the statistical identity of modalities and learning a moderate but significant contribution to the environmental contribution by means of a global scalar gate.

5 Results and Discussion

In this chapter the three predictive maintenance models that have been developed in this research are given a systematic assessment. According to the evaluation strategy as outlined in Chapter 3, findings will be presented in two phases. To measure general predictive capability, first, overall test-set performance is investigated. Second, a targeted winter analysis is carried out, taking only the sequences with end timestamps in the month of December, January or February.

The chapter then interprets the findings in an offshore maintenance viewpoint focusing on balance between detection and practicality of operation. The trade-off involving determining actual faults and alerts minimization are of particular interest because such a balance directly determines the maintenance planning, costs efficiency, and reliability of the system.

5.1 Comparative Analysis of Overall Test-Set Performance

Table 1 gives the summarized held-out test-set results of the three model pipelines. The baseline SCADA-only model achieved 83.29 percent accuracy, 49.87 percent precision, 94.18 percent recall, and 65.21 percent F1-score. The data-level fusion model had a 84.70 percent accuracy, 52.51 percent precision, 83.34 percent recall and had the best overall balance with 64.42 percent F1-score and architecture-level fusion model had a 87.43 percent accuracy, 57.63 percent precision, 92.06 percent recall and 70.88 percent F1-score. Broad classification correctness is reported to be represented by accuracy, however, precision, recall and F1-score tend to be more informative about this imbalanced anomaly-detection task. Precision can show the alert reliability, recall is an ability of the model to address real cases of fault and F1-score gives a composite picture of the two objectives.

Table 1. Comparison of Model Performance on Overall Test-Set

Model	Accuracy	Precision	Recall	F1-score
SCADA-only baseline	0.8329	0.4987	0.9418	0.6521
Data-level fusion	0.8470	0.5251	0.8334	0.6442
Architecture-level fusion	0.8743	0.5763	0.9206	0.7088

The architecture-level fusion model recorded the highest overall correctness with an accuracy of 87.43%. The data-level fusion and SCADA-only baseline model achieved accuracy score respectively 84.70% and 83.29%. This trend indicates that the addition of meteorological context can enhance the broad classification performance, but it alone cannot be used to evaluate the model’s suitability in an imbalanced predictive-maintenance environment.

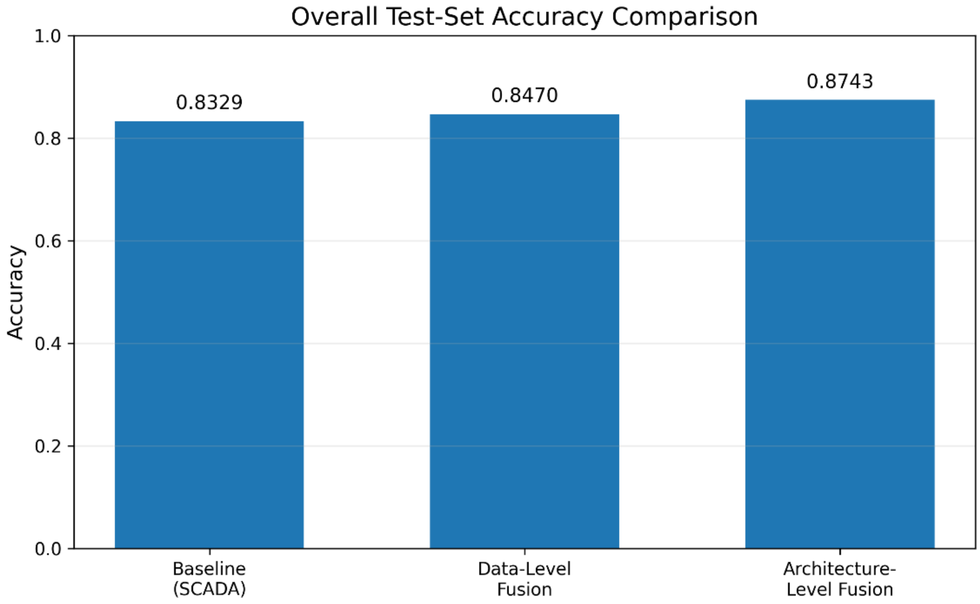

Figure 14. Overall Test-Set Accuracy Comparison

Fusion models enhanced the reliability of positive alerts. The baseline model yielded 49.87 percent precision indicating that its positive classifications had a high percentage of redundant warning. This precision value was boosted by data-level fusion model to 52.51 percent and architecture-level fusion model to 57.63 percent meaning that the dual-branch model resulted in the most reliable overall alert stream.

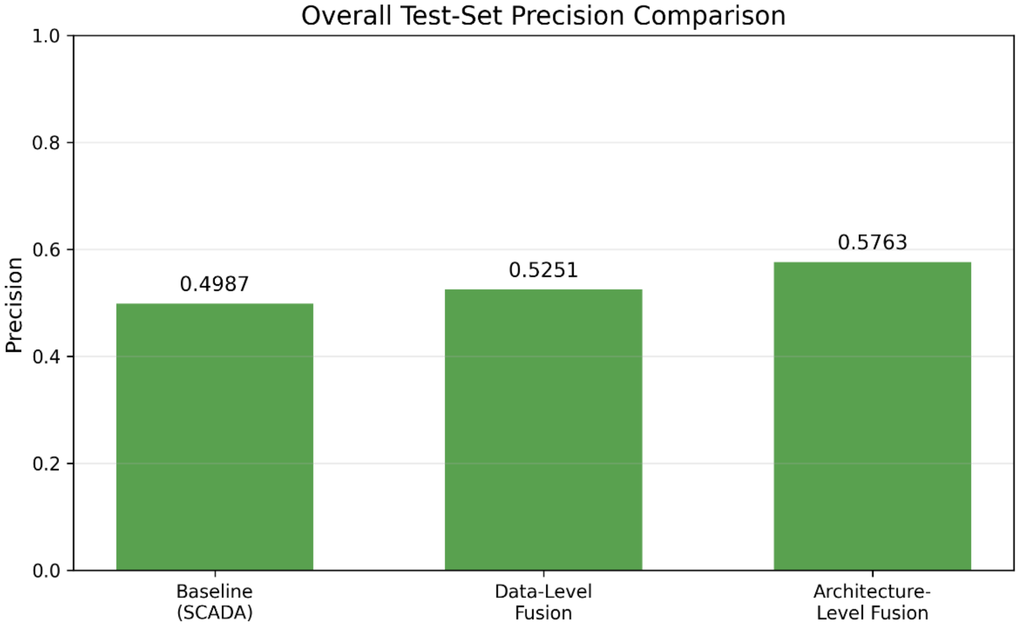

Figure 15. Overall Test-Set Precision Comparison

SCADA-only baseline had the highest share of real anomalies with a recall of 94.18%. The data-level fusion revealed less real fault-related cases and dropped to 83.34%. On the other hand, the architecture-level fusion restored a significant portion of the lost detection activity with 92.06%. This demonstrates that organized fusion retained much of the fault sensitivity of the baseline and it minimized undue faulty warning behavior.

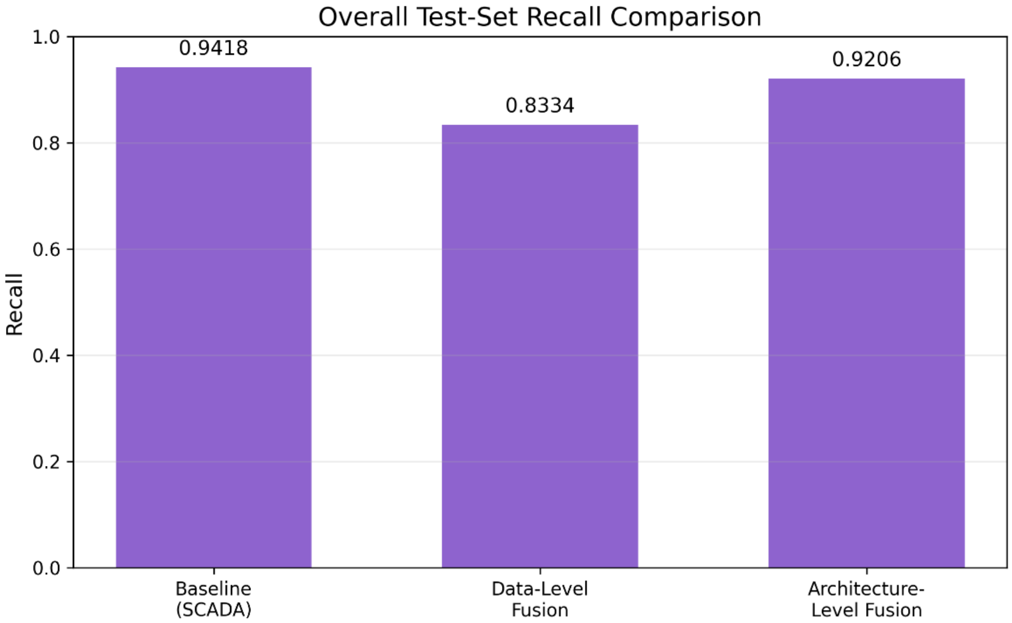

Figure 16. Overall Test-Set Recall Comparison

The architecture-level fusion showed the best combined precision-recall balance with an F1-score of 70.88%. The baseline achieved a score of 65.21% and data-level fusion got 64.42%. Even though the data-level model provided better precision value, its lower capability to identify real anomalies undermined its overall balance. The architecture level model thus provided the most comprehensive overall performance profile.

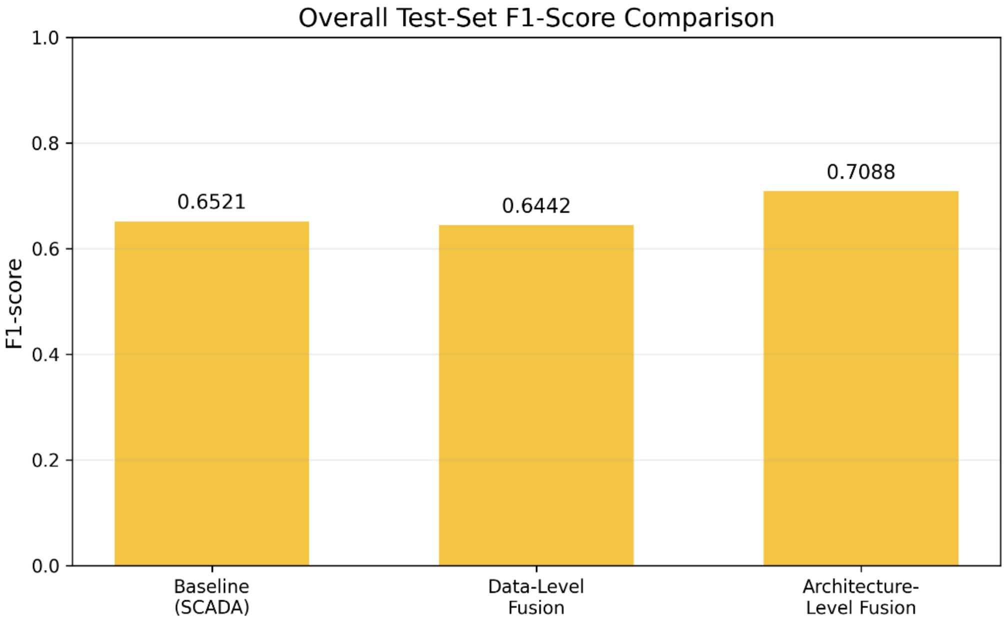

Figure 17. Overall Test-Set F1-Score Comparison

5.2 Comparative Analysis of Winter Test-Set Performance

Table 2 shows the winter test-set results. The SCADA-only model in this subset obtained an accuracy of 84.96 percent, a precision of 64.72 percent, a recall of 95.00 percent and an F1-score of 76.99 percent. The data-level fusion model achieved 90.33 percent accuracy, 81.26 percent precision, 82.49 percent recall, and 81.87 percent F1-score in terms of seasonal alert reliability. The fusion model at the architecture-level came up with the best winter balance of 91.84 percent accuracy, 79.95 percent precision, 92.34 percent recall, and 85.70 percent F1-score. This seasonal analysis is the core of the thesis, since winter is the season when we have limited access to offshore and service choices are most vulnerable to the missed failures and avoiding alerts. The winter subset thus offers a more operationally significant evaluation of whether ERA5-based meteorological fusion enhances model efficacy under severe conditions.

Table 2. Comparison of Model Performance on Winter Test-Set

Model	Accuracy	Precision	Recall	F1-score
SCADA-only baseline	0.8496	0.6472	0.9500	0.7699
Data-level fusion	0.9033	0.8126	0.8249	0.8187
Architecture-level fusion	0.9184	0.7995	0.9234	0.8570

The fusion-based models had significantly increased the overall classification correctness during winter. The best seasonal accuracy of 91.84% was reached through architecture-level fusion, then 90.33% and 84.96% accuracy were achieved by both data-level fusion and SCADA-only fusion model respectively. The increased distance in winter as compared to the full test set implies that environmental information is highly relevant when conditions become harsher.

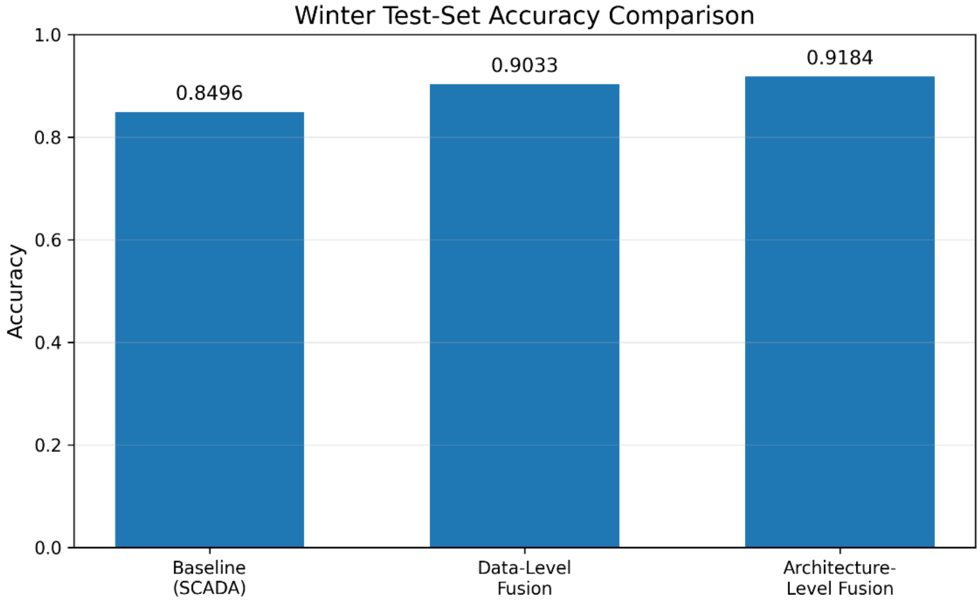

Figure 18. Winter Test-Set Accuracy Comparison

The best enhancement in winter is in alert reliability. The SCADA-only model got 64.72% precision with data-level fusion rising sharply to 81.26%. The architecture-level fusion was much closer with a value of 79.95% which is an indication that the dual-branch model was able to take away a significant portion of the false-alarm load in the baseline model but did not incur the more severe loss of detection sensitivity found in the data-level fusion model.

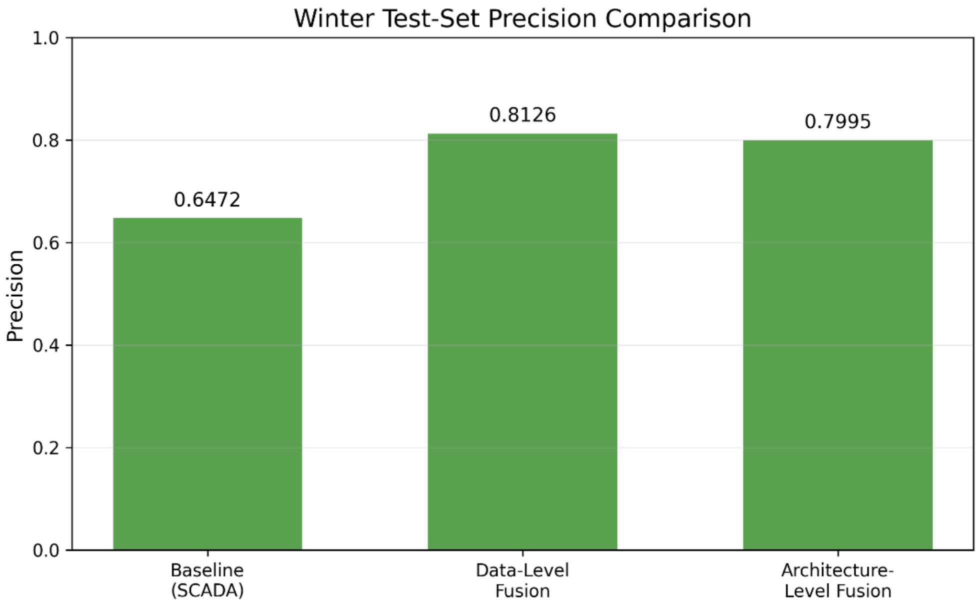

Figure 19. Winter Test-Set Precision Comparison

The SCADA-only baseline model was very sensitive in capturing the actual winter anomalies with recall of 95.00 percent. The data-level fusion dropped to 82.49 percent and this implies that more real fault-based cases were overlooked. Architecture level fusion had a far more robust detection rate of 92.34 percent which shows that it retained much of the safety-oriented sensitivity of the baseline but did not worsen the quality of the alerts.

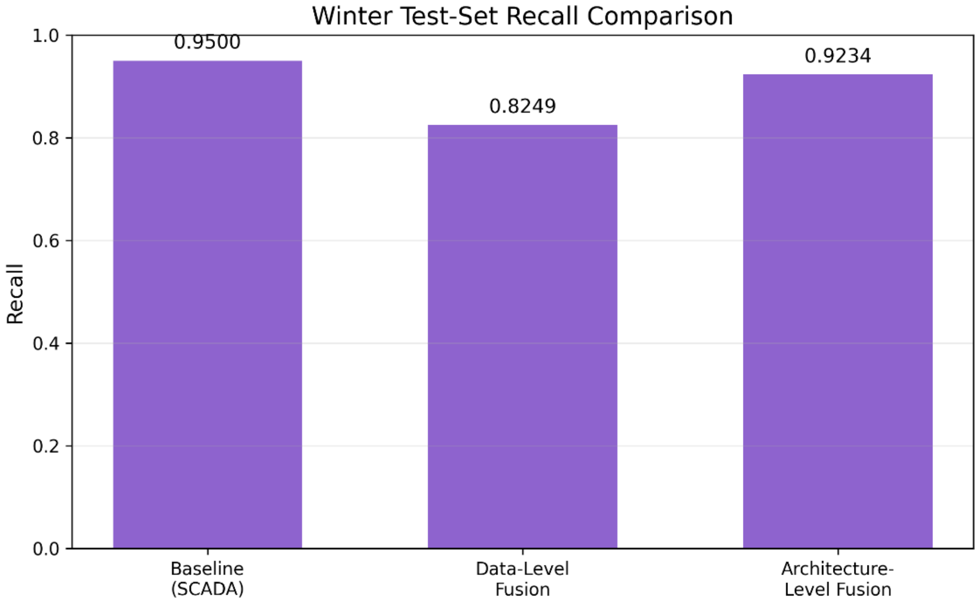

Figure 20. Winter Test-Set Recall Comparison

The benefit of the architecture-level design is confirmed by the winter F1-score. Its score of 85.70% exceeded both data-level fusion (81.87%) and the baseline (76.99%). This finding is particularly significant since the F1-score delivers a sense of reliability of warnings and the successful capture of faults, which is exactly the balance that is needed in the context of offshore maintenance assistance.

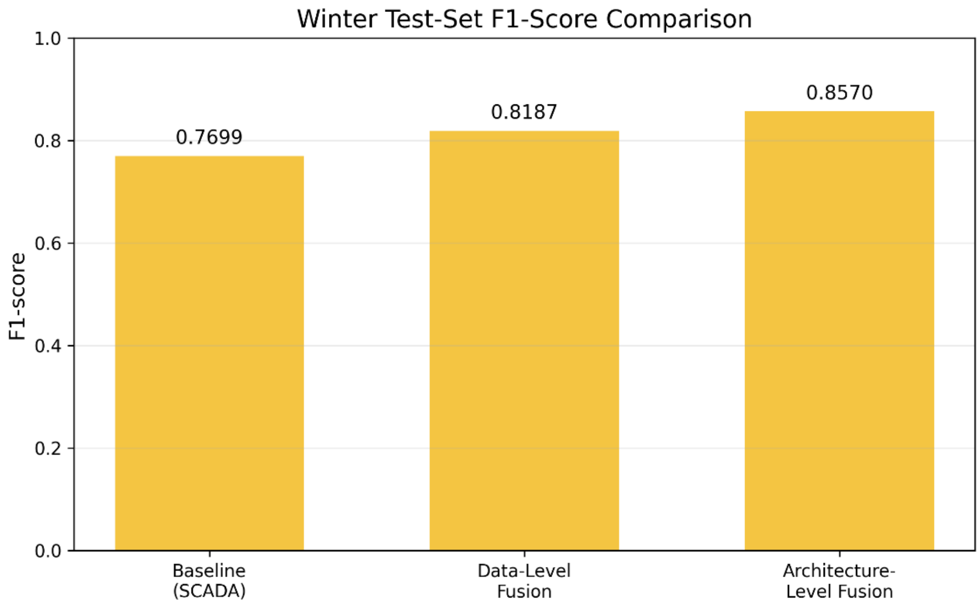

Figure 21. Winter Test-Set F1-Score Comparison

5.3 Confusion Matrix Analysis

Winter confusion matrices are more operative in displaying model behavior than metric values. They display the way each of the models spread its predictions among normal and fault-like conditions, and how the pragmatic implications of every fusion strategy are observed.

With the baseline SCADA-only model, winter sequences were used to show the normal operation, and the model was able to detect correctly 15,635 as normal and falsely identified 3,585 as anomalous. On the fault side, it was able to identify 6,576 actual anomaly intervals and only had 346 misses. This confirm that the baseline was very sensitive to the development of faults, and it also resulted in a heavy warning load. Practically such a behavior is defensive on a failure detection perspective, but is inefficient in terms of planning since numerous normal winter conditions are being considered suspicious.

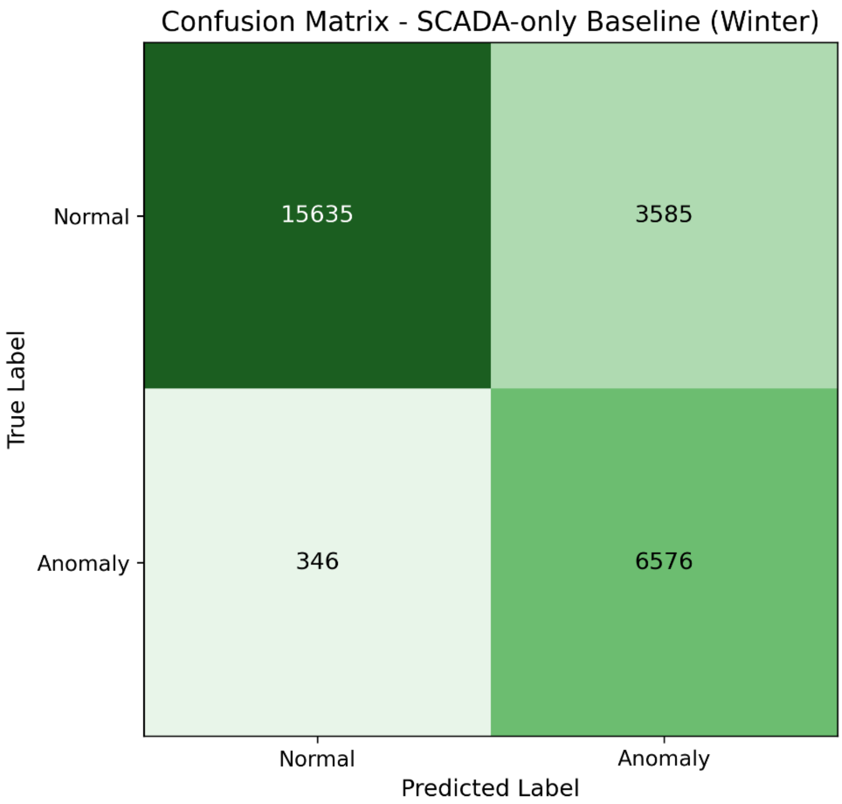

Figure 22. SCADA-only Baseline (Winter) Confusion Matrix

The fusion model (data-level fusion) caused a shift in pattern of prediction to a more alert selective type. It has accurately identified 17,904 normal winter intervals and minimized false alarms to 1,317, a significant improvement over the baseline. Yet, such a tighter boundary of decisions was paid with its price: the model identified 5,710 fault-associated intervals, and 1,212 real anomalies were missed. It means that feature-level meteorological fusion assisted the model in avoiding lots of unnecessary warnings, and, additionally, it reduced the number of actual fault signals quite significantly. In offshore applications, this is not an ideal trade-off as missed winter faults can be costly when there is a short access window.

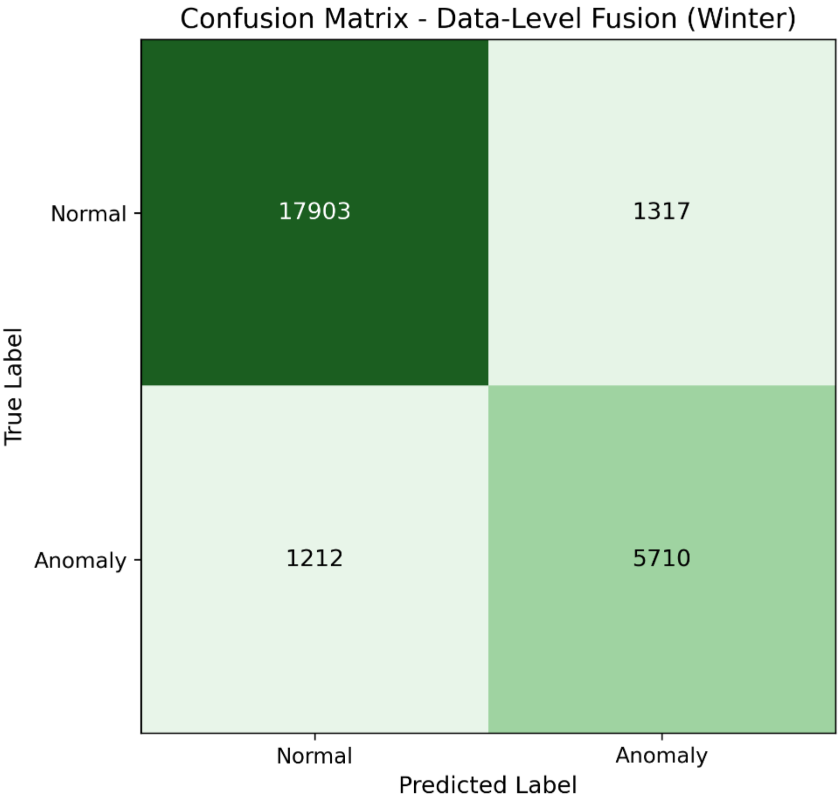

Figure 23. Data-Level Fusion (Winter) Confusion Matrix

The most balanced confusion matrix was obtained using the architecture-level fusion model. It was able to identify 17,617 normal winter intervals correctly and only 1,603 false alerts, yet identified 6392 true anomaly periods and only missed 530. It eliminated 1,982 false alarms in the wintertime as compared to the SCADA-only baseline and only 184 true detections were dropped. This exchange is operationally advantageous: the model maintained the majority of the baseline's fault visibility while significantly decreasing the number of alerts that could have otherwise impeded maintenance planning.

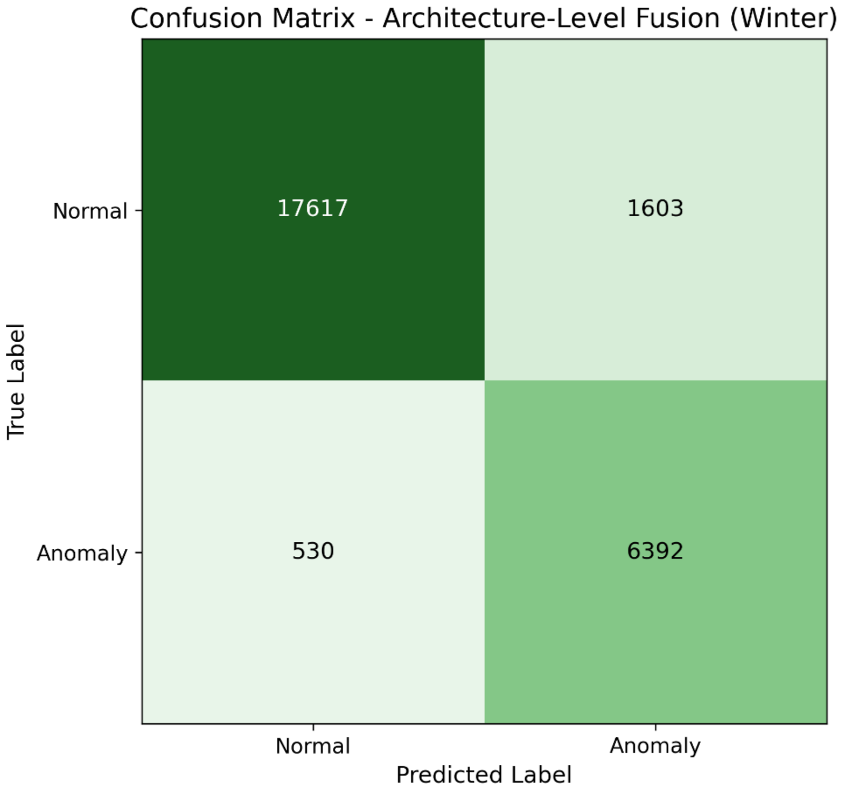

Figure 24. Architecture-Level Fusion (Winter) Confusion Matrix

5.4 Training Dynamics and Model Stability

The training histories carry good evidence on convergence behavior and stability of the models. The baseline model was able to learn quickly. Its training loss was lower in epoch 1 which was 0.2511 whereas its monitoring-validation loss was optimum in epoch 6 with value 0.0712. Validation F1 still achieved high results later in training, but the loss started rising after the optimal checkpoint, leading to the early termination and recovery of the optimal monitoring-validation state.

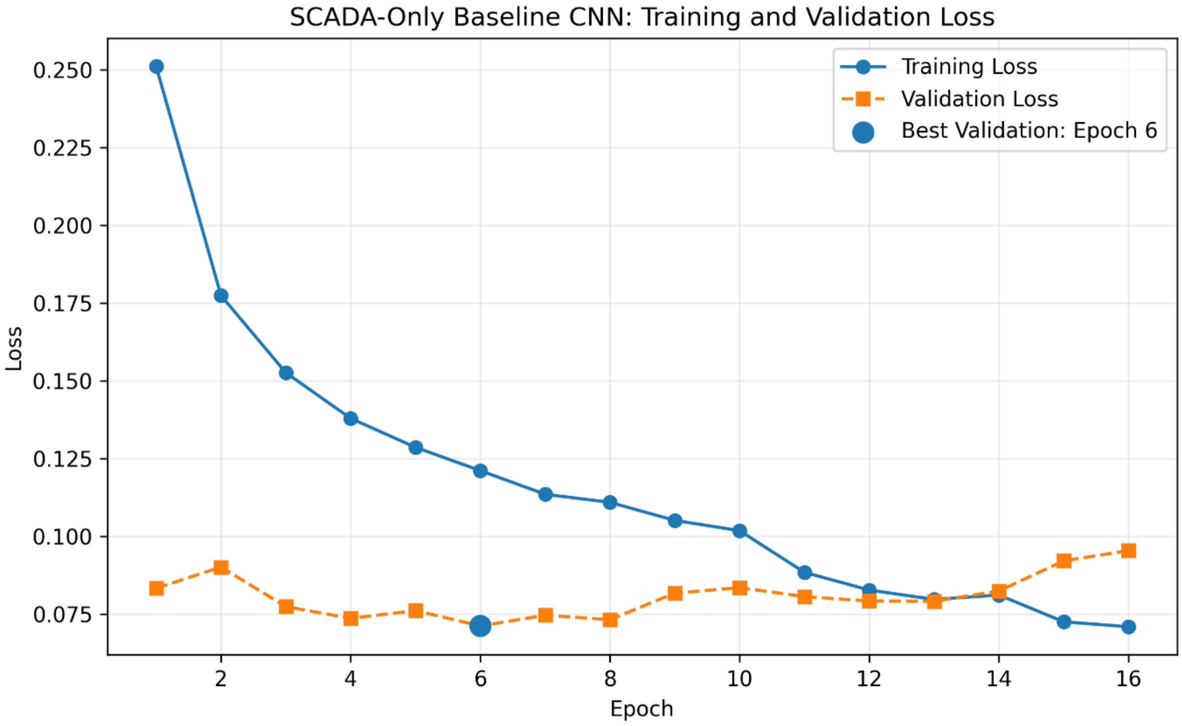

Figure 25. Training and Validation Loss Curve of SCADA-only Baseline Model

At the first epoch, data-level fusion model started with an initial training loss of 0.3317 and monitoring-validation loss of 0.1353. Its lowest validation loss was 0.1141 on the 11th epoch, then it started varying between roughly 0.1202 to 0.1430 until it was early stopped at epoch 21. This trend suggests that early feature-level mixing made a more challenging optimization environment than the baseline. The model continued to learn helpful patterns, but the validation pathway was not as smooth as expected which is in line with the representational complexity that is brought about by combining both SCADA and ERA5 variables prior to feature learning.

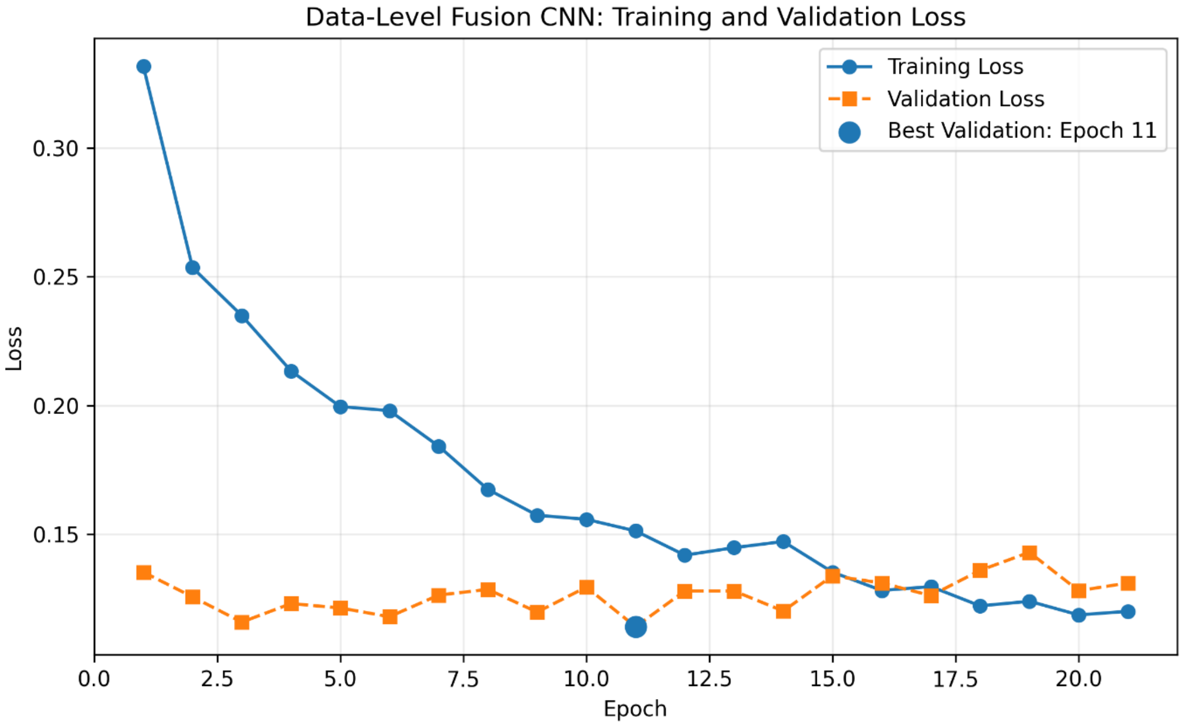

Figure 26. Training and Validation Loss Curve of Data-Level Fusion Model

The architecture-level fusion model had the longest training path, ended at epoch 27. Its training loss decreased from 0.3806 to 0.1205 whereas the epoch 17 had the best monitoring-validation loss of 0.1042. This model began with the maximum loss since it possessed two branches and a learnable environmental gate, although it ended up with a more desirable validation region compared to the data-level fusion model. The extended training process is thus a connotation of extra representational flexibility but not the instability.

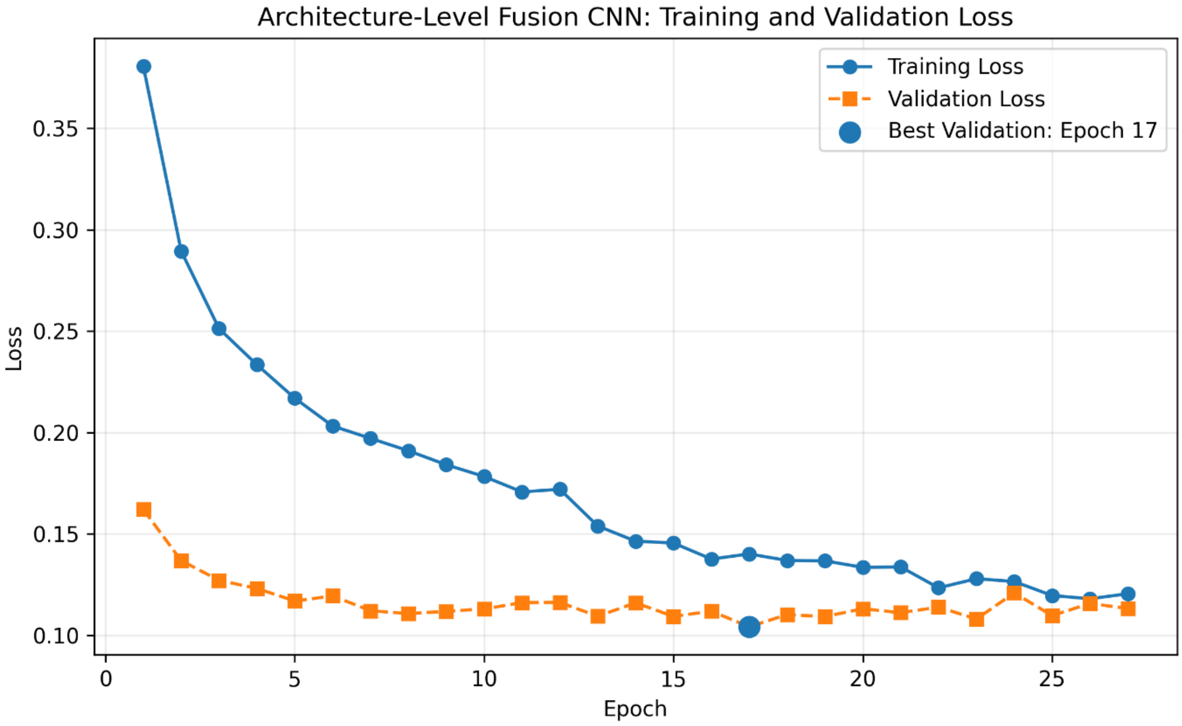

Figure 27. Training and Validation Loss Curve of Architecture-Level Fusion Model

Additionally, this view is supported by monitoring metrics. As an illustration, the architecture-level model achieved monitoring-validation accuracy rates of approximately 98.4-98.7% in later epochs with high precision and sensitivity to anomalies.

5.5 Discussion of Model Performance

The overall and winter performance indicates that SCADA signals do have powerful predictive information regarding the detection of anomalies. This is reflected by high detection rate of the baseline model, particularly during winter. Nevertheless, this model also generated a lot of unneeded alerts so that SCADA-based learning will consider a large percentage of the unusual operating states as a fault.

Fusion at the feature level enhanced the usefulness of warnings by including environmental context, yet it was not entirely beneficial. Data-level fusion model was made selective, particularly during winter and it significantly decreased false alarms. But along with this enhancement, the number of overlooked anomalies was well observed. It implies that merely adding ERA5 variables to the input space might not supply the model with tremendous guidance structure that can distinguish between environmental variability and turbine-internal degradation indications.

The architecture-level model was more effective since it did not entangle heterogeneous sources of data early. It maintained the power of learning sequences of turbine operations by first learning individual SCADA and weather representations and then incorporating meteorological data into narrow decision boundaries. This is what made it come up with the optimal F1-score in both the full test set and in the winter subset.

5.6 Why the Winter Results Matter Most

The most significant evaluation context of this thesis is the winter subset since offshore maintenance is most limited in winter. The severity of weather, the restricted accessibility of a location, lack of sufficient daylight, and increased uncertainty render any maintenance decision more imperative. A model that works well in general, but fails in winter would not have much practical use on the problem under consideration in this case.

The findings of winter results indicate that meteorological context was the most useful in the places where it was supposed to be. Architecture-level fusion model was found to have 91.84% accuracy and F1-score of 85.70% during winter making it beat both the alternatives in terms of overall seasonal performance. This will outline the main thesis statement: ERA5 information is not a simple supplementary source of features but an effective context layer to analyze SCADA behavior at environmentally challenging settings.

Another theme of winter comparison discloses why the fusion model is important. At the data level, fusion was helping to increase the reliability of alerts, but was also overlooking too many actual anomalies. Architecture level fusion allowed maintaining a high fault-capture rate and minimized false warnings. This makes the architecture level model the most operationally significant output for offshore winter maintenance.

5.7 Precision–Recall Tradeoff: Offshore Interpretation

Though SCADA-only baseline model obtained high recall of 0.9500, it has lower precision of 0.6472. This indicates that it detects nearly all real faults, but also generates a lot of false alarms. Within an offshore environment, each false alarm may suggest unneeded planning intricate, consideration of vessel disposal, crew mobilization and operational distraction. So, SCADA-only baseline model is very protective in terms of their safety but very expensive in terms of their logistics.

The data-level fusion model is highly geared in the direction of alert reliability. Its precision increases to 0.8126 and false positives decrease significantly to 1317 from 3585. Nonetheless, there is a drop in recall to 0.8249, and false negatives have increased dramatically to 1,212. This implies that there would be a substantial number of true winter faults events that would remain unnoticed.

The best operational balance is reached with the architecture-level fusion model. Its precision is 79.95, which is much higher than the baseline, and only slightly less than the data-level fusion model. Simultaneously, its recall is still high at 92.34% with only 530 missed faults that is critical to safety-critical applications.

$$\frac{3585-1603}{3585} \times 100 \approx 55.3 \qquad (11)$$

When a true fault goes unreported (false negative) in offshore wind maintenance, the consequences are often catastrophic and expensive and therefore, high recall cannot be compromised. The architecture-level fusion model halved ($\approx 55\%$) winter false positives compared to the baseline and only increased missing fault rates by 5 percent.

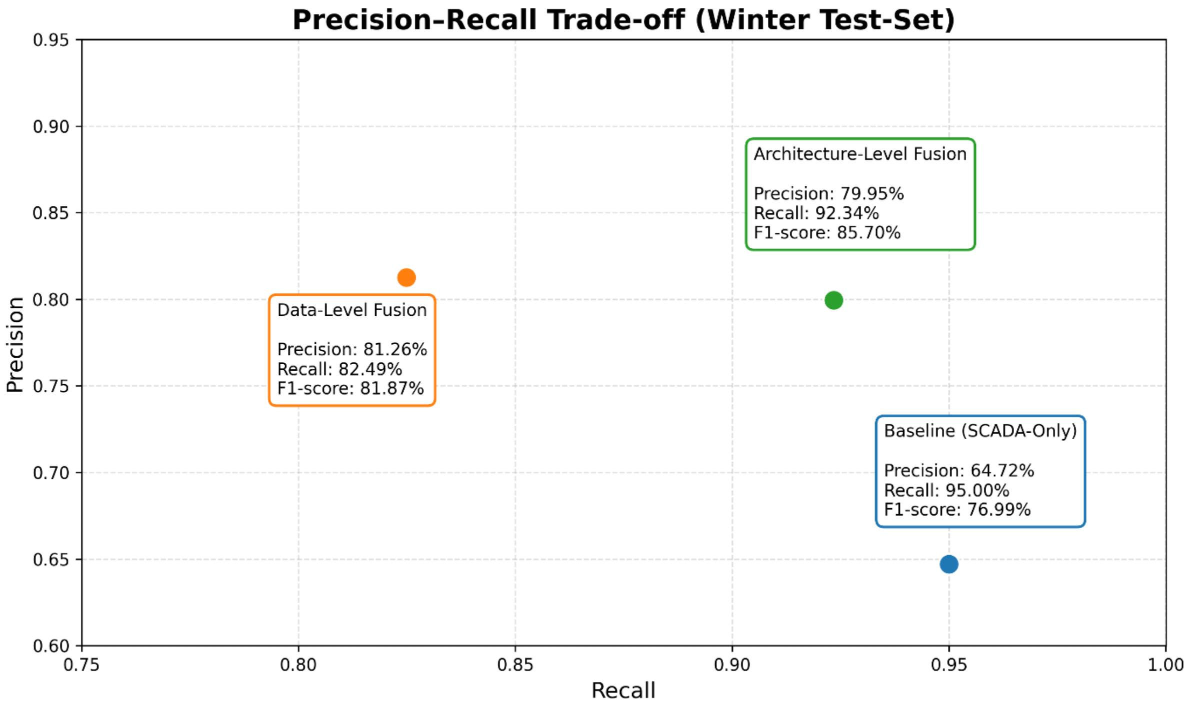

Figure 28. Precision-Recall Trade-Off Across Models in the Winter Test-Set

This is highly impacting trade-off in terms of operations. The fusion model at the architecture level managed to regain most of the precision advantage due to meteorological context and still maintain the high fault sensitivity required by predictive maintenance.

5.8 Discussion of the Architecture-Level Advantage

The high performance of the architecture level fusion model can be explained using the way it organized the interplay of the turbine-operational information and the environmental context. Instead of fusing these inputs at the feature level, the model proceeded to fuse these two data at a later stage of representation learning. In this design, each modality could make its contributions in a manipulable and interpretable way.

The learned environmental gate gives a key insight into this behaviour. As it was observed during training, the gate fell very fast at the beginning with an initial value of 1.0 and then with a rate of much faster reduction approximately towards 0.470. Beyond this point the gate continued to regulate more slowly and eventually approached a value of about 0.238, which is approximately a quarter of its original size.

This convergence pattern gives a more subtle explanation to how the model behaves compared to a mere statement of objective of environmental usefulness. The model did not identify an essential or a negligible role of environmental context in the formation of intermediate regime and significant but secondary role of ERA5 variables as an information source. Practically, this implies that the role of turbine-internal dynamics in the classification process remained dominant, like the meteorological inputs being a selective modifier.

This controlled integration is especially critical to offshore wind environments where abnormal SCADA configurations could be as a result of either actual fault or externally induced operating stress. A combination of attenuating but not eliminating the environmental contribution effectively reduced the chances of overreacting to the weather-related fluctuations and yet retained the aspect of sensitivity to the fault-related deviations.

The benefit of the architecture-level approach is therefore not merely due to adding more data but learning how much influence that data should have. This trade off between operational signals and environmental context justifies the fact that the model was able to achieve a better precision-recall tradeoff than the SCADA-only and data-level fusion alternatives.

5.9 Practical Offshore Maintenance Meaning

The results of the classification directly influence the offshore maintenance activities. In reality, a false alarm cannot be simply considered a statistical mistake. It may result in unneeded planning effort, engineering focus and the unnecessary consideration of vessel dispatching or the mobilization of crew. The large number of such alerts might also decrease operator confidence in the model.

An anomaly that has not been detected achieves a differing yet equally significant result. When no actual pattern of fault is identified, then probably the chance of an early intervention could be lost. Delayed identification can augment risk of downtime and complicate to accommodate right scheduled repairs that add up to already limiting access windows in winter.

This is why the most appropriate model may not necessarily consist of the highest recall rate and the highest precision rate alone. The model is the one that has a good detection of actual faults, yet has an easily manageable load on the alert. It minimized unnecessary winter warnings significantly without sacrificing high capability to alert of the actual anomalies, allowing it to be the most appropriate candidate to be used in maintenance decision support.

5.10 Winter Alert Export and Operational Extension

Winter prediction of architecture-level fusion model was further mapped in structured alert table. The dataset that was exported consisted of 7,995 positive winter alerts (TP+FP=6,392+1,603), and each alert was associated with a timestamp, turbine ID, true label, predicted probability and selected ERA5 variables (i.e. wind speed, air temperature, and sea surface temperature).

The key importance of this export should be to bridge the machine-learning model to a practical maintenance-planning workflow. An input to a downstream optimization model, e.g., a mixed-integer linear programming scheduler (MILP), can take the alert dataset as an input, where maintenance tasks, vessel trips, and access windows can be considered collectively.

In this regard, architecture-level fusion model is not only the most perfoming classifier in the thesis. It is also the most useful front end to a working decision-support pipeline as its better alert reliability makes fewer unnecessary candidate tasks and maintains high fault visibility.

5.11 Limitations

Although it is possible to infer that the suggested modeling framework and experimental results are quite strong evidence regarding the benefits of structured meteorological fusion, it is necessary to discover and interpret the results in the context of the current study and its limitations. The methodological design, data selection and modeling assumptions present some limitations that can affect both the generalizability, and the operational interpretation of the results.

First, the analysis includes a chosen sample of turbines of the CARE-style Wind Farm C data. Though this selection was required as a way of making sure that an experimental pipeline had a controlled and reproducible behavior; the models were trained and evaluated on a small sampling of offshore turbine behavior. Therefore, the results might not be generalized to other wind farms that have other turbine configurations, sensor configurations or otherwise and environmental conditions.

Second, even though ERA5 reanalysis data is valuable and consistent in providing environmental context at the macro-scale level. Consequently, they might not completely capture turbine-level microclimatic phenomena such as localized wind gusts, variations in the intensity of icing, or changes in sea-state. These smaller scale effects can be particularly important when there is extreme winter periods where the local effects may cause the turbine to behave in a manner that has not been fully represented in the fused input.

Third, only five ERA5 variables were used by the environmental branch. Though these were significant and relevant, a more detailed representation of the environment could further enhance classification, particularly by adding more direct offshore condition indicators.

Fourth, although the global scalar gating mechanism offers a more interpretable picture of the environmental influence as compared to more complex fusion strategies, it provides coarse-grained picture of environmental influence. In particular, it does not provide feature-level or time step level contributions of the meteorological variables.

Lastly, the research is about anomaly detection but not optimization of end-to-end maintenance. Though the architecture-level fusion model can be used to generate structured winter alert outputs that can be used in downstream scheduling, the thesis does not directly quantify operational performance measures such as a reduction in vessel deployments, maintenance costs or downtime. The real effect of a better classification performance is therefore seen as indirectly observed and not actually measured.

6 Conclusion and Future Scope

This thesis provides an in-depth analysis of the importance of the environmental context in the improvement of offshore wind turbine fault detection using various approaches of environmental data integration. The research focuses specifically on the use of environmental fusion during winter, when it is most challenging to maintain and operate offshore wind farms. Through the development and comparison of three different predictive maintenance models (SCADA-only, data-level fusion, and architecture-level fusion with a learnable global scalar gate), it is shown that the use of environmental data could be advantageous for the model's performance with the right choice of fusion approach. The architecture-level fusion model proved to be the ideal solution, striking a good balance between detection performance and practicality.

The subsections below summarize the most important findings, discuss the theoretical and practical contribution of the thesis, answer the research questions in greater detail and describe the future working directions. Moreover, research recommendations are given on how future research can be developed on the findings of this study to make the contributions helpful in real life operation of offshore wind turbines.

6.1 Summary of the Main Findings

There is a clear evolution shown in the empirical results in terms of the impact of environmental information on the performance of fault detection with each modeling strategy.

The SCADA-only model sets a solid baseline and demonstrates that there is enough information in turbine-operational data for anomaly detection. The high recall shows that the model is sensitive to faults, but the high false-positive rate suggests that SCADA-only models are likely to trigger alarms on environmentally influenced variations. This makes them less useful in offshore maintenance environment.

The addition of environmental features through data-level fusion improves precision (through lower false positives) but reduces recall. This indicates that meteorological variables can distinguish between fault events and nuisance variations due to external factors. However, this improvement comes with a cost in recall which suggests that feature-level fusion at an early stage can result in a more conservative model and miss some true fault events.

The architecture-level fusion model attains the most balanced result. By separately processing SCADA and environmental data and using a learnable gating function to merge the two information sources, the model retains the benefit of learning from SCADA data while selectively blending the environmental information. This leads to enhanced precision while retaining relatively high recall, especially in winter. It thus offers the best operational gains of the three models.

6.2 Contributions of the Study

The results of this study are not limited to the comparison of models, but they provide theoretically justified and practically valuable contributions. These contributions represent the dual purpose of developing predictive modeling methods and recognizing the practical conditions of the offshore wind farms maintenance.

6.2.1 Theoretical Contribution

This thesis contributes to predictive maintenance in offshore wind power systems in multiple ways. At a methodological level, the thesis shows that the fusion approach of multimodal data is a crucial determinant of the model performance. The findings reveal that combining environmental variables is not enough; instead, structural fusion mechanisms are needed to make the full use of the available contextual information.

Regarding modelling approaches, the proposed architecture-level fusion provides an efficient and transparent alternative for multimodal learning strategies. A global scalar gating mechanism is used to enable the model to control the impact of environmental data without creating undue architectural complexity. This enhances transparency as well as the robustness of the model.

Then, at the assessment point of view, focus on seasonal analysis is another key feature. The emphasis on winter conditions demonstrates the necessity for more context-aware approaches to fault detection where the models' performance may depend on the seasonal context. This perspective offers an improved understanding of how the environmental can be used to enhance fault detection under difficult conditions.

6.2.2 Practical Contribution

The practical impact of the thesis relates to improving efficiency in the maintenance of offshore wind turbines. The architecture-level fusion approach could significantly save costs by reducing false alarms, often causing unnecessary maintenance actions such as mobilizing personnel and ships. This increases the cost-effectiveness of maintenance planning by enabling maintenance to prioritize the real faults while filtering out less important anomalies.

The model's evaluation for winter months adds to its real practical value. Winter is a critical time of the year for maintenance tasks on offshore wind farms as conditions are harsh and access to systems is difficult. This study shows that winter-time maintenance is enhanced by the fusion of weather data that provides improved fault detection performance over SCADA-only baseline model which may be influenced by noise. The fact that the architecture-level fusion approach achieves a high recall with lower false positives during winter is a step forward in increasing reliability and profitability of offshore wind farms.

This research results in exportable winter alerts, which can be used for maintenance optimization. Through alert information export, this research provides a foundation for real-world operation integration, helping operational scheduling and resource management tasks. Overall, this integration of predictive maintenance models with maintenance strategies will help to enhance offshore wind farm operation.

6.3 Addressing the Research Questions

Through a systematic evaluation, three models were analysed to answer the three research questions. The first research question, which examined the performance of the SCADA-only model in winter anomaly detection, was answered by showing that while the model was successful in detecting faults, it had a low precision (many false alarms). It showed that, although SCADA data can be used to detect information about faults, they need to be supplemented for improved performance.

The second research question, which investigated the impact of adding meteorological data through the data-level fusion model, revealed that although weather data improved precision by reducing false alarms, it reduced recall. So, early fusion strategies, while useful in boosting the precision, could reduce the recall, particularly in winter when situations are more complex.

The third research question was related to the architecture-level fusion model. This model, which uses two branches for SCADA and weather data, and then combines them with a learnable scalar gate, gives the best performance in terms of both the overall and winter anomaly detection. It offers the best combination of precision and recall and is the most suitable model for offshore wind farms.

6.4 Recommendations and Future Directions

The results of this thesis can be used to establish several recommendations on future research directions, as well as practical advice on how to advance predictive maintenance in offshore wind systems.

Among the opportunities, one is the incorporation of the exported winter alerts into maintenance optimisation models, including mixed-integer linear programming (MILP) frameworks of maintenance scheduling. This would enable the further research to measure the benefits of operational change, such as possible decreases in the number of vessel trips, crew mobilization, maintenance expenses, and more firmly establish a connection between predictive modeling and decision-making.

Moreover, one should consider additional testing of the architecture-level fusion model on a larger range of turbines and on different wind farms. It would be better to broaden the scope of assessment for better understanding of how the model would generalize to different operational contexts and turbine design to increase its relevance in practice. Another shortcoming that can be enhanced is the integration of more environmental factors, including the height of the waves, the intensity of turbulence, and local weather conditions. These variables can provide a more complete picture of offshore conditions and potentially aid in more precise fault detection, especially in extreme or quickly evolving weather conditions.

Then more sophisticated methods of fusion, such as attention-based models or feature-wise gating, are worth investigating to further improve data modality interaction. Nonetheless, these developments are required to be done in a manner that they are not only interpretable but also computationally efficient particularly when an application is being deployed in an operations context.

Lastly, one could consider the extension of the existing binary classification system to a risk-sensitive maintenance scoring system. Future systems could be more responsive to maintenance decision-making by adding severity or importance levels to predictions to allocate resources more efficiently and appropriately.

Overall, the results of this work show when the methods are applied with the correct modeling strategies, it introduces the effects of environmental context into predictive maintenance that can significantly increase the quality of decisions used in offshore wind systems. A combination of structured data fusion and attention to winter operational conditions enables the research to the path of more credible fault detection that can be consistent with the actual maintenance limitations. In addition to the improvements at model-level, the work is directed to a broader change in the way predictive insights can assist the process of maintenance planning, resources distribution, and reliability of the system. In this regard, the thesis can make a contribution to the development of methodology and improving financial sustainability of offshore wind performance by creating more informed and effective maintenance practices.

References

Attouri, K., Mansouri, M., Kouadri, A., Al-Hmouz, R., & Hossen, A. (2026). Intelligent Multimodal Monitoring of Renewable Energy Systems Using AI Fusion Frameworks. *IEEE Access*, *14*, 5421–5431. https://doi.org/10.1109/ACCESS.2026.3650967

Badarinath, K., Hoebeke, P., Schillebeeckx, D., & Yazicioglu, H. (2024). Data-driven wind turbine sensor health validation. *Journal of Physics: Conference Series*, *2767*(3), 032032. https://doi.org/10.1088/1742-6596/2767/3/032032

Bilgili, M., & Alphan, H. (2022). *Global Growth in Offshore Wind Turbine Technology*. In Review. https://doi.org/10.21203/rs.3.rs-1202466/v1

Brena, R. F., Aguileta, A. A., Trejo, L. A., Molino-Minero-Re, E., & Mayora, O. (2020). Choosing the Best Sensor Fusion Method: A Machine-Learning Approach. *Sensors*, *20*(8), 2350. https://doi.org/10.3390/s20082350

Buckland, M., & Gey, F. (1994). The relationship between Recall and Precision. *Journal of the American Society for Information Science*, *45*(1), 12–19. https://doi.org/10.1002/(SICI)1097-4571(199401)45:1%3C12::AID-ASI2%3E3.0.CO;2-L

Carroll, J., McDonald, A., & McMillan, D. (2016). Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy*, *19*(6), 1107–1119. https://doi.org/10.1002/we.1887 de Boer, D., & Xydis, G. (2024). A review on operation & maintenance strategies for offshore wind farms. *Ships and Offshore Structures*, *19*(10), 1563–1570. https://doi.org/10.1080/17445302.2023.2256594

Dibaj, A., Valavi, M., & Nejad, A. R. (2024). Unsupervised anomaly detection of permanent-magnet offshore wind generators through electrical and electromagnetic measurements. *Wind Energy Science*, *9*(11), 2063–2086. https://doi.org/10.5194/wes-9-2063-2024

Dowell, J., Zitrou, A., Walls, L., Bedford, T., & Infield, D. (2013). Analysis of wind and wave data to assess maintenance access to offshore wind farms. In *Safety, Reliability and Risk Analysis*. CRC Press.

Faulstich, S., Hahn, B., & Tavner, P. J. (2011). Wind turbine downtime and its importance for offshore deployment. *Wind Energy*, *14*(3), 327–337. https://doi.org/10.1002/we.421

Fox, H., Pillai, A. C., Friedrich, D., Collu, M., Dawood, T., & Johanning, L. (2022). A Review of Predictive and Prescriptive Offshore Wind Farm Operation and Maintenance. *Energies*, *15*(2), Article 2. https://doi.org/10.3390/en15020504

Gück, C., Roelofs, C. M. A., & Faulstich, S. (2024). CARE to Compare: A Real-World Benchmark Dataset for Early Fault Detection in Wind Turbine Data. *Data*, *9*(12), 138. https://doi.org/10.3390/data9120138

Hassan, M. (2025). *AI-Based Conditional Monitoring & Predictive Maintenance for Offshore Wind Farms*. https://osuva.uwasa.fi/handle/11111/12195

Hayes, L., Stocks, M., & Blakers, A. (2021). Accurate long-term power generation model for offshore wind farms in Europe using ERA5 reanalysis. *Energy*, *229*, 120603. https://doi.org/10.1016/j.energy.2021.120603

Helbing, G., & Ritter, M. (2018). Deep Learning for fault detection in wind turbines. *Renewable and Sustainable Energy Reviews*, *98*, 189–198. https://doi.org/10.1016/j.rser.2018.09.012

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Irawan, C. A., Ouelhadj, D., Jones, D., Stålhane, M., & Sperstad, I. B. (2017). Optimisation of maintenance routing and scheduling for offshore wind farms. *European Journal of Operational Research*, *256*(1), 76–89. https://doi.org/10.1016/j.ejor.2016.05.059

Jankauskas, M., Serackis, A., Šapurov, M., Pomarnacki, R., Baskys, A., Hyunh, V. K., Vaimann, T., & Zakis, J. (2023). Exploring the Limits of Early Predictive Maintenance in Wind Turbines Applying an Anomaly Detection Technique. *Sensors*, *23*(12), 5695. https://doi.org/10.3390/s23125695

Koltsidopoulos Papatzimos, A. (2020). *Data-driven operations and maintenance for offshore wind farms: Tools and methodologies*. https://hdl.handle.net/1842/36710

Kou, L., Li, Y., Zhang, F., Gong, X., Hu, Y., Yuan, Q., & Ke, W. (2022). Review on Monitoring, Operation and Maintenance of Smart Offshore Wind Farms. *Sensors*, *22*(8), 2822. https://doi.org/10.3390/s22082822

Leahy, K., Hu, R. L., Konstantakopoulos, I. C., Spanos, C. J., Agogino, A. M., & O'Sullivan, D. T. J. (2018). Diagnosing and PredictingWind Turbine Faults from SCADA Data Using Support Vector Machines. *International Journal of Prognostics and Health Management*, *9*(1). https://doi.org/10.36001/ijphm.2018.v9i1.2692

Lei, J., Liu, C., & Jiang, D. (2019). Fault diagnosis of wind turbine based on Long Short-term memory networks. *Renewable Energy*, *133*, 422–432. https://doi.org/10.1016/j.renene.2018.10.031

Lin, T., Ren, Z., Zhu, L., Zhu, Y., Feng, K., Ding, W., Yan, K., & Beer, M. (2025). A Systematic Review of Multisensor Information Fusion for Equipment Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *74*, 1–48. https://doi.org/10.1109/TIM.2025.3529577

Lin, W.-H., Wang, P., Chao, K.-M., Lin, H.-C., Yang, Z.-Y., & Lai, Y.-H. (2021). Wind Power Forecasting with Deep Learning Networks: Time-Series Forecasting. *Applied Sciences*, *11*(21), 10335. https://doi.org/10.3390/app112110335

Liu, Z., & Hui, J. (2024). Advancing predictive maintenance: A deep learning approach to sensor and event-log data fusion. *Sensor Review*, *44*(5), 563–574. https://doi.org/10.1108/SR-03-2024-0183

Maldonado-Correa, J., Martín-Martínez, S., Artigao, E., & Gómez-Lázaro, E. (2020). Using SCADA Data for Wind Turbine Condition Monitoring: A Systematic Literature Review. *Energies*, *13*(12), 3132. https://doi.org/10.3390/en13123132

Maron, J., Anagnostos, D., Brodbeck, B., & Meyer, A. (2022). Artificial intelligence-based condition monitoring and predictive maintenance framework for wind turbines. *Journal of Physics: Conference Series*, *2151*(1), 012007. https://doi.org/10.1088/1742-6596/2151/1/012007

Martini, M., Guanche, R., Losada, I. J., & Vidal, C. (2017). Accessibility assessment for operation and maintenance of offshore wind farms in the North Sea. *Wind Energy*, *20*(4), 637–656. https://doi.org/10.1002/we.2028

Masoumi, M. (2023). Machine Learning Solutions for Offshore Wind Farms: A Review of Applications and Impacts. *Journal of Marine Science and Engineering*, *11*(10), 1855. https://doi.org/10.3390/jmse11101855

Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *2020 11th International Conference on Information and Communication Systems (ICICS)*, 243–248. https://doi.org/10.1109/ICICS49469.2020.239556

Pandit, R. K., Kolios, A., & Infield, D. (2020). Data-driven weather forecasting models performance comparison for improving offshore wind turbine availability and maintenance. *IET Renewable Power Generation*, *14*(13), 2386–2394. https://doi.org/10.1049/iet-rpg.2019.0941

Pandit, R., & Wang, J. (2024). A comprehensive review on enhancing wind turbine applications with advanced SCADA data analytics and practical insights. *IET Renewable Power Generation*, *18*(4), 722–742. https://doi.org/10.1049/rpg2.12920

Peng, D., Liu, C., Desmet, W., & Gryllias, K. (2023). Condition Monitoring of Wind Turbines Based on Anomaly Detection Using Deep Support Vector Data Description. *Journal of Engineering for Gas Turbines and Power*, *145*(091009). https://doi.org/10.1115/1.4062768

Ren, Z., Verma, A. S., Li, Y., Teuwen, J. J. E., & Jiang, Z. (2021). Offshore wind turbine operations and maintenance: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, *144*, 110886. https://doi.org/10.1016/j.rser.2021.110886

Rinaldi, G., Thies, P. R., & Johanning, L. (2021). Current Status and Future Trends in the Operation and Maintenance of Offshore Wind Turbines: A Review. *Energies*, *14*(9), 2484. https://doi.org/10.3390/en14092484

Santiago, R. A. da F., Barbosa, N. B., Mergulhão, H. G., Carvalho, T. F. de, Santos, A. A. B., Medrado, R. C., Filho, J. B. de M., Pinheiro, O. R., & Nascimento, E. G. S. (2024). Data-Driven Models Applied to Predictive and Prescriptive Maintenance of Wind Turbine: A Systematic Review of Approaches Based on Failure Detection, Diagnosis, and Prognosis. *Energies*, *17*(5), 1010. https://doi.org/10.3390/en17051010

Schröder, L., Dimitrov, N. K., Verelst, D. R., & Sørensen, J. A. (2022). Using Transfer Learning to Build Physics-Informed Machine Learning Models for Improved Wind Farm Monitoring. *Energies*, *15*(2), 558. https://doi.org/10.3390/en15020558

Schrotenboer, A. H., uit het Broek, M. A. J., Jargalsaikhan, B., & Roodbergen, K. J. (2018). Coordinating technician allocation and maintenance routing for offshore wind farms. *Computers & Operations Research*, *98*, 185–197. https://doi.org/10.1016/j.cor.2018.05.019

Soares, P. M. M., Lima, D. C. A., & Nogueira, M. (2020). Global offshore wind energy resources using the new ERA-5 reanalysis. *Environmental Research Letters*, *15*(10), 1040a2. https://doi.org/10.1088/1748-9326/abb10d

Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., & Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*, *133*, 620–635. https://doi.org/10.1016/j.renene.2018.10.047

Tang, M., Zhao, Q., Ding, S. X., Wu, H., Li, L., Long, W., & Huang, B. (2020). An Improved LightGBM Algorithm for Online Fault Detection of Wind Turbine Gearboxes. *Energies*, *13*(4), 807. https://doi.org/10.3390/en13040807

Tapoglou, E., Forster, R. M., Dorrell, R. M., & Parsons, D. (2021). Machine learning for satellite-based sea-state prediction in an offshore windfarm. *Ocean Engineering*, *235*, 109280. https://doi.org/10.1016/j.oceaneng.2021.109280

Tautz-Weinert, J., & Watson, S. J. (2017). Using SCADA data for wind turbine condition monitoring – a review. *IET Renewable Power Generation*, *11*(4), 382–394. https://doi.org/10.1049/iet-rpg.2016.0248

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2022). *The Computational Limits of Deep Learning* (arXiv:2007.05558). arXiv. https://doi.org/10.48550/arXiv.2007.05558

Turnbull, A., & Carroll, J. (2021). Cost Benefit of Implementing Advanced Monitoring and Predictive Maintenance Strategies for Offshore Wind Farms. *Energies*, *14*(16), 4922. https://doi.org/10.3390/en14164922

Tutivén, C., Encalada-Dávila, Á., Vidal, Y., & Benalcázar-Parra, C. (2023). Detecting bearing failures in wind energy parks: A main bearing early damage detection method using SCADA data and a convolutional autoencoder. *Energy Science & Engineering*, *11*(4), 1395–1411. https://doi.org/10.1002/ese3.1398

Udo, W., & Muhammad, Y. (2021). Data-Driven Predictive Maintenance of Wind Turbine Based on SCADA Data. *IEEE Access*, *9*, 162370–162388. https://doi.org/10.1109/ACCESS.2021.3132684

Ulmer, M., Jarlskog, E., Pizza, G., Manninen, J., & Goren Huber, L. (2020). Early Fault Detection Based on Wind Turbine SCADA Data Using Convolutional Neural Networks. *PHM Society European Conference*, *5*(1), 9. https://doi.org/10.36001/phme.2020.v5i1.1217

Wang, Z., Guo, Y., & Wang, H. (2022). Review on Monitoring and Operation-Maintenance Technology of Far-Reaching Sea Smart Wind Farms. *Journal of Marine Science and Engineering*, *10*(6), 820. https://doi.org/10.3390/jmse10060820

Wrzask, K., Kołakowska, A., Jasik, P., Syty, P., Dembski, J., & Wiszniewski, B. (2025). SCADA-Based Offshore Wind Turbine Monitoring: A Review of Methods of Addressing Marine Environmental Challenges. *Polish Maritime Research*, *nr 4*. https://doi.org/10.2478/pomr-2025-0061

Xu, J., Luo, Y., Wu, G., Wang, W., Zhang, Z., & Kanapathipillai, A. (2026). Multi-Source Data-Driven CNN–Transformer Hybrid Modeling for Wind Energy Database

Reconstruction in the Tropical Indian Ocean. *Remote Sensing*, *18*(2), 226. https://doi.org/10.3390/rs18020226

Yacouby, R., & Axman, D. (2020). Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. In S. Eger, Y. Gao, M. Peyrard, W. Zhao, & E. Hovy (Eds.), *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems* (pp. 79–91). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.eval4nlp-1.9

Zgraggen, J., Ulmer, M., Jarlskog, E., Pizza, G., & Huber, L. G. (2021). Transfer Learning Approaches for Wind Turbine Fault Detection using Deep Learning. *PHM Society European Conference*, *6*(1), 12–12. https://doi.org/10.36001/phme.2021.v6i1.2835

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., & Si, Y. (2018). A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost. *IEEE Access*, *6*, 21020–21031. https://doi.org/10.1109/ACCESS.2018.2818678

Zhang, S., Robinson, E., & Basu, M. (2023). Wind Turbine Predictive Fault Diagnostics Based on a Novel Long Short-Term Memory Model. *Algorithms*, *16*(12), 546. https://doi.org/10.3390/a16120546

Zhang, Y., Kehtarnavaz, N., Rotea, M., & Dasari, T. (2024). Prediction of Icing on Wind Turbines Based on SCADA Data via Temporal Convolutional Network. *Energies*, *17*(9), 2175. https://doi.org/10.3390/en17092175

Zhao, H., Liu, H., Hu, W., & Yan, X. (2018). Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renewable Energy*, *127*, 825–834. https://doi.org/10.1016/j.renene.2018.05.024

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, *115*, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050