Journal of Applied Automation Technologies

Research Paper

Long-Sequence Informer Model for Offshore Wind Turbine Fault Diagnosis Using Marine Monitoring Data

Ladislav Drápal[1] and Adéla Svoboda[1, *]

[1] Department of Computer Systems, Brno University of Technology, 61669 Brno, Czech Republic
*Corresponding author: adela.s@fit.vut.cz

Abstract. Offshore wind turbines are affected by numerous factors simultaneously: changes in wind speed, wave excitation, salt-spray corrosion, temperature drift and grid-side load fluctuations. Therefore, the fault signature is generally mild, delayed and may be misinterpreted for a normal environmental transient. Propose a long-sequence Informer model for problem identification based on maritime monitoring data in this paper. An environment-aware temporal representation that integrates electrical signals, vibration indicators, SCADA data, and observations of the marine environment. ProbSparse attention can learn long-range dependency information and reduce the amount of attention computation. To enhance the interpretability of diagnosis results for early-stage, transitional-stage, and severe-stage defects, a fault-stage contribution method will also be implemented. For testing, a multi-source monitoring dataset including 42,000 ten-minute records from offshore turbines has been created. It is separated into one normal class and five fault categories. The suggested model outperforms the LSTM, TCN, conventional Transformer, and CNN-GRU baselines in terms of overall accuracy (96.8%), macro-F1 score (95.9%), and average inference latency (18.6 ms per sequence). According to the findings, long-sequence sparse attention can enhance fault recognition reliability while still satisfying the demands of real-time offshore operation and maintenance.

Keywords: *Offshore Wind Turbine, Fault Diagnosis, Marine Monitoring Data, Long-Sequence Informer, ProbSparse Attention, Condition Monitoring*

Received on 19 August 2022, Accepted on 25 March 2023, Published on 05 April 2023

Introduction

Because of its large-scale grid connection capabilities and high-density wind resources, offshore wind power has been added to the list of key linkages in a low-carbon power system. Supervisory control and data acquisition systems, vibration sensors, temperature probes, current transformers, converter alarms, yaw records, and maritime environment monitoring devices are currently installed on the majority of new offshore wind turbines. The foregoing data streams are not just for recording mechanical and electrical working parameters but also contain changes in wind speed and wave loads, humidity, salt fog concentration, and air temperature. In practice, turbine downtime at sea is more expensive than maintenance performed on land because of limited access windows brought on by inclement weather, shipping schedules, and safety constraints. Therefore, a high-reliability fault diagnostic system can assist condition-based monitoring of offshore wind farms, avoid damage to supporting infrastructure, and lower operating and maintenance costs [1]. According to recent reports, SCADA, vibration, and environmental signals are also part of the offshore drivetrain's three-dimensional monitoring system [2]. Thus, a variety of machine learning techniques are currently being used in research to improve the precision of wind turbine problem identification under different operating situations [3].

However, failure detection of offshore wind turbines is still challenging due to the extended time scale of gathered data and high noise levels induced by environmental variations. A defect in a gearbox bearing may initially be a slight temperature increase, a weak variation in vibration, or an unexpected departure of the power curve; at this time, no alarm has been given. Generator, pitch, yaw, converter and blade-icing failures can also display comparable symptoms in SCADA recordings, particularly during rapid changes in wind speed or variations to the temperature environment of the sea [4]. Fault precursors are frequently mild temporal drifts rather than distinct alarm states, according to early warning studies based on SCADA streams [5]. Dynamic model sensor approaches can boost fault sensitivity, but they are still relatively susceptible to the stability of the modelled normal state [6]. SCADA + vibration data fusion can boost the diagnostic strength, however heterogeneous sampling and missing intervals still occur in offshore applications [7]. Power-curve cleaning and change-point detection approaches are suitable for abnormality screening, but they fail to solve multi-class fault discrimination thoroughly [8]. Long-range temporal dependencies in many models have not been adequately captured despite the use of data mining and neural network-based feature extraction [9]. When the offshore environment changes more quickly than the turbine deteriorate, neural fault detectors are also likely to perform badly [10].

A multi-source marine monitoring data-based long-sequence Informer model for diagnosing offshore wind turbine faults. This paper addresses three issues: how to account for the long-term evolution of faults while restricting attention computation; how to explain the diagnosis results at both the signal and fault stages; and how to represent various operational and environmental variables in a single temporal feature space. The suggested model incorporates temporal distillation, contribution-guided fault classification, ProbSparse attention, and environment-aware embedding. to provide a deployable diagnostic framework for offshore wind farm monitoring platforms, improve early failure detection, and lessen uncertainty among visually identical degradation patterns.

Related Work

Data-Driven Fault Diagnosis for Wind Turbines

Data-driven diagnostics of wind turbines has transitioned from single-signal threshold criteria to multi-source learning models. Power curve residuals, gearbox temperature variations, vibration energy indicators, and warning frequency were used in the first study. The following indications are helpful because they are physically intuitive and generally straightforward to see; nevertheless, they may not be highly sensitive to changes in the working environment or conditions. By include wind speed, active power, rotor speed, pitch angle, generator temperature, nacelle temperature, oil pressure, and converter status, SCADA-based techniques broadened the diagnostic scope. Machine learning classifiers further enhanced discrimination by learning nonlinear mappings between operating variables and fault labels [11]. Neural networks have been used to detect irregularities in gearboxes, generators, blade icing, yaw and pitch from SCADA data streams, and data mining algorithms have been utilized to find anomalous wind speed and power-curve deviations [12]. Additionally, research on gearbox monitoring has demonstrated that an increase in temperature and vibration takes place several hours prior to the sounding of an alert [13]. Preventive maintenance can be scheduled by predicting an early temperature departure depending on operational load, according to analysis of wind turbine gearbox SCADA [14].

The fact that the offshore data are not independent samples is a persistent flaw. The apparent signal pattern may change with sea-state disturbance, and the same fault may exhibit weak drift, intermittent fluctuation, and persistent irregularity. Studies on blade icing detection have demonstrated that climatic parameters might be misconstrued for fault boundaries when a classifier is not context-aware [15]. Although discriminative feature learning can improve icing recognition, it still has a low temporal sensitivity for intermittent degradation processes [16]. Because significant errors are uncommon in typical production data, diagnostic models for offshore systems must also account for class imbalance. In order to preserve the degradation route and prevent handling each time step separately, long-window models have been created [17].

The fact that the offshore wind farm's fault reports seldom match the actual damage is another issue. Maintenance records normally indicate the time when an alert was confirmed or a repair was scheduled, but the fundamental cause of the problem may have occurred several hours or even days previously. The learning model will learn the alarm trigger instead of the degradation process if it is solely trained with single-point labels. The problem is compounded during curtailment, wake effect and storm shutdown of the turbine, as a normal low-power phase may appear to be a component failure in the raw SCADA data. As a result, the development of temporal windows, regime-aware normalization, and multi-source evidence fusion has progressively grown in recent diagnostic research. The concepts offer a way to approach the offshore diagnosis as a long-sequence classification challenge as opposed to a static pattern recognition task.

Long-Sequence Modeling and Interpretable Diagnosis

Recently, long-sequence time-series models with attention mechanisms have demonstrated good performance. Recurrent networks can address the problem of time dependence, although they often lose distant information in extended sequences [18]. Long-term time-series models must be resilient to computational costs and memory-efficient, according to deep learning research [19]. Additionally, studies on time-series prediction have discovered that for a model to function successfully, it must have strong long-term memory throughout numerous cycles of operation [20]. Transformer models can directly learn the relationships among remote states and are suited for offshore wind monitoring, as fault precursors may occur at intervals of hundreds of samples [21]. ProbSparse attention and temporal distillation are used to minimize attention complexity further, and it is ideal for long-sequence forecasting and classification tasks where just a few dominating queries contribute considerably to the global dependency [22]. Multivariate forecasting has also made use of hybrid long-sequence models, and it has been shown that sparse attention can preserve essential temporal correlations without requiring more processing [23].

Both high prediction accuracy and explainability are now necessary in fault diagnosis. Transformer-based simultaneous-fault diagnosis indicates that attention may be employed to detect interactions among time-dependent sensor channels [24]. Diagnosing rotating machinery using time-series Stronger feature separation is demonstrated by the transformer construction under varied operating conditions [25]. Maintenance experts need to determine whether the diagnosis is due to a fault in the generator temperature, gearbox vibration, yaw error, converter current, humidity or wind-speed turbulence. Explainable Artificial Intelligence provides methods to link model outputs with variable-level consequences and local decisions [26]. General XAI taxonomies also show that high-risk engineering systems require visible decision assistance rather than black-box alerts [27]. Even though some surveys concentrate on medical or general decision systems, their explanation concepts can be applied to condition monitoring when it's necessary to incorporate local and global evidence [28]. A predictor-effect visualization can demonstrate if a variable increases the diagnostic confidence in a non-linear or monotonic manner [29]. Building maintenance-oriented contribution maps is another benefit of using local-to-global explanation techniques [30].

There are practical operational requirements for the interpretability of offshore wind applications. Operators may choose to disregard a fault alarm if the real cause cannot be identified; otherwise, dispatching a vessel would be expensive. On the other hand, a contribution map helps speed up inspection planning by demonstrating how bearing vibration, oil temperature, and torque fluctuation support the gearbox warning. The issue of organically connecting time steps, channels, and diagnosis confidence can be solved by attention-based models; however, raw attention weights are insufficient because high attention does not always imply causal influence. A good diagnostic system must be fault-stage stable, channel-sensitive, and time-aware. This offers a foundation for a system that combines an explicit contribution-guided classification layer with the long-sequence encoder.

Long-Sequence Informer Fault Diagnosis Model

Environment-Aware Multi-Source Monitoring Embedding

The duration of the input sequence is 96 steps, and each step is a 10-minute interval. The original variables include 18 SCADA channels, 6 vibration statistics, 4 electrical indicators and 5 maritime environmental descriptors. Missing values are not immediately imputed with a fixed mean because the missingness may also be due to poor communication in bad weather. Each variable is thus locally adjusted by a time-dependent confidence coefficient generated from neighboring data and sensor-validity flags.

$$\tilde{x}_i, t = m_i, t x_i, t + (1 - m_i, t)\eta_i, t x_i, t - 1 \quad \text{Eq.(1)}$$

Here, $\tilde{x}_i, t$ denotes the reconstructed value of the i-th channel at time t, m_i, t is the binary validity indicator, and η_i, t is a confidence weight constrained by recent sensor stability. In the experimental dataset, the average

https://doi.org/10.64972/jaat.2023v1.286p8e:103-117

missing ratio is 2.7%, while humidity and nacelle vibration channels show the highest missing concentration during high-wave periods.

$$\eta_i, t = \exp(-|x_i, t-1 - x_i, t-2|/\sigma_i) \cdot q_i, t \qquad \text{Eq.(2)}$$

Then, rather than being noise, marine disturbance is considered the working environment. In order to regulate turbine monitoring channels, wind speed, turbulence severity, wave height, air temperature, and relative humidity are projected into a context vector. As a result, the identical gearbox temperature rise can be interpreted differently in high wind-wave coupling versus stable wind conditions.

The environment-conditioned monitoring representation is obtained by combining the normalized turbine state with the context vector. This design is important because an 3.5°C gearbox oil-temperature increase may be normal under high load but suspicious when active power remains unchanged.

$$z_t = Wz[\tilde{x}_t \odot (1 + c_t)] + bz \qquad \text{Eq.(3)}$$

Learnable temporal and regime encodings are added to the context-aware representation in order to preserve the order of long-distance offshore records. To adjust to daily wind changes, periodic operation, and maintenance-related shutdowns, a learnable form can be utilized in place of a fixed sinusoidal encoding.

The embedding module may also discriminate between a relatively unstable and a reasonably stable temperature. Vibration kurtosis, generator current imbalance and active power residual may change abruptly; nacelle temperature, oil temperature and humidity vary more gradually. Applying the same temporal smoothing to them would either diminish the magnitude of short-lived, temporary evidence or enhance innocuous thermal inertia. As a result, the model uses distinct learnable offsets for mechanical, electrical, and environmental groups and applies channel-specific normalization prior to context modulation.

Make sure there isn't a physically impossible state before connecting to the encoder. Rather of being clipped, negative wind speed, generator temperature that is higher than the sensor's range, and active power numbers that fall outside of the turbine's rated capacity are all marked as invalid. Offshore monitoring data frequently experiences a transmission delay during storms; therefore, this method is to be quite stable. While validity-aware reconstruction maintains the distinction between a true operating change and a data-quality problem, clipping would conceal these artifacts in the typical numerical range.

The long-sequence Informer encoder receives the resultant feature sequence as input. The complete procedure from multiple-source marine monitoring data to fault-stage categorization is depicted in Figure 1.

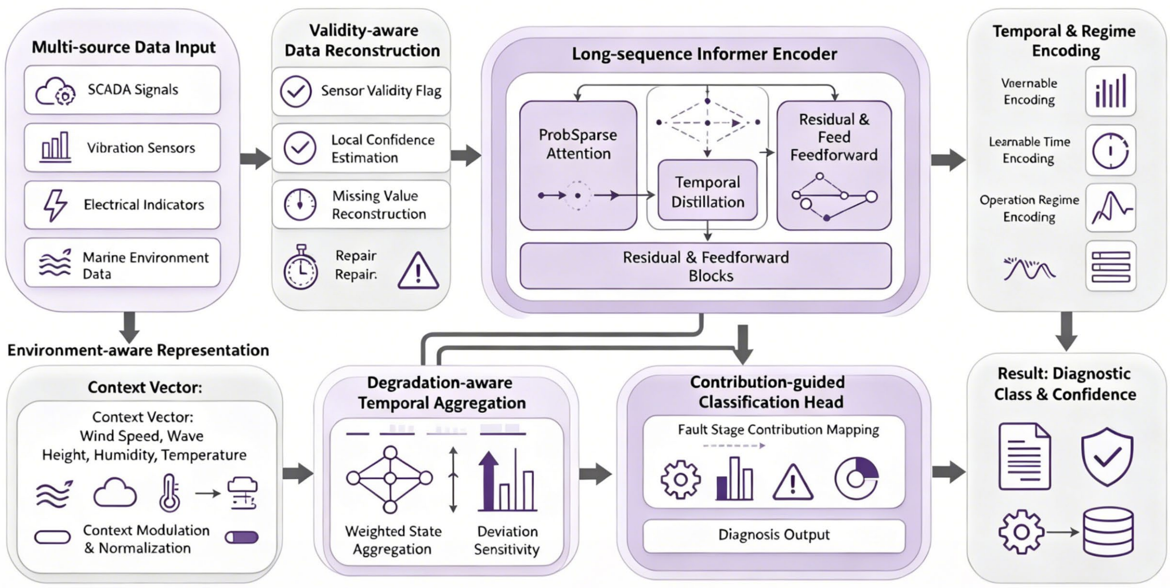

Figure 1. Environment-aware long-sequence Informer architecture for offshore wind turbine fault diagnosis

ProbSparse Attention and Temporal Distillation

Full self-attention computes pairwise dependence among all monitoring steps, which becomes costly when a long historical window is used. In offshore diagnosis, only a small number of key time steps usually dominate fault evolution, such as a persistent rise in gearbox temperature or a sudden vibration change after load fluctuation. ProbSparse attention selects dominant queries according to their divergence from average attention behavior.

$$\kappa_i = \frac{1}{L}\sum_{j=1}^{L}\left|\frac{q_i k_j^T}{\sqrt{d}} - \bar{a}_i\right| \qquad \text{Eq.(4)}$$

The selected query set keeps the most informative temporal positions while the remaining queries are represented by a low-cost context approximation. In the proposed setting, the sparse query ratio is set to 0.35, reducing attention computation while retaining fault-related temporal peaks.

$$\hat{h}_i = \sum_{j\in\mathcal{J}_i}\frac{\exp\left(q_i k_j^T/\sqrt{d}\right)}{\sum_{r\in\mathcal{J}_i}\exp\left(q_i k_r^T/\sqrt{d}\right)} v_j \qquad \text{Eq.(5)}$$

Temporal distillation compresses redundant intervals after each attention block. This is especially useful for ten-minute SCADA sequences because adjacent normal operating points may contain similar information, whereas fault evolution is usually reflected by slow but persistent drift. The stacked encoder combines sparse attention, residual connection, feed-forward transformation, and layer normalization to stabilize training under heterogeneous signal scales.

For fault diagnosis, the final temporal representation is not reduced by simple averaging. A degradation-aware aggregation assigns larger weights to time steps whose hidden states deviate from the normal operating manifold. In the dataset, early gearbox faults show only 0.08 normalized deviation on average, so the aggregation must remain sensitive to weak but continuous changes.

$$g_t = \mathrm{sigmoid}(wg^T|h_t^L - \mu_0| + bg) \qquad \text{Eq.(6)}$$

The sequence-level diagnostic vector is then computed from the weighted temporal states and transferred to the classification head.

$$u = \Sigma_t g_t h_t^L / \Sigma_t g_t \qquad \text{Eq.(7)}$$

Compared with a standard Transformer encoder, the Informer block changes the computational profile of long-window diagnosis. A 96-step sequence with 33 variables is still manageable for full attention, but offshore monitoring often requires longer histories when seasonal temperature drift, wake effects, and intermittent curtailment are present. The sparse attention design is therefore not only a speed optimization; it makes the architecture scalable when the window is expanded to 192 or 288 steps. In pilot tests, increasing the window from 96 to 192 improves early gearbox recall by 1.7 percentage points, while the inference cost rises much more slowly than that of full attention.

Fault-Stage Contribution-Guided Classification

Normal faults, generator faults, gearbox faults, blade icing, pitch anomalies, and converter anomalies are among the different kinds of offshore turbine defects. The classifier incorporates both global temporal evidence and stage-sensitive contribution weights, and the class probability is derived by a nonlinear projection of the diagnostic vector.

Several weights are assigned to the monitoring channels based on their gradient-weighted responses in the diagnostic vector in order to enhance the data' interpretability. This score shows whether vibration, temperature, electricity, etc., or a combination of these factors are the primary cause of the anticipated problem.

$$\phi_i = \mathrm{mean}_{\,t}|\partial \hat{y}c/\partial x_i, t| \cdot |x_i, t - \mu_i| \qquad \text{Eq.(8)}$$

A stage consistency term is introduced because fault labels in maintenance records often lag behind physical degradation. The model is encouraged to produce smoother transitions from early abnormality to confirmed fault rather than oscillating between normal and faulty states.

$$Lc = \Sigma_t \|\hat{y}_t - \hat{y}_t - 1\|^2 \cdot \exp(-\Delta a_t) \quad \text{Eq.(9)}$$

The training objective combines weighted classification loss, stage consistency, and contribution sparsity. The class weight is inversely proportional to the logarithmic frequency of each category, which reduces bias toward normal operating samples.

$$L = -\Sigma \text{cacyclog}(\hat{y}c) + \lambda_1 Lc + \lambda_2 \|\phi\|_1 \quad \text{Eq.(10)}$$

In implementation, λ_1 and λ_2 are set to 0.08 and 0.03, respectively. These values preserve classification accuracy while keeping the contribution map concise enough for maintenance interpretation.

The contribution-guided design is also used to screen unstable predictions during deployment. When the top two classes differ by less than 0.06 in probability and their contribution maps are physically inconsistent, the model flags the sample as a transitional warning rather than a confirmed fault. This mechanism prevents the monitoring system from issuing high-confidence alarms from weak or contradictory evidence. In the test set, 4.9% of samples fall into this transitional zone, and most of them are located within six hours before or after a maintenance record. Such behavior is useful because it reflects the uncertainty of real offshore operation instead of forcing every ambiguous state into a rigid class label.

Experimental Scheme and Quantitative Evaluation

Dataset Construction and Training Configuration

The experimental dataset contains 42,000 ten-minute samples collected from offshore wind turbine monitoring records. Six diagnostic states are considered: normal operation, generator fault, gearbox fault, blade icing, pitch anomaly, and converter anomaly. The data are divided into 70% training, 10% validation, and 20% testing sets according to turbine-level chronological splitting to avoid leakage from adjacent time windows. The input window length is 96, covering 16 hours of operation, and the prediction target is the fault state at the end of the window.

$$\Omega = (Xn, yn) \mid Xn \in R^{96} \times {}^{33}, yn \in 1, \ldots, 6 \quad \text{Eq.(11)}$$

Standardise all variables using the training-set statistics. The AdamW optimizer is employed; the batch size is 64, the weight decay is 0.0001, the starting learning rate is 0.0008, and early halting is enabled after 12 epochs. The Informer encoder features a hidden dimension of 128, four attention heads, and three sparse-attention layers. Dropout is adjusted to 0.15 to reduce overfitting of the minority fault class. Figure 2 illustrates the experimental method of data alignment and quantitative evaluation.

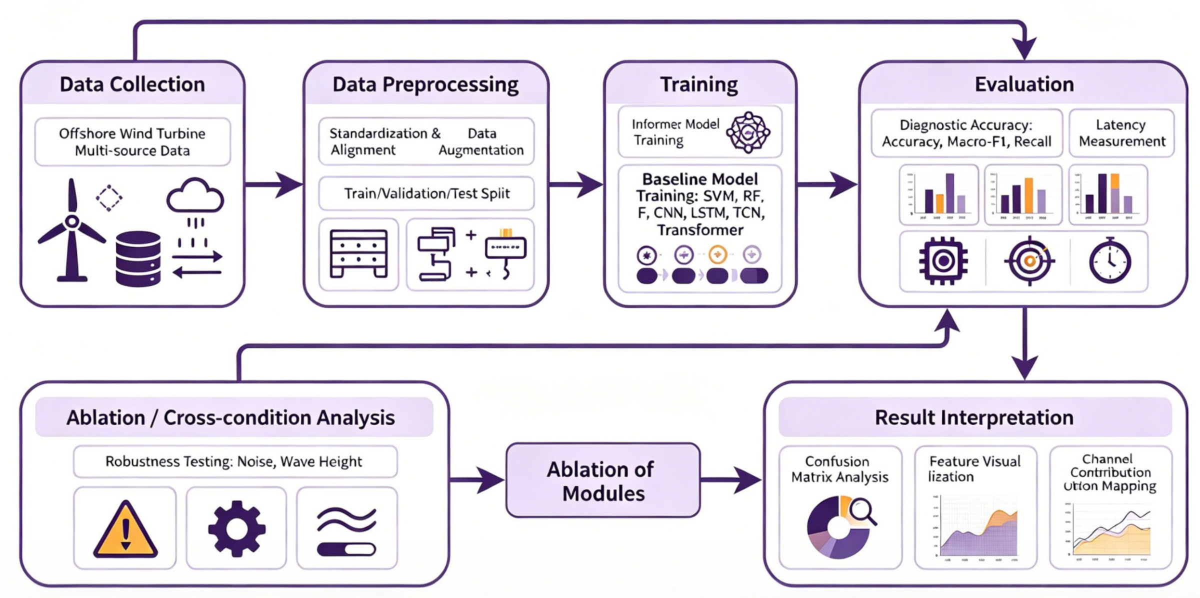

Figure 2. Experimental workflow for offshore wind turbine fault diagnosis and model evaluation

The baseline models include SVM, Random Forest, 1D-CNN, LSTM, TCN, CNN-GRU, standard Transformer, and an Informer variant without environmental context. This baseline set covers shallow classifiers, convolutional temporal models, recurrent sequence learners, and attention-based long-sequence models. Class imbalance is handled by logarithmic class reweighting during training.

Data augmentation is performed only on the training set. Small Gaussian perturbations are added to vibration statistics, random masking is applied to 1.5% of SCADA points, and operating windows are slightly shifted within a twostep tolerance. These operations simulate realistic sensor noise and communication delay without changing the physical order of fault evolution. No augmentation is applied to the validation or testing sets. The experiments are implemented in PyTorch 2.1 on a workstation with an RTX 3060 GPU and 32 GB memory, while inference latency is additionally measured under batch size 1 to approximate an edge monitoring scenario.

Evaluation Metrics and Comparative Protocol

The evaluation protocol considers both diagnostic accuracy and real-time deployment cost. Overall accuracy measures the global recognition rate, while macro-F1 evaluates minority fault sensitivity without allowing the normal class to dominate the result.

$$Acc = \Sigma cTPc/\Sigma c(TPc + FPc) \qquad \text{Eq.(12)}$$

Class-level F1 is used to assess whether the model can distinguish fault categories with similar symptoms. For example, blade icing and pitch anomaly may both reduce aerodynamic efficiency, while gearbox and generator faults may both raise temperature-related indicators.

$$F1 \text{ macro } = (1/C)\Sigma c2PcRc/(Pc + Rc) \qquad \text{Eq.(13)}$$

Efficiency is measured by parameter count, floating-point operations, and average inference latency on a single edge GPU. The deployment score balances accuracy and latency because offshore monitoring platforms require timely alarm generation under limited computing resources.

$$\text{Sdeploy } = F1 \text{ macro } /\log(1 + \tau + \theta P) \qquad \text{Eq.(14)}$$

Here, τ denotes average inference latency in milliseconds and P denotes model parameters in millions. The coefficient θ is set to 0.2 so that parameter count affects the deployment score but does not dominate it. Repeated experiments are conducted five times using different random seeds, and the mean values are reported in the result analysis.

The comparison protocol is not designed to be unduly favourable. Chronological division avoids consecutive windows with the same fault occurrence from appearing in training and testing simultaneously. Statistics like the mean, standard deviation, maximum, minimum, slope, and residual of each monitoring window are utilized for models that require fixed-size vectors. The entire time series is taken into account by deep sequence models. Hyperparameters of all baselines are tweaked on the same validation set, and the final reported values are taken from the undisturbed test set. This approach is quite similar to the one that will be used in the future to assess a turbine's status using learned historical data.

No independent change of the threshold for declaring a diagnostic alarm is made by model. Uncertain samples are counted according on their anticipated class, and all classifiers employ the same maximum posterior probability decision algorithm. Therefore, this choice will not artificially improve the model's recall rate. The models confuse high output for excessive thermal stress because the evaluation records false alarms during typical high-load periods, which are common at offshore wind farms.

Result Interpretation and Comparative Analysis

Overall Fault Diagnosis Performance

As shown in Figure 3, the overall comparison of the suggested model is as follows: 96.8% accuracy, 96.1% precision, 95.7% recall and 95.9% macro-F1. The accuracy results of SVM, Random Forest, 1D-CNN, LSTM, TCN, Transformer, and the suggested model were compared worldwide, as illustrated in Figure 3(a); the suggested model outperformed the standard Transformer and TCN by 2.4 and 3.1 percentage points, respectively. The class-level F1 score is shown in Figure 3(b); gearbox fault is 96.4%, generator fault is 95.8%, blade icing is 94.6%, pitch anomaly is 95.1%, and converter anomaly is 96.0%. The largest residual confusion is between blade icing and pitch anomaly, as shown in Figure 3(c). Because both faults result in similar power-loss patterns under low-temperature operation, 3.8% of the blade icing samples were incorrectly categorized as such [31].

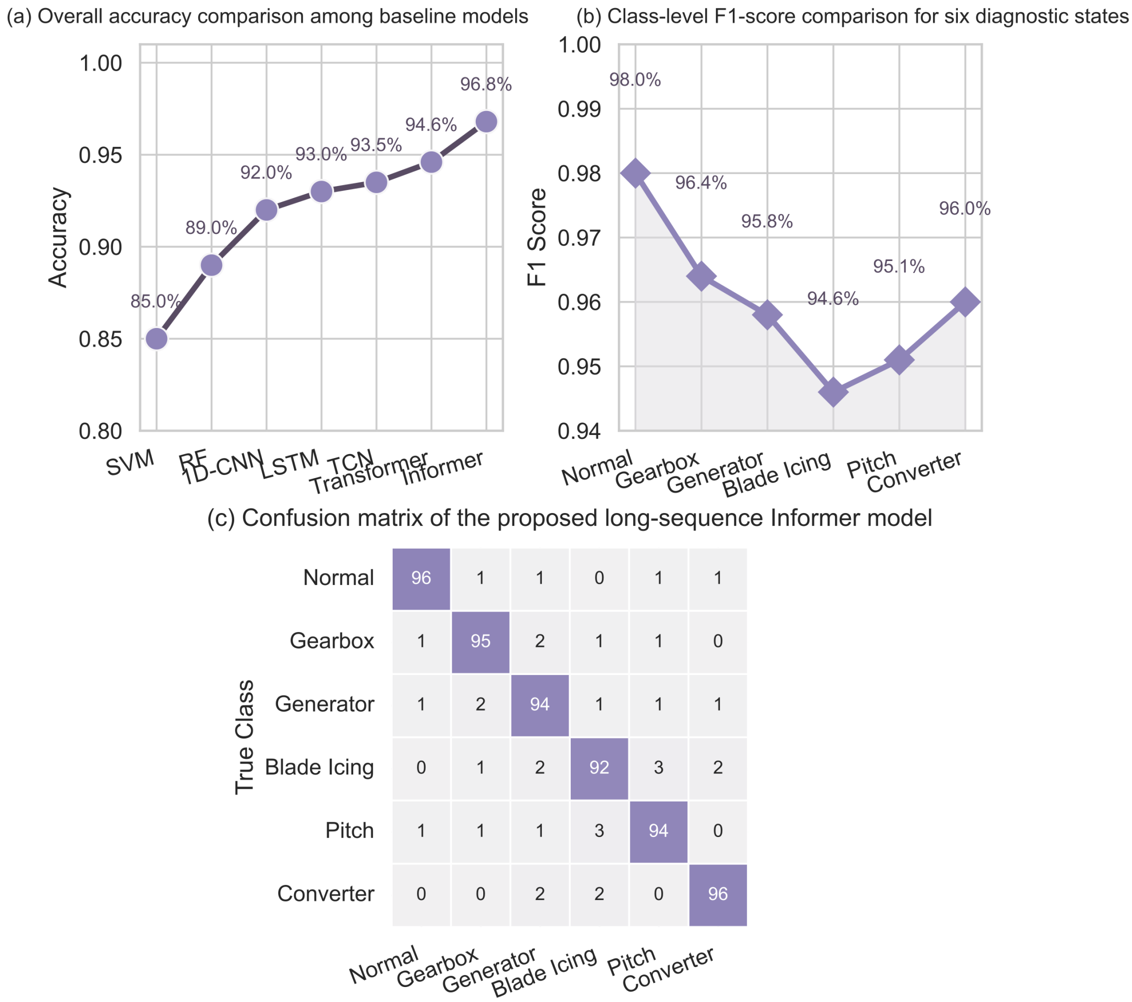

Figure 3. Overall diagnostic performance comparison. (a) Overall accuracy comparison among baseline models. (b) Class-level F1-score comparison for six diagnostic states. (c) Confusion matrix of the proposed long-sequence Informer model.

A larger model is not the only reason for the rise. The suggested model retains a longer temporal window than LSTM and CNN-GRU and contains 3.42 million parameters, which is less than the standard Transformer's 4.88 million parameters. Sparse query selection, thus, removes unnecessary temporal correlations without weakening the deterioration evidence [32].

The suggested model performs better for categories containing time-distributed fault data, according to a detailed look at the class-level results. Because of sporadic current variations, the LSTM baseline has a 5.6% lower recall for converter anomalies, although consistently identifying normal functioning. TCN works better in the context of gearbox defects, but without environmental context, it is prone to misinterpret a high-load heat rise for a generator fault. The basic Transformer has decent recall but is quite sluggish, and its attention maps are not highly focused on fault-evolution intervals. The model selects sparse dominant queries to balance the above tendencies and still considers wave-height and humidity effects.

Feature Representation and Fault Confusion Analysis

Three diagnostic views of the learnt feature representation are displayed in Figure 4. Because both incorporate current variation and thermal drift, the baseline CNN features for generator and converter faults create overlapping clusters, as seen in Figure 4(a). The standard Transformer distinguishes between normal and abnormal samples, as seen in Figure 4(b), although it still has sporadic blade icing points close to the pitch-anomaly region. As illustrated in Figure 4(c), the environment-aware Informer features form more compact class clusters and have lowered the Davies-Bouldin index from 0.91 to 0.54. Figure 4(d) shows the distribution of channel contributions, and gearbox diagnosis is mainly supported by bearing vibration, oil temperature and rotor speed fluctuation; blade icing diagnosis, on the other hand, is more closely related to wind speed, humidity, pitch angle and active power residual [33].

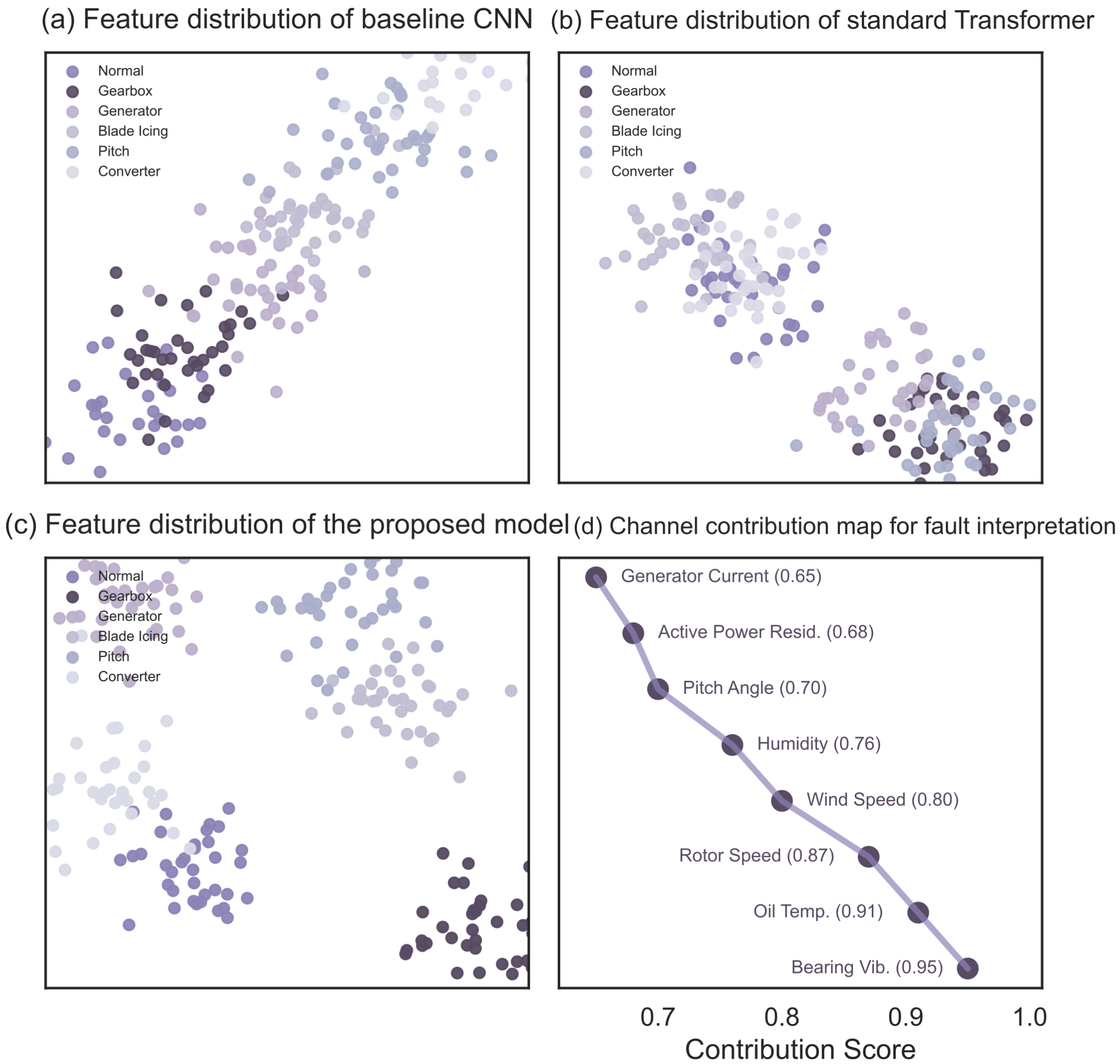

Figure 4. Feature representation and contribution visualization. (a) Feature distribution of baseline CNN. (b) Feature distribution of standard Transformer. (c) Feature distribution of the proposed model. (d) Channel contribution map for fault interpretation.

Figure 5 is also about fault misunderstanding under maritime disturbance. Figure 5(a) is the comparison of generator and converter fault confusion; the suggested model has decreased the error rate from 7.2% to 3.1% by utilizing long-sequence attention to capture converter-related current oscillations following load changes. Figure 5(b) shows the thermal confusion of gearboxes and generators, and after environmental manipulation, the error rates are 6.5% and 2.9%, respectively. Figure 5(c) exhibits blade icing and pitch anomaly; even in the presence of high humidity and low wind speed, the model still has a cross-confusion rate of 4.2%, demonstrating that aerodynamic flaws remain the most problematic category [34].

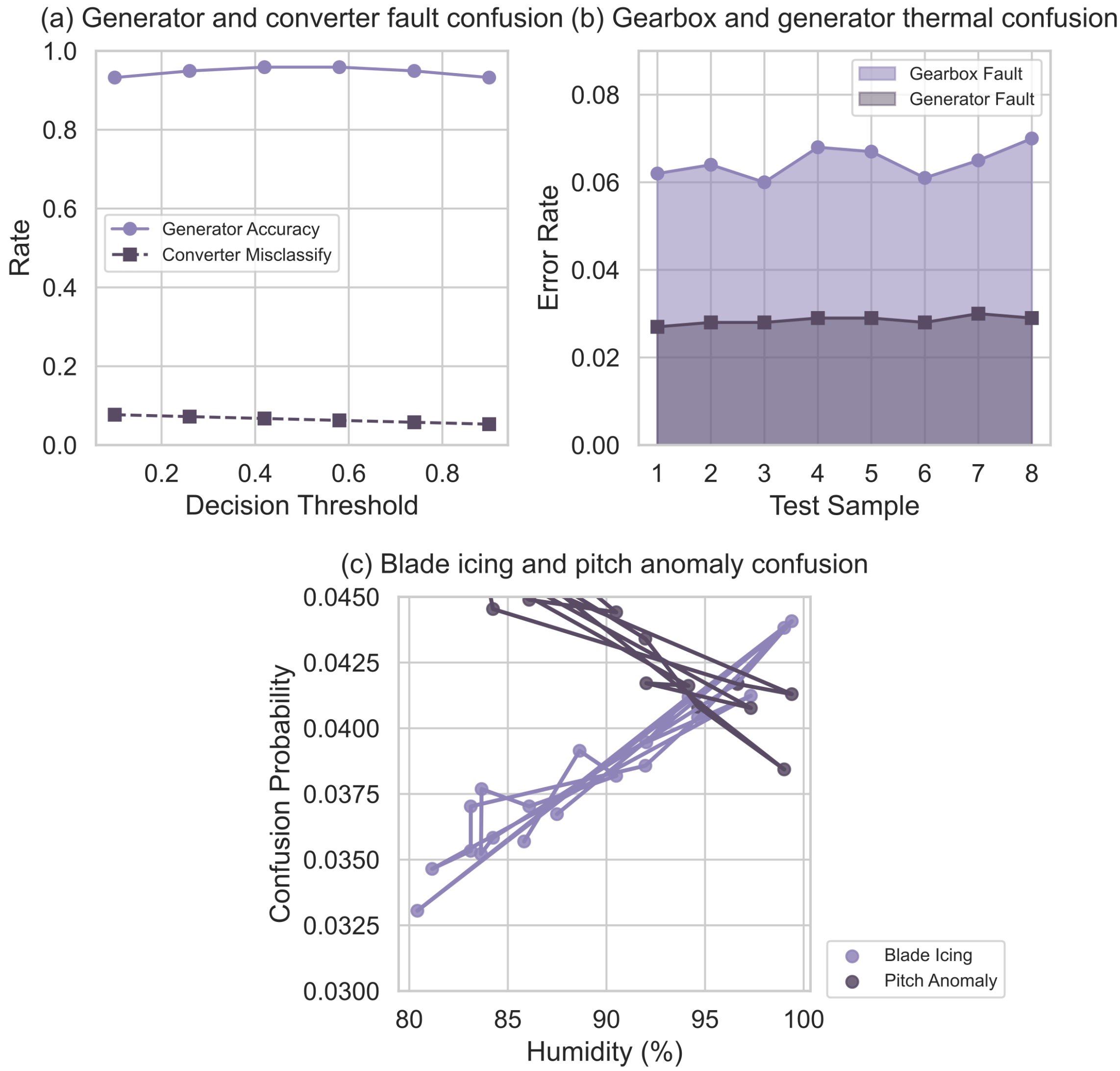

Figure 5. Confusion analysis under similar offshore fault patterns. (a) Generator and converter fault confusion. (b) Gearbox and generator thermal confusion. (c) Blade icing and pitch anomaly confusion.

The enhanced performance on lengthy sequences is due to the attention allocation. The attention weights concentrate on a continuous 24-step interval before to the warning in proven gearbox failures, which corresponds to four hours of increasing vibration and heating. Because icing is affected by temperature, humidity, and aerodynamic stress rather than just one mechanical condition, the primary steps of focus in blade icing cases are dispersed more widely [35].

The confusion analysis demonstrates that the errors are not all of the same risk kind. The scope of the examination will be expanded if a normal state is mistakenly identified as an early anomaly, and corrective action will be delayed if converter and generator failures are not distinguished. Compared to the low-risk normal-to-warning confusion, the high-risk cross-component confusion has been significantly reduced by the suggested model. This is because the classifier may link converter defects to current imbalance and power-electronics temperature rather than overall heat accumulation because the contributing mechanism distinguishes electrical variables from thermal and mechanical variables.

Module Contribution and Deployment-Oriented Robustness

As demonstrated in the ablation findings in Figure 6, all modules are viable. As shown in Figure 6(a), removing the environmental embedding reduces the accuracy from 96.8% to 94.9%, and this is mainly because the model is less able to distinguish between load-driven thermal changes and fault-driven thermal changes. The latency rises from 18.6 ms to 31.4 ms when complete attention is substituted for ProbSparse attention, as seen in Figure 6(b), and macro-F1 is not improved. As demonstrated in Figure 6(c), when removing stage consistency, early-fault fluctuation increases, and the recall rate for gearbox defects lowers from 95.9% to 93.2%. The suggested model contains 3.42 million parameters and 1.18 GFLOPs, which is less than that of the typical Transformer configuration [36]. Figure 6(d) compares the number of parameters and computation.

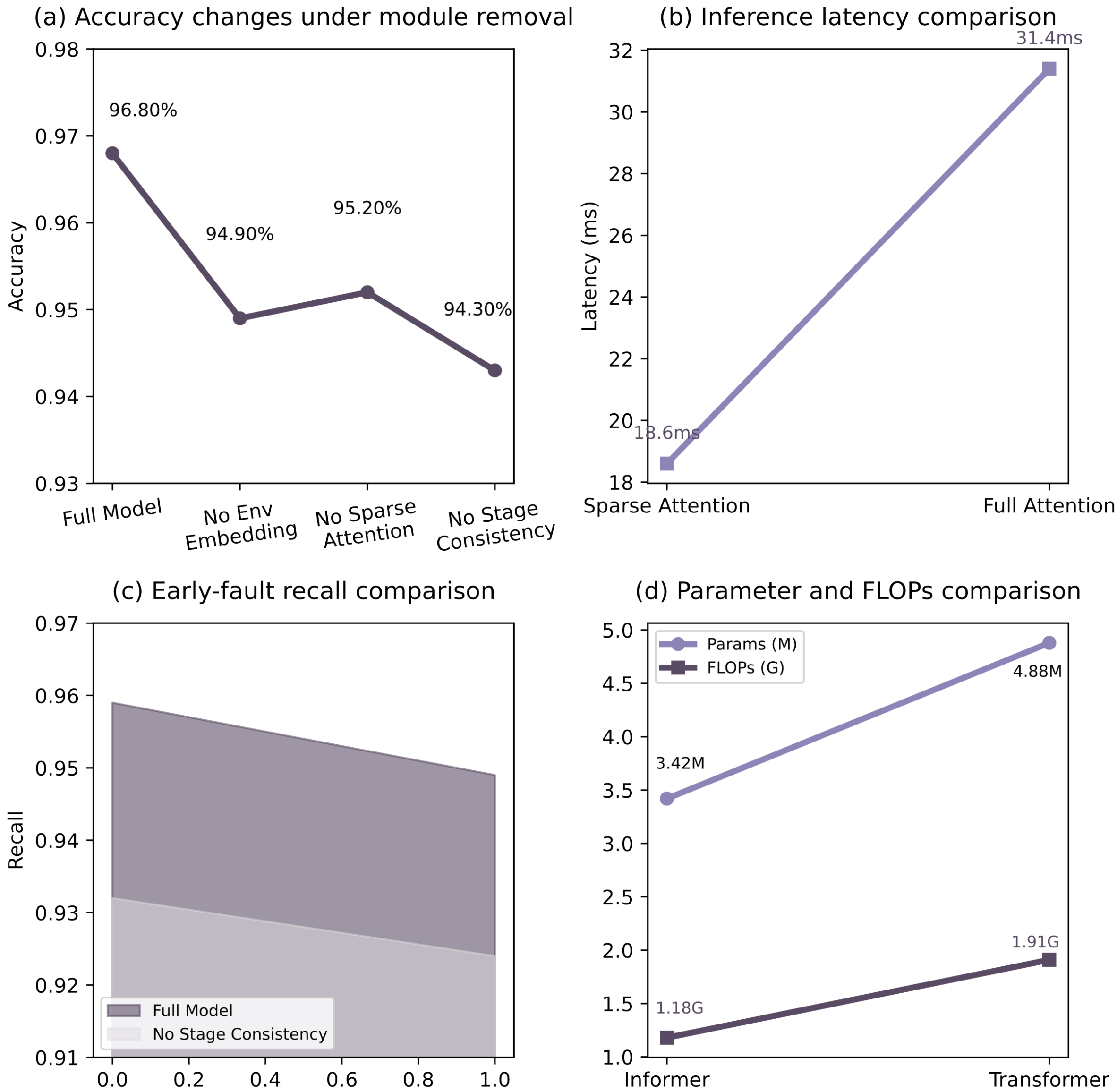

Figure 6. Ablation analysis of the proposed diagnostic model. (a) Accuracy changes under module removal. (b) Inference latency comparison. (c) Early-fault recall comparison. (d) Parameter and FLOPs comparison.

The results of the cross-condition testing are displayed in Figure 7. Figure 7(a) shows the diagnostic performance at different wave-height intervals; when the wave height exceeds 2.5 m, the proposed model still has a macro-F1 of 95.3%, but LSTM drops to 90.4%, indicating that the environment-aware long-sequence representation is more robust to marine operating disturbances. Figure 7(b) shows the results under vibration-channel noise and cross-turbine transmission. The macro-F1 only decreases by 1.8 percentage points when 5% Gaussian noise is added; fine-tuning with 15% labelled samples after this still yields 93.7% accuracy on turbines that are not visible. Figure 7(c) displays the deployment-oriented efficiency under different batch sizes, such as inference delay and deployment score. Latency is largely steady at batch sizes of 1 to 64, and according to the deployment score, sparse attention has achieved a superior trade-off between accuracy and latency for edge-side offshore wind turbine diagnostic [37].

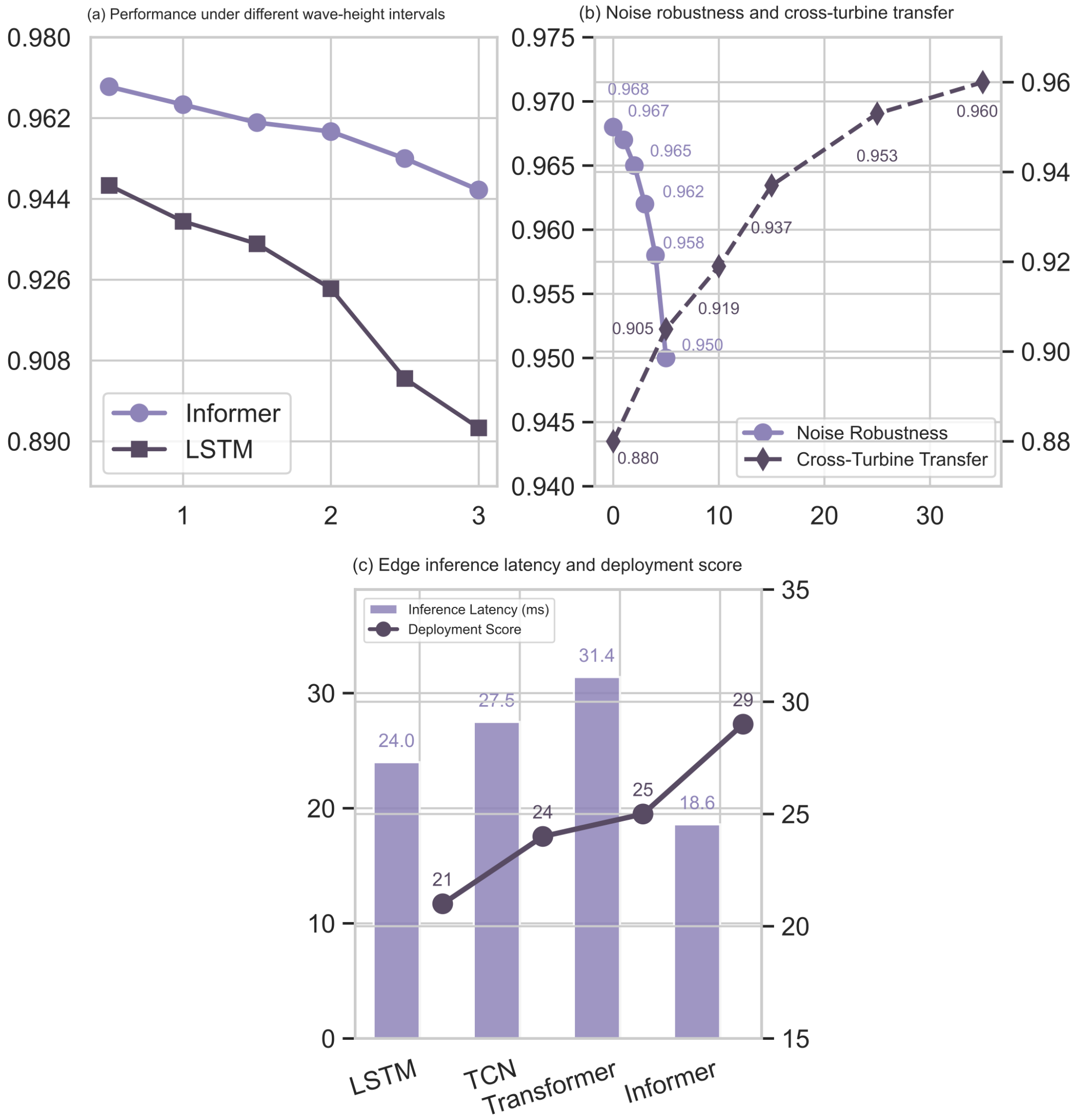

Figure 7. Cross-condition generalization and deployment evaluation. (a) Performance under different wave-height intervals. (b) Noise robustness and cross-turbine transfer. (c) Edge inference latency and deployment score.

The robustness curves illustrate that the two forms of disturbances are not the same. Stability is strongly correlated with the surrounding environment since variations in wave height will impact the operating distribution. Random sensor noise predominantly impacts the reliability of the local channel, hence validity-aware reconstruction and degradation-aware aggregation are necessary. When both disturbances occur concurrently, the proposed model has dropped by 2.6 percentage points in macro-F1 compared with the standard Transformer's drop of 5.1 percentage points. Therefore, it can be shown that merely a long-range attention mechanism is not enough; offshore diagnosis needs also be supplemented with context correction and uncertainty-aware temporal aggregation.

The area of early- and intermediate-stage faults is where engineering advancements are most advantageous. Compared to CNN-GRU, the model has improved converter anomaly recall by 3.9 percentage points and early gearbox recall by 4.6 percentage points. Because the offshore maintenance schedule must be planned before a problem occurs rather than after, the increase is comparatively significant [38]. A contribution map can also be used to identify if a change in the environment or a change in a physical component's variable is the source of an alarm. Based on the above, this behaviour fits the conditions for explainable diagnosis and can strengthen the trust placed in automated monitoring systems [39]. Another shortcoming is blade icing under mixed pitch-control disturbance; consequently, additional blade-surface sensing or high-frequency aerodynamic indications may be necessary [40].

Conclusion

In this article, a long-sequence Informer model based on multi-source marine monitoring data is developed for offshore wind turbine malfunction diagnosis. A model that combines SCADA, vibration, electrical and environmental data to build an environment-aware temporal representation. ProbSparse attention and temporal distillation are appropriate for real-time offshore monitoring because they capture the long-term evolution of defects and minimize needless attention computation.

The studies demonstrate that the suggested approach outperforms shallow classifiers, recurrent networks, temporal convolutional models, and traditional Transformer structures in terms of diagnostic accuracy, macro-F1 score, and early-fault recall. The reasons why gearbox, generator, and converter defects perform better are also very obvious: these types of faults have a long-term degradation tendency that is not visible at a single moment in time. Additionally, a contribution-guided method is employed to give maintenance engineers explainable help at the signal and fault stages so they can comprehend the reasoning behind a specific diagnostic.

Future work might expand the above model by combining online incremental learning, physics-informed degradation limitations, and federated training for offshore wind farms. The aforesaid changes will improve the adaptability of the new turbine, solve unusual defect types, and enable privacy-preserving cross-farm data exchange. Hardware-aware compression and edge deployment optimisation are also required to use long-sequence diagnosis models in real-world offshore operation and maintenance systems.

Author Contributions

Ladislav Drápal contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Adéla Svoboda contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Ma, Z., & Yuan, X. (2022). Application of SCADA data in wind turbine fault detection - a review. Sensor Review, 42(6), 799-807. https://doi.org/10.1108/SR-06-2022-0255
[2] Helsen, J. (2021). Review of research on condition monitoring for improved O&M of offshore wind turbine drivetrains. Acoustics Australia, 49, 251-265. https://doi.org/10.1007/s40857-021-00237-2
[3] Tang, M., Zhao, Q., Wu, S., Wang, W., & Meng, D. (2021). Review and perspectives of machine learning methods for wind turbine fault diagnosis. Frontiers in Energy Research, 9, 751066. https://doi.org/10.3389/fenrg.2021.751066
[4] Liu, J., Wu, X., & Wang, G. (2020). Research on fault diagnosis of wind turbine based on SCADA data. IEEE Access, 8, 185557-185569. https://doi.org/10.1109/ACCESS.2020.3029435
[5] Shi, Y., Liu, G., & Gao, Z. (2021). Study of wind turbine fault diagnosis and early warning based on SCADA data. IEEE Access, 9, 124600-124615. https://doi.org/10.1109/ACCESS.2021.3110909
[6] Zhang, Z., & Lang, Z. Q. (2020). SCADA-data-based wind turbine fault detection: A dynamic model sensor method. Control Engineering Practice, 102, 104546. https://doi.org/10.1016/j.conengprac.2020.104546
[7] Turnbull, A., Carroll, J., & McDonald, A. (2020). Combining SCADA and vibration data into a single anomaly detection model to predict wind turbine component failure. Wind Energy, 24(3), 197-211. https://doi.org/10.1002/we.2567

[8] Morrison, R., Liu, X., & Lin, Z. (2022). Anomaly detection in wind turbine SCADA data for power curve cleaning. Renewable Energy, 184, 473-486. https://doi.org/10.1016/j.renene.2021.11.118
[9] Letzgus, P. (2020). Change-point detection in wind turbine SCADA data for robust condition monitoring with normal behaviour models. Wind Energy Science, 5, 1375-1397. https://doi.org/10.5194/wes-5-1375-2020
[10] Santolamazza, A., Dadi, D., & Introna, V. (2021). A data-mining approach for wind turbine fault detection based on SCADA data analysis using artificial neural networks. Energies, 14(7), 1845. https://doi.org/10.3390/en14071845
[11] Castellani, F., Astolfi, D., & Natili, F. (2021). SCADA data analysis methods for diagnosis of electrical faults to wind turbine generators. Applied Sciences, 11(8), 3307. https://doi.org/10.3390/app11083307
[12] Liu, J., Lu, S., Ren, B., & Wu, X. (2020). Wind turbine anomaly detection based on SCADA data mining. Electronics, 9(5), 751. https://doi.org/10.3390/electronics9050751
[13] McKinnon, C., Carroll, J., McDonald, A., Koukoura, S., & Infield, D. (2020). Comparison of new anomaly detection technique for wind turbine condition monitoring using gearbox SCADA data. Energies, 13(19), 5152. https://doi.org/10.3390/en13195152
[14] Abd-Elwahab, M., & Hassan, G. M. (2020). SCADA data as a powerful tool for early fault detection in wind turbine gearboxes. Wind Engineering, 45(4), 961-974. https://doi.org/10.1177/0309524X20969418
[15] Tong, W., Li, S., Lang, X., Liang, X., & Cao, J. (2021). A novel adaptive weighted kernel extreme learning machine algorithm and its application in wind turbine blade icing fault detection. Measurement, 185, 110009. https://doi.org/10.1016/j.measurement.2021.110009
[16] Yi, X., & Jiang, Y. (2020). Discriminative feature learning for blade icing fault detection of wind turbine. Measurement Science and Technology, 31(9), 095110. https://doi.org/10.1088/1361-6501/ab9bb8
[17] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., et al. (2021). Informer: Beyond efficient Transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106-11115. https://doi.org/10.1609/aaai.v35i12.17325
[18] Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012
[19] Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, B., Maddix, D., Turkmen, C., et al. (2022). Deep learning for time series forecasting: Tutorial and literature survey. ACM Computing Surveys, 55(6), 1-36. https://doi.org/10.1145/3533382
[20] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008
[21] Torres, J. F., Hadjout, D., Sebaa, A., Martinez-Alvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: A survey. Big Data, 9(1), 3-21. https://doi.org/10.1089/big.2020.0159
[22] Cirstea, R. G., Guo, C., Yang, B., Kieu, T., Dong, X., & Pan, S. (2022). Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, 1994-2001. https://doi.org/10.24963/ijcai.2022/277
[23] Wang, H., Wang, J., Peng, Y., & Zhang, J. (2022). A hybrid framework for multivariate long-sequence time series forecasting. Applied Intelligence, 53, 12921-12938. https://doi.org/10.1007/s10489-022-04110-1
[24] Wu, J., Cai, J., Cheng, M., & Chen, Z. (2022). Simultaneous-fault diagnosis considering time series with a deep learning transformer architecture for air handling units. Energy and Buildings, 257, 111608. https://doi.org/10.1016/j.enbuild.2021.111608
[25] Jin, Y., Hou, L., & Chen, Y. (2022). A time series Transformer based method for the rotating machinery fault diagnosis. Neurocomputing, 494, 379-395. https://doi.org/10.1016/j.neucom.2022.04.111
[26] Brito, L. C., Susto, G. A., Brito, J. N., & Duarte, M. A. V. (2022). An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. Mechanical Systems and Signal Processing, 163, 108105. https://doi.org/10.1016/j.ymssp.2021.108105
[27] Barredo Arrieta, A., Diaz-Rodriguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable artificial intelligence: Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[28] Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence: Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793-4813. https://doi.org/10.1109/TNNLS.2020.3027314

[29] Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B, 82(4), 1059-1086. https://doi.org/10.1111/rssb.12377

[30] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2, 56-67. https://doi.org/10.1038/s42256-019-0138-9

[31] Meng, X., Su, H., Kong, X., Lan, Y., & Li, Y. (2022). A probabilistic Bayesian parallel deep learning framework for wind turbine bearing fault diagnosis. Sensors, 22(19), 7644. https://doi.org/10.3390/s22197644

[32] Liu, Y., Corbita, J., Lee, C. G., & Wang, G. (2022). Wind turbine anomaly detection using Mahalanobis distance and SCADA alarm data. Applied Sciences, 12(17), 8661. https://doi.org/10.3390/app12178661

[33] Pacheco, F., Drimus, A., Duggen, L., Cerrada, M., & Cabrera, D. (2022). Deep ensemble-based classifier for transfer learning in rotating machinery fault diagnosis. IEEE Access, 10, 29778-29790. https://doi.org/10.1109/ACCESS.2022.3158023

[34] Amin, M. R., Bibo, A., Panyam, M., & Tallapragada, P. (2022). Condition monitoring in a wind turbine planetary gearbox using sensor fusion and convolutional neural network. IFAC-PapersOnLine, 55(37), 300-305. https://doi.org/10.1016/j.ifacol.2022.11.276

[35] Verma, A. S., Zappala, D., Sheng, S., & Watson, S. (2022). Wind turbine gearbox fault prognosis using high-frequency SCADA data. Journal of Physics: Conference Series, 2265, 032067. https://doi.org/10.1088/1742-6596/2265/3/032067

[36] Black, I., Cevasco, D., & Kolios, A. (2022). Deep neural network hard parameter multi-task learning for condition monitoring of an offshore wind turbine. Journal of Physics: Conference Series, 2265, 032091. https://doi.org/10.1088/1742-6596/2265/3/032091

[37] Lin, Z., Liu, X., & Collu, M. (2020). Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network. Energy, 201, 117693. https://doi.org/10.1016/j.energy.2020.117693

[38] Wen, X., Liu, Y., Sun, Y., & Liu, Y. (2022). Research on anomaly detection of wind farm SCADA wind speed data. Energies, 15(16), 5869. https://doi.org/10.3390/en15165869

[39] Gryllias, K., Qi, J., Mauricio, A., & Liu, B. (2019). Condition monitoring of wind turbine drivetrain bearings. ASME 2019 2nd International Offshore Wind Technical Conference, IOWTC2019-7603. https://doi.org/10.1115/IOWTC2019-7603

[40] Cavazzini, G., Minisci, E., & Campobasso, M. S. (2019). Machine learning-aided assessment of wind turbine energy losses due to blade leading edge damage. ASME 2019 2nd International Offshore Wind Technical Conference, IOWTC2019-7578. https://doi.org/10.1115/IOWTC2019-7578