EAI Endorsed Transactions
on Scalable Information Systems Research Article **EAI.EU**

A Reliable Data Fusion and Predictive Maintenance Framework of Industrial Internet of Things Using Explainable Artificial Intelligence: Improving Resilience and Fast Recovery in Future Manufacturing Systems

Yiting Bai[1,*]

[1]School of Information Science and Technology, Dalian Maritime University, Dalian 116033, China

Abstract

INTRODUCTION: Due to the development of the Industrial Internet of Things (IIoT), it has been possible to make the predictive maintenance of the manufacturing industry using data. Nevertheless, recent developments in manufacturing systems are prone to higher levels of disruptions caused by sensor degradation, data anomalies and equipment failures, which threaten production continuity and supply chain resilience. Conventional fault predictor models have no clear data quality evaluation and readability rendering them ineffective in aiding quick recovery.
OBJECTIVES: In order to overcome these challenges, the present paper introduces TEF-Net as a reliable data fusion and predictive maintenance approach based on explainable artificial intelligence, in particular, to improve the resiliency of manufacturing systems and their ability to recover quickly in the case of data quality issues.
METHODS: The method starts by building four data quality indicators (completeness, stability, consistency across time, correlation consistency), and produces sensor trust measures through TrustNet. Such trust measures are included in dynamic graph calibration and graph attention fusion to reduce low-quality data and combine multiple source characteristics. Fault probabilities are then extracted by a Transformer-GRU architecture that uses degradation-related temporal features, and explainability is provided using trust scores, attention weights, and feature contributions.
RESULTS: The experimental findings prove that TEF-Net has an accuracy of 0.947, a recall of 0.932, a F1-score of 0.936, and an AUC of 0.982, which is better than comparative models. In multi-error disturbance environments which are used to mimic manufacturing disruptions in practice, TEF-Net retains an F1-score of 0.884, which is much less degraded than base approaches, which is a direct indicator of its ability to make manufacturing systems resilient.
CONCLUSION: The results support the statement that TEF-Net is much more accurate, strong, and understandable in predicting IIoT maintenance compared to other models and can be considered as a reliable basis of AI when it comes to predicting failures early and recovering them quickly in the future.

Keywords: IoT, explainable AI, trusted data fusion, fault prediction, graph attention network

Received on 14 July 2026, accepted on 10 August 2026, published on 21 September 2026

doi: 10.4108/eetsis.14025

1. Introduction

The last few years have seen the world manufacturing systems experiencing unprecedented disruptions due to equipment failure, volatility in supply chains and poor quality of sensor data [1]. Anticipation, resistance, and quick recovery to these disruptions, which is defined as manufacturing resilience, has become a strategic necessity of

*Corresponding author. Email: 3404449856@qq.com

Industry 4.0 and the new Industry 5.0 paradigm [2]. Predictive maintenance, which can be implemented through the Industrial Internet of Things (IIoT) is one of the pillars of such resilience because it allows actively detecting faults and scheduling maintenance based on the faults conditions [3]. Multi-source sensor data, such as vibration, temperature, current, pressure, rotational speed, and load, are constantly gathered by IIoT systems to offer data support to AI-based fault prediction [4]. But, actual industrial settings are dynamic in nature and sensor information is frequently corrupted by noise pollution, missing values, sensor drift, and communication errors [5] resulting in poor quality of input data and accordingly decreased predictive accuracy of fault predicting models. Such data quality problems do not just reduce prediction precision but also slow down the process of detecting faults and increase the time of the recovery of the system which directly reduces manufacturing resilience [6].

Recently, deep learning techniques have been extensively used in industrial fault prediction problems [7]. Neural networks including CNN, LSTM, GRU, Transformer, and graph neural networks can also learn time relationships and sensor interconnections automatically [8]. Nevertheless, the two significant defects of the current approaches are: firstly, many models take the reliability of the input data as a given fact without explicitly assessing the quality of sensor data, which makes it challenging to reduce the effects of poor quality data on the predicted results; secondly, deep models tend to be highly black-boxed[9], and this restricts the ability to understand the underlying rationale of fault prediction outputs[10], and hence their use in high-reliability industrial operation and maintenance settings; finally, and most importantly, current approaches do not explicitly design manufacturing resilience, i.e., they do not measure how prediction accuracy changes during disruptions or offer explainable recommendations that would help to make quick recovery choices[11].

In order to overcome the mentioned problems, the present paper suggests a Trustful IoT Data Fusion and Fault Prediction model named TEF-Net, which is implemented using explainable artificial intelligence [12]. The approach takes multi-source sensor time-series data as input, initially tests sensor credibility with the help of data quality characteristics and TrustNet[13], then incorporates this credibility in dynamic graph construction and graph attention fusion procedures[14] to obtain trustworthy multi-source data fusion and lastly, predicts the fault probability through a Transformer-GRU framework[15] and offers an explainable analysis with credibility metrics, attention outcomes, and feature contribution scores[16].TEF-Net has three mechanisms that directly support manufacturing resiliency by explicitly modeling data trustworthiness and quantifying prediction reliability under disturbances: (1) early identification of data-quality anomalies that can cause equipment failure, (2) strong prediction performance that persists despite sensor degradation and (3) explainable outputs, which allow maintenance practitioners to make informed recovery choices. To give a visual overview of the suggested approach, Figure 1 provides a summary of the complete TEF-Net process, starting with the acquisition of the multi-source sensor data, followed by the evaluation of trustworthiness, trust-calibrated graph fusion, and temporal fault prediction and ending with the explainable maintenance-oriented outputs. Unlike a standard sequential pipeline in which data-quality screening is detached from downstream modeling, TEF-Net propagates each trust score into graph-edge calibration and attention, so reliability estimation directly changes the representation received by the temporal predictor. Recent studies have emphasized trustworthy predictive maintenance [17] and lightweight Transformer-based fault diagnosis [18], while TEF-Net specifically couples trust estimation, graph fusion, temporal prediction, and explanation in one reliability-aware path.

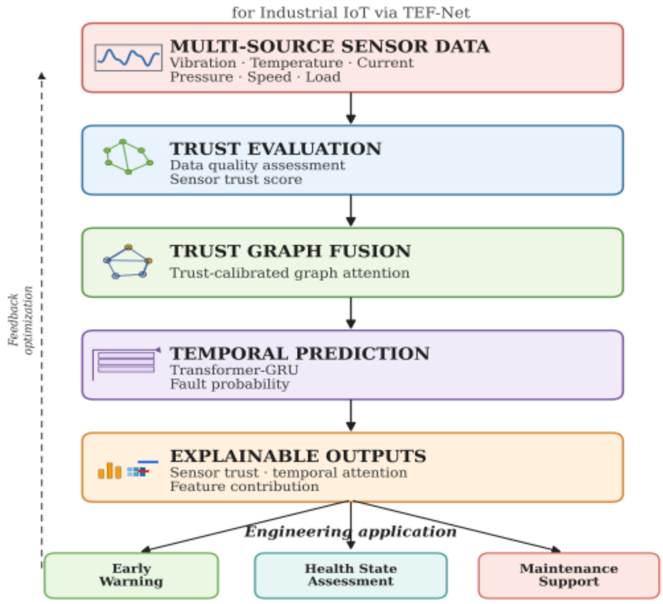


Figure 1. General structure of TEF-Net as a trustworthy multi-source sensor fusion with explainable fault prediction in IIoT-based predictive maintenance

2. TEF-Net Model Design

The TEF-Net structure has been built with manufacturing resiliency as its top priority. Instead of viewing fault prediction as a standalone classification problem, the framework takes into consideration the entire resilience lifecycle, including anticipation (anticipating the occurrence of data quality degradation prior to the occurrence of equipment failure), withstand (preserving the accuracy of prediction during sensor perturbation) and recovery (offering explainability to help in determining maintenance behaviors). The model architecture of trust assessment, graph-based fusion and temporal prediction is organized systematically in order to facilitate these resilience functions. TrustNet detects unreliable sensors, which could signal the forthcoming failures or problems with data infrastructure; reliable graph attention makes sure that poor information provided by individual sensors does not spoil the fused multi-source representation; and the Transformer-GRU predictor offers both precise predictions and comprehensible explanations to expedite post-failure recovery. All the sub-modules listed below are directly part of the predictive maintenance framework of resilience.

2.1 Problem Definition

The monitoring of the IloT-based equipment consists of the collection of various types of sensor data, including vibration, temperature, current, pressure, rotational speed, and load. Assume that there are N sensors in the system. The measurement of the i-th sensor at time t is represented by:

Accordingly, the data collected by all sensors at time t can be expressed as:

$$X_t=[x_t^1,x_t^2,\cdots,x_t^N] \tag{1}$$

The model input is a sliding time window of length L that is used to predict whether a fault will occur in the future:

$$X_{t-L+1:t}=[X_{t-L+1},X_{t-L+2},\cdots,X_t] \tag{2}$$

The model outputs the fault probability at the future time step:

$$\hat{y}_{t+\Delta}=f(X_{t-L+1:t}) \tag{3}$$

where Δ represents the prediction horizon and the output variable lies within the range [0, 1]. In case the predicted value is higher than a predetermined threshold, the equipment is considered to be at risk of a possible failure. In the framework of manufacturing resilience, the prediction horizon Δ becomes vital, because shorter horizons allow giving an early warning and preventive intervention before the disruption increases, whereas the base of the decision rule, based on thresholds, gives a precise signal of maintenance activities which minimize down-time of production and speed up restoration of the system. The resulting problem formulation directly contributes to the goal of resilience by converting raw sensor information into time-sensitive, actionable fault alerts. The resulting overall architecture of the proposed TEF-Net model is illustrated in Figure 2, as can be seen based on the aforementioned problem formulation. The model combines data trustworthiness evaluation, trust-based graph attention combination, and Transformer-GRU-based prediction of temporal faults to provide reliable and explicable predictive maintenance.

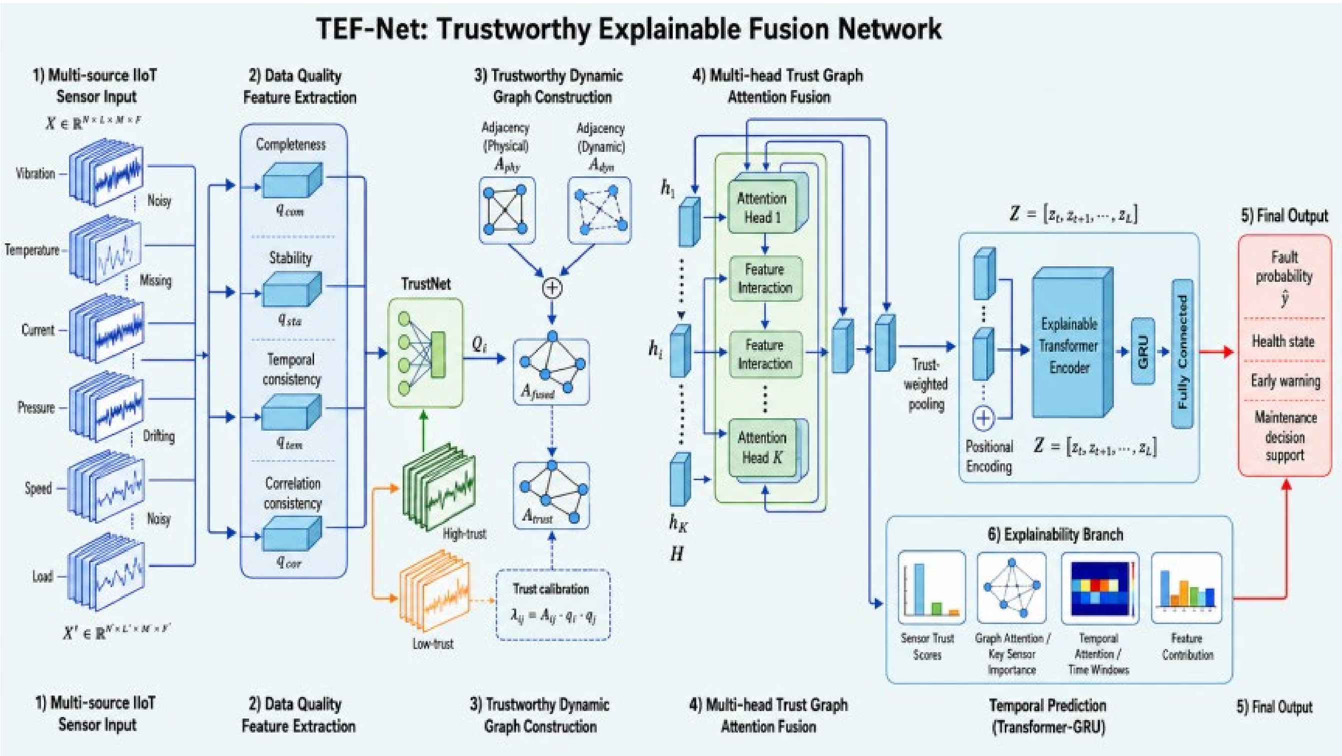


Figure 2. TEF-Net Model

2.2 Data Trustworthiness Assessment

The industrial field data are not necessarily accurate. The sensor data can be corrupted by missing data, drift, sudden change, noise pollution, or communication failures [19,20]. False alarm or failure to detect will happen when these data are directly used to feed a prediction model. This study hence initially finds a moving trust level of all sensors. The i-th sensor has its data quality measured in four dimensions;

$$Q_i=\left[q_i^{com},q_i^{sta},q_i^{tem},q_i^{cor}\right] \quad (4)$$

and the four components are the indicators of data completeness, stability, consistency in time and correlation with the existing sensor and the related sensors respectively.

Completeness is used to measure the missing ratio of sensor data:

$$q_i^{com}=1\text{-}m_i \quad (5)$$

where m_i represents the absence ratio of sensor i in the given time interval. Stability is a way of measuring if the present data variance is a deviation of the normal operating conditions:

$$q_i^{sta}=exp\left(-\frac{|\sigma_i^{cur}\text{-}\bar{\sigma}_i^{ref}|}{\bar{\sigma}_i^{ref}+\epsilon}\right) \quad (6)$$

where σ_i^{cur} is the standard deviation of the current window and σ_i^{ref} is the reference standard deviation representing the historical normal fluctuation level.

Temporal consistency is applied to check whether the present value is in line with the past temporal trend of the same sensor:

$$q_i^{tem}=exp(-|x_t^i\text{-}\hat{x}_t^i|) \quad (7)$$

where the estimated value has been predicted based on historical observations. If there is a big difference between the actual and estimated values then it means that the available data can be abnormal.

Correlation consistency is used to evaluate whether the coupling relationship between sensors changes abnormally:

$$q_i^{cor}=exp\left(-\frac{1}{|N(i)|}\sum_{j\in N(i)}\left|\rho_{ij}^{cur}\text{-}\rho_{ij}^{his}\right|\right) \quad (8)$$

where the current correlation and historical correlation are the relationship of sensors i and j in the current window and under normal historical conditions respectively.

The quality features mentioned above are then submitted to a trustworthiness learning network to give the sensor trust score:

$$c_i=\sigma\left(W_2\delta\left(W_1Q_i+b_1\right)+b_2\right) \quad (9)$$

in which the trust score is limited to the range [0,1]. Higher values mean that the existing measurements provided by sensor i are more trustworthy, and lower values imply that sensor i can be influenced by noise, drift or other irregularities. The trust score is a key resilience facilitator because it enables the system to identify data quality degradation at early stages, which can often be before actual equipment failures, allowing the system to give a predictive signal of possible disruptions. Additionally, sensors with low-trust are automatically down-weighted in later fusion processes to make sure any anomalous readings of individual channels do not ruin the general state estimation of the system. The mechanism is a representation of the resilience of manufacturing systems, since it preserves situational awareness in case subsets of sensors are damaged due to noise, drift, or communication failures.

In Eq. (9), TrustNet is a two-layer multilayer perceptron that maps the four-dimensional quality vector Q_i through a hidden nonlinear layer to one scalar output; the final sigmoid explicitly normalizes the trust score c_i to [0,1]. The same TrustNet mapping is shared across sensors, and its parameters are jointly optimized with the predictor using the losses defined at the end of Section 2.4.

2.3 Trustworthy Graph Attention-Based Data Fusion

The different sensors in the industrial system cannot be independent [21]. As an illustration, if the motor current rises, it can cause the temperature to increase or when a bearing vibration is abnormal there might be a change in the local temperature. Thus, the multi-source sensor system is defined as a graph structure [22]:

$$G=(V,E,A) \quad (10)$$

where V is the set of sensors, E is the set of connections between sensors, and A is the adjacency matrix.

The adjacency matrix consists of two parts:

$$A=\lambda A^{phy}+(1\text{-}\lambda)A^{dyn} \quad (11)$$

where the physical adjacency is a physical relationship defined by equipment structure or process knowledge, and the dynamic adjacency is a relationship that is learned based on data correlations. Specifically, the dynamic matrix is recomputed for each window from normalized node-feature affinities:

$$A_{ij}^{\text{dyn}}=\text{softmax}_j\left(\frac{h_i^T h_j}{\sqrt{d}}\right) \quad (11a)$$

where d is the node-feature dimension.

In order to enhance data trustworthiness, the graph structure is adjusted based on the acquired trust scores:

$$\tilde{A}_{ij}=A_{ij}c_ic_j \quad (12)$$

when the calibrated edge weight is decreased in case of low-trust scores of either sensor. The mechanism avoids abnormal sensors to have an undue effect on other nodes in graph-based fusion [23].

The design of a trustworthy graph attention mechanism is based on the trust-calibrated graph. Attention score between node i and node j is defined as follows:

$$e_{ij}=LeakyReLU\left(a^T\left[W\text{h}_i W\text{h}_j c_i c_j \tilde{A}_{ij}\right]\right) \quad (13)$$

where and are the node features, and are the node trust scores and the calibrated edge weight is the trust-aware relationship between nodes.

After normalization, the attention weight is obtained as:

$$\alpha_{ij}=\frac{exp(e_{ij})}{\sum_{k\in N(i)}exp(e_{ik})} \tag{14}$$

The node feature is then updated by aggregating information from neighboring nodes:

$$\mathrm{h}_i^{'}=\sigma\left(\sum_{j\in N(i)}\alpha_{ij}W\mathrm{h}_j\right) \tag{15}$$

By using this method, the model adaptively integrates multi-source information based on sensor associations and data reliability. The graph fusion mechanism of trust calibration is especially useful when disruptions happen. In case a sensor has sudden drift or communication failure, the low trust score will spread across the calibrated edge weights (Equation 12) and attention scores (Equation 13) and therefore, the bad information cannot affect other nearby nodes. Such an ability to isolate faults stops failures in individual sensors from spreading to misdiagnosis of entire systems, which is an essential feature of robust manufacturing systems that need to remain aware of their operations even with partial sensor malfunctioning. Graph structure also allows quick adjustment: sensor trust scores are updated dynamically with new incoming data and the fusion process constantly recalibrates to represent the existing reliability environment, which facilitates quick recovery by making sure that recovered sensors are quickly returned into the monitoring network.

2.4 Temporal Fault Prediction

Following reliable graph attention fusion, the model gets the equipment state representation in every time window:

$$Z=[z_{t-L+1},z_{t-L+2},\cdots,z_t] \tag{16}$$

The faults of industrial machines usually come with two types of temporal properties. One of them is long term degradation, i.e., wear, aging and fatigue. Another property is short term abnormality, i.e., sudden increase in vibration, fluctuations in current, and a sharp increase in temperature. Hence, to predict faults, a Transformer-GRU architecture is applied [24]. The Transformer is employed to find long-range temporal relationships:

$$H=Transformer(Z+PE) \tag{17}$$

and PE is a positional encoding. Transformer has the ability to learn the relationships between various historical time steps and is appropriate to use when the trend of the gradual degradation is to be captured.

Subsequently, GRU is used to further model continuous state transitions:

$$s_t=GRU(H_t,s_{t-1}) \tag{18}$$

GRU enhances the model's ability to describe short-term dynamic variations and state transitions.

Finally, the fault probability is output through a fully connected layer:

$$\hat{y}=sigmoid(W_o s_t+b_o) \tag{19}$$

Here the output is the predicted probability of the occurrence of a fault in the equipment in the future. The module can make use of the Transformer to encode long-term degradation trends and the GRU to represent ongoing dynamism and hence boost the results of fault prediction. The Transformer-GRU architecture also provides additional manufacturing resilience as it has the ability to record both gradual degradation and sudden anomaly. Long-term degradation patterns (e.g., wear, fatigue) are registered by the Transformer attention on long historical periods, allowing predicting failures before they disrupt operations. At the same time, the recurrent structure of the GRU records the dynamics of the short-term variations (e.g., sharp increase in vibrations, fast increase in temperatures), which enables the system to quickly find out and react to the sudden disruptive events. This combination guarantees that the predictive maintenance system is effective across the whole range of disruption timescales - slow degrading to sudden failure - thus ensuring the maximum possible window of opportunity to act preventively and the minimum possible recovery time when failures do happen.

During joint training, the fault-prediction branch is optimized by binary cross-entropy:

$$L_{\text{cls}}=-\frac{1}{S}\sum_{n=1}^{S}\left[y_n\ln\hat{y}_n+(1-y_n)\ln(1-\hat{y}_n)\right] \tag{20}$$

To make the learned trust score consistent with the four quality indicators, the quality target is defined as:

$$\bar{q}_{n,i}=\frac{q_{n,i}^{\text{com}}+q_{n,i}^{\text{sta}}+q_{n,i}^{\text{tem}}+q_{n,i}^{\text{cor}}}{4} \tag{21}$$

The corresponding Trust Loss is:

$$L_{\text{trust}}=\frac{1}{SN}\sum_{n=1}^{S}\sum_{i=1}^{N}\left(c_{n,i}-\bar{q}_{n,i}\right)^2 \tag{22}$$

The overall objective is:

$$L=L_{\text{cls}}+\beta L_{\text{trust}} \tag{23}$$

where β controls the contribution of the Trust Loss.

For post-hoc feature attribution, SHAP (SHapley Additive exPlanations) is applied to the trained TEF-Net [25]. For a prediction $f(x)$, the explanation satisfies:

$$f(x)-\mathbb{E}[f(X)]=\sum_{j=1}^{M}\phi_j \tag{24}$$

where ϕ_j is the SHAP contribution of feature j. A positive ϕ_j shifts the output toward the fault class, whereas a negative value shifts it toward the normal class; these signed sensor-level contributions are subsequently evaluated in the interpretability analysis.

With the trust-calibrated fusion mechanism, temporal predictor, training objectives, and explanation procedure defined, the following section evaluates predictive accuracy, calibration, robustness, and interpretability under controlled data-quality disturbances.

3. Experimental Section

3.1 Dataset Introduction

The experiments were conducted using the MetroPT-3 dataset, a publicly available real-world predictive maintenance dataset provided by the UCI Machine Learning Repository. The dataset was collected from the Air Production Unit (APU) compressor of a metro train under actual operational conditions and contains 1,516,948 multivariate time-series observations with 15 sensor signals. These signals consist of seven analog measurements, including compressor and reservoir pressures, motor current, and oil temperature, together with eight digital signals describing the operating states of valves and other APU components. The dataset covers operational data collected between February and August 2020 and provides maintenance reports specifying the occurrence intervals of equipment failures. These characteristics make MetroPT-3 suitable for evaluating multi-source sensor fusion and predictive maintenance models under realistic industrial operating conditions.

The raw sensor variables were first checked for abnormal values and then normalized to reduce the influence of different measurement scales on model training. Since MetroPT-3 does not directly provide sample-wise fault labels, the failure intervals documented in the official maintenance reports were used to construct the prediction targets. Specifically, the continuous sensor sequences were divided into sliding windows of length L, and a sample was assigned a fault label of 1 when an officially reported failure occurred within the subsequent prediction horizon Δ; otherwise, it was assigned a normal-operation label of 0. This labeling strategy is consistent with the predictive-maintenance objective of identifying potential equipment failures before their occurrence. The official dataset provides the corresponding failure records, allowing the original unlabeled time series to be used for failure prediction and anomaly-detection evaluation.

To ensure reliable evaluation and avoid temporal information leakage, the generated time-series samples were chronologically divided into training, validation, and test sets, with samples from the same continuous operating segment not shared between the training and test sets. The original MetroPT-3 data contain no missing values; therefore, noise, missing-data, drift, pulse anomaly, and combined disturbance scenarios were additionally introduced only during the robustness evaluation stage to reproduce common data-quality degradation conditions in practical IIoT environments. These controlled disturbances were used to evaluate whether TEF-Net can maintain reliable sensor fusion and fault prediction when part of the sensing infrastructure becomes degraded or unreliable.

3.2 Comparative Experiments and Performance Analysis

Figure 3 shows the comparison between the results of GAT-GRU, Transformer-GRU, and the proposed TEF-Net in various evaluation measures. Figure 3(a) is a radar chart showing the overall performance of six metrics that are Accuracy, Precision, Recall, F1-score, AUC, and MCC; Figure 3(b) also indicates the particular changes of trends in every single metric.

Calibration is additionally evaluated using Expected Calibration Error (ECE):

$$\mathrm{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{n} |\mathrm{acc}(B_m) - \mathrm{conf}(B_m)| \qquad (25)$$

where B_m denotes the m-th confidence bin, $\mathrm{acc}(B_m)$ is its empirical accuracy, and $\mathrm{conf}(B_m)$ is its mean predicted confidence. Lower ECE indicates better agreement between prediction confidence and observed correctness.

The radar plot profile of TEF-Net can be seen in Figure 3(a) which is located on the outermost edge and indicates that the performance of TEF-Net is better than that of GAT-GRU and Transformer-GRU. In particular, TEF-Net has the maximum values across all three models: Accuracy of 0.947, Precision of 0.940, Recall of 0.932, F1-score of 0.936, AUC of 0.982, and MCC of 0.895. TEF-Net also has a unique benefit in terms of F1-score and AUC as shown in Figure 3(b); its F1-score is higher by 0.916 to 0.936 and its AUC by 0.970 to 0.982 compared to Transformer-GRU; likewise, it is higher by 0.910 to 0.936 and its MCC by 0.849 to 0.895 compared to GAT-GRU. These findings indicate that TEF-Net does not only promote better overall performance in classification, but also contributes to the model stability in categorization and defect detection.

The main cause of the performance enhancement of TEF-Net is its integration of sensor credibility estimation before the fault prediction, which was also used to perform dynamic graph calibration, graph attention fusion, and weighted pooling. Accordingly, TEF-Net is more resistant to the impact of low-quality sensor data on prediction results than traditional graph temporal models. TEF-Net performance means better manufacturing resiliency. In the same conditions of the data, increased accuracy and F1-score implies less missing faults (not to experience unexpected disruptions) and less false alarms (not to have unnecessary production stoppages). An AUC of 0.982 is a clear indication that TEF-Net has superior discriminative power at every level of decision threshold, hence offering maintenance planners more flexibility between the cost of preventive action and the possibility of disruption. Importantly, as demonstrated in later robustness experiments, TEF-Net has much lower performance loss in the presence of disturbances as compared to baseline models- this performance resilience is the quantified representation of the method to preserve the situational awareness and fast recovery in case of stresses on manufacturing systems. In order to add a more thorough quantitative analysis, Table 1 also presents the fault prediction performance of TEF-Net and representative baseline models in terms of Accuracy, Precision, Recall, F1-score, AUC, MCC, and FPR.

Table 1. Comparison of fault prediction performance across different models

model	Accuracy	Precision	Recall	F1-score	AUC	MCC	FPR
SVM	0.841	0.826	0.802	0.814	0.884	0.671	0.122
RF	0.866	0.852	0.831	0.841	0.910	0.721	0.105
XGBoost	0.887	0.874	0.858	0.866	0.932	0.766	0.089
LightGBM	0.892	0.881	0.862	0.871	0.938	0.776	0.085
1D-CNN	0.875	0.861	0.846	0.853	0.921	0.746	0.097
TCN	0.901	0.889	0.876	0.882	0.945	0.795	0.078
LSTM	0.884	0.869	0.852	0.860	0.930	0.762	0.091
GRU	0.890	0.876	0.861	0.868	0.936	0.774	0.083
BiLSTM	0.898	0.885	0.871	0.878	0.944	0.791	0.080
CNN-LSTM	0.907	0.895	0.881	0.888	0.951	0.808	0.072
Transformer	0.914	0.902	0.891	0.896	0.958	0.823	0.067
GCN	0.895	0.883	0.865	0.874	0.941	0.784	0.082
GAT	0.909	0.897	0.884	0.890	0.954	0.812	0.071
ST-GCN	0.918	0.908	0.897	0.902	0.961	0.832	0.064
GAT-GRU	0.925	0.916	0.904	0.910	0.966	0.849	0.059
Transformer-GRU	0.930	0.921	0.912	0.916	0.970	0.860	0.054
TEF-Net	**0.947**	**0.940**	**0.932**	**0.936**	**0.982**	**0.895**	**0.039**

Table 1 indicates that the TEF-Net performs better than the rest of the models in terms of overall performance, highest Accuracy, Precision, Recall, F1-score, AUC, and MCC and lowest FPR which shows that it is more accurate in its predictions and controls false-alarm levels. As standard non-graph fusion references without the TrustNet/GAT mechanism, the LSTM and CNN-LSTM baselines obtain F1-scores of 0.860 and 0.888, respectively, compared with 0.936 for TEF-Net; the corresponding gains are 0.076 and 0.048. Consistently, removing GAT in the ablation study reduces the F1-score to 0.894, a decrease of 0.042 from the complete model, directly quantifying the contribution of graph-based sensor interaction modeling.

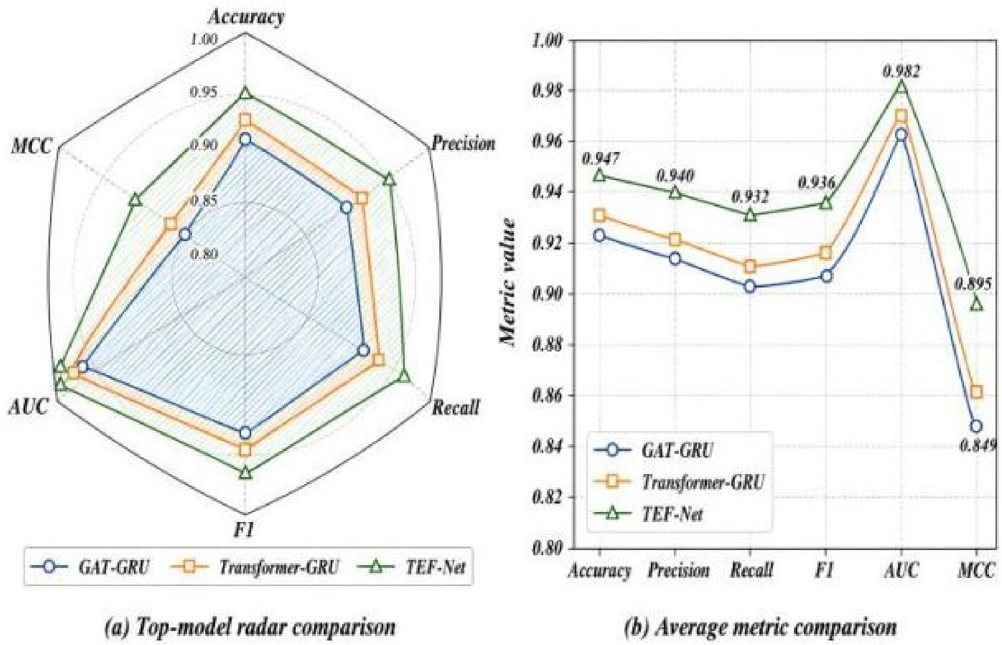
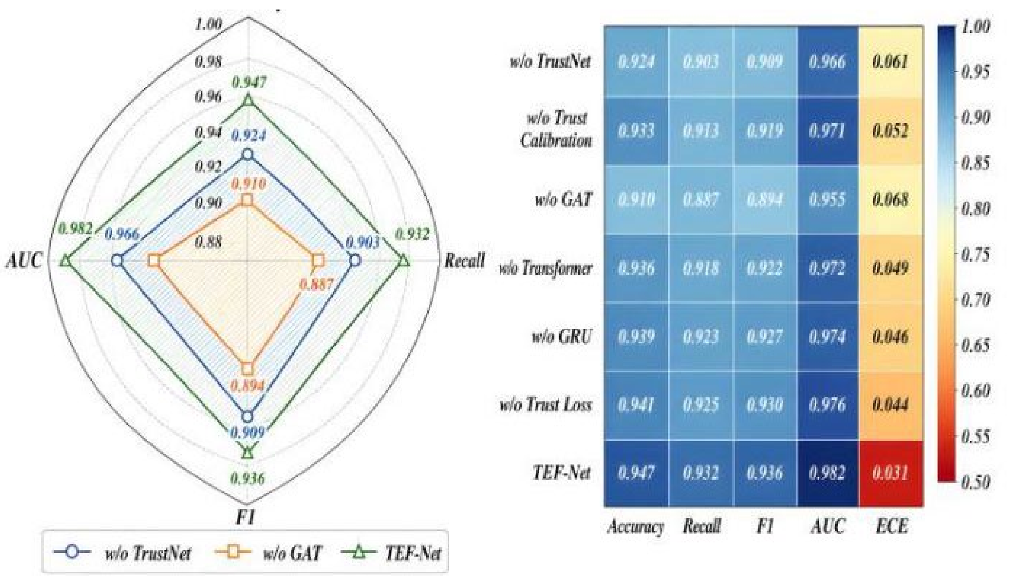


Figure 3. Comparison of main performance metrics across different models

3.3 Ablation Experiment and Module Validity Analysis

The results of TEF-Net ablation experiment are provided as Figure 4. As can be seen in Figure 4(a), the biggest radar chart area is observed with the complete TEF-Net, which means that this model has better overall results than models without important modules. Deletion of TrustNet led to a decrease of the F1-score by 0.936 to 0.909 and deletion of GAT caused its further decrease to 0.894, which proves that trust modeling and graph attention fusion are crucial to the performance of the model with GAT having the strongest effect on multi-source sensor correlation modeling.

The heatmap in Figure 4(b) confirms the effectiveness of every module. The full TEF-Net attains the highest accuracies of 0.947, recall of 0.932, F1-score of 0.936, and an AUC of 0.982, with the lowest ECE of 0.031. Conversely, deletion of TrustNet raises ECE to 0.061 and removal of trust graph calibration decreases F1-score to 0.919, which shows that trustworthiness does not only add value to prediction performance but also to the reliability of model outcomes. The ablation experiments also show the resiliency role of each module. Large performance degradation on removing TrustNet (F1-score drops to 0.909) means that trust evaluation is important not only in order to be accurate but also to provide predictability stability in case of data quality changes - one of the most important enablers of manufacturing resilience. Likewise, elimination of GAT (F1-score drops to 0.894) indicates that it is important to model inter-dependencies between sensors in order to preserve coherent system level awareness in case of abnormal behavior of individual sensors. These results establish the fact that the two modules help the system to resist the data-quality fluctuations and predict in a reliable manner.



(a) Ablation radar comparison *(b) Ablation study heatmap*

Figure 4. Results of TEF-Net ablation experiments

3.4 Robustness and Interpretability Analysis

Figure 5 shows the strength of the model in different data disturbance situations. According to Figure 5(a) in clean, Noise, Missing, Drift, Pulse and Multi-error conditions TEF-Net scored F1 scores of 0.936, 0.918, 0.907, 0.895, 0.912 and 0.884 respectively all higher than GAT-GRU and Transformer-GRU. Remarkably, during the Multi-error condition, TEF-Net had an F1-score of 0.884 which is indicative of the fact that it performs well even in the case of multiple anomalies. Figure 5(b) also shows that as the rate of missing and the level of noise increases the performance of TEF-Net decreases but the overall trend is declining gradually which means that the model is highly adaptable to low-quality data.

The strong outcomes have immediate implications on manufacturing resiliency and value of deployment. The slight performance loss with disturbances (only a difference of 0.052 between Clean and Multi-error in terms of the F1-score, as opposed to 0.091 in GAT-GRU and 0.087 in Transformer-GRU) quantitatively indicates that TEF-Net has a certain level of resilience, i.e., the capability to provide a useful predictive maintenance despite the sensor infrastructure being disrupted. When applied to actual manufacturing settings in which sensor failure, signal breakdowns, and data corruption are unavoidable, this performance resilience means less unplanned production halts, lower emergency repair expenditures, and decreased time to recover systems. The steady F1-score of 0.884 in case of multi-error implies that, even in the most severe situation of disruptions, the system would be able to properly detect about 88 percent of faults, which could give the plant operators enough confidence in making decisions at the moments most critical.

From a computational perspective, for each window the graph-attention stage scales approximately with $O(|E|d)$, Transformer self-attention with $O(L^2d)$, and the GRU with $O(Ld^2)$, where $|E|$ is the number of retained sensor edges and d is the latent feature width. Consequently, self-attention becomes the main cost for long windows, whereas sparse graph construction limits the GAT overhead. For resource-constrained deployment, reduced attention width, sparse edges, pruning, or quantization can be used; highly constrained endpoints may execute TEF-Net at an edge gateway rather than assuming unrestricted on-device inference.

The interpretability results of TEF-Net are depicted in Figure 6. As it can be seen in Figure 6 (a), Vibration, Temperature, and Speed have the highest credibility scores and Current has the lowest, which indicates that the model is capable of distinguishing between the reliability of sensor data. The figure 6 (b) shows an increased attention of the model to later time intervals especially the 8th -10th intervals, which suggests that the model focuses on critical degradation intervals before failure events. Figure 6(c) indicates that the Vibration and Temperature add positively to the prediction of failures, and the Speed and Pressure negatively do so. On the whole, TEF-Net interprets its predictions in three dimensions such as sensor credibility, important time windows, and contributions of features.

The three explanation channels have distinct physical meanings. A low trust score indicates that a sensor channel itself is currently unreliable; temporal attention identifies the time interval carrying the strongest degradation evidence; and the sign and magnitude of a SHAP value quantify the direction and strength of a physical variable's contribution to the fault probability. Thus, the positive vibration and temperature contributions in Figure 6(c) support mechanical/thermal degradation evidence, whereas negative speed and pressure contributions oppose the fault decision. Attribution strength can be summarized by $|\varphi_j|$, while the signed value retains the direction of influence.

The explainability outputs are no longer just academic findings but also have real world applications in making recovery decisions about manufacturing operations. Upon a fault warning, maintenance staff may quickly check on the credibility results to determine which sensors reported trustworthy information and what sensors gave doubtful information, compare the temporal attentional trends to find the most important period of time at which degradation occurred, and examine the feature contribution scores to see which physical parameters were responsible to make the prediction. The multi-dimensional explanation allows diagnosing the root cause of the problem in minutes instead of hours, which dramatically reduces the detect-analyze-respond-recover cycle which characterizes manufacturing resilience. These interpretable outputs could be used in deployment situations by incorporating them into maintenance dashboards and decision support systems so that operators can get actionable intelligence to fill the gap between predicted AI predictions and actual maintenance choices made by humans. Also, the trust scores act as an ongoing health indicator of the sensor infrastructure itself and can be used to proactively calibrate or replace sensors prior to quality degradation of data affecting overall system reliability, a closed loop resilience mechanism that goes beyond predicting faults to cover the whole data acquisition pipeline.

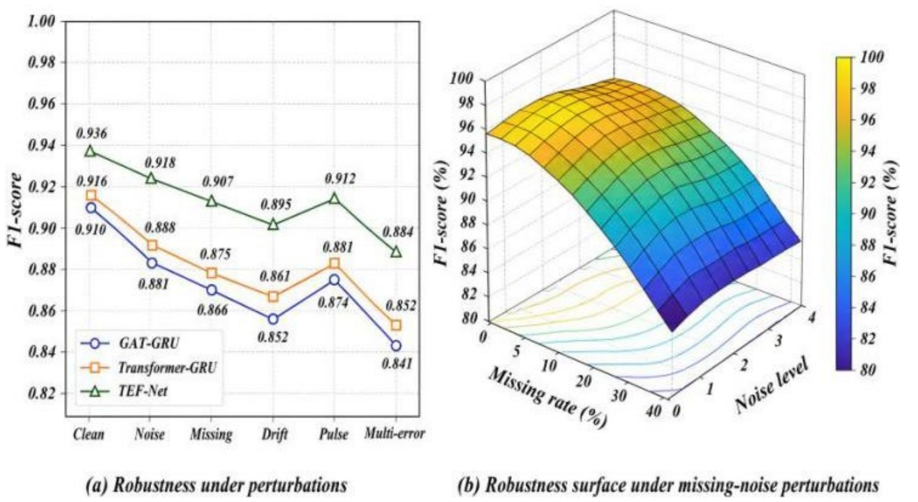


Figure 5. Robustness analysis under data disturbance conditions

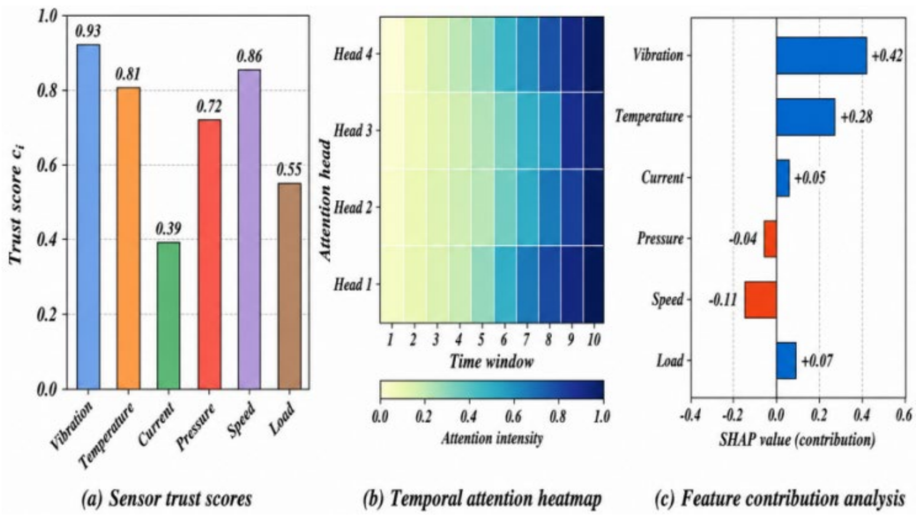


Figure 6. TEF-Net Interpretability Analysis Chart

4. Conclusion

With the development of manufacturing systems towards Industry 4.0 and Industry 5.0, it has become a strategic necessity to be able to foresee, endure, and regain quickly after interruptions. In this paper, the author tackles some of the most serious issues in predictive maintenance that involve the use of the IIoT, such as inconsistent quality of multi-source sensor data, inability to resist disturbances, and the absence of interpretability in prediction outcomes, with the presentation of the TEF-Net, a reliable data fusion and predictive maintenance algorithm using explainable artificial intelligence.

There are three interconnected contributions that form the core innovations of TEF-Net. To begin with, sensor credibility evaluation based on TrustNet explicitly measures the quality of data which allows detecting the early signs of data deterioration and eliminating low-quality data prior to its corruption of the prediction. Secondly, the trust-calibrated graph attention fusion mechanism represents inter-dependencies between sensors but removes the influence of anomalous sensors so that system-level awareness is sustained in case of failure of individual sensors. Thirdly, the Transformer-GRU architecture can be used to represent both long-term degradation and short-term anomalies whereas the explainability outputs, i.e., credibility scores, temporal attention, and feature contributions, are multi-dimensional and offer practical implications to maintenance decision-making.

The experimental findings show that TEF-Net can reach accuracy of 0.947, recall of 0.932, F1-score of 0.936, and AUC of 0.982, which is higher than comparative models such as GAT-GRU and Transformer-GRU. Ablation experiments prove that TrustNet and GAT are important, and their deletion decreases the F1-score by 0.909 and 0.894, respectively. TEF-Net has a quantitative validation in its ability to enable resilience in manufacturing systems as it retains an F1-score of 0.884 under multi-error disturbance conditions with much less loss in performance compared to baselines.

TEF-Net offers concrete benefit to the entire resilience lifecycle of industrial manufacturing, including early fault warning by way of trust-based anomaly detection, strong prediction that is not affected by sensor perturbations, and explainable results which speed up the process of identifying the root cause and recovery. The approach offers a reliable AI framework that can be used in predictive maintenance and this will lead to a reduction in unplanned downtime, increase in the life of the equipment, and improve overall manufacturing resilience.

The integration of TEF-Net with digital twin platforms will be investigated in future work to facilitate real-time simulation-based what-if analysis of maintenance strategies and in conjunction with edge computing architectures to minimize inference latency and offer closed-loop adaptive recovery on the factory floor. Such extensions will enhance the resilience properties of manufacturing systems even more by shifting them out of the passive response to fault toward self-healing and autonomous production environments.

References

[1] Alexopoulos K, Anagiannis I, Nikolakis N, Chryssolouris G. A quantitative approach to resilience in manufacturing systems[J]. International Journal of Production Research, 2022, 60(24): 7178-7193.

[2] Zonta T, da Costa C A, da Rosa Righi R, de Lima M J, da Trindade E S, Li G P. Predictive maintenance in the Industry 4.0: A systematic literature review[J]. Computers & Industrial Engineering, 2020, 150: 106889.

[3] Carvalho T P, Soares F A A M N, Vita R, Francisco R P, Basto J P, Alcalá S G S. A systematic literature review of machine learning methods applied to predictive maintenance[J]. Computers & Industrial Engineering, 2019, 137: 106024.

[4] Dalzochio J, Kunst R, Pignaton E, Binotto A, Sanyal S, Favilla J, Barbosa J. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges[J]. Computers in Industry, 2020, 123: 103298.

[5] Aivaliotis P, Georgoulias K, Chryssolouris G. The use of Digital Twin for predictive maintenance in manufacturing[J]. International Journal of Computer Integrated Manufacturing, 2019, 32(11): 1067-1080.

[6] Çınar Z M, Nuhu A A, Zeeshan Q, Korhan O, Asmael M, Safaei B. Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0[J]. Sustainability, 2020, 12(19): 8211.

[7] Achouch M, Dimitrova M, Ziane K, Sattarpanah Karganroudi S, Dhouib R, Ibrahim H, Adda M. On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges[J]. Applied Sciences, 2022, 12(16): 8081.

[8] Pech M, Vrchota J, Bednář J. Predictive Maintenance and Intelligent Sensors in Smart Factory: Review[J]. Sensors, 2021, 21(4): 1470.

[9] Ayvaz S, Alpay K. Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time[J]. Expert Systems with Applications, 2021, 173: 114598.

[10] Çakır M, Güvenç M A, Mistikoğlu S. The experimental application of popular machine learning algorithms on predictive maintenance and the design of IIoT based condition monitoring system[J]. Computers & Industrial Engineering, 2021, 151: 106948.

[11] Nordal H, El-Thalji I. Modeling a predictive maintenance management architecture to meet industry 4.0 requirements: A case study[J]. Systems Engineering, 2021, 24(1): 34-50.

[12] Cummins L, Sommers A, Ramezani S B, Mittal S, Jabour J, Seale M, Rahimi S. Explainable Predictive Maintenance: A Survey of Current Methods, Challenges and Opportunities[J]. IEEE Access, 2024, 12: 57574-57602.

[13] Gawde S, Patil S, Kumar S, Kamat P, Kotecha K, Alfarhood S. Explainable Predictive Maintenance of Rotating Machines Using LIME, SHAP, PDP, ICE[J]. IEEE Access, 2024, 12: 29345-29361.

[14] Saeed W, Omlin C. Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities[J]. Knowledge-Based Systems, 2023, 263: 110273.

[15] Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P, Qian B, Wen Z, Shah T, Morgan G, Ranjan R. Explainable AI (XAI): Core Ideas, Techniques, and Solutions[J]. ACM Computing Surveys, 2023, 55(9): 194.

[16] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A Survey of Methods for Explaining Black Box Models[J]. ACM Computing Surveys, 2018, 51(5): 93.

[17] Ucar A, Karakose M, Kırımça N. Artificial Intelligence for Predictive Maintenance Applications: Key Components,

Trustworthiness, and Future Trends[J]. Applied Sciences, 2024, 14(2): 898.

[18] Dong Y, Jiang H, Mu M, Wang X. Multi-sensor data fusion-enabled lightweight convolutional double regularization contrast transformer for aerospace bearing small samples fault diagnosis[J]. Advanced Engineering Informatics, 2024, 62: 102573.

[19] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead[J]. Nature Machine Intelligence, 2019, 1(5): 206-215.

[20] Azamfar M, Singh J, Bravo-Imaz I, Lee J. Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis[J]. Mechanical Systems and Signal Processing, 2020, 144: 106861.

[21] Cui J, Xie P, Wang X, Wang J, He Q, Jiang G. M2FN: An end-to-end multi-task and multi-sensor fusion network for intelligent fault diagnosis[J]. Measurement, 2022, 204: 112085.

[22] Lin T, Zhu Y, Ren Z, Huang K, Gao D. CCFT: The Convolution and Cross-Fusion Transformer for Fault Diagnosis of Bearings[J]. IEEE/ASME Transactions on Mechatronics, 2024, 29(3): 2161-2172.

[23] Duan H, Chen G, Yu Y, Du C, Bao Z, Ma D. DyGAT-FTNet: A Dynamic Graph Attention Network for Multi-Sensor Fault Diagnosis and Time–Frequency Data Fusion[J]. Sensors, 2025, 25(3): 810.

[24] Lee S, Kim Y, Choi H J, Ji B. Uncertainty-Aware Fault Diagnosis of Rotating Compressors Using Dual-Graph Attention Networks[J]. Machines, 2025, 13(8): 673.

[25] Lundberg S M, Lee S I. A Unified Approach to Interpreting Model Predictions[C]. Advances in Neural Information Processing Systems, 2017, 30: 4765-4774.