
Anomaly Pattern Mining for Predictive Maintenance of Industrial Machinery Using Sensor Time-Series Data

Dr.S.Sivakumar, Mrs. T. Gayathri and Mrs. S. Praveena

Department of Computer Science (AI&DS)

Sengunthar Arts and Science College

Tiruchengode

Namakkal -637205

ssivakkumarr@yahoo.com

ABSTRACT

Industrial machinery continuously generates multivariate sensor signals that contain important information about operational states, degradation trends, and impending failures. However, conventional predictive maintenance techniques based on fixed thresholds or isolated anomaly detection often struggle to capture nonlinear sensor interactions, long-range temporal dependencies, and evolving anomaly patterns. To address these limitations, this study proposes APM-TCN (Adaptive Anomaly Pattern Mining with Temporal Convolutional Networks), a predictive maintenance framework for discovering abnormal operating patterns from sensor time-series data. The framework integrates Variational Autoencoder (VAE)-based latent representation learning, Temporal Convolutional Networks (TCN), self-attention, adaptive anomaly scoring, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise)-based temporal pattern clustering. Initially, the VAE normalizes noisy sensor observations and transforms them into compact latent representations. The TCN subsequently learns short- and long-range temporal dependencies, while self-attention emphasizes sensor characteristics associated with progressive equipment degradation. An adaptive anomaly score combines reconstruction error, temporal prediction error, and attention-derived information to identify deviations from normal operating behavior. DBSCAN then groups related anomaly observations to reveal recurring degradation signatures, which are subsequently utilized for failure-risk prediction. Experimental evaluation demonstrates 97.2% anomaly detection accuracy, 95.8% precision, 96.4% recall, and 96.1% F1-score, together with 92.6% early-failure prediction accuracy. APM-TCN further achieves a 31.5% reduction in false alarms and a 39.2% improvement in early-warning capability compared with conventional monitoring approaches. The findings demonstrate that adaptive temporal anomaly pattern mining can provide reliable early indications of machinery degradation, supporting proactive maintenance planning, reducing unplanned downtime, and improving industrial equipment reliability.

Keywords: Predictive Maintenance, Anomaly Pattern Mining, Industrial Machinery, Sensor Time-Series Data, Variational Autoencoder, Temporal Convolutional Network, Self-Attention, DBSCAN, Anomaly Detection, Failure Prediction, Condition Monitoring, Deep Learning.

1. Introduction

Industrial equipment continuously generates multivariate sensor inputs for load, temperature, vibration, pressure, and rotating speed. These sensor streams support condition monitoring and predictive maintenance because they may show equipment degradation before

severe failure through progressive changes in temporal behaviour. Because of nonlinearity, noise, high dimensionality, and lack of stationarity, industrial sensor data can be difficult to distinguish from normal operational variations and meaningful degradation trends. Inadequate pattern recognition may cause machine faults, unexpected downtime, greater maintenance costs, and lower production reliability. Thus, modern industrial systems must include an intelligent technique to recognise unusual temporal activity and transform it into early maintenance information.

Recent research has examined machine learning and deep learning for RUL estimates and equipment prognostics. Deep-learning networks and relevance vector machines may predict data-driven degradation from rotating-machinery observations to learn failure-related features [1]. To capture degradation-related temporal dependencies, researchers have examined temporal convolutional networks (TCNs) using gated TCN structures for RUL estimation [3] and dual-task TCNs with multi-channel attention [2]. Attention and multi-scale temporal convolution capture degradation characteristics at different temporal resolutions [4]. Transformer-based representations and reweighting techniques have also been used to estimate turbofan RUL [5]. These approaches improve temporal representation and prediction accuracy by estimating RUL instead of discovering recurrent anomaly patterns to support predictive maintenance.

Another potential research topic is multivariate time series anomaly detection. Adversarial learning using variational autoencoders may identify anomalies in complex multivariate sequences [6]. Researchers have used missing-data management and collaborative feature connections to discover abnormalities in industrial robot data [7]. Time-frequency analysis and TCN-attention processes have been used to identify cyber-industrial anomalies [8]. Researchers have also studied contrastive learning using time-frequency representations for multivariate time-series anomaly detection [9]. Many studies have highlighted issues with online model-based anomaly detection, including fluctuating temporal behaviour, multivariate dependencies, and the need for trustworthy online monitoring [10]. These tests show that anomaly representation and detection have improved, but most approaches still cannot locate clusters of anomalies at distinct periods.

New processes in predictive-maintenance research increase industrial diagnostics and failure prediction. Research on knowledge-graph-assisted anomaly identification links abnormalities to root-cause investigation [11]. Unsupervised learning can detect faults directly from multivariate industrial pump sensor data [12]. Continuous learning has been used to anticipate failure in non-stationary condition-monitoring scenarios where sensor distributions may change [13]. Researchers have also used uncertainty-aware recurrent learning to anticipate rotating equipment RUL [14]. Time-frequency contrastive learning can train discriminative representations without reconstruction-based anomaly detection [15]. Despite these advances, adaptive anomaly scoring, nonlinear latent representation learning, explicit mining of recurrent anomaly patterns for predictive maintenance, and temporal dependency extraction form a coherent approach that has not been fully examined.

This study proposes APM-TCN, a predictive-maintenance framework that may identify degradation tendencies in industrial sensor time-series data. The proposed architecture uses self-attention, adaptive anomaly scoring, TCN-based temporal modelling, variational representation learning, and density-based anomaly-pattern grouping. This study developed an integrated framework for anomaly detection and predictive maintenance, including compact nonlinear latent representations from multivariate sensor data, an adaptive anomaly score that combines reconstruction error with temporal prediction residuals, density-based clustering to identify recurring degradation patterns, and early failure-risk estimation. Experiments using the

NASA C-MAPSS turbofan engine dataset evaluate the framework's anomaly detection, degradation pattern recognition, and predictive maintenance capabilities.

2. Literature Review

Predictive maintenance research increasingly uses sensor time-series data to learn degradation characteristics and improve failure prediction and RUL evaluation. Existing techniques indicate that deep-learning models may capture complex nonlinear correlations in machinery signals, depending on temporal representation, feature extraction, and operating conditions. A multi-dimensional attention mechanism and feature-sequence dimensional sample convolution and interaction network (MFSSCINet) was trained to grasp sensor-temporal sequence interactions to predict machinery RUL [16]. This approach calculates RUL and doesn't seek repetitive abnormalities, but it helps extract degradation-related information from multivariate measurements. Similarly, Anomaly-informed remaining useful life estimation (AIRULE) uses deep learning to incorporate anomalous information into RUL estimates [17]. Including anomaly data improves deterioration awareness, although the technique focuses on RUL estimation rather than temporal anomaly pattern mining.

Attention-based machinery prognostics approaches have also been widely researched. Power-machinery RUL prediction uses a dual-attention deep-learning framework to prioritise degradation-sensitive sensor data [18]. One-dimensional deep convolutional autoencoder (1D-DCAE) high-quality representation learning in rolling-bearing RUL prediction has also improved degradation feature extraction [19]. Continuous Wavelet Transform (CWT) for feature extraction and a Channel-Temporal Mixed MLP (CT-MLP) layer may help explain bearing degradation [20]. These approaches demonstrate the importance of time and feature selection. Their primary objective is to produce a continuous RUL estimate, not to distinguish between normal operation and emerging abnormal patterns; however they occasionally include guided degradation information.

This might be investigated using feature-engineering and sensor-fusion methodologies. Tool RUL prediction has been studied utilising wavelet-based feature extraction and machine learning-based sensor fusion [21]. These strategies may integrate sensor data to identify deteriorating equipment quality. Signal processing or human representations may become less effective if sensor distributions alter or unanticipated degradation patterns arise. A deep reinforcement-learning system based on ResNet and a soft actor-critic approach (SAC-ResNet framework) has also been used to predict rolling-bearing RUL [22]. Because of its processing complexity and incentive-design requirements, reinforcement learning may not be well suited for continuous anomaly monitoring.

Current research includes representation-learning, physics-informed, and probabilistic methods. An improved degradation-augmented attention framework (DA-RUL) for estimation may boost the influence of deterioration-related features [23]. Deep state-space models (Kalman-DVAE (K-DVAE) use probabilistic temporal dynamics for RUL estimation to portray machinery health states [24]. A novel Physics-Informed Self-Supervised Diagnosis (PI-SSD) framework using latent ordinary differential equations and Transformer encoders has shown that data-driven temporal representations and physical knowledge can diagnose spinning equipment [25]. These approaches manage uncertainty, generalise, and temporal modelling better, although they focus more on diagnosis and RUL estimation than on discovering clusters of recurring abnormalities. Anomaly detection using embeddings is critical for Industry 5.0 applications. Researchers have studied embedding models for multivariate time-series anomaly detection to learn compact representations from industrial sensor streams [26]. Representation-

based anomaly detection in complex industrial environments may reduce the need for manual feature creation.

In contrast, representation learning cannot categorise recurring deterioration signals or demonstrate temporal relationships between individual defects. RUL prediction, attention-based degradation modelling, sensor fusion, RL, probabilistic state estimation, physics-informed diagnostics, and anomaly detection via embeddings have all shown promise. However, predictive maintenance, including latent representation learning, temporal dependency modelling, adaptive anomaly scoring, and explicit grouping of repeated anomaly patterns, is lacking. To address this issue, we provide the APM-TCN framework, which identifies temporal abnormality patterns and uses them to detect and assess equipment failure risks.

3. Proposed Methodology

The APM-TCN architecture uses multivariate sensor time-series data to provide an integrated predictive-maintenance pipeline for identifying aberrant operating behaviour and predicting equipment failures shown in Figure 1. The structure begins with C-MAPSS engine sensor data. It includes engine IDs, operating cycles, operational characteristics, and sensor readings. Preprocessing normalises raw sensor data using Min-Max or Z-score transformation, resolves missing values, and filters noise. This preprocessing phase standardises sensor sequences for learning representations and temporal analysis.

A Variational Autoencoder (VAE) learns compact latent representations of machinery operating state from normalised sensor sequences. The VAE encoder converts input signals into a latent vector Z while the decoder reconstructs sensor observations. The difference between the original and reconstructed signals produces the reconstruction error E_{rec} , which indicates abnormal behavior that the learned normal operating distribution cannot adequately represent. The TCN employs dilated convolutional layers to detect temporal correlations on multiple time scales after receiving the latent representation. The temporal prediction error E_{temp} is the difference between actual and anticipated sensor behaviour over time, created by the TCN.

The self-attention feature-refinement module improves TCN temporal features. Query, Key, and Value representations compute sensor-property and degradation-related temporal attention weights. A further source of anomaly information is the attention score A_{att} . The reconstruction error, temporal prediction error, and attention score are subsequently combined in the adaptive anomaly scoring module using $S(t) = \alpha E_{rec}(t) + \beta E_{temp}(t) + \gamma A_{att}(t)$. A dynamic threshold $\tau(t)$ is then applied, where observations satisfying $S(t) > \tau(t)$ are identified as anomalous, while observations satisfying $S(t) \leq \tau(t)$ are considered normal.

The anomaly data is given to the DBSCAN-based temporal pattern clustering module. DBSCAN detects sparse data as noise or outliers and groups dense observations by abnormal feature-space density. Because the resulting groups represent normal, deterioration, failure, and noise/outlier patterns, it can identify repeated degradation signatures in time-related anomaly data. Based on these clustered patterns, the failure-risk prediction module evaluates the progression from normal operation to degradation and failure. This module forecasts how failure risk will fluctuate during the operating cycle and provides statistics to predict impending failure.

Transforming projected risk information enables proactive maintenance planning, early warning for high-risk equipment, and issue diagnostics. This reduces unplanned downtime and improves equipment reliability. The APM-TCN architecture integrates anomaly detection and predictive maintenance by creating a continuous analytical pathway from raw sensor

observations to preprocessing, latent representation, temporal modelling, attention refinement, adaptive anomaly scoring, DBSCAN pattern discovery, failure-risk prediction, and maintenance decision.

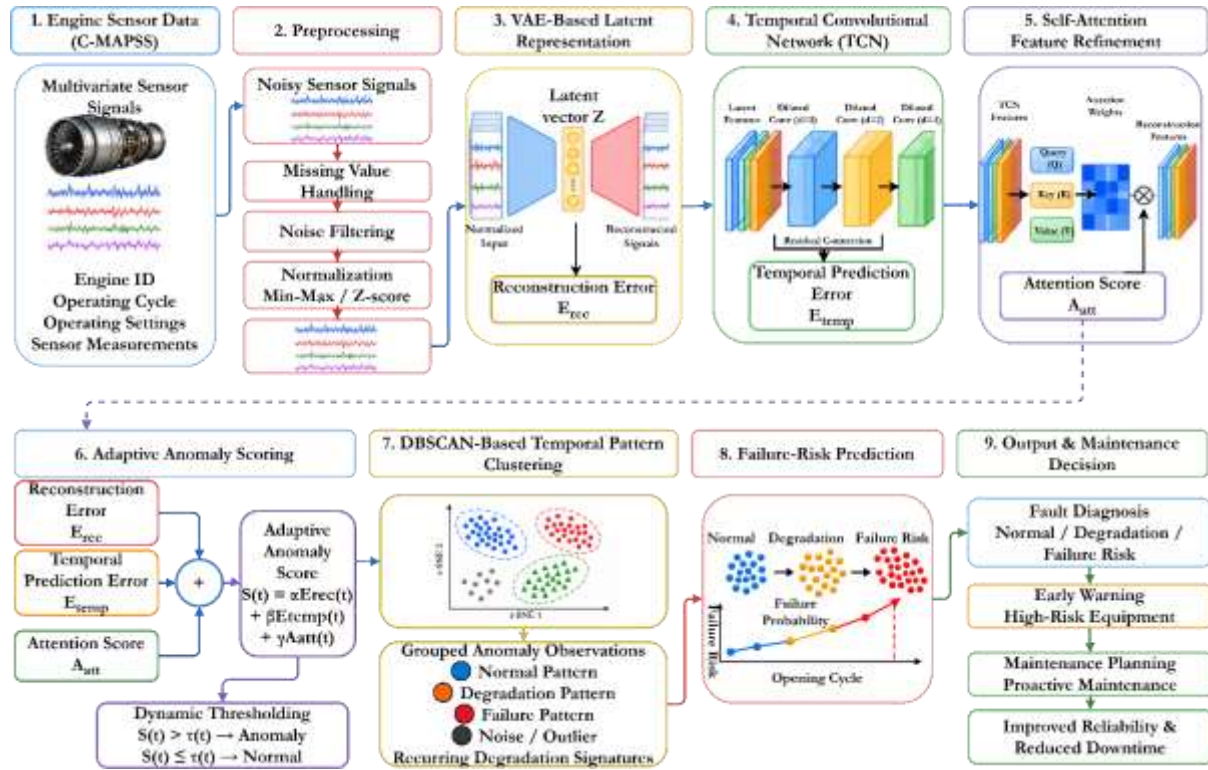


Figure 1: Proposed Methodology

$$X^p = \frac{\mathcal{F}(M \odot X + (1-M) \odot \tilde{X}) - \mu}{\sigma + \epsilon} \quad (1)$$

Equation (1) represents preprocessing of the multivariate C-MAPSS sensor data, including missing-value replacement, noise filtering, and sensor-wise normalization. Here, X is the raw sensor matrix, M is the validity mask, $\tilde{X}$ represents estimated missing values, $\mathcal{F}(\cdot)$ denotes noise filtering, μ and σ are the sensor-wise mean and standard deviation, ϵ is a small numerical-stability constant, and X^p is the resulting preprocessed sensor representation supplied to the VAE.

$$q_\phi(z | X^p) = \mathcal{N}(z; \mu_z(X^p), \text{diag}(\sigma_z^2(X^p))) \quad (2)$$

Equation (2) represents the VAE-based latent representation learning stage, where the preprocessed multivariate sensor data are transformed into a compact probabilistic latent space. Here, $q_\phi(z | X^p)$ denotes the approximate posterior distribution learned by the VAE encoder, X^p represents the preprocessed sensor matrix obtained from Equation (1), z denotes the latent representation of the machinery operating state, $\mu_z(X^p)$ represents the latent mean vector estimated from the input data, $\sigma_z^2(X^p)$ denotes the corresponding latent variance vector, $\text{diag}(\cdot)$ represents the diagonal covariance matrix constructed from the latent variances, and ϕ denotes the learnable parameters of the VAE encoder. The resulting latent representation captures the essential characteristics of the sensor observations while reducing dimensionality and preserving the variability of machinery operating conditions.

$$\mathbf{z} = \mu_z(\mathbf{X}^p) + \sigma_z(\mathbf{X}^p) \odot \varepsilon_z, \quad \varepsilon_z \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (3)$$

Equation (3) represents the VAE reparameterization and latent sampling process, which generates a stochastic latent representation while maintaining differentiability during model training. Here, $\mathbf{z}$ denotes the sampled latent representation of the machinery sensor sequence, $\mu_z(\mathbf{X}^p)$ represents the latent mean vector estimated by the VAE encoder, $\sigma_z(\mathbf{X}^p)$ denotes the latent standard deviation vector, ε_z represents a random noise vector sampled from the standard normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{0}$ denotes the zero vector, $\mathbf{I}$ represents the identity matrix, and $\odot$ denotes element-wise multiplication. This formulation enables the VAE to sample informative latent representations from the learned probability distribution while allowing the model parameters to be optimized through gradient-based learning. The resulting latent representation $\mathbf{z}$ is subsequently provided to the temporal modeling stage for learning machinery degradation patterns.

$$\mathcal{L}_{VAE} = \frac{1}{TD} \|\mathbf{X}^p - D_\psi(\mathbf{z})\|_F^2 + \lambda D_{KL}[q_\phi(\mathbf{z} | \mathbf{X}^p) \| p(\mathbf{z})] \quad (4)$$

Equation (4) represents the VAE training objective used to learn a compact and informative latent representation of the preprocessed C-MAPSS sensor data. Here, $\mathcal{L}_{VAE}$ denotes the overall VAE loss, $\mathbf{X}^p$ represents the preprocessed sensor matrix, $D_\psi(\mathbf{z})$ denotes the reconstructed sensor representation generated by the decoder, $\mathbf{z}$ is the latent representation learned by the encoder, T represents the number of temporal observations, D denotes the number of sensor variables, $\|\cdot\|_F^2$ represents the squared Frobenius norm used to measure reconstruction error, $D_{KL}(\cdot \| \cdot)$ denotes the Kullback–Leibler divergence between the learned latent distribution and the prior distribution, $q_\phi(\mathbf{z} | \mathbf{X}^p)$ represents the encoder-derived latent distribution, $p(\mathbf{z})$ denotes the predefined latent prior distribution, and λ controls the contribution of the latent-distribution regularization term. The reconstruction term preserves important sensor characteristics, while the KL-divergence term encourages a structured latent space suitable for subsequent temporal degradation analysis.

$$\mathbf{h}_t^{(l)} = \sigma\left(\sum_{k=0}^{K-1} W_k^{(l)} \mathbf{h}_{t-d_l k}^{(l-1)} + \mathbf{b}^{(l)}\right) \quad (5)$$

Equation (5) represents the Temporal Convolutional Network (TCN)-based extraction of temporal features from the latent sensor representations, enabling APM-TCN to capture both short-term and long-term degradation dependencies in machinery operation. Here, $\mathbf{h}_t^{(l)}$ denotes the temporal feature representation at time step t and TCN layer l , $\mathbf{h}_{(t-d_l k)}^{(l-1)}$ represents the feature at the preceding layer with temporal offset determined by the dilation factor, $W_k^{(l)}$ denotes the learnable convolution kernel at position k in layer l , K represents the convolution kernel size, d_l denotes the dilation factor of layer l , $\mathbf{b}^{(l)}$ represents the layer bias, and $\sigma(\cdot)$ denotes the nonlinear activation function. The dilated temporal convolution expands the receptive field without requiring excessively deep networks, allowing the framework to identify evolving sensor patterns and long-range dependencies associated with progressive machinery degradation.

$$\mathbf{H}^{att} = \sum_{t=1}^T \alpha_t \mathbf{h}_t, \quad \alpha_t = \frac{\exp(\mathbf{w}^T \mathbf{h}_t)}{\sum_{j=1}^T \exp(\mathbf{w}^T \mathbf{h}_j)} \quad (6)$$

Equation (6) represents the attention-weighted temporal feature refinement stage, where the TCN-generated temporal representations are assigned adaptive importance based on their relevance to machinery degradation. Here, $\mathbf{H}^{att}$ denotes the resulting attention-refined temporal representation, $\mathbf{h}_t$ represents the TCN feature vector at time step t , T denotes the total number of temporal observations in the input sequence, α_t denotes the normalized attention weight assigned to the t -th temporal feature, and $\mathbf{w}$ represents a learnable attention parameter.

vector. The term $w^T h_t$ measures the relevance of each temporal feature, while the exponential normalization ensures that the attention weights remain positive and collectively represent the relative importance of the temporal observations. Consequently, H^a emphasizes informative degradation-related patterns while reducing the influence of less relevant temporal observations.

$$S(t) = \alpha E_{\text{rec}}(t) + \beta E_{\text{temp}}(t) + \gamma A_{\text{att}}(t), \quad \alpha + \beta + \gamma = 1 \quad (7)$$

Equation (7) represents the adaptive anomaly score generation stage of APM-TCN, where complementary information from sensor reconstruction, temporal prediction, and attention-based feature analysis is integrated to quantify abnormal machinery behavior. Here, $S(t)$ denotes the overall anomaly score at time step t , $E_{\text{rec}}(t)$ represents the reconstruction error between the original and VAE-reconstructed sensor observations, $E_{\text{temp}}(t)$ denotes the temporal prediction error obtained from the TCN-based temporal modeling process, and $A_{\text{att}}(t)$ represents the attention-derived anomaly information associated with the temporal feature representation. The parameters α , β , and γ denote the adaptive weighting coefficients assigned to the reconstruction, temporal, and attention components, respectively, while the constraint $\alpha + \beta + \gamma = 1$ ensures a normalized contribution of the three components. The resulting score $S(t)$ provides an integrated representation of sensor deviation, temporal inconsistency, and degradation-relevant feature importance for identifying anomalous operating states.

$$\tau(t) = \mu_S + \kappa \sigma_S(t), \quad \mathcal{A}(t) = \begin{cases} 1, & S(t) > \tau(t) \\ 0, & S(t) \leq \tau(t) \end{cases} \quad (8)$$

Equation (8) represents the dynamic thresholding mechanism used to distinguish anomalous operating states from normal machinery behavior based on the adaptive anomaly score. Here, $\tau(t)$ denotes the time-dependent anomaly threshold, μ_S represents the mean anomaly score estimated from the reference operating data, $\sigma_S(t)$ denotes the local standard deviation of the anomaly score around time step t , and κ is a sensitivity coefficient controlling the deviation required for an observation to be considered anomalous. The variable $S(t)$ represents the adaptive anomaly score obtained from Equation (7), while $\mathcal{A}(t)$ denotes the binary anomaly decision, where a value of 1 indicates an anomalous operating state and 0 indicates normal behavior. By adapting the threshold according to the observed score distribution, this mechanism can accommodate changes in machinery operating conditions and reduce dependence on a fixed anomaly threshold.

$$\mathcal{C}_k = \left\{ h_i^{\text{att}} \mid \|h_i^{\text{att}} - h_j^{\text{att}}\|_2 \leq \varepsilon, \quad |\mathcal{N}_\varepsilon(h_i^{\text{att}})| \geq \text{MinPts} \right\} \quad (9)$$

Equation (9) represents the DBSCAN-based temporal anomaly pattern clustering stage, where attention-refined temporal representations are grouped by density and similarity to reveal recurring machinery degradation patterns. Here, $\mathcal{C}_k$ denotes the k -th temporal anomaly cluster, h_i^{att} and h_j^{att} represent attention-refined feature vectors corresponding to different temporal observations, $\|\cdot\|_2$ denotes the Euclidean distance between two feature representations, ε represents the neighborhood radius used to determine the proximity of observations, $\mathcal{N}_\varepsilon(h_i^{\text{att}})$ denotes the set of neighboring observations within the ε -radius of h_i^{att} , and MinPts denotes the minimum number of neighboring observations required to form a dense region. Observations satisfying the density condition are assigned to the same cluster, whereas isolated observations that do not satisfy the required density can be treated as noise or isolated anomaly patterns. The resulting clusters provide recurring temporal degradation signatures that can subsequently support failure-risk prediction and maintenance decision-making.

$$P_{\text{fail}}(t) = \sigma(w_f^T [H_t^{\text{att}} \parallel c_t \parallel s_t] + b_f) \quad (10)$$

Equation (10) represents the failure-risk prediction stage of APM-TCN, where the attention-refined temporal representation, clustered anomaly information, and current anomaly score are integrated to estimate the probability of impending machinery failure. Here, $P_{\text{fail}}(t)$ denotes the predicted failure probability at time step t , $\sigma(\cdot)$ represents the sigmoid activation function that converts the model output into a probability value between zero and one, H_t^{att} denotes the attention-refined temporal feature representation, c_t represents the feature vector associated with the temporal anomaly cluster identified through DBSCAN, s_t denotes the adaptive anomaly score or its corresponding feature representation, $\parallel$ denotes vector concatenation, w_f represents the learnable failure-risk prediction weight vector, and b_f denotes the prediction-layer bias. By jointly utilizing temporal degradation features, recurring anomaly-pattern information, and the adaptive anomaly score, the resulting $P_{\text{fail}}(t)$ provides an integrated estimate of the likelihood of impending machinery failure and supports early-warning and predictive maintenance decisions.

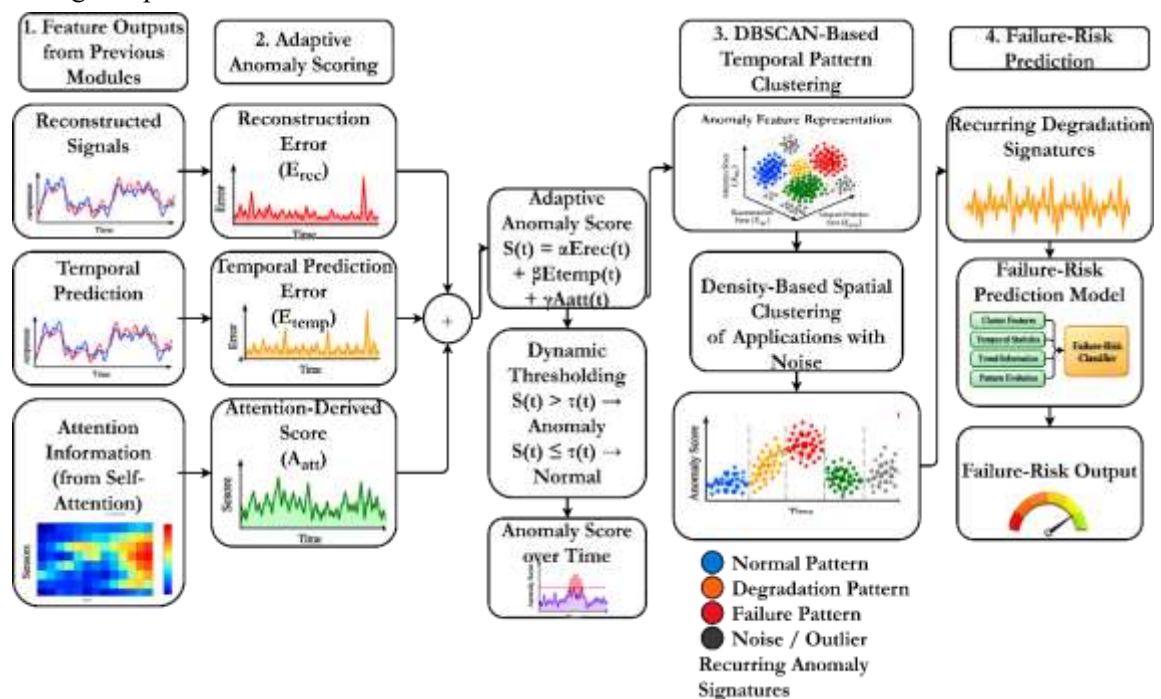


Figure 2. Internal Mechanism of APM-TCN for Adaptive Anomaly Scoring, Temporal Pattern Clustering, and Failure-Risk Prediction

The APM-TCN framework's internal processing mechanism is shown in the figure using feature outputs from the VAE, TCN, and self-attention modules. Reconstructed signals, temporal prediction outputs, and attention data serve as supplemental feature sources in the first stage. The VAE's reconstructed signals quantify the error between the original and reconstructed sensor readings, whereas the TCN's temporal prediction outputs specify the error between the actual and expected temporal behaviour. Attention-derived self-attention module information highlights important operation patterns and deteriorating conduct. It concerns sensor characteristics.

The second stage performs adaptive anomaly scoring by combining the reconstruction error E_{rec} , temporal prediction error E_{temp} , and attention-derived score A_{att} . These components are integrated to calculate the adaptive anomaly score as $S(t) = \alpha E_{\text{rec}}(t) + \beta E_{\text{temp}}(t) + \gamma A_{\text{att}}(t)$, where α , β , and γ control the contribution of the respective information sources. Next, a dynamic threshold distinguishes normal operating behaviour from abnormalities.

Observations with anomaly scores below the time-dependent threshold are normal; those above the threshold are abnormal. DBSCAN is used to represent anomalous characteristics in the third phase. Clustering organises anomalous data by density and similarity in the combined feature space. This lets the system distinguish between normal, degradation, failure, and noise/outlier functioning patterns over time. Instead of addressing each anomalous data point independently, clusters provide repeated anomaly signatures that capture machinery behaviour changes over time.

Recurrent deterioration features are used to forecast failure probability in the fourth stage. The failure-risk classifier is trained on cluster-level features, time-series data, trend information, and pattern evolution to forecast equipment failure risk. The failure-risk output below may improve proactive maintenance planning and equipment failure detection. Figure 2 depicts how APM-TCN organizes anomaly data into temporal patterns and uses them to estimate failure risk. Algorithm 1 expresses the APM-TCN for Adaptive Anomaly Pattern Mining and Failure-Risk Prediction

Algorithm 1: APM-TCN for Adaptive Anomaly Pattern Mining and Failure-Risk Prediction

Input: C-MAPSS sensor time-series data X
Output: Anomaly labels A , anomaly clusters C , and failure-risk P_{fail}
1: Initialize VAE, TCN, attention, and failure-risk parameters.
2: Preprocess X and obtain X^p .
3: While the training criterion is not satisfied, do
4: Encode X^p and obtain μ_z and σ_z .
5: Generate latent representation z .
6: Reconstruct the sensor sequence using the VAE decoder.
7: Update VAE parameters using the reconstruction and KL losses.
8: Extract temporal features using the TCN.
9: Calculate attention weights and obtain H^{att} .
10: Calculate $E_{rec}(t)$, $E_{temp}(t)$, and $A_{att}(t)$.
11: Calculate the adaptive anomaly score:
$S(t) = \alpha E_{rec}(t) + \beta E_{temp}(t) + \gamma A_{att}(t)$
12: Update the network parameters.
13: End While
14: Calculate the dynamic anomaly threshold $\tau(t)$.
15: For each time step t do
16: If $S(t) > \tau(t)$ then
17: Set $A(t) = 1$.
18: Else
19: Set $A(t) = 0$.
20: End If
21: End For
22: Apply DBSCAN to the detected anomalous representations.
23: Obtain temporal anomaly clusters C .
24: For each detected anomaly cluster do
25: Extract the corresponding degradation representation.
26: Estimate the failure risk $P_{fail}(t)$.
27: If $P_{fail}(t) \geq \tau_{f}$ then
28: Generate an early-warning maintenance alert.
29: End If
30: End For

31: Return A, C, and P fail.

The APM-TCN method initialises the VAE, TCN, self-attention, and failure-risk prediction components with the normalised representation X^p using C-MAPSS sensor time-series data. VAE encoders transform preprocessed sensor sequences into mean and variance latent representations during training. The VAE decoder then reconstructs the latent feature Z from these representations. VAE parameters are improved using reconstruction and KL-divergence losses to simulate equipment operational conditions. The TCN processes the learned latent data to extract temporal degradation features, while the attention mechanism focuses on the most important temporal information. The adaptive anomaly score ($S(t)$) is calculated by integrating reconstruction error, temporal prediction error, and attentional information. Using a dynamic threshold, the method distinguishes normal and unusual operating conditions. DBSCAN clusters the aberrant representations to uncover recurring degradation patterns. Finally, an early maintenance alert is generated when the predicted risk exceeds the threshold, estimating failure risk $P_{\text{fail}}(t)$; when the degradation representations are from recognised clusters. Thus, the system uses a unified predictive-maintenance architecture for latent representation learning, temporal feature extraction, adaptive anomaly detection, anomalous-pattern grouping, and failure-risk computation.

4. Results and Discussion

Experimentally testing the APM-TCN framework's ability to detect anomalies, identify anomalous patterns, and forecast early failure for industrial equipment using multivariate sensor time-series data. The assessment procedure considers recall, accuracy, precision, false-alarm reduction, and early-warning skills. Compare the results with baseline methods to study adaptive anomaly scoring, temporal feature learning, and anomaly-pattern grouping. We'll use statistics to demonstrate APM-TCN's ability to recognise unusual operating conditions and provide reliable early warnings for predictive maintenance.

a. Dataset Description:

The NASA C-MAPSS Turbofan Engine Dataset was used to evaluate the proposed APM-TCN system because it provides multivariate sensor time-series data showing gradual degradation and run-to-failure behaviour. The dataset includes four subsets—FD001, FD002, FD003, and FD004—produced under different operating conditions and failure types. This observation covers the engine's unique identification, operating cycle, three operational parameters, and sensor data. Training trajectories run from system startup to failure, whereas test trajectories end shortly before failure and provide ground-truth useful-life values. Because of fault configurations, degradation trajectories, and operating conditions, early failure prediction, temporal dependency learning, and anomaly-pattern recognition were examined under different operating conditions [27].

b. Experimental Setup:

The recommended framework was evaluated on a workstation with 64 GB RAM, an NVIDIA RTX 4090 GPU, and an Intel Core i9 CPU to run computationally intensive deep-learning applications. Testing was done on Windows 11 (64-bit) using cuDNN, PyTorch 2.4, CUDA 12.1, and Python 3.11 for GPU-accelerated model training. NumPy, Pandas, Scikit-learn, SciPy, and Matplotlib handled data preparation, statistical analysis, clustering, assessment, and visualisation. Time-series data from NASA C-MAPSS sensors were tabulated before being transformed into fixed-length temporal sequences. PyTorch incorporated VAE, TCN, self-attention, adaptive anomaly-scoring, and DBSCAN. APM-TCN and the baseline approaches were compared using similar hardware and software to make repeatability easy.

c. Anomaly Detection Accuracy

The Anomaly Detection Accuracy statistic assesses the APM-TCN framework's capacity to distinguish usual and abnormal operating states in industrial equipment sensor time-series data. With a detection rate of 97.2% and a recognition rate of 98.2%, the framework can reliably assess the operating condition of 97.2% of tested samples. Once trained representations identify deviations from expected machinery activity, the APM-TCN framework detects anomalies via temporal feature extraction and adaptive anomaly-score construction. High accuracy indicates the framework's ability to differentiate normal and non-standard engine states across all evaluated trajectories. This statistic is useful for assessing APM-TCN's classification capabilities since it considers normal-state and anomaly predictions. Divide the number of evaluated observations by the total number of correctly categorised observations to calculate anomaly detection accuracy. It considers TP for properly diagnosed anomalies and TN for correctly recognised normal observations; FP and FN for incorrect classifications. The math expression is (11):

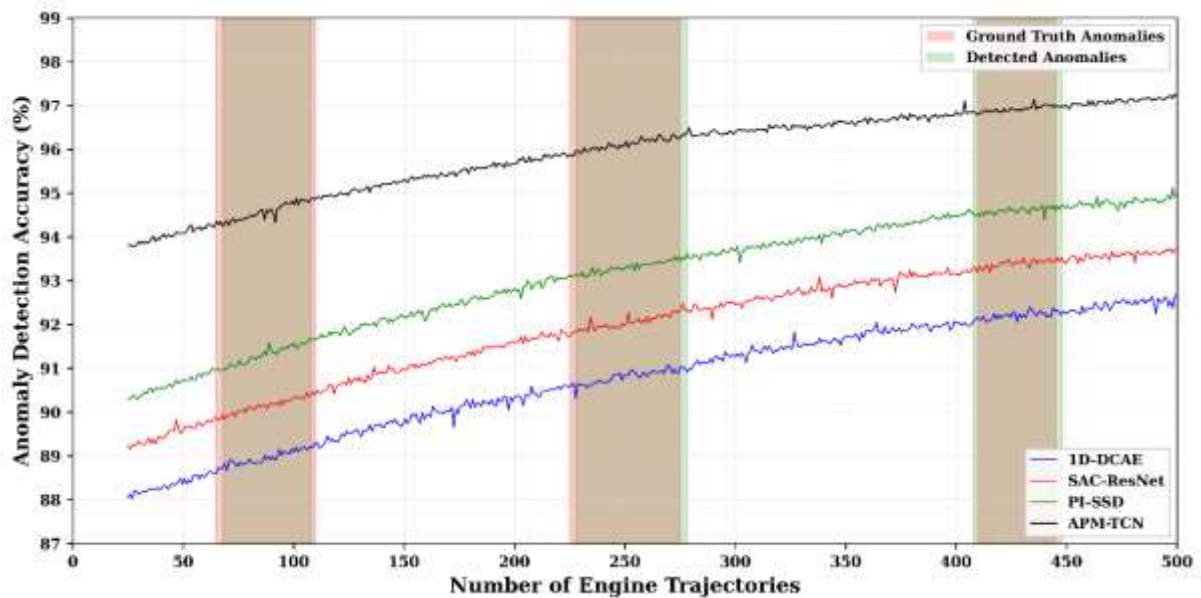


Figure 3: Anomaly Detection Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (11)$$

Figure 3 and equation (11) Anomaly Detection Accuracy. For APM-TCN, the reported 97.2% anomaly detection accuracy indicates a high overall detection capability. However, accuracy alone does not reveal whether errors are concentrated in anomaly detection or normal-state recognition; therefore, precision, recall, and F1-score are additionally used to provide a more detailed evaluation of the framework's anomaly-detection performance.

d. Precision

Precision measures how well the proposed APM-TCN architecture classifies observations as aberrant and generates anomaly alerts. To clarify, accuracy assesses the reliability of anomaly signals by measuring how many of the predicted anomalies are actual. APM-TCN's 95.8% accuracy shows that most anomalous events reflect true abnormal operating conditions. This is particularly important in predictive maintenance, since many false anomaly alarms can lead to unnecessary inspections, maintenance, equipment shutdowns, and higher operating

costs. APM-TCN reduces false alarms by maintaining high accuracy, showing that its adaptive anomaly-scoring technique can distinguish real from spurious sensor abnormalities. Both accurate and faulty findings are utilised to determine accuracy. TP represents the number of anomalous observations APM-TCN identified, whereas FP represents the number of regular observations mislabeled as anomalies. The math formula is (12):

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \quad (12)$$

The proposed APM-TCN 95.8% accuracy suggests that most anomaly alarms are true abnormalities. Validating the framework's ability to reduce false-positive anomaly detections is essential for reliable condition monitoring and predictive-maintenance implementation high precision value supports 97.2% anomaly detection accuracy. Table 1 expresses the Precision.

Table 1: Precision

Number of Engine Trajectories	1D-DCAE Precision (%)	SAC-ResNet Precision (%)	PI-SSD Precision (%)	APM-TCN Precision (%)
100	88.5 ± 0.7	89.8 ± 0.6	90.9 ± 0.5	92.1 ± 0.4
200	90.1 ± 0.6	91.3 ± 0.5	92.3 ± 0.4	93.5 ± 0.3
300	91.3 ± 0.5	92.5 ± 0.4	93.5 ± 0.4	94.5 ± 0.3
400	92.5 ± 0.4	93.7 ± 0.4	94.7 ± 0.3	95.3 ± 0.2
500	93.7 ± 0.4	94.9 ± 0.3	95.6 ± 0.2	95.8 ± 0.2

e) Recall

Recall is an APM-TCN framework's ability to recognise true anomalous operating conditions in industrial equipment sensor data. Precision measures anomaly-alert accuracy, whereas recall measures the framework's ability to detect true abnormalities. APM-TCN detects 96.4% of sensor sequence abnormalities, reducing the chance of unknown machinery anomalies. This characteristic is critical in predictive maintenance because if problems are not discovered quickly, progressive degradation may occur, leading to unexpected equipment failure or greater maintenance demands. With self-attention and adaptive anomaly scoring, the VAE-TCN architecture learns temporal representations with good recall, allowing it to capture a large proportion of anomalous temporal patterns. Recall is calculated using TP and FN. TP represents correctly identified anomalies, whereas FN represents actual abnormalities misclassified as normal. The math formula is (13):

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \quad (13)$$

With a 96.4% recall rate, APM-TCN remains susceptible to atypical operating conditions. To minimise missed failure-related patterns, high recall is crucial because false-negative detections may delay maintenance. APM-TCN's 95.8% accuracy and recall suggest it balances finding many true problems with limiting false alarms. Without considering accuracy or F1-score, higher recall can increase false-positive detections. Table 2 express the Precision.

Table 2: Recall

Number of Engine Trajectories	1D-DCAE (%)	SAC-ResNet (%)	PI-SSD (%)	APM-TCN (%)
100	89.6 ± 0.6	90.7 ± 0.5	91.6 ± 0.4	92.5 ± 0.4
200	91.2 ± 0.5	92.2 ± 0.4	93.1 ± 0.4	93.8 ± 0.3
300	92.4 ± 0.4	93.4 ± 0.3	94.3 ± 0.3	94.8 ± 0.3
400	93.6 ± 0.3	94.6 ± 0.3	95.4 ± 0.2	95.7 ± 0.2

500	94.8 ± 0.3	95.8 ± 0.2	96.2 ± 0.2	96.4 ± 0.2
-----	------------	------------	------------	-------------------

f. F1-score

Anomaly detection systems with false alarms and missed abnormalities can be assessed using the F1-score, which balances recall and precision. Its F1-score of 96.1% reveals that the proposed APM-TCN framework detects abnormal operating conditions without incorrect anomaly classifications. Unlike accuracy, which depends on the percentage of normal samples, F1-score directly evaluates anomaly detection precision and recall. For industrial predictive maintenance, systems that are highly accurate but have limited memory may miss key deterioration events, while systems that are less accurate but generate many false alarms may lead to unnecessary inspections. Finally, the F1-score reveals that APM-TCN reliably detects abnormalities while lowering false-positive and false-negative errors. The harmonic mean of recall and accuracy determines the F1-score, which is lower when one metric is substantially lower than the other as explored mathematically in (14):

Table 3: F1-score

Number of Engine Trajectories	1D-DCAE	SAC-ResNet	PI-SSD	APM-TCN
100	88.9 ± 0.6	90.1 ± 0.5	91.0 ± 0.4	92.2 ± 0.4
200	90.6 ± 0.5	91.6 ± 0.4	92.5 ± 0.4	93.6 ± 0.3
300	91.9 ± 0.4	92.8 ± 0.3	93.7 ± 0.3	94.5 ± 0.3
400	93.1 ± 0.3	94.0 ± 0.3	94.8 ± 0.2	95.4 ± 0.2
500	94.3 ± 0.3	95.2 ± 0.2	95.8 ± 0.2	96.1 ± 0.2

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

Precision and Recall are the percentages of real anomalies predicted and detected, respectively. The harmonic-mean approach counts both metrics toward the final score, rather than letting one compensate for the other. Thus, APM-96.1% TCN's F1-score balances detection completeness and prediction reliability, allowing it to identify abnormal temporal patterns in industrial equipment sensor data.

g) False-Alarm Rate (%)

The false-alarm rate reduction measures how well the proposed APM-TCN system reduces false anomaly alerts compared to a baseline. Industrial predictive maintenance false alarms result from misidentifying typical equipment operating characteristics as abnormal. In experiments, APM-TCN reduces superfluous anomaly notifications by 31.5% compared to the comparator approach. Reducing false alarms is essential to avoid unnecessary inspections, maintenance interventions, operational interruptions, and inefficient use of maintenance resources. APM-TCN uses adaptive anomaly scoring, temporal feature learning, and anomaly-pattern analysis to increase maintenance warning reliability. It can identify persistent undesirable behaviour from multivariate sensor measurements after normal modifications. Compare the baseline method's false-alarm rate to the proposed frameworks to assess the reduction in equation (15a):

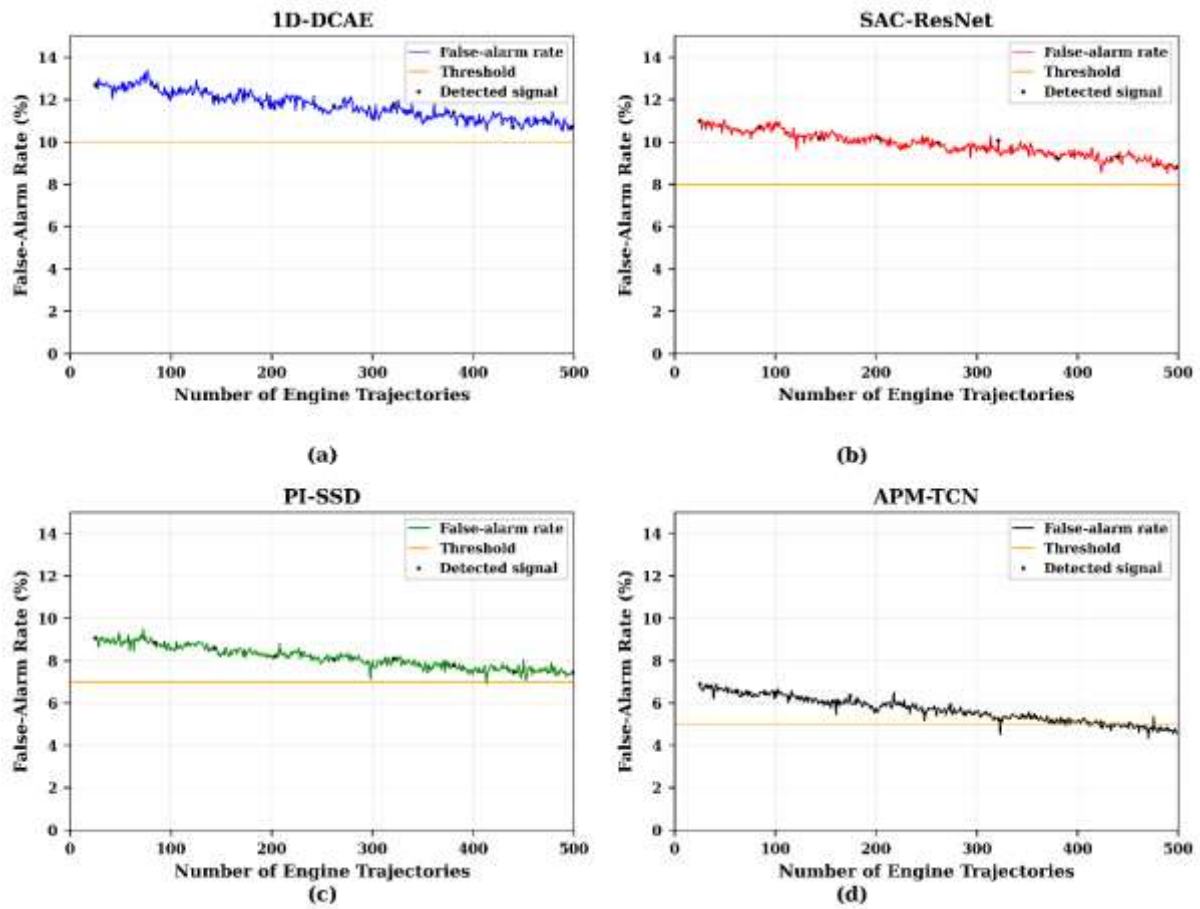


Figure 4: False-Alarm Rate (%)

$$\text{False-Alarm Rate} = \frac{FP}{FP+TN} \times 100 \quad (15a)$$

Figure 4 and equation (15a) define the vii) False-Alarm Rate (%). Where FP denotes the number of normal observations incorrectly identified as anomalies, while TN represents the number of normal observations correctly classified as normal. The corresponding reduction is calculated as (15b):

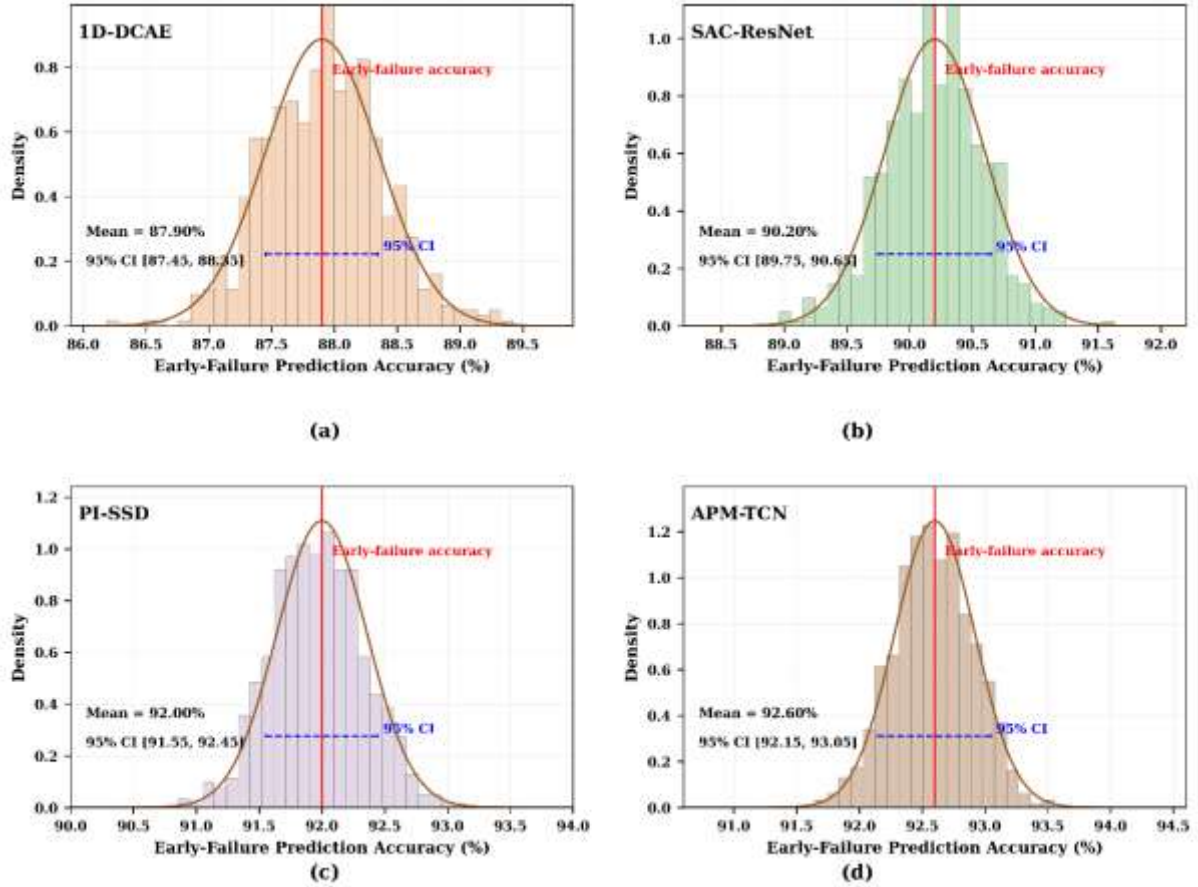
$$\text{False-Alarm Reduction} = \frac{FAR_{\text{Baseline}} - FAR_{\text{APM-TCN}}}{FAR_{\text{Baseline}}} \times 100 \quad (15b)$$

where FAR_{Baseline} represents the false-alarm rate of the selected baseline method and $FAR_{\text{(APM-TCN)}}$ represents the false-alarm rate obtained by APM-TCN. Therefore, the reported 31.5% reduction indicates that APM-TCN decreases false anomaly alerts by 31.5% relative to the specified baseline under the experimental evaluation.

h) Early-Failure Prediction Accuracy

Early-failure prediction accuracy measures how well the proposed APM-TCN framework detects patterns of equipment deterioration early enough to prevent failure. Early detection is crucial in predictive maintenance because it allows maintenance actions to be scheduled before equipment reaches a critical operational condition. In 92.6% of the prediction scenarios

analysed, APM-TCN successfully identified the assessed early-failure situations, with an early-failure prediction accuracy of 92.6%. The framework's temporal feature extraction and anomaly-pattern analysis capabilities enable this by capturing progressive changes in multivariate sensor readings rather than relying solely on anomalous data. As a result, the framework can foretell when equipment may start to degrade, giving you more time to arrange



maintenance before things go south. The accuracy of early failure predictions is defined as the fraction of examined early failure prediction situations that were successfully anticipated calculated in equation (16):

Figure 5: Early-Failure Prediction Accuracy (%)

$$\text{Early-Failure Prediction Accuracy} = \frac{N_{\text{correct predictions}}}{N_{\text{total prediction cases}}} \times 100 \quad (16)$$

Figure 5 and equation (16) discuss early-failure prediction accuracy. Where N_{correct} (" predictions) represents the number of degradation or failure events correctly identified within the predefined early-warning period, and N_{total} (" prediction " " cases) represents the total number of evaluated prediction cases. Thus, the reported 92.6% early-failure prediction accuracy indicates that APM-TCN reliably identifies impending machinery failure under the defined experimental conditions. This metric complements the anomaly-detection measures because it evaluates not only whether abnormal behavior is detected, but also whether the detected degradation provides sufficiently early information for proactive maintenance planning. Table 4 shows the Ablation Study.

Table 4: *Ablation Study*

Configurati on	VA E	TC N	Self- Attenti on	Adapti ve Anoma ly Scorin g	DBSCA N	Accura cy (%)	Precisi on (%)	Reca ll (%)	F1- Scor e (%)
APM-TCN w/o VAE	✗	✓	✓	✓	✓	94.8 ± 0.4	93.9 ± 0.4	94.6 ± 0.3	94.2 ± 0.3
APM-TCN w/o TCN	✓	✗	✓	✓	✓	95.1 ± 0.3	94.2 ± 0.3	94.9 ± 0.3	94.5 ± 0.3
APM-TCN w/o Self- Attention	✓	✓	✗	✓	✓	95.6 ± 0.3	94.8 ± 0.3	95.3 ± 0.3	95.0 ± 0.2
APM-TCN w/o Adaptive Scoring	✓	✓	✓	✗	✓	95.3 ± 0.3	94.5 ± 0.3	95.1 ± 0.3	94.8 ± 0.3
APM-TCN w/o DBSCAN	✓	✓	✓	✓	✗	96.0 ± 0.2	95.2 ± 0.2	95.8 ± 0.2	95.5 ± 0.2
Full APM- TCN	✓	✓	✓	✓	✓	97.2 ± 0.2	95.8 ± 0.2	96.4 ± 0.2	96.1 ± 0.2

Table 4 shows that removing components reduces APM-TCN's anomaly detection performance. Removing the VAE reduces compact representation extraction from multivariate sensor data because it decreases the model's ability to learn latent representations. Performance decreases when the TCN is removed because it cannot capture temporal dependencies across operational cycles. Removing self-attention and adaptive anomaly scoring reduces the ability to identify typical variations from abnormal operating patterns. Anomaly-pattern clustering organises time-connected aberrant data because deleting DBSCAN reduces it considerably. APM-TCN has 97.2% accuracy, 95.8% precision, 96.4% recall, and 96.1% F1-score.

5. Conclusion

In conclusion recommend using APM-TCN to predict when industrial equipment will require repair by examining sensor data from multiple factors over time. Through VAE-based latent representation learning, TCN-based temporal feature extraction, self-attention, adaptive anomaly scoring, dynamic thresholding, and DBSCAN-based anomalous pattern grouping, the system identifies machinery degradation. The APM-TCN model integrates reconstruction deviations, temporal prediction errors, and attention-derived information for anomalous operational behaviour and failure-risk assessment. The framework may provide early warnings, reduce equipment failures, and enable maintenance planning. Future work will verify APM-TCN across industrial datasets, improve uncertainty-aware failure-risk prediction, and merge online and ongoing learning to account for changing operational conditions for real-time predictive-maintenance applications.

REFERENCES

- [1]. Li, L., Xu, J., & Li, J. (2023). Estimating remaining useful life of rotating machinery using relevance vector machine and deep learning network. *Engineering Failure Analysis*, 146, 107125.
- [2]. Fu, S., Lin, L., Wang, Y., Guo, F., Zhao, M., Zhong, B., & Zhong, S. (2024). MCA-DTCN: A novel dual-task temporal convolutional network with multi-channel attention for first prediction time detection and remaining useful life prediction. *Reliability Engineering & System Safety*, 241, 109696.
- [3]. He, S., Du, M., Jiang, X., Zhang, W., & Wang, C. (2024). VAEAT: Variational AutoEncoder with adversarial training for multivariate time series anomaly detection. *Information Sciences*, 676, 120852.
- [4]. Qin, Y., Gan, F., Xia, B., Mi, D., & Zhang, L. (2024). Remaining useful life estimation of bearing via temporal convolutional networks enhanced by a gated convolutional unit. *Engineering Applications of Artificial Intelligence*, 133, 108308.
- [5]. Xu, Z., Zhang, Y., & Miao, Q. (2024). An attention-based multi-scale temporal convolutional network for remaining useful life prediction. *Reliability Engineering & System Safety*, 250, 110288.
- [6]. Correia, L., Goos, J. C., Klein, P., Bäck, T., & Kononova, A. V. (2024). Online model-based anomaly detection in multivariate time series: Taxonomy, survey, research challenges and future directions. *Engineering Applications of Artificial Intelligence*, 138, 109323.
- [7]. Yang, B., Long, W., Zhang, Y., Xi, Z., Jiao, J., & Li, Y. (2024). Multivariate time series anomaly detection: Missing data handling and feature collaborative analysis in robot joint data. *Journal of Manufacturing Systems*, 75, 132–149.
- [8]. Tu, F.-F., Liu, D.-J., Yan, Z.-W., Jin, X.-B., & Geng, G.-G. (2024). STFT-TCAN: A TCN-attention based multivariate time series anomaly detection architecture with time-frequency analysis for cyber-industrial systems. *Computers & Security*, 144, 103961.
- [9]. Kim, G., Choi, J. G., & Lim, S. (2024). Using transformer and a reweighting technique to develop a remaining useful life estimation method for turbofan engines. *Engineering Applications of Artificial Intelligence*, 133, 108475.
- [10]. Rathinam, V., Krishnasamy, V., Rajheshwari, K. C., Pethuraj, M. S., & Baskar, S. (2026). Design of a Fault Tolerance System for Industrial Controllers Utilizing Fractional Ensemble Learning. *International Journal of Information Technology & Decision Making*, 1-40.
- [11]. Klein, P., Malburg, L., & Bergmann, R. (2025). Combining informed data-driven anomaly detection with knowledge graphs for root cause analysis in predictive maintenance. *Engineering Applications of Artificial Intelligence*, 145, 110152.
- [12]. Shaikh, F., Ahmed, B. S., & Swerin, A. (2025). Unsupervised detection of faults in industrial pumps from multivariate time series. *Machine Learning with Applications*, 22, 100784.
- [13]. Benatia, M. A., Hafsi, M., & Ben Ayed, S. (2025). A continual learning approach for failure prediction under non-stationary conditions: Application to condition monitoring data streams. *Computers & Industrial Engineering*, 204, 111049.
- [14]. He, J., Ma, Z., Liu, Y., & Yang, Z. (2026). Remaining useful life prediction of rotating machine via long short-term memory network with uncertainty quantification. *Engineering Applications of Artificial Intelligence*, 164, 113280.
- [15]. Zhang, W., Li, X., Li, J., Ma, J., Kong, P., Zhang, S., & Liu, Y. (2025). A time–frequency contrastive learning model for anomaly detection in multivariate time series. *Complex & Intelligent Systems*, 11(12), 475.
- [16]. Cen, Z., Hu, S., Hou, Y., Chen, Z., & Ke, Y. (2024). Remaining useful life prediction of machinery based on improved Sample Convolution and Interaction Network. *Engineering Applications of Artificial Intelligence*, 135, 108813.
- [17]. Kamat, P., Kumar, S., Patil, S., & Kotecha, K. (2024). Anomaly-informed remaining useful life estimation (AIRULE) of bearing machinery using deep learning framework, *MethodsX* 12 (2024) 102555, doi: 10.1016. J. MEX.

- [18]. Wang, F., Liu, A., Qu, C., Xiong, R., & Chen, L. (2025). A deep-learning method for remaining useful life prediction of power machinery via dual-attention mechanism. *sensors*, 25(2), 497.
- [19]. Wang, C., Jiang, W., Shi, L., & Zhang, L. (2025). Rolling bearing remaining useful life prediction using deep learning based on high-quality representation. *Scientific reports*, 15(1), 8228.
- [20]. Jin, Z., Chen, C., Syntetos, A., & Liu, Y. (2025). Enhanced bearing RUL prediction based on dynamic temporal attention and mixed MLP. *Autonomous Intelligent Systems*, 5(1), 2.
- [21]. Assafo, M., & Langendoerfer, P. (2025). Tool remaining useful life prediction using feature extraction and machine learning-based sensor fusion. *Results in Engineering*, 107297.
- [22]. Luu, T. T., & Huynh, D. A. (2025). A ResNet-based deep reinforcement learning framework using soft actor-critic for remaining useful life prediction of rolling bearings. *Results in Engineering*, 106739.
- [23]. Abbas, Z., Wang, H., Hussain, M., Song, X., Ali, Z., & Cao, W. (2026). Remaining useful life estimation based on improved attention mechanism. *Results in Engineering*, 29, 108813.
- [24]. Star, M., & McKee, K. (2025). Deep state space models for remaining useful life estimation. *Applied Soft Computing*, 114003.
- [25]. Amin, M. A., Ahsan, M. S., Maua, J., Ahmed, M., & Nur, K. (2026). Physics-informed self-supervised diagnosis of rotating machinery using latent ODEs and transformer encoders. *Plos one*, 21(2), e0339239.
- [26]. Colombi, L., Vespa, M., Belletti, N., Brina, M., Dahdal, S., Tabanelli, F., ... & Vignoli, M. (2025). Embedding Models for Multivariate Time Series Anomaly Detection in Industry 5.0: L. Colombi et al. *Data Science and Engineering*, 1-17.
- [27]. <https://www.kaggle.com/datasets/behrad3d/nasa-cmaps>