machines

Article

Noise-Conditioned Denoising Autoencoder with Temporal Attention for Bearing RUL Prediction

Zhongtian Jin [1], Chong Chen [2], Aris Syntetos [3] and Ying Liu [1,*]

[1] Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK
[2] Guangdong Provincial Key Laboratory of Cyber-Physical System, Guangdong University of Technology, Guangzhou 510006, China
[3] PARC Institute of Manufacturing Logistics and Inventory, Cardiff Business School, Cardiff University, Cardiff CF10 3EU, UK
* Correspondence: liuy81@cardiff.ac.uk

Abstract

Bearings are important elements of mechanical systems and the correct forecasting of their remaining useful life (RUL) is key to successful predictive maintenance. Nevertheless, noise interference during different operating conditions is also a significant problem in predicting their RUL. Existing denoising-based RUL prediction models often show degraded performance when exposed to heterogeneous and non-stationary noise, resulting in unstable feature extraction and reduced generalisation. To address the challenge of heterogeneous and non-stationary noise in bearing RUL prediction, this study proposes a hybrid framework that combines a noise-conditioned convolutional denoising autoencoder (NC-CDAE) and a temporal attention transformer (TAT). The NC-CDAE adaptively suppresses diverse noise types through conditional modulation, while the TAT captures long-term temporal dependencies to enhance degradation trend learning. This synergistic design improves both the noise robustness and temporal modelling capability of the system. To further validate the model under varying conditions, synthetic datasets with different noise intensities were generated using a conditional generative adversarial network (cGAN). Comprehensive experiments show that the proposed NC-CDAE + TAT framework achieves lower and more stable errors than state-of-the-art methods, reducing RMSE by up to 23.6% and MAE by 18.2% on average and maintaining consistent performance (an RMSE between 0.155 and 0.194) across diverse conditions.

Keywords: deep learning; remaining useful life; prognostic and health management; transformer network

Check for updates

Academic Editors: Hongrui Cao, Jianping Xuan and Yongqiang Liu

Received: 9 December 2025
Revised: 5 January 2026
Accepted: 6 January 2026
Published: 8 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In industrial machinery, bearings are both crucial and common in industries like rail, aerospace, and wind. Nevertheless, industrial applications have many bearings that work under rather severe conditions, being more prone to corrosion, wear, or other types of degradation [1,2]. Proper forecasting of the remaining useful life (RUL) of bearings is hence very important [3]. However, in different working conditions, the occurrence of various noise interferences tends to the introduction of errors in RUL forecasting. Not only does reliable and efficient prediction of life assist in predicting maintenance of machines and avoiding unexpected downtime [4], but it also helps to reduce economic losses and safety risks. Therefore, predicting the robustness of models to various forms of interference is an essential component of RUL prediction.

The current methods of bearing signal denoising and RUL prediction are traditional signal processing methods and deep learning-based methods. Indicatively, the simplest deep learning denoising methods, including denoising autoencoders (DAEs), convolutional autoencoders (CAEs) [5], and recurrent neural network-based denoising models, have been used to pre-process raw vibration data. Although such methods can minimise noise to a reasonable degree and enhance further feature extraction, they have several drawbacks. DAEs can have difficulties predicting local structure on highly non-stationary signals, CAEs can be subject to changes in noise distributions, and RNN-based models can be computationally expensive and unstable in long sequences [6]. Consequently, single-method solutions tend not to be very robust to a wide range of noise.

Due to the high pace of development of deep learning, modularised model design has emerged as a promising path to solving complex problems. Regarding the case of RUL prediction, a denoising step before feature extraction has been demonstrated to improve the prediction accuracy. Convolutional denoising autoencoders (CDAEs) are capable of good preservation of local spatial correlations and wavelet thresholding can maintain valuable time-frequency data [7]. Variational Autoencoders (VAEs) [8] and noise distribution-learnable generative adversarial network (GAN)-based denoising techniques have also been considered, which provide flexibility in learning noise distributions. Although each of them have their own benefits, including strong local feature learning (CDAEs), adaptive multi-resolution analysis (wavelet thresholding) [9], and generative robustness (VAEs/GANs), all of them have difficulties in ensuring consistent robustness under various levels of noise and different operational conditions. This drawback highlights the significance of noise adaptability as a defining factor for measuring denoising models in RUL tasks.

In practical bearing prognostics, vibration signals are inevitably affected by heterogeneous and non-stationary noise caused by varying operating conditions, environmental disturbances, and sensor-related interference. A common drawback of the current data-driven RUL prediction algorithms is that their performance cannot remain steady under circumstances where the noise properties during testing are not the same as those in training and, as a result, it is challenging to attain a high level of prediction accuracy and a high level of robustness simultaneously. Most of the current methods employ fixed or intuitively acquired denoising methods, which can either insufficiently suppress noise or excessively smooth degradation-related features, particularly when different noise levels are used. To overcome such difficulties, this paper will use a noise-sensitive, time-sensitive RUL prediction model that serves to trade-off accuracy and robustness in realistic scenarios of noise. The specifics of the model are as follows:

- A noise-conditioned convolutional denoising autoencoder (NC-CDAE) is introduced to explicitly estimate the noise influence level prior to signal reconstruction, enabling adaptive denoising that avoids unnecessary computational effort and reduces the risk of over- or under-denoising.
- By conditioning the denoising process on estimated noise characteristics, the framework preserves degradation-sensitive features while adapting to different noise conditions.
- A temporal attention transformer (TAT) is employed to model the intrinsic temporal dependencies of bearing degradation, allowing the network to selectively focus on fault-relevant time segments and capture long-term degradation trends that are essential for remaining useful life prediction.

Through the integration of noise-conditioned denoising and a TAT, the proposed framework provides a principled solution to the accuracy–robustness trade-off that commonly limits existing RUL prediction approaches under complex noise environments.

The rest of this paper is organised as follows. Section 2 is a review of the related literature on RUL prediction and the application of synthetic data (SDG) and denoising techniques in predictive tasks. Section 3 presents the suggested methodology, both the model architecture and the temporal attention mechanism. Section 4 represents the datasets, ablation studies, and parameter settings. Section 5 gives the results of the experiment and further analysis. Lastly, Section 6 wraps up the paper and gives future research directions.

2. Related Works

2.1. Denoising Methods for Degradation Signal Processing

Denoising methods are a popular strategy to make RUL prediction more accurate for lithium-ion batteries. Models based on neural networks and their variations can be used extensively in such tasks [10]. As an example, a denoising transformer neural network (DTNN) has been proposed to predict lithium-ion battery RUL and has proved to be capable of handling complicated degradation signals [11]. Moreover, a more recent method of data-driven decomposition, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), was used to break down battery capacity degradation into multi-scale component sequences [12]. These deconstructed characteristics have also been used to enhance accuracy in the prediction of RUL in battery systems [13]. In general, it has been observed that the predictive power and strength of battery degradation modelling improve greatly when denoising methods are added.

Regarding bearings, denoising has been one of the main areas of study in the field of RUL prediction [14,15]. Some of the most popular methods include autoencoder-based methods that have the potential to learn representative latent features. Convolutional denoising autoencoders (CDAEs) are demonstrated to be effective in preserving local fault-related patterns [7,16], whereas stacked denoising autoencoders (SDAs) with thresholding strategies have been found to balance sensitivity and robustness in various operating conditions [17]. Even more advanced architectures, including adaptive denoising residual networks (AD-ResNets) [18] and recurrent denoising autoencoders (R-DAEs) [19], are also shown to be highly performing by incorporating temporal modelling capabilities or adaptive shrinkage processes.

In addition to autoencoders, classical time-frequency denoising techniques (wavelet thresholding [20,21] and empirical wavelet transform [22]) have been widely applied to obtain stable indicators of degradation. A more recent hybrid method jointly trains wavelet-based denoising images using CNNs or transformer encoders [23] and tries to combine handcrafted multi-resolution analysis with feature-based image generation. These works underline the significance of the combination of denoising and feature learning.

Over the past few years, denoising approaches to mechanical degradation cues have gradually evolved to be no longer based on explicit signal filtering, but on representation-conscious and adaptive learning processes. Alternatively, rather than simply reducing noise in the raw signal space, several studies seek to reduce the effects of noise by improving the strength of feature representations. Indicatively, Spirto et al. [24] contrasted time-frequency image-based CNNs with Symmetrical Dot Pattern (SDP)-based CNNs to prove that different signal-to-image mappings can reach similar diagnostic accuracy but use less computation, hence enhancing robustness and real-time application under noisy conditions. In addition to representation transformation, a learnable interpretable wavelet Kolmogorov–Arnold convolutional LSTM, which added a learnable wavelet kernel to convolutional layers to learn noise-resilient spatial features and attention-enhanced temporal modelling, which addresses non-stationary and noisy operating conditions, was proposed by Chen et al. [25]. More recently, Guo et al. [26] suggested an end-to-end RUL prediction scheme that integrates an autoencoder-based soft-thresholding denoising block and a multiscale temporal attention transformer, allowing denoising thresholds and temporal relationships to be adaptively adjusted according to signal properties.. These papers suggest that there is an apparent shift in the direction of adaptive and structure-conscious denoising, in which noise resilience is increasingly realised in the design of a model, not necessarily by hard-and-fast filtering choices.

Overall, current denoising methods of degradation signal processing have shown that the use of learning-based representations is much more effective to enhance the performance of RUL prediction in the presence of noise. The classical signal processing techniques and traditional autoencoder-based denoisers work well in reducing noise when the statistical behaviour of the noise is mostly stable, but their effectiveness has been found to fall in real-world situations of non-homogeneous types of noise, non-stationary operating environments, and a variety of signal-to-noise ratios. More recent advancements demonstrated that stronger robustness may be encouraged by incorporating domain-sensitive representations, adaptive thresholding, or attention-centred mechanisms into the denoising process. However, in most current approaches, noise adaptation is implicit and it is performed by the choice of representations or the learning of parameters, not by explicitly training the denoising behaviour based on learned perceptions of noise. This weakness demonstrates that it is necessary to have denoising models which can model noise explicitly as a condition-dependent variable, thus inspiring the construction of noise-conditioned denoising models, like the NC-CDAE presented in this paper.

2.2. Deep Learning Frameworks for RUL Prediction

The development of deep learning techniques has largely enabled vibration-based RUL predictions, taking advantage of their ability to model dynamic degradation of vibration signals with little processing. Recurrent models like LSTM and GRU are well-used since they are good at learning temporal variations [27,28] and better versions like DOS-ELM use adaptive forgetting factors to overcome degradation that is of a non-stationary nature [29]. The CNN-based or convolution-recurrent hybrid models [30–32] are capable of both local fault transient and longer-term behaviour and they perform highly in some situations. Also popular is time-frequency feature learning, where inputs are time-frequency transformed to an STFT or a wavelet and the result is used as an input to the CNNs to enhance their ability to handle non-stationary behaviours [20,21,23,33].

Although these improvements have been achieved, it is observed that most of these architectures degrade their performance when subjected to heterogeneous operating conditions or in complicated, noisy environments. This has resulted in the search for more expressive sequence models. Recently, transformer-based architectures, which have the advantage of a long-range dependency modelling capability, have attracted considerable interest in RUL prediction [34–41]. Better predictive stability has been shown in frequency-aware attention [42], multi-scale hierarchical transformers [40], and hybrid convolution–attention structures [43–45]. Nevertheless, conventional self-attention remains prone to instability when exposed to noisy inputs and most transformer-based RUL models do not have a specific denoising or noisy input adaptation mechanism.

The other area of research is synthetic data generation (SDG). Due to the expensive and time-intensive nature of high-quality degradation datasets, synthetic augmentation is also an attractive method to enhance robustness and decrease overfitting. Multiple-mode or condition-dependent degradation patterns have also been modelled with GAN-based synthetic data generation, demonstrating potential [46–48]. However, the application of conditional GANs to explicitly recreate different noise regimes is not well studied in bearing prognostics. This opens the possibility of integrating SDG with the noise-robust learning of features to yield more generalizable RUL models.

In conclusion, in the available literature, there are two major conclusions. First, denoising is necessary and is particularly successful in cases when the noise properties are different under different conditions. Second, attention-based temporal modelling is a prospective path that can be used to model both short- and long-term cues of degradation, although its resiliency to various noise distributions remains insufficient. Based on such insights, the paper presents a hybrid framework in which a noise-conditioned denoiser (NC-CDAE) is combined with the TAT model, with the assistance of cGAN-generated multi-noise data, in order to guarantee stability under operating conditions.

3. Methodology

This section describes the proposed methodology for RUL prediction with noisy and varying operating conditions. The benchmark datasets XJTU-SY and PRONOSTIA are first used as the baseline and a conditional generative adversarial network (cGAN) is proposed to augment the data and mimic various noise scenarios in a more realistic way than either Gaussian perturbations or fixed SNR. Figure 1 provides the network structure. The added signals are then processed through the NC-CDAE, which adaptively denoises while retaining important degradation patterns. Finally, the TAT is used to capture long-range dependencies and make accurate RUL estimations. Together, these elements create a powerful and noise-resistant predictive maintenance system.

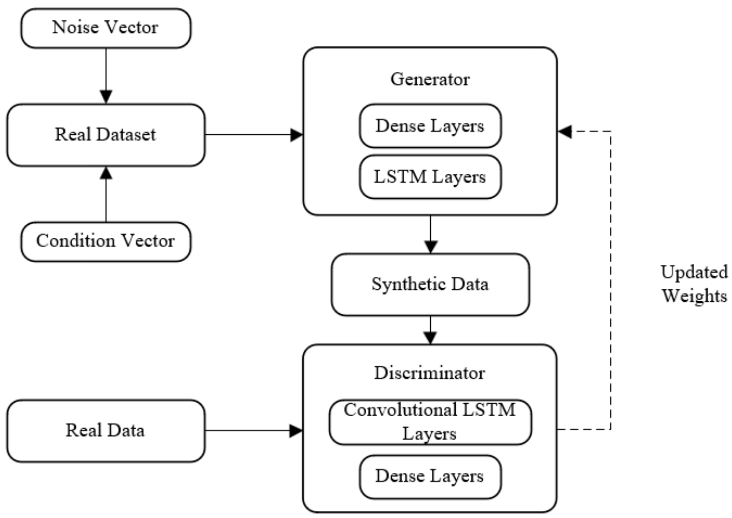

Figure 1. cGAN for the Synthetic Data.

3.1. Data Acquisition and Preprocessing

To overcome the data shortage issue and simulate different noise levels, a conditional generative adversarial network (cGAN) is adopted for synthetic bearing vibration signal generation in this paper. Figure 1 shows the cGAN for the synthetic data. The framework is composed of two major components, a generator and a discriminator. The generator takes, as input, a random noise vector z along with a condition vector c, where c describes varying levels of noise (e.g., low, medium, and high). Through a series of dense and LSTM layers, the generator generates time series signals as output in the same dimension as the real bearing data. The discriminator, in turn, takes the real and generated samples along with the condition vector and attempts to determine their authenticity. In the process of adversarial training, the generator learns that it is making realistic signals under specified noise conditions, while the discriminator improves at telling which are real and which are synthetic. This iterative process leads to a synthetic dataset that closely matches the statistical properties of the original data but with controllable noise variability.

3.2. NC-CDAE

The NC-CDAE is used as the denoising module in our framework. In contrast to the general CDAE, the proposed model uses a noise-conditioning module to ensure the model can adapt to the various noise distributions that occur in the adversarial simulation stage. In terms of the denoising activity, this design helps improve effectiveness at variable operation conditions or noise intensities.

The architecture of the NC-CDAE comprises an encoder, a bottleneck with noise conditioning, and a decoder. The encoder progressively reduces the temporal resolution of the input signal, X_{noisy}, extracting hierarchical feature maps. An auxiliary noise estimation branch is introduced at the top of the encoder to estimate a noise descriptor, $\hat{\sigma}$:

$$Sigma\hat{\sigma} = h(X_{noisy}) \tag{1}$$

where $\hat{\sigma}$ represents either the estimated noise variance or a vector encoding the noise type. This descriptor is then injected into the bottleneck layer through feature-wise linear modulation (FiLM).

In the bottleneck, the latent representation f is modulated according to the estimated noise descriptor as follows:

$$f' = \gamma(\hat{\sigma})\cdot f + \beta(\hat{\sigma}) \tag{2}$$

The noise modulation functions $\gamma(\cdot)$ and $\beta(\cdot)$ are applied to two separate two-layer multilayer perceptrons (MLPs) in the proposed NC-CDAE. Both MLPs are composed of an input layer, one ReLU-activated hidden layer, and an output layer with the same number of channels as the latent features. In particular, the noise descriptor $\hat{\sigma}$ is initially run through the MLPs to obtain the scaling and shifting parameters that are involved in the feature-wise linear modulation (FiLM) process. The dimension of the noise features is 16. This value has been chosen as a compromise between representation capacity and computational complexity and is based on early experiments. It was identified that a lower-dimensional descriptor was not sufficient to represent heterogeneous noise properties, but higher dimensions only offered marginally better performance with a correspondingly more complex model. Noise descriptors of 16 dimensions were thus considered a stable and effective setting for any experiment.

The specific mechanism of noise modulation is depicted in Figure 2. The encoder is a two-layer one-dimensional convolutional network that has a kernel size of 5 and a stride of 2, sequentially taking out the hierarchical temporal features of the noisy input signal. The encoder output includes a dedicated noise estimation branch that applies global average pooling followed by a fully connected layer, producing a 16-dimensional noise descriptor $\hat{\sigma}$. The noise descriptor is then input to two separate two-layer MLPs using ReLU activation to produce the FiLM modulation parameters γ and β. These parameters are adaptive parameters that adjust the latent features in the bottleneck through feature-wise linear modulation. The decoder is a structure that reflects the encoder structure through transposed convolution layers to restore the denoised signal. This architecture will allow adaptive denoising in heterogeneous noise environments, as well as having a lightweight structure.

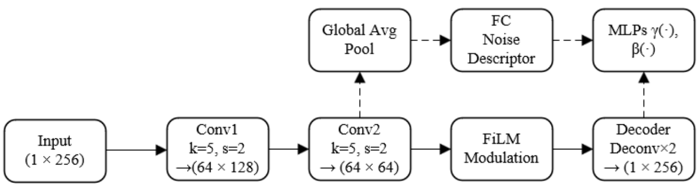

Figure 2. Architecture of the NC-CDAE.

The model is trained with a combined objective. The main term is an L_1 reconstruction loss to ensure signal fidelity and an auxiliary term supervises the noise estimation branch when ground truth noise information is available:

$$L = \left\| \hat{X}_{clean} - X_{clean} \right\|_1 + \lambda \|\hat{\sigma} - \sigma\|_1 \tag{3}$$

By incorporating noise conditioning, the NC-CDAE can handle multiple noise environments in a unified framework. It avoids over-smoothing while preserving the impulsive features that are critical for fault-related signal analysis.

3.3. Temporal Attention Transformer for RUL Prediction

After denoising, the cleaned signals are fed to the temporal attention transformer (TAT), which is designed to model short-term and long-term dependencies in vibration signals. This stage allows the model to identify degradation patterns and better forecast fault progression.

Input embedding: The denoised signals are divided into overlapping windows of a length of 256 samples with a stride of 128. Each window is passed through a linear layer to obtain a feature embedding of 128 dimensions. A sinusoidal positional encoding is applied to maintain temporal order so that the transformer is able to differentiate between different time steps.

Self-attention layer: A multi-head self-attention module with 8 heads and 128 hidden dimensions is used as the self-attention module for each embedding sequence. The self-attention mechanism enables the model to give more weight to fault-relevant transients while deflating irrelevant background fluctuations. This feature makes the transformer more robust than purely recurrent architectures.

Feed-forward layer. The output of the attention module goes through a feed-forward network with two linear layers and ReLU activation. The chosen covert size is 256, which is large enough to provide enough modelling power without too much computational cost. Layer normalisation and residual connections are used to stabilise training and improve convergence.

Stacked layers: The TAT is made up of 4 stacked transformer layers. Stacking multiple layers allows the model to learn temporal features in a hierarchical manner: shallow layers capture local vibration features, whereas deeper layers capture long-range dependencies and degradation features.

Prediction head: Global average pooling is performed to aggregate the contextualised features and pass them over a fully connected layer to generate the final output. In the case of classification tasks, this head returns class probabilities for each type of fault. For regression tasks like this, it gives us an estimated RUL.

The overall network architecture of the proposed NC-CDAE + TAT framework is illustrated in Figure 3. The combination of the NC-CDAE and the TAT yields complementary advantages. The NC-CDAE performs adaptive noise shaping of the input signals, allowing the fault transients that are frequently masked in noise to be maintained. The TAT then extends this with the ability to capture temporal dependencies and selective attention to fault-critical time steps. Combined, the two modules constitute a strong end-to-end framework that increases fault detection accuracy, increases prediction stability, and adapts well to changing noise conditions.

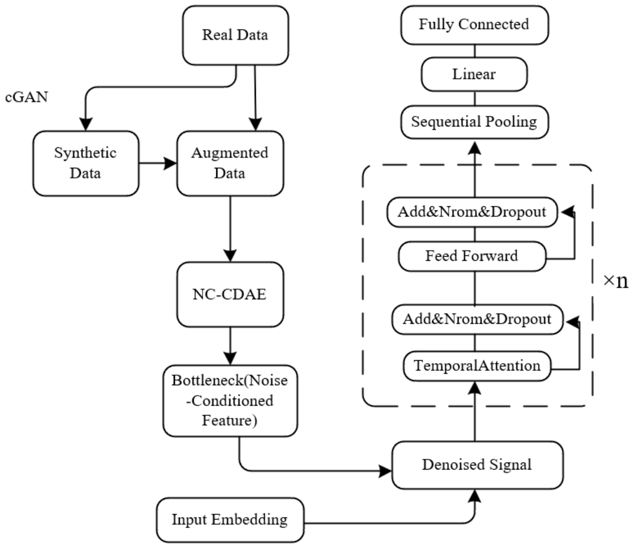

Figure 3. Network structure.

3.4. Training and Evaluation

The proposed framework is trained and tested on synthetic noisy datasets created by extending XJTU-SY and PRONOSTIA with a conditional GAN that adds various noise conditions beyond the typically assumed Gaussian noise. The NC-CDAE + TAT framework presented has several hyperparameters that control the temporal resolution, modelling capacity, and computational efficiency. These parameters were chosen according to the characteristic timescale of bearing degradation, as opposed to an automated network optimisation algorithm. In particular, the vibration signals were divided into sliding windows of 256 samples with a stride of 128 samples, which offers a compromise between transient fault-related signal localities and the possibility to model the longer periods of degradation evolution between successive windows with a stride of 128 samples. When carrying out prognostics, long-term temporalities gain more influence towards the later degradation phase, and the chosen length of the window enables the temporal model to identify long-term temporalities, without unduly decreasing the training samples.

In the TAT, there are four layers of transformers with eight attention heads that create a hierarchical temporal feature learning with complementary attention to the various degradation-related patterns at the various timescales. To achieve a trade-off between the efficacy of the denoising and the complexity of the model, the NC-CDAE module used four convolutional layers with a kernel size of 5 and 64 filters, and a noise features dimension of 16 in order to code the noise attributes without over-parameterisation. The optimiser used was the Adam optimiser and the learning rate was 0.0005 with a batch size of 64 and 80 training epochs. No specific hyperparameter optimisation step was presented; rather, the chosen network setup was later confirmed by ablation trials of network depth and attention head count, as presented in Section Ablation of Network Depth and Attention Heads.

4. Experimental Study

4.1. Data Description

In order to assess the performance of the proposed framework, the experiments are carried out on two popular bearing degradation datasets: XJTU-SY and PRONOSTIA. In addition, both datasets deliver run-to-failure vibration signals gathered under different speeds and loads and as a result, they serve as ideal benchmarks for RUL prediction.

Since the noise condition used in the real world is more complex than what the datasets directly provide, a conditional generative adversarial network (cGAN) is used to augment the original signals. The generator adds several types of noise, including Gaussian noise, impulse noise, and mixed noise with different signal-to-noise ratios (SNRs). This method produces synthetic noisy data by maintaining the degradation trend of the original data while emulating realistic noisy environments. These synthetic datasets are used for training and evaluation, thus ensuring that the model is evaluated on different noise conditions.

4.2. Modal Setup

The vibration signals are firstly normalised and divided into fixed-length windows of 256 samples with a stride of 128 samples. Each window is considered a sample to be trained for the model. The NC-CDAE module is set up with four convolutional layers with a kernel size of five and 64 filters. The noise descriptor dimension of the conditioning branch is set to 16. The TAT module consists of 4 transformer layers, where each layer has 8 attention heads, a model dimension of 256, and a feedforward dimension of 1024. Dropout is set to 0.2 in order to prevent overfitting.

Training is performed with the Adam optimiser and an initial learning rate of 0.0005, a batch size of 64, and for 80 epochs. The decayed learning rate is applied to the learning rate after every 20 epochs. All experiments were implemented using PyTorch 1.10.0 and executed on a workstation equipped with an NVIDIA GeForce RTX GPU (NVIDIA Corporation, Santa Clara, CA, USA).

Synthetic Data Generation Parameter Study

Although the main interest of this work is concentrated on denoising and RUL prediction, the quality of the synthetic data generated by the conditional GAN (cGAN) is also important to develop reliable datasets. To make sure that the generated data is similar enough to the real signals, we have conducted a small parameter study on the cGAN. Four representative configurations were tested, with different latent noise dimensions, conditional vector sizes, generator and discriminator learning rates, and discriminator layer numbers. These hyperparameters were selected because they have a direct impact on the diversity and fidelity of the generated sequences. The quality of the generated samples was assessed from two points of view.

- The distributional similarity, measured by the PCA overlap percentage and KL divergence between the real and synthetic feature distributions.
- The downstream relevance, assessed by the RMSE and the MAE, was evaluated when the synthetic data were used to train the prediction pipeline and validated on the real data.

4.3. Ablation Experiments

In addition to cross-model comparisons, ablation studies are designed to probe the contribution of the individual components in the proposed framework.

- Without Noise Conditioning (CDAE + TAT). This setting removes the noise descriptor modulation, reducing the denoiser to a conventional CDAE. The purpose is to investigate whether conditioning on noise information is critical for adaptability to varying noise distributions.
- Without Denoising (TAT only). Here, raw noisy signals are directly input into the TAT. This experiment is designed to explore whether temporal attention alone can compensate for noise or whether explicit denoising remains necessary.
- Without Attention (NC-CDAE + GRU). In this variant, the temporal predictor is simplified by replacing the transformer with a GRU. This setup allows us to analyse the

relative importance of self-attention in capturing long-range dependencies compared to recurrent sequence modelling.

These ablation studies are not only intended to validate the necessity of each component but also to provide insights into the design space of noise-aware denoising and temporal modelling for prognostics.

Ablation of Network Depth and Attention Heads

To further investigate the trade-off between model complexity and denoising/feature extraction ability, we designed an ablation study by varying two structural hyperparameters. First, the number of convolutional layers in the NC-CDAE was set to 3, 4, and 5, which directly affects the network's denoising capacity and reconstruction quality. Second, the number of attention heads in the TAT module was set to 4 and 8, which controls the granularity of temporal dependency modelling and influences computational cost. Combining these factors yields six experimental settings: (3, 4), (3, 8), (4, 4), (4, 8), (5, 4), and (5, 8), where the first value denotes the NC-CDAE layers and the second indicates the TAT heads.

All other training protocols remained identical to Section 4.3 to ensure fair comparison, including the data augmentation strategies, optimiser settings, and evaluation metrics. Each configuration was trained and evaluated three times, and average results were reported in terms of the RMSE, MAE, parameter size, and inference latency. This design enables us to quantify the contribution of network depth and attention width to both performance and efficiency.

4.4. Compare with Different Models

To demonstrate the effectiveness of the proposed NC-CDAE + TAT framework, we compare it with three representative baselines from traditional machine learning, signal processing, and deep learning:

- Traditional Machine Learning (PCA + RF): This baseline applies Principal Component Analysis (PCA) for noise reduction, followed by a Random Forest (RF) regressor for RUL prediction.
- Time-Frequency Transformation + CNN (STFT + CNN): This method employs Short-Time Fourier Transform (STFT) to transform vibration signals into time-frequency representations, which are then processed by a convolutional neural network for RUL prediction.
- GRU + Attention: This baseline applies a Gated Recurrent Unit (GRU) network for sequence modelling, enhanced with an attention mechanism to highlight critical time steps.
- CNN-TAT: A model that uses convolutional layers for feature extraction but without a denoising module.
- CDAE + TAT: A pipeline that employs the conventional CDAE as a denoiser before the TAT predictor.

These comparisons highlight the diversity of the baseline models considered, ranging from traditional CNNs to advanced hybrid architectures. This comprehensive evaluation ensures a fair and robust assessment of the proposed approach's performance relative to existing methods.

5. Results and Discussion

5.1. Synthetic Data Quality Evaluation

Table 1 shows that larger latent and conditional dimensions gave better distributional similarity (a higher PCA overlap and a lower KL divergence) and downstream prediction accuracy (a reduced RMSE). A–D denote different cGAN parameter configurations evaluated in the synthetic data generation study. The quality of the synthetic data was improved in all indicators, which proves that the selected cGAN hyperparameters have a substantial impact on the quality of the synthetic data. In the settings that were tested, the best overall balance was configuration C, which had maximum PCA overlap and minimum prediction errors. Based on these findings, configuration C is taken as the default for further experiments.

Table 1. Synthetic data quality and prediction results under different cGAN settings.

Param Setting	Noise Dim	Cond. Dim	LR ($\times 10^{-4}$)	Disc. Layers	PCA Overlap (%)	KL Divergence	RMSE	MAE
A	64	16	5	3	82.5	0.124	0.185	0.162
B	128	16	5	3	85.2	0.102	0.179	0.157
C	128	32	5	4	88.7	0.095	0.172	0.149
D	256	32	3	4	90.1	0.081	0.169	0.146

To further improve the completeness and transparency of the synthetic data analysis, Figure 4 presents a representative real vibration signal selected from the bearing degradation dataset together with synthetic signals generated by the conditional GAN under different noise conditions. The real signal reflects the typical oscillatory behaviour and long-term degradation trend observed in run-to-failure bearing data. Based on this reference, the cGAN produces multiple noisy variants conditioned on predefined noise levels, resulting in synthetic signals with gradually increasing noise intensity. Importantly, despite the added noise, the temporal alignment, oscillatory structure, and degradation-related trend remain consistent across all synthetic samples. This confirms that the cGAN does not arbitrarily alter the underlying signal characteristics but instead introduces realistic noise perturbations while preserving fault-relevant information.

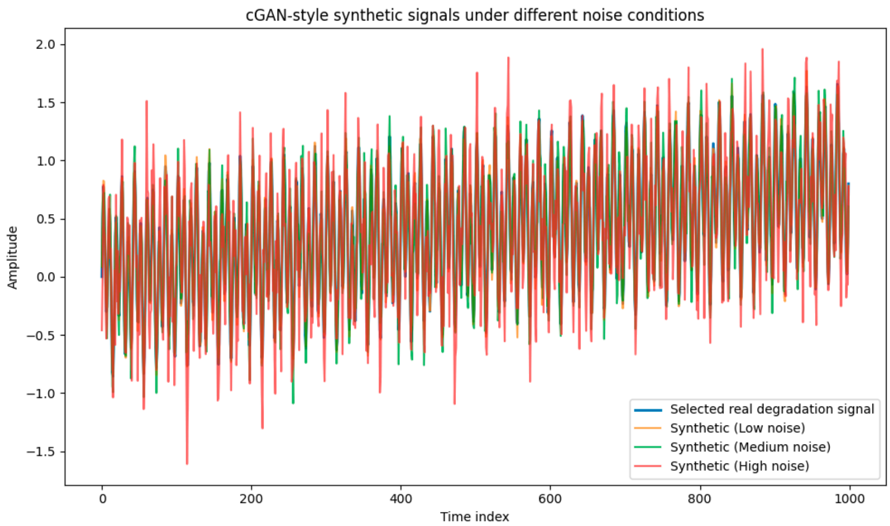

Figure 4. Real bearing vibration signals and cGAN-generated synthetic signals under different noise conditions.

Along with the quantitative analysis presented in Table 1, the intuitive validation of the cGAN model is presented in Figure 4. Although measures like PCA are similar and the KL divergence is used to evaluate the distributional similarity in the feature space, the visual representation in time shows that the signals generated can be physically interpreted and used in the prognostic modelling. The existence of the same predictable patterns of degradation when subjected to varying conditions of noise confirms that the synthetic data can be effectively used to support downstream denoising and RUL prediction tasks. The above observations reveal that the cGAN has a good compromise between diversity of noise and signal fidelity, hence justifying its application in augmenting training and testing data when diverse noise conditions exist.

5.2. Ablation Experiment

To facilitate conducting a clear and consistent assessment of the proposed framework in varying noise conditions, six artificial datasets are developed and employed during the experimental study. These datasets are created with the help of the cGAN when applied to the actual bearing vibration signals of two reference datasets, i.e., XJTU-SY [49] and PRONOSTIA [50]. Datasets 1–3 are acquired by dropping noise onto the vibration signals of three bearing operating conditions in the XJTU-SY dataset, and Datasets 4–6 are acquired in the same way using three bearing operating conditions of the PRONOSTIA dataset. On a dataset, the cGAN generates noisy versions whose noise property is heterogeneous and maintains the original trend of degradation of the original signals. Each dataset is split at random into 80% training and 20% testing, respectively, which is used in all experimental setups. The data structure used in this type of dataset construction allows a systematic comparison of model performance across various source datasets and noise conditions and all the later tables and figures in this section use these six datasets unless otherwise indicated.

Table 2 reports the RMSE and MAE values for six experimental runs under different ablation settings, bold values indicate the best (lowest) RMSE or MAE for each dataset. The full model ("Original," i.e., NC-CDAE + TAT) consistently achieves the lowest errors across all six groups, with an RMSE ranging from 0.155 to 0.205 and an MAE ranging from 0.137 to 0.177. Removing noise conditioning and reverting to a conventional CDAE + TAT configuration leads to a moderate performance drop, while using TAT alone further increases both the RMSE and the MAE. The weakest performance is observed when the transformer is replaced by a GRU ("NC-CDAE + GRU"), where errors rise significantly (e.g., a RMSE above 0.25 in several cases). These results indicate that both noise conditioning in the denoiser and temporal attention in the predictor are necessary for optimal performance.

Table 2. Ablation experiments for six synthetic datasets (Datasets 1–3 are derived from XJTU-SY [49] and Datasets 4–6 are derived from PRONOSTIA [50]).

Metric		1	2	3	4	5	6
RMSE	Original	**0.173**	**0.199**	**0.205**	**0.163**	**0.155**	**0.184**
	CDAE + TAT	0.192	0.226	0.224	0.202	0.164	0.197
	TAT only	0.214	0.259	0.247	0.219	0.175	0.213
	NC-CDAE + GRU	0.279	0.337	0.286	0.259	0.199	0.252
MAE	Original	**0.166**	**0.175**	**0.177**	**0.149**	**0.137**	**0.167**
	CDAE + TAT	0.181	0.210	0.211	0.186	0.146	0.183
	TAT only	0.194	0.231	0.223	0.198	0.163	0.201
	NC-CDAE + GRU	0.215	0.306	0.263	0.233	0.185	0.226

Figure 5 presents an RMSE comparison across the six synthetic datasets under different ablation settings. Figure 6 shows the corresponding MAE comparison for the same datasets. The differences in performance seen can also be explained in the framework of noise distribution adaptation and long-term dependency modelling. In a traditional CDAE, denoising is implicitly learned on the training noise distribution and its fixed reconstruction behaviour is suboptimal in novel noisy conditions, which can be either residual noise or over-smoothing of features of degradation. In comparison, the denoiser proposed is noise conditioned on a noise signal and actively changes its latent features using a feature-wise linear transformation to address noise distribution disparity and stabilise feature representations across noise regimes. In terms of temporal modelling, the GRU-based predictor is based on recurrent state propagation and is prone to poor effective memory and information decay on long horizons, particularly when the input is noisy. However, the self-attention mechanism in the TAT enables direct interactions across distant time steps and selectively highlights fault-relevant segments. This attribute is especially useful in the later degradation phase, in which long-term dependencies prevail over the RUL evolution, which is why the NC-CDAE + TAT setup always performs better.

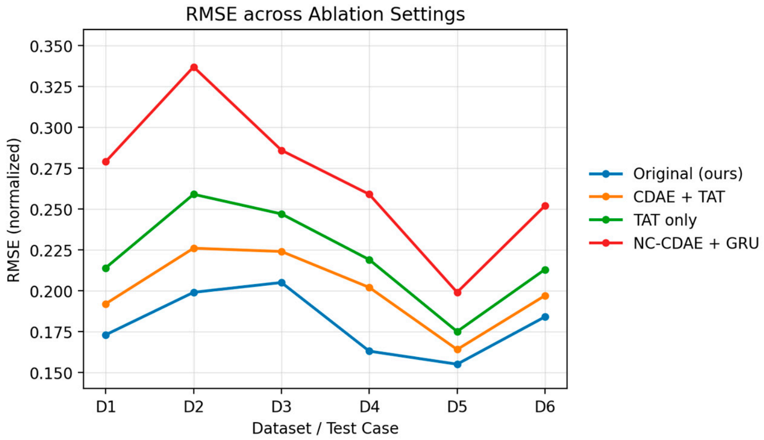

Figure 5. RMSE comparison for the six synthetic datasets under different ablation settings.

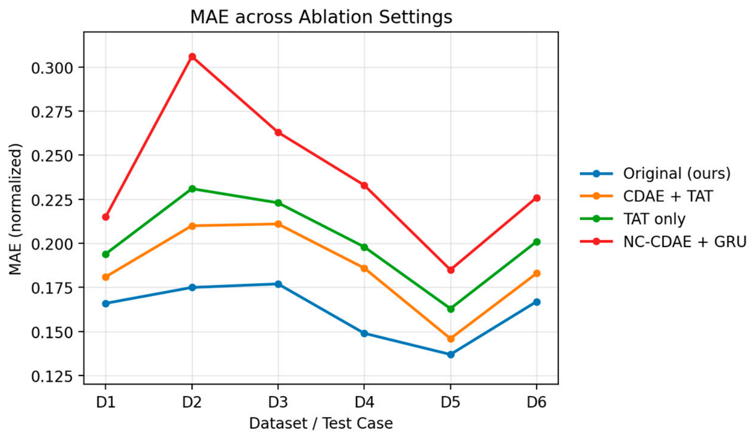

Figure 6. An MAE comparison for the six synthetic datasets.

The comparison brings out two important insights. First, the better results of the complete model prove that the noise-conditioned denoiser is efficient in reducing various noise and retaining degradation-relevant features that are important to predict RUL correctly. The deterioration of the CDAE + TAT and TAT-only conditions indicate that denoising is not as adaptive as conditioning and conditioning is not as adaptive as denoising. Second, the significant error reduction in the NC-CDAE + GRU condition indicates the significance of multi-head self-attention for capturing long-range dependencies and targeting fault-relevant time steps. Collectively, the results confirm the rationale of the combination of NC-CDAE and TAT, that the synergistic effect of noise-sensitive denoising and temporal attention is critical towards realising powerful performance across different noise conditions.

5.3. Ablation Experiment on Network Depth and Attention Heads

The results in Table 3 highlight two main trends. First, increasing the number of NC-CDAE layers from three to five generally improves denoising quality, as reflected in the lower RMSE and MAE, but the marginal gains diminish when moving from four to five layers, while computational cost continues to grow. Second, using eight attention heads in the TAT consistently outperforms four heads in terms of accuracy, especially under more challenging noise conditions, though it comes at the expense of higher latency and more FLOPs. Overall, the (4, 8) configuration provides the best trade-off between performance and efficiency, which aligns with the default setting used in the main experiments. For resource-constrained environments, (4, 4) may be preferable due to its lower computational burden, while (5, 8) offers slightly better robustness in highly noisy scenarios at the cost of heavier computation.

Table 3. Results of ablation of NC-CDAE depth and TAT attention heads.

Combination (Layers, Heads)	RMSE	MAE	Parameters (M)	Relative FLOPs	Avg. Latency (ms)
(3, 4)	0.206	0.188	2.915	0.741	1.002
(3, 8)	0.197	0.181	3.324	0.893	1.121
(4, 4)	0.191	0.173	3.550	0.925	1.083
(4, 8)	0.184	0.167	3.971	1.003	1.227
(5, 4)	0.188	0.171	4.254	1.139	1.205
(5, 8)	0.183	0.165	4.733	1.280	1.362

In addition to the general decrease in the RMSE and MAE, a deeper analysis of the prediction performance at various stages of degradation gives additional information about the performance benefits of the suggested framework. At the initial degradation phase, the frequency noise can often overwhelm the vibration signal and it is hard to identify weak fault-related vibration patterns using conventional models. The noise-conditioning of the NC-CDAE assists in the reduction of irrelevant variations and the maintenance of finer degradation signals, leading to more consistent early-stage predictions. The longer-term temporal dependencies develop as the degradation moves to the middle and later phases, as there are non-linear and cumulative trends of fault development. During this stage, the self-attention mechanism of the TAT allows the adaptive weighting of temporally distant but degradation-relevant features and this is why the later stages of RUL prediction show enhanced stability and lower error variance.

Moreover, the proposed framework has rather stable performance when using small sample training conditions in contrast to baseline models. This strength is explained by the fact that noise suppression and temporal modelling are decoupled: the NC-CDAE reduces input variability by its conditioning on noise attributes, whereas the attention-based predictor predicts salient degradation regions as opposed to looking at dense sequential patterns. This makes the model less susceptible to the lack of data, and thus, it can be more advantageous in real-life prognostic situations with limited labelled run-to-failure information.

5.4. Comparison with Different Models

Table 4 presents the RMSE and MAE values of the different methods across six test cases, bold values indicate the best (lowest) RMSE or MAE for each dataset. The proposed framework consistently achieves competitive or superior performance, with the lowest RMSE and MAE values in most groups, such as 0.135/0.127 in Test 5 and 0.194/0.177 in Test 6. Compared with the CNN-TAT and CDAE + TAT conditions, which both exceed a 0.24 RMSE in several cases, the proposed model maintains more stable error levels across all test groups. The PCA + RF model performs relatively well in some lower-noise scenarios (e.g., Test 1, RMSE = 0.184), but its performance fluctuates more strongly in later cases. The STFT + CNN model provides competitive results in Tests 4 and 5

(RMSE = 0.168 and 0.147), indicating that time-frequency representations can be effective, but performance deteriorates under other conditions. The GRU + Attention model shows moderate accuracy but is consistently outperformed by our method. Overall, the data confirm that the proposed NC-CDAE + TAT achieves lower and more stable error rates compared to both traditional and deep learning baselines.

Table 4. The detailed results for the six synthetic datasets. (Datasets 1–3 are derived from XJTU-SY [49] and Datasets 4–6 are derived from PRONOSTIA [50]).

Test	CNN-TAT		CDAE + TAT		PCA + RF		STFT + CNN		GRU + Attention		Proposed (Ours)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Dataset 1	0.297	0.277	0.278	0.265	0.184	0.161	0.167	0.155	0.174	0.161	**0.168**	**0.154**
Dataset 2	0.305	0.278	0.280	0.252	0.244	0.221	0.211	0.190	0.237	0.211	**0.199**	**0.184**
Dataset 3	0.327	0.299	0.321	0.293	**0.199**	**0.164**	0.203	0.174	0.228	0.199	0.216	0.197
Dataset 4	0.387	0.369	0.449	0.421	0.172	0.167	**0.168**	**0.145**	0.193	0.144	0.190	0.166
Dataset 5	0.245	0.195	0.242	0.231	0.187	0.158	0.147	0.123	0.141	0.132	**0.135**	**0.127**
Dataset 6	0.214	0.189	0.183	0.161	0.205	0.172	0.198	0.180	0.200	0.189	**0.194**	**0.177**

Figure 7 compares the RMSE performance of different models across the six synthetic datasets. Figure 8 illustrates representative predicted RUL trajectories of the proposed method and baseline models under different noise conditions. The comparative results highlight the strengths and weaknesses of different approaches. Traditional PCA + RF is computationally efficient but limited in capturing complex degradation patterns, leading to inconsistent performance. The STFT + CNN model leverages time-frequency features to improve robustness in non-stationary cases, but its handcrafted transformation and higher computational burden reduce its generalizability. The GRU + Attention model is strong in sequence modelling but lacks an explicit denoising stage, making it vulnerable to noise contamination. In contrast, the proposed method integrates noise-conditioned denoising with temporal attention, allowing it to adaptively suppress diverse noise while focusing on fault-relevant time steps. This synergy explains why it consistently achieves the best balance between accuracy and robustness across different test conditions.

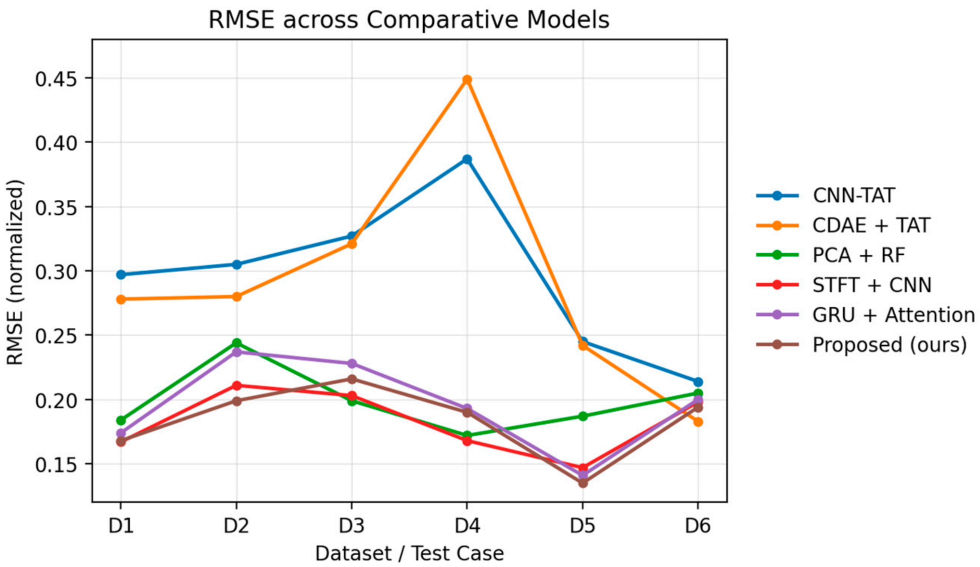

Figure 7. An RMSE comparison for the six synthetic datasets.

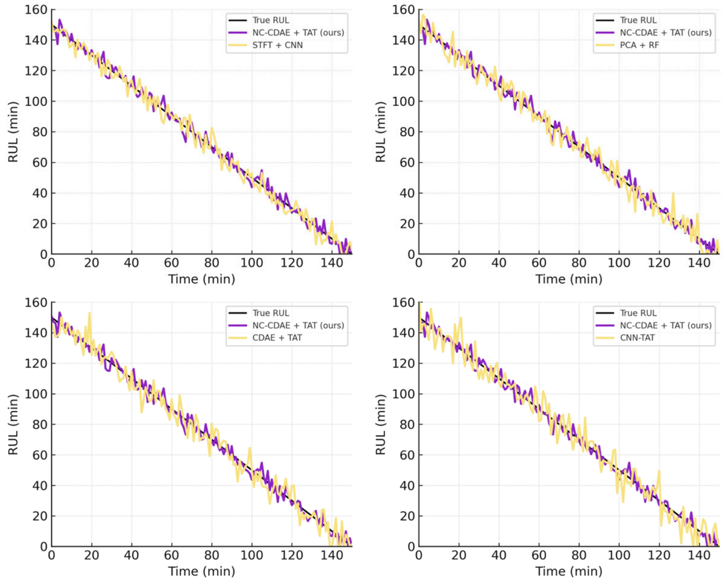

Figure 8. A comparison of the predicted RUL trajectories between the proposed method and baseline models under representative noise conditions.

In a further initial assessment under small sample conditions (with 20–25% of the initial training data only) all the baseline models showed a notable degradation. As an illustration, the PCA + RF and STFT + CNN models recorded a mean improvement in the RMSE of 18.7% and 14.3%, respectively, whereas the GRU + Attention model experienced erratic convergence with occasional prediction drift during the late stage of degradation. In comparison, the suggested NC-CDAE + TAT model had a relatively small RMSE increment of 6.1% and the predicted RUL curves of the model were smooth and trend-following. Small-sample prediction is not a major scope of this work; however, these findings suggest that the proposed framework still holds rather good stability even in conditions of data scarcity.

5.5. Discussion

The experimental results demonstrate that the suggested NC-CDAE + TAT framework attains consistent and precise RUL prediction in various cases of noise. It was revealed that the cGAN-based augmentation is necessary; more similar configurations tended to improve downstream values, which proves that the quality of the synthetic data is a direct condition of the robustness of the model. The experiments concerning ablation also showed that noise conditioning and temporal attention are inseparable, as the removal of each of the modules resulted in a considerable decrease in performance and the interchange of the self-attention module with the GRU produced the lowest results. These results explain why adaptability is the result of two complementary functions: adaptive denoising and long-range temporal modelling.

The trade-off between accuracy and efficiency was brought out by the parameterised ablation of NC-CDAE depth and the number of heads in the TAT. Though more profound networks and broader attention enhanced the accuracy of prediction, there was a decrease in returns after moderate sizes and the (4, 8) configuration was found to be the most balanced. Lastly, these findings were supported by cross-model comparisons: PCA + RF, STFT + CNN, and GRU + Attention performed competitively in some limited conditions, but varied widely in their performance when exposed to complex noise. On the contrary, the NC-CDAE + TAT offered fewer and more consistent errors, making its robustness aspect its most important advantage.

Besides the accuracy of prediction, computational complexity and real-time interruptiveness are also important factors for deployment in industry. According to Table 3, the proposed NC-CDAE + TAT framework is of moderate model size and inference latency with a trade-off between performance and efficiency. The framework can respond to the needs of standard industrial condition monitoring systems, where diagnostics based on vibrations are carried out at a fixed rate, as the framework requires an average of milliseconds to make an inference. The architecture proposed is simple, with a limited depth and width of attention, as compared to more complex and deeper attention-based models, and yet the model is capable of noise-robust RUL prediction. All these attributes indicate that the proposed approach is effective not only in terms of accuracy but also expedient in the context of real-time or near-real-time predictive maintenance services.

Beyond the quantitative results, this study also provides several conceptual insights. By explicitly conditioning the denoising process on noise characteristics, the proposed NC-CDAE framework demonstrates that noise can be treated as informative contextual input rather than merely a nuisance to be suppressed. This perspective enables the model to adapt its feature representations across different noise regimes, contributing to more robust RUL prediction. Moreover, the conditioning mechanism can be naturally extended beyond noise to incorporate other operational factors, such as load or temperature, offering a flexible pathway for further improving model adaptability. The ablation analysis on network depth and attention heads further suggests that practical deployment does not always require maximum model complexity, but rather an appropriate balance between accuracy, robustness, and computational efficiency. Together, these findings highlight the potential of noise-aware and context-conditioned architectures as a promising direction for future prognostic model design.

6. Conclusions

A noise-robust bearing RUL prediction framework was proposed in this paper that combines an NC-CDAE with a TAT. The paramount input of the suggested solution is that noise is explicitly modelled as a condition-dependent variable as opposed to its emergence as an implicit nuisance parameter. The framework, by conditioning the denoising process on perceived noise properties, as well as pairing it with an attention-based temporal model, overcomes a long-standing problem in the field of data-driven prognostics, which is the desirable balance between prediction accuracy and resilience to non-stationary and heterogeneous noise.

Experiments carried out on datasets with the addition of cGAN-generated multi-noise signals show that the proposed NC-CDAE + TAT framework outperforms representative traditional machine learning, signal processing, and deep learning baselines. The findings indicate that they not only had reduced average prediction errors, but also that their performance variability over various noise regimes is significantly lower. Ablation experiments further substantiate the need for the two components: noise conditioning increases adaptability to the different noise intensities, whereas temporal attention increases the modelling of the long-range dependencies of degradation, especially in the later stages of life of the bearing when degradation dependencies become more significant.

In addition to its empirical performance, the suggested framework also provides a number of conceptual insights. To begin with, the noise-conditioning process enables noise characteristics to be explicitly incorporated as contextual information, allowing the model to adapt its denoising behaviour dynamically across different noise conditions. This view creates an opportunity to transfer the conditioning mechanism beyond noise to other factors of operation, like load, speed, or temperature. Second, the temporal attention mechanism offers a flexible means of capturing short-term fault-related transients and long-term degradation trends, which is in line with the inherent nature of RUL prediction as a time-series regression problem.

Although these are the benefits, this study has a few limitations that indicate how future research can be conducted. Vibration signals are considered the central concern of the current framework and the inclusion of multi-sensor data (temperature, acoustic emission, or electrical signals) might provide additional benefits for the strength of prediction. Also, the noise diversity of the cGAN-based synthetic data generation could be enhanced; however, the realism of the simulated noise pattern may be further confirmed through field data from industrial settings. Modelling-wise, future work might consider lightweight or scarified attention mechanisms to make them more computationally efficient to deploy in real-time. In addition, engineers can further use the noise-conditioned design to jointly model various operating conditions, which could offer a more holistic solution in prognostics when the industrial environment is highly variable.

In general, the proposed NC-CDAE + TAT framework has methodological and practical implications for noise-robust bearing RUL prediction. This work presents a principled and generalizable noise-sensitive denoising model with a TAT by clearly combining the former with the latter to make predictions in the context of realistic and challenging operating conditions in predictive maintenance.

Author Contributions: Z.J. contributed to the conceptualization, methodology design, and writing of the manuscript. C.C. and A.S. provided validation and critical revisions. Y.L. supervised the project, contributed to writing, and provided project resources. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study are publicly available and can be accessed from the XJTU-SY rolling element bearing dataset and the PRONOSTIA bearing dataset.

Acknowledgments: The authors would like to acknowledge the publicly available datasets used in this study and the support of Cardiff University for providing the necessary infrastructure.

Conflicts of Interest: The authors declare that they have no conflicts of interest or competing interests.

References

1. Chen, Y.; Peng, G.; Zhu, Z.; Li, S. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. *Appl. Soft Comput.* **2020**, *86*, 105919. [CrossRef]
2. Ji, P.; Yu, T.; Zhang, Y.; Gong, W.; Yu, Q.; Li, L. Deep learning prediction of amplitude death. *Auton. Intell. Syst.* **2022**, *2*, 26–36. [CrossRef]
3. Chen, C.; Wu, D.; Liu, Y. Recent advances of AI for engineering service and maintenance. *Auton. Intell. Syst.* **2022**, *2*, 38. [CrossRef]
4. Fu, H.; Liu, Y. A deep learning-based approach for electrical equipment remaining useful life prediction. *Auton. Intell. Syst.* **2022**, *2*, 34. [CrossRef]
5. Gu, Y.; Cao, J.; Song, X.; Yao, J. A denoising autoencoder-based bearing fault diagnosis system for time-domain vibration signals. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 9790053. [CrossRef]
6. Alonso, S.; García-Benito, J.; Martínez, M.; Rodríguez, P.; Pérez, R. Gap imputation in related multivariate time series through recurrent neural network-based denoising autoencoder. *Int. Comput. Aided Eng.* **2024**, *31*, 157–172. [CrossRef]
7. Yao, X.; Liu, Y.; Zhang, Q.; Li, H.; Chen, J. RUL prediction method for rolling bearing using convolutional denoising autoencoder and bidirectional LSTM. *Meas. Sci. Technol.* **2024**, *35*, 035111. [CrossRef]
8. Im, D.-I.; Kim, D.; Lee, S.; Park, J. Denoising criterion for variational auto-encoding framework. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Volume 31, pp. 10777–10783. [CrossRef]
9. Xu, T.; Li, H.; Song, Z.; Zhao, F.; Wang, J.; Liu, Y. Multi-resolution LSTM-based prediction model for remaining useful life of aero-engine. *IEEE Trans. Veh. Technol.* **2024**, *73*, 1931–1941. [CrossRef]
10. Xu, M.; Zhang, Q.; Li, Y.; Chen, H.; Huang, Z. STDD-net: A spatio-temporal degradation disentanglement network for RUL prediction of rolling bearings. *Meas. Sci. Technol.* **2025**, *36*, 106216. [CrossRef]

11. Han, Y.; Zhao, L.; Sun, Q.; Wang, T.; Li, D.; Zhou, J. Remaining useful life prediction of lithium-ion batteries by using a denoising transformer-based neural network. *Energies* **2023**, *16*, 6328. [CrossRef]
12. Zhou, K.; Zhang, Z. Remaining useful life prediction of lithium-ion batteries based on data denoising and improved transformer. *J. Energy Storage* **2024**, *100*, 113749. [CrossRef]
13. Xia, T.; Zhang, Y.; Liu, J.; Wang, H.; Chen, Z. An accurate denoising lithium-ion battery remaining useful life prediction model based on CNN and LSTM with self-attention. *Ionics* **2023**, *29*, 5315–5328. [CrossRef]
14. Fu, S.; Li, X.; Wang, Y.; Zhang, H.; Liu, Z. A bearing fault diagnosis method based on wavelet denoising and machine learning. *Appl. Sci.* **2023**, *13*, 5936. [CrossRef]
15. Liu, L.; Song, X.; Chen, K.; Hou, B.; Chai, X.; Ning, H. An enhanced encoder–decoder framework for bearing remaining useful life prediction. *Measurement* **2021**, *170*, 108753. [CrossRef]
16. Yuliani, A.R.; Pardede, H.F.; Ramdan, A.; Zilvan, V.; Yuwana, R.S.; Amri, M.F.; Kusumo, R.B.S.; Pramanik, S. Robust remaining useful life prediction of lithium-ion battery with convolutional denoising autoencoder. *J. Mechatron. Electr. Power Veh. Technol.* **2024**, *15*, 93–104. [CrossRef]
17. Yang, L.; Xu, C.; Liao, Y.; Duan, R. Stack denoising autoencoder and state-space model based bearing RUL prediction method. In Proceedings of the 2022 International Conference on Sensing, Measurement & Data Analytics in the Era of Artificial Intelligence (ICSMD), Harbin, China, 30 November–2 December 2022; IEEE: New York, NY, USA, 2022; pp. 1–6. [CrossRef]
18. Chen, Y.; Zeng, X.; Huang, H. Fault diagnosis of rolling bearings based on adaptive denoising residual network. *Processes* **2025**, *13*, 151. [CrossRef]
19. Shang, Y.; Lin, T.-R.; Luo, H. A degradation trend analysis of rolling element bearings using an improved fuzzy C-mean clustering and a recurrent denoising auto-encoder. *Measurement* **2025**, *256*, 118285. [CrossRef]
20. Lekkas, G.; Vrochidou, E.; Papakostas, G.A. Time–frequency transformations for enhanced biomedical signal classification with convolutional neural networks. *BioMedInformatics* **2025**, *5*, 7. [CrossRef]
21. Chegini, S.N.; Bagheri, A.; Najafi, F. Application of a new EWT-based denoising technique in bearing fault diagnosis. *Measurement* **2019**, *144*, 275–297. [CrossRef]
22. Mishra, C.; Samantaray, A.K.; Chakraborty, G. Rolling element bearing fault diagnosis under slow speed operation using wavelet de-noising. *Measurement* **2017**, *103*, 77–86. [CrossRef]
23. Huang, P.; Wang, Y.; Gu, Y.; Qiu, G. A bearing RUL prediction approach of vibration fault signal denoise modeling with Gate-CNN and Conv-transformer encoder. *Meas. Sci. Technol.* **2024**, *35*, 066104. [CrossRef]
24. Spirto, M.; Melluso, F.; Nicolella, A.; Malfi, P.; Cosenza, C.; Savino, S.; Niola, V. A comparative study between SDP-CNN and time–frequency CNN-based approaches for fault detection. *J. Dyn. Monit. Diagn.* **2025**. [CrossRef]
25. Chen, J.; Li, T.; He, J.; Liu, T. An interpretable wavelet Kolmogorov–Arnold convolutional LSTM for spatio-temporal feature extraction and intelligent fault diagnosis. *J. Dyn. Monit. Diagn.* **2025**, *4*, 183–193. [CrossRef]
26. Guo, F.; Fu, W.; Qin, J.; Ma, Q.; Kang, Y. Remaining useful life prediction for bearings using multiscale temporal-enhanced deformable attention with soft thresholding. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3554112. [CrossRef]
27. Wang, Y.; Wang, Y. A denoising semi-supervised deep learning model for remaining useful life prediction of turbofan engine degradation. *Appl. Intell.* **2023**, *53*, 22682–22699. [CrossRef]
28. Yang, H.; Wang, P.; An, Y.; Shi, C.; Sun, X.; Wang, K.; Zhang, X.; Wei, T.; Ma, Y. Remaining useful life prediction based on denoising technique and deep neural network for lithium-ion capacitors. *eTransportation* **2020**, *5*, 100078. [CrossRef]
29. Berghout, T.; Mouss, L.-H.; Kadri, O.; Saïdi, L.; Benbouzid, M. Aircraft engines remaining useful life prediction with an adaptive denoising online sequential extreme learning machine. *Eng. Appl. Artif. Intell.* **2020**, *96*, 103936. [CrossRef]
30. Hui, J.; Huang, J.; Zhang, F.; Zhang, Y.; Tian, Y. Research on remaining life prediction of machine tool spindle bearing combining neural network and weighted average de-noising method. In Proceedings of the International Symposium on Robotics, Artificial Intelligence, and Information Engineering (RAIIE 2022), Hohhot, China, 15–17 July 2022; SPIE: Bellingham, WA, USA, 2022; Volume 12454, p. 124542L. [CrossRef]
31. Zhao, B.; Yuan, Q. A novel deep learning scheme for multi-condition remaining useful life prediction of rolling element bearings. *J. Manuf. Syst.* **2021**, *61*, 450–460. [CrossRef]
32. Wang, H.; Yang, J.; Wang, R.; Shi, L. Remaining useful life prediction of bearings based on convolution attention mechanism and temporal convolution network. *IEEE Access* **2023**, *11*, 24407–24419. [CrossRef]
33. Wang, Q.; Xu, F. A novel rolling bearing fault diagnosis method based on adaptive denoising convolutional neural network under noise background. *Measurement* **2023**, *218*, 113209. [CrossRef]
34. Tayade, A.; Patil, S.; Phalle, V.; Kazi, F.; Powar, S. Remaining useful life (RUL) prediction of bearing by using regression model and principal component analysis (PCA) technique. *Vibroeng. Procedia* **2019**, *23*, 30–36. [CrossRef]
35. Kamat, P.V.; Sugandhi, R.; Kumar, S. Deep learning-based anomaly-onset aware remaining useful life estimation of bearings. *PeerJ Comput. Sci.* **2021**, *7*, e795. [CrossRef] [PubMed]

36. Yang, F.; Li, X.; Wang, M.; Zang, H.; Pang, W.; Wang, M. WaveForM: Graph enhanced wavelet learning for long sequence forecasting of multivariate time series. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 10754–10761. [CrossRef]
37. Li, Z.; Rao, Z.; Pan, L.; Xu, Z. MTS-Mixers: Multivariate time series forecasting via factorized temporal and channel mixing. *arXiv* **2023**, arXiv:2302.04501. [CrossRef]
38. Siddique, M.F.; Ahmad, Z.; Ullah, N.; Kim, J. A hybrid deep learning approach: Integrating short-time Fourier transform and continuous wavelet transform for improved pipeline leak detection. *Sensors* **2023**, *23*, 8079. [CrossRef] [PubMed]
39. Zhu, J.; Chen, N.; Peng, W. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3208–3216. [CrossRef]
40. Hu, J.; Wen, Q.; Ruan, S.; Liu, L.; Liang, Y. Twins: Revisiting non-stationarity in multivariate time series forecasting. *arXiv* **2024**, arXiv:2406.03710. [CrossRef]
41. Mo, Y.; Wu, Q.; Li, X.; Huang, B. Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit. *J. Intell. Manuf.* **2021**, *32*, 1997–2006. [CrossRef]
42. Huang, X.; Zhang, P.; Shi, W.; Dong, S.; Wen, G.; Lin, H.; Chen, X. Frequency Hoyer attention based convolutional neural network for remaining useful life prediction of machinery. *Meas. Sci. Technol.* **2021**, *32*, 125108. [CrossRef]
43. Zeng, F.; Li, Y.; Jiang, Y.; Song, G. A deep attention residual neural network-based remaining useful life prediction of machinery. *Measurement* **2021**, *181*, 109642. [CrossRef]
44. Jiang, M.; Wang, K.; Sun, Y.; Chen, W.; Xia, B.; Li, R. MLGN: Multi-scale local-global feature learning network for long-term series forecasting. *Mach. Learn. Sci. Technol.* **2023**, *4*, 045059. [CrossRef]
45. Liu, Z.; Zhang, L.; Xiao, Q.; Huang, H.; Xiong, G. Performance degradation assessment of railway axle box bearing based on combination of denoising features and time series information. *Sensors* **2023**, *23*, 5910. [CrossRef]
46. El Emam, K.; Hoptroff, R. *The Synthetic Data Paradigm for Using and Sharing Data*; Cutter Consortium: Arlington, MA, USA, 2019. Available online: https://www.cutter.com/article/synthetic-data-paradigm-using-and-sharing-data-503526 (accessed on 5 January 2026).
47. Fonseca, J.; Bação, F. Tabular and latent space synthetic data generation: A literature review. *J. Big Data* **2023**, *10*, 92. [CrossRef]
48. Lambay, A.; Liu, Y.; Morgan, P.; Ji, Z. Generating synthetic datasets with deep learning models for human physical fatigue analysis. *Machines* **2025**, *13*, 235. [CrossRef]
49. Wang, B.; Lei, Y.; Li, N.; Guo, L. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Trans. Reliab.* **2018**, *67*, 401–412. [CrossRef]
50. Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Morello, B.; Zerhouni, N.; Varnier, C. PRONOSTIA: An experimental platform for bearings accelerated life test. In Proceedings of the IEEE International Conference on Prognostics and Health Management (PHM), Denver, CO, USA, 18–21 June 2012; IEEE: Denver, CO, USA, 2012; pp. 1–8.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.