

Received 10 January 2025, accepted 7 March 2025, date of publication 17 March 2025, date of current version 25 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3551772

 SURVEY

Bearing Prognostics Using the PRONOSTIA Data: A Comparative Study

HARIOM DHUNGANA^{ID}, (Member, IEEE), THORSTEIN RYKKJE^{ID}, (Member, IEEE),
AND ALEXANDER S. LUNDERVOLD^{ID}

Western Norway University of Applied Sciences, 5063 Bergen, Norway

Corresponding author: Hariom Dhungana (hdhu@hvl.no)

ABSTRACT Vibration-based condition monitoring is a well-developed and established field in maintenance engineering. Failure prognosis is a crucial aspect of condition monitoring, particularly Remaining Useful Life (RUL) prediction. Over the last decade, the PRONOSTIA bearing dataset has become the standard reference for testing these prognostic algorithms. However, the lack of standardized comparisons makes it difficult to objectively assess the relative performance of different methods. This paper systematically compares three established data-driven artificial intelligence approaches: classical machine learning, deep learning, and transfer learning approaches. We analyze their application to the PRONOSTIA dataset, providing detailed discussions of their relative strengths, limitations, and achieved performance. While this study primarily serves as a benchmark for testing new prognostic algorithms in vibration monitoring, we hope the insights will also broadly apply to other condition-monitoring techniques.

INDEX TERMS Condition monitoring, data-driven approach, RUL prediction, vibration data, machine learning, deep learning, transfer learning.

I. INTRODUCTION

Rolling element ball bearings, mechanical devices consisting of spherical rolling elements held between circular inner and outer rings, are among the most common components found in rotating machinery [1]. They help support rotating shafts and reduce friction between these shafts and stationary machine parts. The bearings are commonly used in machines where the shafts need to rotate smoothly with minimal friction, and their failure ranks as one of the leading causes of machine breakdowns. More than half of rotary machine system faults are due to the bearings, which also commonly cause failures in equipment like high-voltage induction motors, wind turbine gearboxes, and induction motors, as discussed in [2]. Therefore, it is crucial to develop effective methods to evaluate the current health status of bearings and predict their Remaining Useful Life (RUL). RUL, defined as the time until the end of useful life, is a critical metric in Prognostics and Health Management (PHM). Accurate RUL

predictions guide maintenance decisions, reducing downtime and costs while improving reliability.

Estimating the RUL is essential for minimizing risks and enhancing reliability in condition monitoring. Risk and reliability are closely related in this context: lower risk leads to higher reliability, and higher risk results in lower reliability [3]. By accurately predicting the RUL, risks can be effectively managed and mitigated, thereby improving the overall reliability of the monitored equipment. RUL prediction was used to assess bearing performance during condition management. The bearing's operational lifespan in industrial systems depends on rotation speed, applied load, and average life expectancy. The bathtub curve is crucial in reliability practice, illustrating how bearing failure rates change over time [4]. Like other industrial equipment, bearing performance follows this curve, which includes three phases: burn-in period (infant mortality), useful life (random failure), and wear-out period (wear-out failure) [5]. During the burn-in period, failure rates decrease after initial use. In the useful life phase, failure rates increase due to random factors affecting performance. In the wear-out phase, failure rates rise rapidly as the bearing exceeds its design lifespan.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiajie Fan^{ID}.

Vibration-based condition monitoring has become a well-established field with robust prognostic techniques. Despite its maturity, researchers continue to explore ways to enhance these methods, aiming for improved accuracy and reliability. This ongoing advancement is significantly supported by publicly available reference datasets, which provide a valuable resource for evaluating and validating newly developed algorithms. The PRONOSTIA-bearing or FEMTO-bearing dataset, released over a decade ago, has become a standard benchmark for evaluating prognostic algorithms [6]. This dataset provides accelerated degradation data, allowing researchers to test their methods on realistic failure scenarios.

Bearing degradation is a complex, highly nonlinear process influenced by many factors, such as structural, operational, and environmental attributes. Thus, predicting the RUL of bearings is a crucial and challenging issue in condition management systems. Accurate RUL prediction allows for economical operation by avoiding unnecessary maintenance and replacement and minimizing downtime for critical systems. This review provides a timely and comprehensive examination of the latest trends in bearing RUL prediction, focusing on data-driven models and evaluating them based on error and accuracy scores using PRONOSTIA data.

A. RECENT ADVANCES AND CHALLENGES

Recent studies highlight advances in bearing prognostics using machine learning and hybrid methods. Wei et al. [7] proposed an attention-aware graph convolutional neural network that achieved superior performance by capturing both temporal and spatial dependencies in vibration signals. Zhuang et al. [8] introduced a temporal partial domain adaption framework that effectively handles varying operational conditions. These methods highlight a shift toward architectures capable of addressing real-world complexities. Multi-source adversarial distillation domain adaptation network captures source-specific representation showed more precise prediction results by avoiding the limitation of direct single-domain adaptation [9]. He et al. [10] proposed a data augmentation method that enhances time-frequency features using dynamic attention for RUL prediction.

Most research emphasizes accuracy, computational resources, and time for online RUL prediction are equally important [11]. CNN-based models have longer training times due to feature extraction compared to RNNs [12]. The FELM model excels in RUL prediction accuracy for small sample environments, while the LSTM-UQ outperforms traditional LSTM and CNN models with comparable training times. The FFCGRU model further reduces training and prediction times, making it suitable for real-time applications [13].

Despite its widespread use, a comprehensive and standardized comparison of different prognostic approaches using PRONOSTIA is missing, and several critical challenges in bearing prognostics remain unaddressed:

- 1) The lack of standardized evaluation frameworks makes it difficult to compare different prognostic approaches objectively
- 2) The impact of operational conditions on prediction accuracy is not fully understood
- 3) The integration of domain knowledge with AI-based approaches remains limited
- 4) The practical implementation of prognostic systems faces challenges in real-world industrial settings

B. PROGNOSTIC BASICS

Failure prognostics is a crucial aspect of condition-based maintenance (CBM), focusing on predicting the RUL of bearings using condition observation data. RUL, defined as the estimated time until a bearing reaches the end of its functional lifespan, is a pivotal metric in PHM. An equipment failure prognostic program typically involves four technical processes: data acquisition, constructing health indicators (HI), dividing health stages (HS), and predicting RUL [14]. By following ISO 13381 for machine health prognostics and RUL prediction, organizations can enhance their maintenance strategies, reduce downtime, and extend bearing life [15]. Additionally, beginners in research can gain valuable insights into the prognostics process, historical data requirements, and various RUL estimation techniques. The standard also provides guidelines for performance evaluation and implementation. However, while these rules offer a solid foundation, they should not constrain innovation and creative thinking in the field.

There is no standard method for categorizing RUL prediction approaches, leading to various classifications and naming conventions. This lack of uniformity can confuse readers trying to understand these different approaches. We follow the meanings and coverage of four different categories: physics-based approaches, statistics-based approaches, AI-based approaches, and hybrid approaches as in [14]. The scope of these prognostic techniques, including datasets focusing on data-driven AI approaches, is shown in Figure 1. The upper section highlights widely used publicly available datasets for RUL prediction, such as XJTU, IMS, and others. Emerging studies leverage multi-source domain transfer learning to demonstrate superior prediction accuracy by capturing greater variability and minimizing domain-specific biases.

Physics-based models describe bearing degradation through mathematical representations of failure mechanisms, linking parameters such as material properties and stress levels. These parameters are identified through experiments, finite element analysis, or other techniques. However, physics-based approaches are rarely used to predict bearing RUL due to the complexity of failure mechanisms under varying and severe conditions.

Statistics-based approaches use conditional probability distributions to estimate RUL based on empirical data, independent of physics-based assumptions [16]. They effectively

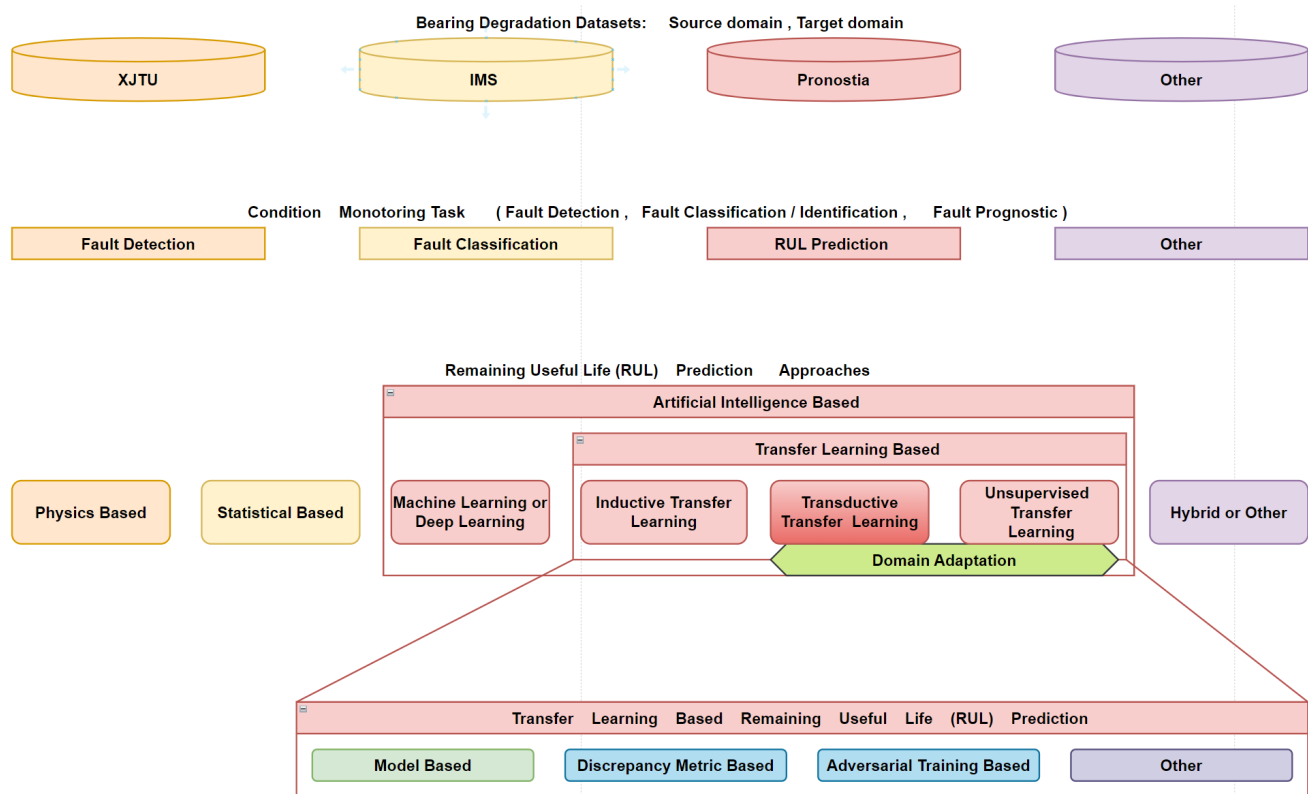


FIGURE 1. Various types of RUL prediction approaches and the most commonly used open-source dataset for bearing degradation.

describe the uncertainty of the degradation process and its impact on RUL prediction. Some of the works using statistical approaches for bearing RUL prediction are shown in Table 1. Moreover, these approaches require deep knowledge and experience but still carry uncertainty.

Data-driven AI approaches learn bearing degradation patterns directly from observational data, using machine learning techniques rather than relying on physics-based or statistical models. These methods can predict the RUL by mapping condition monitoring data to RUL value without needing expert knowledge or prior experience [24]. They require minimal historical data, are less complex and expensive, and offer a balanced trade-off between complexity, cost, precision, and applicability. They are increasingly popular for complex machinery prognostics, with the most publications on bearing RUL estimation in the past decade. Furthermore, AI approaches can be divided into three subcategories: traditional machine learning-based, deep learning-based, and transfer learning-based. This classification will help in understanding and applying different AI methods more effectively.

Each category has its own limitations in RUL prediction. A hybrid approach aims to combine the strengths of different methods to address these limitations [25]. A physics-based RUL prediction approach was combined with a data-driven AI approach to improve accuracy in [26]. The status degradation function was used to describe the bearing degradation

process, while a convolutional autoencoder deep learning method was employed for data denoising, dimensionality reduction, and feature extraction in RUL prediction.

C. MOTIVATION AND OUR CONTRIBUTION

Failure analysis is crucial for ensuring the safe operation of bearings, with fault mode identification and RUL prediction as key components. By identifying failure modes and causes, we can propose effective measures to enhance bearing service life and prevent potential accidents. A comprehensive overview of the different bearing parts, their types, probable damage locations, and causes of failure is presented in [27]. Incipient faults in one component can easily spread to other components through frequent contact, leading to the simultaneous occurrence of various fault patterns. If RUL is accurately predicted following the initial fault detection [28], sufficient time is provided to ensure safety, enhance reliability, and develop an effective maintenance plan. Moreover, safety is paramount in modern industrial processes as bearing degradation and failure pose significant hazards. This has driven the adoption of real-time monitoring and RUL prediction to prevent sudden failures and safeguard life and property [29].

Condition monitoring (CM), decision-making, and RUL prediction are closely interconnected in asset management. Condition monitoring involves collecting real-time data on

TABLE 1. Summary of statistical prognostic technique.

	Basics	Applications
ARIMA [17]	It combines autoregressive (AR) terms, differencing (I), and moving average (MA) terms to model temporal dependencies	Used to forecast future RUL values by modeling the historical time series data of system degradation
Random Coefficient	Coefficients in regression equations are allowed to vary randomly across different entities between different instances of the same process.	Can handle variability in degradation patterns among different units of the same system
Wiener Process (Brownian Motion) [18], [19]	It is a continuous-time stochastic process with independent increments and normally distributed increments	Can model the stochastic nature of failure processes over time
Gamma Process [20]	It is a type of stochastic process with gamma-distributed increments and is used in reliability engineering to model the degradation.	Can model the gradual deterioration of system performance and predict the RUL based on observed degradation.
Gaussian Process [21]	It is a non-parametric approach to regression that models distributions over functions	Can provide probabilistic forecasts of RUL by modeling the underlying distribution of degradation paths.
Markov Model [22]	It uses state transitions to model systems that change over time with probabilities depending only on the current state, not past states.	First models the system's state transitions and estimates the time until the system transitions to a failure state.
Kalman Filter [23]	It is an algorithm that uses a series of measurements observed over time to estimate unknown variables, with applications in linear dynamic systems	Can be used to estimate the state of a system and predict RUL by updating predictions as new measurements become available
Proportional Hazard Model	They are used in survival analysis to examine the effect of various factors on the time until an event occurs.	Can analyze the impact of covariates on the survival time and predict the RUL based on these factors

equipment health to identify potential issues [30]. This data informs decision-making, enabling timely actions to prevent failures. These decisions then guide maintenance planning, optimizing schedules, resources, and interventions to ensure reliability and cost-efficiency [31]. Effective decision-making relies on comprehensive insights derived from fault detection, identification, quantification, and prognosis to ensure accurate assessment and timely action [32]. Prognosis or forecasting, regardless of the term used, predicts future outcomes based on current data and trends [33]. Forecasting supports planning in industrial maintenance by predicting future equipment degradation conditions, resource needs, and failures, enabling proactive interventions.

To address the challenges outlined above, this study systematically compares three prominent data-driven AI approaches for bearing RUL prediction: traditional machine learning, deep learning, and transfer learning, specifically focusing on the PRONOSTIA dataset. Our key contributions are as follows:

- 1) We establish a standardized framework for evaluating and comparing the performance of different prognostic algorithms on the PRONOSTIA dataset, addressing the current lack of such a framework in the literature. This framework considers various metrics, including prediction accuracy (using error and score as defined by the IEEE PHM 2012 Prognostic Challenge), computational cost, and model complexity.
- 2) We conduct a detailed analysis of each AI approach (traditional machine learning, deep learning, and transfer learning) as applied to the PRONOSTIA dataset. This includes examining the impact of different feature engineering techniques (time-domain, frequency-domain, time-frequency domain), the role of signal denoising, and the influence of various model architectures and hyperparameters on prediction accuracy. We analyze performance across all available operating conditions present in the PRONOSTIA data.
- 3) Based on our findings, we provide practical guidance for selecting the most appropriate AI approach depending on specific application requirements, such as desired prediction accuracy, available computational resources, and the amount of labeled training data. We highlight the strengths and limitations of each method in the context of the PRONOSTIA dataset. Furthermore, we offer insights into the challenges of determining the optimal RUL prediction time and discuss the trade-offs between early and late prediction.
- 4) By thoroughly analyzing a wide range of AI-based prognostic models and identifying the current state-of-the-art performance on the PRONOSTIA dataset (to the best of our knowledge), this study establishes a clear benchmark for future research. We identify limitations of existing methods and suggest promising directions for further advancements in bearing RUL prediction, particularly in the areas of transfer learning across different operating conditions and bearings, model interpretability, and industrial implementation.

This review aims to be a valuable resource for researchers and practitioners working on bearing RUL prediction. By offering a structured and comprehensive overview of existing techniques, along with a comparative analysis using the PRONOSTIA dataset, we hope to facilitate further progress in this critical area. Table 3 highlights the gaps addressed by our work compared to other published reviews. We focus on accuracy, model interpretability, and computational efficiency. We also address the challenge of determining the optimal RUL prediction time, balancing the need for early warning with the desire for accurate forecasts. Finally, we present the latest optimal accuracy achieved on the PRONOSTIA dataset, establishing a baseline for future improvements and suggesting directions to surpass this current benchmark.

We conducted a thorough review using the Scopus database to ensure a comprehensive and relevant literature search.

TABLE 2. Search Strings and Research Articles Identified in Scopus.

Search string	Journal papers	Conferences	Total	Similarity	Selected
RUL and Prediction	655	308	963	563	50
RUL and ML	98	75	173	57	15
RUL and DL	206	77	283	75	27
RUL and TL	39	10	49	38	25
RUL and PRONOSTIA	567	141	708	343	34

TABLE 3. Comparison of the existing literature and our review. * indicates topics covered in this work.

	Datasets	Feature Engineering	Traditional Machine Learning	Deep Learning	Transfer Learning	Comparative study	Perspective
[36]		*	*				*
[35]				*	*		
[37]							*
[27]					*		
[38]		*	*				
[34]		*				*	
[14]		*	*	*			
[39]	*			*	*		
[40]					*		*
This review	*	*	*	*	*	*	*

We employed various search strings, detailed in Table 2, to identify pertinent articles published after the release of the PRONOSTIA dataset in 2012. The authors note the inclusion of 1176 papers using the PRONOSTIA dataset, published in Mechanical Systems and Signal Processing between 2012 and early 2024.

A benchmark study for bearing diagnostics is already available [34]; however, a comparative study on RUL prediction remains currently unavailable. A comprehensive evaluation of deep learning-based fault diagnosis models was provided as a benchmark study in [35]. No comprehensive study has explored data-driven approaches for RUL prediction in bearings. To our knowledge, this is the first comprehensive literature review that thoroughly explains various AI-based RUL prediction techniques for a common benchmark. Table 3 compares existing literature reviews, highlighting the unique aspects of our work, while Table 2 details the search strings used and the number of articles identified.

II. DATASETS AND EVALUATION METRIC

To evaluate and benchmark RUL prediction methods, numerous studies have employed publicly available datasets. Examples include open-source datasets like IMS bearing [41] and XJTU-SY [42], as well as closed-source datasets [43]. Among these, the PRONOSTIA accelerated degradation dataset has become a widely recognized benchmark for prognostic research. It provides vibration and temperature signals collected across the full lifespan of bearings during accelerated degradation experiments.

The PRONOSTIA dataset, developed by the FEMTO-ST Institute, captures bearing degradation under controlled conditions without pre-seeded faults. Figure 2 illustrates the

experimental platform used to collect this data. Introduced during the IEEE International Conference on PHM 2012, the dataset has since been extensively used to evaluate prognostic algorithms [6].

The PRONOSTIA experiments aimed to simulate accelerated degradation within a short timeframe by applying a radial force exceeding the bearings’ maximum dynamic load capacity. The rotational speed was maintained constant during each test, while vibration and temperature signals were continuously recorded using two DYTRAN 3035B accelerometers and a platinum RTD PT100 PROSENSOR thermocouple. The degradation process concluded when the vibration signal amplitude exceeded 20g, marking the bearing’s end of life.

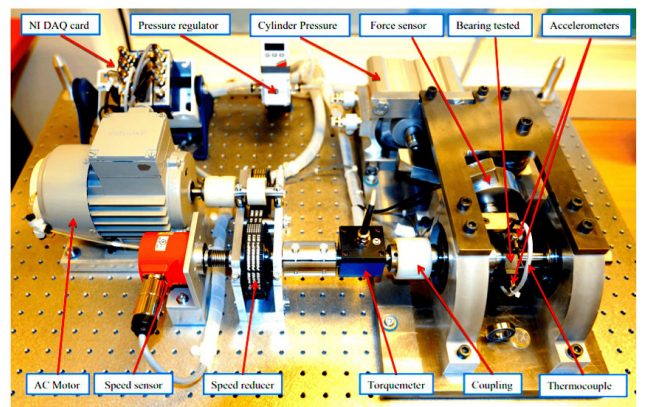


FIGURE 2. Overview of the PRONOSTIA experimental platform.

A total of 17 experiments were conducted under three distinct operating conditions, detailed in Table 4. Each condition varies in terms of load and rotational speed, enabling analysis across different stress scenarios.

TABLE 4. Operating condition of various experiments.

	Condition 1	Condition 2	Condition 3
Load (Newton) / Speed (RPM)	4000 / 1800	4200 / 1650	5000 / 1500
Training sets	Bearing1_1 Bearing1_2	Bearing2_1 Bearing2_2	Bearing3_1 Bearing3_2
Testing sets	Bearing1_3 Bearing1_4 Bearing1_5 Bearing1_6 Bearing1_7	Bearing2_3 Bearing2_4 Bearing2_5 Bearing2_6 Bearing2_7	Bearing3_3

The division of experiments into training and testing sets ensures that all operating conditions are represented in both phases. Specifically, six experiments were used for training, while the remaining eleven experiments were reserved for testing. This setup reflects real-world challenges, as extracting degradation patterns from training bearings and applying them to testing units is inherently complex.

While the PRONOSTIA dataset has proven valuable for RUL prediction, it has limitations. Only two failure causes—radial load and rotational speed—are considered,

excluding other potential factors like contamination or bearing currents, which are critical in industrial settings. Additionally, the dataset lacks intermediate inspections of rolling surfaces, limiting its ability to correlate physical degradation with signal features.

Evaluation of RUL prediction models require a variety of metrics. The IEEE PHM 2012 Prognostic Challenge used a scoring function to evaluate percentage errors, penalizing overestimation and underestimation differently. Figure 3 visualizes this scoring function. The average accuracy score is calculated as in Equation (1)-(3). Additional metrics are widely used, such as mean absolute error (MAE) [44], mean absolute percentage error (MAPE) [45], mean square error (MSE), root mean square error (RMSE) [46], relative accuracy (RA), cumulative relative accuracy (CRA) [47], and R^2 [8], [24]. These metrics are described mathematically in Equation (4)- (8): Where y_i is the actual RUL and $\hat{y}_i$ is the predicted value. MSE measures the average squared difference between estimated and actual values. It indicates the risk of prediction error and is always non-negative, with values closer to 0 being preferable. Lower MAE, RMSE, and higher R^2 reflect better estimator performance.

$$Er_i = \frac{y_i - \hat{y}_i}{y_i} \times 100. \quad (1)$$

$$A_i = \begin{cases} \exp(-\ln(0.5) \cdot (Er_i/5)), & \text{if } Er_i \leq 0 \\ \exp(+\ln(0.5) \cdot (Er_i/20)), & \text{if } Er_i > 0 \end{cases} \quad (2)$$

$$\text{Score} = \frac{1}{m} \sum_{i=1}^m A_i \quad (3)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \times 100 \quad (5)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

$$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

III. DATA DRIVEN PROGNOSTIC PROCESS

A data-driven prognostic process typically involves four technical steps: data acquisition and signal preprocessing, health indicator construction, detection of anomaly or RUL prediction time, and RUL prediction model. The following subsections describe each step in detail.

A. DATA ACQUISITION AND SIGNAL DENOISING

As previously mentioned, directly measuring vibration signals is challenging due to their intricate structures.

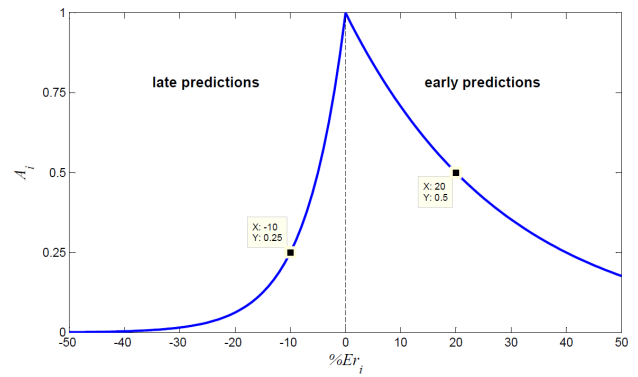


FIGURE 3. The scoring function evaluates RUL estimates based on their percentage error.

The accelerometer, typically mounted on the outer case, captures vibration signals from various sources, some of which introduce noise. During the run-to-failure period, the vibration data exhibits a non-stationary nature.

Data cleaning removes errors from time-series data, improving analysis accuracy. A separation-based algorithm for signal denoising effectively isolates the desired signal from noise by leveraging decomposition, noise estimation, thresholding, and reconstruction steps. Wavelet-based methods effectively enhance signal quality by isolating relevant features from noise by analyzing different frequency components [48]. Time Synchronous Averaging (TSA) is another signal denoising technique widely used in vibration analysis and fault diagnosis of bearings [29]. It enhances the signal by isolating its periodic components, identifying anomalies and specific fault types more easily.

The linear rectification technique denoises a signal by applying a linear transformation that suppresses noise while preserving the underlying signal structure, effectively smoothing spurious local fluctuations [49]. Moving averages smooth short-term fluctuations, highlighting trends for feature extraction [50]. EMD was effective for denoising vibration data in RUL prediction due to its ability to handle non-linear and non-stationary signals [51]. Autoencoder-decoder architectures were used for advanced signal denoising [49], [52]. The encoder extracts features from the noisy signal, and the decoder reconstructs a cleaner version, leveraging deep learning to effectively remove noise. LSTM-Autoencoder was used for reducing noise in raw vibration data in [53]. A two-step signal denoising method: first, the vibration signal is denoised using a local projective noise reduction method, then recurrence plots of the denoised signals are drawn to achieve a higher diagnosis rate was applied in [54]. Hou et al. [55] introduced a cross-transformer network paired with an unsupervised data-cleaning algorithm. A stack denoising autoencoder was used in [56]. While various signal denoising techniques are essential for improving the accuracy and reliability of RUL prediction, the choice of method depends on specific prediction requirements,

the availability of preprocessing models, and resource constraints.

B. HEALTH INDICATOR CONSTRUCTION

Observing bearing damage is challenging due to limited shutdown opportunities and the microscopic scale of early-stage faults, which require specialized instruments. Internal faults within rollers are also hard to identify without disassembly. As a result, constructing reliable HIs is essential for accurate RUL prediction. The construction of effective HIs is critical for bearing failure prognostics, as a well-designed HI simplifies prognostic modeling and improves prediction accuracy. Since HI construction involves fitting the degradation process, feature selection based on criteria such as trend, monotonicity, and robustness provides a more accurate reflection of bearing failure, thereby enhancing predictability. A suitable HI should exhibit monotonicity, accurately reflecting the irreversible degradation over the bearing's operational life [57]. Just like monotonicity, robustness is an inherent characteristic of a HI.

HI construction's fundamental concepts and theories were categorized into statistical parameter-based methods, signal preprocessing-based methods, and machine learning-based methods in [36]. Feature fusion combines multiple features to create a comprehensive health indicator. Various time, frequency, and time-frequency domain features were used for HI construction in [58]. A quadratic function-based deep convolutional auto-encoder was developed for HI construction in [59]. State space model was used to construct probabilistic entropy-based HI, which is less sensitive to high probabilistic distribution data and noise, making it effective under varying load and speed conditions [60]. While various HI construction techniques have pros and cons, feature extraction and selection are crucial for AI-based RUL prediction.

Raw vibration signals acquired from accelerometers contain extensive information regarding bearing degradation. However, these signals alone cannot directly depict the trend of bearing life degradation and often contain redundant data. Therefore, effective feature extraction across multiple domains is performed to derive comprehensive and accurate degradation features, enhancing the understanding of bearing degradation.

1) TIME DOMAIN FEATURES

Different mechanical faults, such as unbalance, misalignment, and bearing defects, often have unique waveforms. Time-domain analysis extracts characteristic features (like maximum, minimum, mean, peak, peak-to-peak interval, standard deviation, crest factor, root mean square, skewness, kurtosis, coefficient of variation, crest factor, clearance factor, waveform factor, pulse factor and so on) to capture dynamic signal changes effectively and are crucial for trend prediction. A detailed description, formulas, and the physical significance of time-domain features were presented in [57].

2) FREQUENCY DOMAIN FEATURES

While time-domain features are valuable for early fault detection, their inability to specify fault types or provide detailed insights into degradation processes hinders a deeper understanding of RUL prediction [61]. Frequency domain features reveal failure trends via spectral data and enhance analysis when combined with statistical methods. Spectrum analysis using fast Fourier transform (FFT) is the most common method. While the power spectrum is widely accepted, other spectra offer advantages in specific cases. The Hilbert transform is useful for envelope analysis, and the cepstrum can detect harmonics and sideband patterns in the power spectrum [62]. Frequency-domain analysis extracts characteristic features from the frequency spectrum, such as spectral peaks, frequency bands, harmonics, sidebands, spectral centroid, spectral variance, spectral skewness, spectral kurtosis, total harmonic distortion, frequency ratios, cepstrum, etc., enhancing the ability to detect, diagnose, forecast the failure. A detailed description, along with formulas and the physical significance of frequency-domain features, were presented in [57].

3) TIME-FREQUENCY DOMAIN FEATURES

A limitation of frequency features is their inability to handle non-stationary signals, common with bearing faults. To address this, time-frequency analysis, which examines signals in both time and frequency domains, offers a more comprehensive signal characterization. Different time-frequency domain features, including Short-Time Fourier Transform (STFT) [63], wavelet transform [48], Hilbert-Huang Transform (HHT) [51], spectrogram and S-transform (combines elements of both wavelet transform and STFT) are used in RUL prediction.

Empirical mode decomposition (EMD) can extract time-frequency features while staying in the time domain, making it easier to analyze [64]. Wavelet analysis is commonly used in bearing RUL prediction for feature extraction, breaking down vibration signals into basic waveforms, and analyzing their spectral characteristics with wavelet packet transform (WPT) and Fourier analysis. Discrete wavelet transform (DWT) decomposes a signal in a multiscale manner to separate high- and low-frequency components. Wavelet-based features can generally be grouped into four main categories: entropy-based, energy-based, spectral-based, and statistical-based [65]. Entropy-based features measure irregularities in vibration signals, aiding fault identification, while energy-based features detect faults by analyzing changes in wavelet subband energy, supporting RUL prediction. Spectral and statistical features are typically applied to the basic waveforms reconstructed from wavelet decomposition coefficients.

Besides these features, other hybrid features are also extracted to reflect the degradation mechanism. Since no single feature captures all information from time-varying signals, multidimensional representation is essential for

effective HI construction. Deep learning has been employed as a feature extraction technique capable of operating directly on raw data. By transforming vibration data into spectrograms, wavelets, or other representations, the ability of CNNs to detect meaningful patterns is significantly enhanced. A deep CNN-based spectrum-principal-energy-vector feature extraction method was used in [66]. CWT-CNN image feature-based HI was used in [48]. A semi-supervised double attention-guided assessment approach was used to incorporate local semantics in feature extraction in [67]. Generative Adversarial Networks (GANs) were used for generalized multiscale feature extraction in [68]. Feature extraction from convolutional autoencoders combined with status degradation functions for RUL prediction showed better accuracy than many state-of-the-art methods in [26].

Feature selection removes irrelevant features, reducing computation time and improving model accuracy [69]. Feature selection methods are commonly classified into filters, wrappers, embedded, and hybrid methods [70]. Filter-based feature selection includes Pearson's Correlation, Spearman's Correlation, Kendall's Tau, Distance Correlation, Mutual Information, Laplacian Score, Monotonicity, Spectrum Graph, and robustness values [71]. Principal Component Analysis (PCA) reduces the dimensionality of a feature set with many correlated variables while retaining as much variation as possible and was used as a feature selection technique in [72]. Relief-based algorithms (RBAs) have gained popularity by effectively balancing computational efficiency with sensitivity to complex patterns of association while also being adaptable to regression features [73]. The wrapper method selects the best features by testing different combinations and using the model's performance to find the best subset without looking inside the model. These techniques are more popular in classification problems rather than regression, and moreover, they are classified based on information gain, Chi-square tests, Fisher score ranking, and the ratio of missing values in the features. Embedded methods combine the strengths of filter and wrapper methods, using algorithms with built-in feature selection. Popular examples include LASSO and RIDGE regression, which use penalization functions to reduce overfitting. SHapley Additive exPlanations (SHAP) is often considered an embedded method in feature selection, as it helps understand each feature's contribution to the model's predictions and explains the output of machine learning models [43].

In deep learning models, features are automatically selected for RUL prediction model, and the integration of attention layers with encoder-decoder architectures further enhances prediction accuracy [7], [74], [75]. To make consistency across selected features, feature scaling or normalization is used to enable accurate and balanced health indicators. To eliminate the negative impact of varying value ranges on RUL prediction, normalization methods (robust scaler) were applied to standardize the range of feature values [76].

C. RUL PREDICTION TIME (RPT)

Bearings only start to degrade after some use, so predicting their RUL before any signs of degradation is both unreliable and unnecessary [53]. HI, division points represent thresholds that mark significant changes in bearing degradation. These points, called RPT, help predict when the bearing will reach a critical failure point. Condition monitoring divides bearing health into two or more stages to track degradation. In a two-stage division, a distinction is made between healthy (nominal operating) and degraded (anomalous) states to trigger the RUL prediction time. Many existing approaches select the RUL prediction time subjectively, which can either include unrelated noise or omit critical degradation information. However, they agree that the first initial occurrence of abnormalities or anomalies during bearing operation is referred to as the basis of RUL prediction time.

An exponential regression-based local degradation model was used to find RUL prediction time in [77]. An alarm bound technique (ABT) using a linear regression model was proposed to detect the onset of bearing degradation, and the gradient of the bearing health index, also determined through linear regression, was used to establish the failure threshold in [78]. A GAN was employed to model the distribution of healthy state data, and a HI is created to predict the initial failure time [79]. Since failure criteria are subjective for bearing in different applications, the RUL prediction time is not universally agreed upon. As a result, multiple RUL prediction times have been using in the PRONOSTIA dataset.

D. PROGNOSTIC MODELS

Prognostic algorithms rely heavily on time series data collected during the bearing degradation. These algorithms can be categorized into three main types based on their model complexity and learning mechanisms: traditional machine learning-based methods, which use established algorithms to analyze and predict based on historical data; deep learning-based methods, which leverage advanced neural networks to learn complex patterns from large datasets automatically; and transfer learning-based methods, which adapt pre-trained models to new but related domains to improve prediction accuracy with limited data.

1) CLASSICAL MACHINE LEARNING

Machine learning algorithms analyze historical data to predict RUL, effectively modeling nonlinear systems. The classical algorithms can generally be classified into supervised and unsupervised. The primary purpose of supervised learning is to find the optimal mapping from inputs to target outputs. In contrast, unsupervised learning algorithms work with input-only data without labeling, aiming to analyze the data distribution to uncover helpful information about its underlying structure. Prediction is a form of regression that forecasts future values in a time series.

The feature-based method is one of the oldest and most widely used approaches for fault prognosis. Researchers have

proposed many classical ML models including support vector regressor (SVR) [50], [72], [80], hidden Markov models (HMM) [22], random forest [43], and self-organizing maps (SOM) [64]. A hybrid approach combining sparse vector machine regressions with exponential degradation models and Fréchet distance achieved superior accuracy compared to state-of-the-art methods [42].

2) DEEP LEARNING (DL)

DL has become a leading approach for RUL prediction due to its ability to model complex patterns through multiple non-linear processing layers that learn hierarchical representations of data. Classical ML-based RUL prediction relies on shallow learning architectures, explicit model equations, and extensive prior knowledge, limiting its effectiveness in the age of big data [45]. DL methods excel at automatic feature extraction and systematic pattern modeling, offering superior accuracy for large datasets [76], [81]. Based on evaluation metrics MAE and RMSE, the DL model consistently outperformed the classical ML models across all proportions of testing data. There is no consensus on the number of categories for classifying the DL model [35], [82]. However, it is categorized into five types: multi-layer perception (MLP) [45], [83], Auto-encoders (AE) [26], [83], Convolutional Neural Networks (CNN) [48], [84], [85], [86], Recurrent Neural Networks (RNN) [58], [87], and generative models [88].

a: MLP-BASED

An MLP, proposed as a prototype for ANN, is a fully connected network with multiple hidden layers. The MLP is commonly used for comparison, usually with one hidden layer and a reasonable number of neurons. The number of hidden layers, weights, biases, and activation functions significantly affect its performance. The training uses backpropagation and regularization methods to help prevent overfitting. However, training an MLP to predict the RUL of bearings is challenging due to the large number of parameters involved. Overfitting, which leads to good performance on the training dataset but poor performance on the testing dataset, is a common issue. Dropout is a simple and effective method to address this and improve the generalization ability of the neural network [89]. Deep NNs share similar depth with deep CNNs and offer a fair comparison, given that deeper architectures are often challenging to train effectively.

b: AUTO-ENCODERS (AE) AND ITS VARIANTS

An AE is an unsupervised learning algorithm that seeks to learn a condensed representation of the input data by encoding it into a lower-dimensional space and then reconstructing the original input from this representation [90], [91]. After training, the encoder component of the autoencoder can be used to extract features from new data, which can then be input into a regression model to predict RUL [92]. The reconstruction error from an autoencoder can

reveal deviations from normal operating conditions, which can serve as a basis for RUL estimation. Convolutional autoencoder along with status degradation model was used to predict RUL [26]. VAEs, an extension of AEs, learn probabilistic latent variables. A fully convolutional variational autoencoder was used for feature extraction, and a domain adaptation method based on the dynamic benchmark was tested for its performance in incremental predictions [93].

c: CNN AND ITS VARIANTS

A CNN is commonly used for tasks involving spatial data, such as images, but it can also be applied to 1D time series vibration data by converting it into 2D representation. Various signal conversion methods can be used, including data matrix transformation, pixel intensity representation, Gramian angular field [94], Markov transition field [95], recurrence Plot [54], WPT, STFT, and so on. A CNN uses convolutional layers and pooling layers to extract spatial features. A regression model can then predict RUL using the trained CNN's output. The STFT-CNN approach significantly reduces the computational burden and memory usage while preserving resolution, resulting in superior accuracy compared to DNN [63]. WPT-based time-frequency representation and multiscale CNN (MSCNN) showed smaller errors compared to a standard CNN in [84]. This improvement is due to combining the final convolutional layer and the final pooling layer, which effectively utilizes local and global features the high-level layers learn. Moreover, MS-CNN integrated dilation has smaller errors than MS-CNN due to the detailed information extracted by low-level dilation layers. Temporal adjacent discrete cosine transformation CNN was proposed to reduce the load and complexity of wavelet-based CNN in [86]. CNN-bootstrap integrated method was proposed in [96], producing superior performance than MSCNN. The deep convolutional autoencoder combined with CNNs (DCAE-CNN) was used in [97]. The Double-CNN model architecture had a more powerful feature extraction capability than a standard CNN, leading to higher prediction accuracy and greater robustness [47]. The Transformer model is a highly versatile and robust architecture that utilizes self-attention and parallel processing to handle sequential data efficiently [98]. The Convolutional Transformer (CoT) showed better than CNN, MSCNN, DCNN, and DSCN in [59].

TCNs combine the strengths of CNNs and RNNs, a variation of CNN for time series modeling, offers a large receptive field with minimal memory requirements [99], [100], [101], [102]. By incorporating dilation, TCNs achieve a larger receptive field without adding more parameters, enabling them to capture long-range dependencies and improve accuracy. A dual competitive attention module with multidimensional competition enhanced the accuracy of RUL prediction [103], [104], [105]. An adaptive degradation stage division and TCN-based piecewise RUL estimation method were proposed, demonstrating superiority over basic CNN

and RNN models in [101]. EMD-TCN were applied in [85], yielding a 10–20% improvement in average scores over traditional convolutional algorithms. The integration of a causal dilated convolution structure with a residual self-attention mechanism in the TCN improved computational efficiency and demonstrated clear superiority in accuracy over both TCN and TCN-SA [102].

d: RNN AND ITS VARIANTS

An RNN contains feedback connections from the hidden or output layers to the preceding layers, enabling it to process dynamic information. RNNs, Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) networks are designed to handle sequential data and are effective at capturing temporal dependencies [49], [106], [107], [108], [109], [110], [111], [112]. Traditional RNNs struggle to capture long-term dependencies due to the vanishing gradient problem; however, gated architectures like LSTM and GRU address this issue by maintaining a more stable gradient. LSTMs can model long-term dependencies in time-series data, such as the progression of machinery degradation over time. Performance degradation assessment model of bearing was built using LSTM recurrent network in [111]. An end-to-end deep framework for RUL estimation, utilizing convolutional and LSTM recurrent units, was demonstrated in [113]. Vanilla LSTM, with a single hidden layer and a prediction layer [114], serves as a basic LSTM configuration. Variants like CNN-LSTM [115], Convolution-LSTM [63] and Bidirectional LSTM [24], [35], [116] were utilized for RUL prediction. BiLSTM captures bidirectional dependencies over long distances and selectively learns to remember and forget information, capturing more context and dependencies in the data. Conv-LSTM integrates CNN layers for feature extraction with LSTMs for sequence prediction, where each LSTM unit includes a convolutional processing of the input, similar to CNN-LSTM. The Encoder-Decoder LSTM addresses sequence-to-sequence prediction by using two sub-models: one encodes the input sequence into a fixed-length vector, and the other decodes this vector to predict the output sequence. Fast Fourier convolution GRU showed better prediction than encoder-decoder RNN [44]. A dual-stage attention GRU (DA-GRU) achieves a lower RMSE than a standard GRU, while a parallel GRU outperforms the DA-GRU. The parallel DA-GRU achieves the lowest RMSE of all [110]. The attention mechanism assigned greater weights to important features and time steps, enabling attention-based LSTM to outperform basic LSTM in prediction accuracy [117].

e: GENERATIVE MODEL

A deep belief network (DBN) is a generative model consisting of multiple layers of stochastic, latent variables. It is typically trained using unsupervised learning algorithms like RBMs. In RUL prediction, a DBN can learn hierarchical representations of time series data, and each layer in the

network captures increasingly abstract features. After training, the features learned by the DBN can be used as input for a supervised learning algorithm, such as a regression model, to predict RUL. The key advantage of DBN is compelling pre-training on large amounts of unlabelled data, which enhances the learning and generalization of RUL prediction models. The deep adversarial methodology achieved superior RUL prediction performance and demonstrated its ability to predict the RUL even with a small percentage of labeled data [88]. GANs generate synthetic vibration data that mimics training data, helping to augment datasets and enhance model robustness. A combined approach involving a diffusion process and a temporal attention-based data augmentation mechanism was employed to enhance prediction accuracy in [118].

f: META-HEURISTIC ALGORITHMS IN DL

Meta-heuristic algorithms, originating in the 1970s with evolutionary-based methods, have since expanded to include physics-based, human-based, swarm-based, and hybrid approaches [119]. Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) are the most widely used for their predictive power and serve as benchmarks for new algorithm development. The integration of DL models with meta-heuristic algorithms offers a powerful approach to RUL prediction by combining the strengths of both techniques. DL models excel in capturing complex patterns and nonlinear relationships from datasets, providing accurate feature extraction, whereas PSO was used in kernel parameter optimization in relevance vector machines [115]. Moreover, they can be applied in various stages of DL-based RUL prediction tasks, such as feature selection, hyperparameter optimization, weight initialization, and data augmentation. Five metaheuristic algorithms named GA, PSO, Ant Colony Optimization (ACO), Teaching-Learning-Based Optimization (TLBO), and Evolutionary Mating Algorithm (EMA)—strategically enhanced feature selection to improve neural network forecasting accuracy in [120]. This hybrid approach enhances predictive performance by optimizing DL models for better generalization and robustness, effectively addressing the complexities of RUL prediction.

3) TRANSFER LEARNING (TL)

Collecting labeled data for RUL predicting models is particularly challenging in industrial settings, where sensor data may be scarce or difficult to label. As a result, classical ML and DL methods often struggle to achieve high accuracy in RUL prediction due to the limited availability of training data. Furthermore, many DL approaches assume that training and test data come from similar distributions, but this assumption is frequently violated in real-world applications, leading to reduced prediction performance. To overcome these limitations, TL has been introduced as a solution, enabling the transfer of knowledge from other data or models. TL leverages its strong feature learning and domain

transfer capabilities to improve model performance, even when limited labeled data is available for the target task [2], [92], [121]. By using information from different sources, TL enhances the generalization ability of RUL models, making them more robust and accurate in industrial contexts.

Feature learning is a specific approach within inductive TL, where the learned features from the source domain are transferred to the target domain. In DL models, this often involves reusing the lower layers of a neural network and fine-tuning the higher layers on the target data [122]. These features encapsulate the knowledge gained from the source domain into the target domain by freezing the lower layers. Domain adaptation techniques reduce the distribution discrepancy between the source and target domains, where data distributions can vary significantly due to different operating conditions or different bearings. Fine-tuning allows the model to adjust its learned representations better to fit the specific characteristics of the small target dataset.

There is no universally accepted standard for categorizing TL. However, commonly used TL techniques in RUL prediction include model-based, discrepancy metric-based, and adversarial training-based methods [2].

Model-based TL methods assume that source and target domains share some parameters or prior distributions of hyperparameters [123]. Fine-tuning transfers network parameters from a source domain to a related target domain. By learning shared parameters and prior knowledge, a well-trained network structure from the source domain can quickly adapt to test data in the target domain. Fine-tuning involves pre-training a network on the source domain, then freezing or gradually unfreezing its hidden layers while adjusting parameters for the target domain. A depth-wise separable convolution recurrent network (DSCRN) is presented for RUL estimation of bearings and compared with MLP, CNN, LSTM, CLSTM, and GRU models, demonstrating superior performance in [123]. Additionally, the transfer strategy across different operating conditions further reduces the NRMSE. However, TDSCRN's use of depth-wise separable convolution and BLSTM involves higher computational costs during model training.

Discrepancy metric-based TL methods aim to minimize the difference in feature distributions between source and target domains [124], [125], [126]. The PRONOSTIA dataset exhibits significant distribution differences across domains due to varying operating conditions. However, these differences can be mitigated by minimizing the distribution distance using specific discrepancy metrics. Once aligned in a common feature space, the RUL prediction model trained on the source domain can effectively generalize to target domain data, aided by calculating typical discrepancy metrics. Domain adaptation technology addresses inconsistencies in feature distribution across different failure behaviors, with TCNN achieving higher accuracy than TMLP and domain-adversarial training of neural networks DANN [126]. An advanced feature representation method

called contractive denoising autoencoder was used to extract deep features from the HHT marginal spectrum alongside Transfer Component Analysis (TCA) [125]. The TBiGRU model, which utilizes feature extraction from DTW and Wasserstein distance, outperforms TMLP, TDANN, TCNN, and TCA models [124]. A cross-domain transfer prediction using TBi-LSTM and multi-kernel maximum mean discrepancy was proposed, demonstrating better performance than TMLP, tDANN, TCNN, and TCA models [127]. A cross-domain transfer prediction method using a TCN with residual self-attention was presented, outperforming DCNN, MSCNN, and BiLSTM models [128]. Deep transfer metric learning for kernel regression was introduced, outperforming DANN, TMLP, TCNN, and TCA models [129].

Adversarial training, commonly used in unsupervised TL, draws inspiration from GANs [79], [122], [130], [131], [132], [133]. Its goal is to align feature space distributions by adjusting feature representations rather than applying geometric transformations. The network consists of three main components: (a) a feature extractor for deep feature extraction from data in both source and target domains, (b) a regressor that uses these features for RUL prediction, and (c) a domain classifier to distinguish the origin domain of the training data.

A deep domain adaptive network (DDAN) based TL model with selective convolutional RNN for feature extraction has been proposed, and it performs more accurately than DDAN, TCNN, TRNN, and TLSTM [130]. A Wasserstein distance-based weighted domain adversarial neural network (WD-WDANN) outperforms state-of-the-art methods like Bi-LSTM and MSCNN [131]. A metric adversarial domain adaptation (MADA) approach has been proposed, demonstrating its superiority over LSTM, DSCRN, TCA, LSTM-DANN, and CADA [132]. A GAN learns the data distribution, enabling a data alignment method to extract entity-invariant features for RUL prediction, which was effectively implemented even without a precise RUL prediction time [79]. An unsupervised domain adaptation method called deep residual LSTM with domain-invariance was applied and achieved a higher score compared to other feature extraction methods like Deep CNN, Deep LSTM, and Deep Residual Network (DRN) [133]. The Sparse Domain Adaptation Network (SDAN) is an adaptive mechanism for extracting domain-invariant features across varying operational conditions. Its denoising capabilities allow SDAN to capture more degradation information, resulting in higher accuracy than 1-D CNN, LSTM, and BiLSTM [122]. Dynamic Domain Adaptation (DDA) with fuzzy set theory for conditional distribution discrepancy loss and a dynamic adaptive factor was proposed for predicting bearing RUL across multiple working conditions. DDA combines the benefits of both marginal and conditional distribution adaptation while mitigating the impact of distribution weight fluctuations on model performance [134].

Besides the three categories mentioned above, various hybrid TL models, including (DTL-SAE [92], TPDAN [135], TMLP [136], LSTM-DANN [137]) have also been used for the PRONOSTIA dataset. The multisource-multitarget domain adaptation transfer learning model enhances the prediction precision and generalization of RUL prediction models [138]. The deep subdomain adaptive regression network demonstrated strong cross-domain generalization for RUL prediction of bearings under various operating conditions using multichannel time-frequency vibration data [139]. A Bayesian dual network with unsupervised domain adaptation was proposed for cross-domain RUL prediction, using different datasets, outperforming MLP-DA, TCA, DANN, and ATNN [140]. A deep feature disentanglement TL network was introduced to extract domain-invariant features, achieving the smallest error in most prediction tasks compared to SOTA methods [141]. A TPDAN for transferable RUL prediction across different working conditions was introduced, demonstrating superior performance compared to the same model without transfer learning, as well as DANN and MMD, with MMD proving more effective than DANN in feature alignment [135]. In comparison to TCN-SA, TCN, TCN-SECA, and TCN-ECA, the multi-stage shrinkage attention temporal convolutional network model not only significantly reduces prediction errors such as MAE and RMSE but also increases SCORE in [142].

IV. RESULTS

This section explores the characteristics of the raw vibration datasets, the construction of health indicators using domain-specific features, and the evaluation of prediction models ranked by accuracy. We discuss experimental results, their implications, and key trends in RUL prediction.

A. DATA ACQUISITION AND SIGNAL DENOISING

We have considered all six training sets and 11 test sets. Each test set includes multiple sub-files recorded every 10 seconds, with each acceleration sub-file containing 2,560 rows of vibration data. This large dataset is sufficient to illustrate the degradation process. The temperature sub-file includes 600 rows of temperature readings. The statistical correlation between vibration characteristics, surface temperatures, and bearing service life suggests that temperature has a lesser impact on degradation monitoring [143]. Consequently, like most researchers, we included data only from the accelerometer, as temperature sensors provide less sensitive information.

Figure 4. displays the complete lifetime vibration signals from horizontal and vertical directions for six training sets of three working conditions; the red plot shows the data collected by the horizontal accelerometer, and the green color shows the data collected by the vertical accelerometer. Despite being under the same working conditions, these signals exhibit significant differences in degradation trends and lifetimes. The amplitudes of vibration signals increase

over time, highlighting the considerable role of vibration signals. The data's limited quantity and instability make predicting the bearings' RUL challenging.

Directly measuring vibration signals is challenging due to their complexity. Accelerometers mounted on the outer case often capture noise from multiple sources. Hence, it is imperative to enhance fault information initially to facilitate more representative feature extraction and accurate fault prognostics. Figure 5. shows the simple vibration data cleaning method, the top graph shows the raw vibration data horizontal accelerometer before failure criteria, the following graph shows a 3-point moving average filter, and the last graph shows the TSA of five samples before the failure. TSA enhances features by averaging the raw signal over several cycles, reducing noise and interference. Moving average filters and TSA effectively reduce noise and enhance signal clarity, enabling more accurate analysis. An unsupervised segmented data cleaning algorithm showed higher prediction accuracy compared to uncleaned data, even with the same hyperparameters across different DL models in [55]. This is because the algorithm effectively removes noise and irrelevant segments, enhancing the model's ability to learn meaningful patterns.

B. HEALTH INDICATOR CONSTRUCTION

HI's are typically constructed by combining features derived from vibration signals. Accurate RUL prediction depends significantly on the accuracy of HI representation. Constructing a HI from feature fusion can comprehensively characterize the degradation process of machines. As previously explained, feature extraction and selection are fundamental in HI construction.

Figure 6. illustrates the variations in different domain features named: time domain, frequency domain, and time-frequency domain features over the entire life cycle, along with spatial representations as 2D images. The top graph shows the time domain features, such as Maximum, Mean, RMS, Variance, Standard Deviation, Power, Peak, P2P, Skewness, Kurtosis, and Crest Factor, which provide insights into the overall behavior of vibration signals. They capture signal morphology and amplitude changes, which are essential for detecting early signs of bearing degradation. The change processes of these features differ, with each providing only partial details about bearing degradation. The second graph shows the frequency domain features like Maximum Frequency, Sum of Frequencies, Mean Frequency, Variance of Frequency, Peak Frequency, Skewness of Frequency, and Kurtosis of Frequency. These features reveal the frequency content of the signals, helping identify fault-related frequencies and patterns and enabling better-trending ability compared to time domain features. Specifically, time-domain features are more sensitive to damage severity, while frequency-domain features are more sensitive to the type of damage. Feature fusion was implied in almost all works in Table 6 because it improves RUL prediction by combining complementary information from

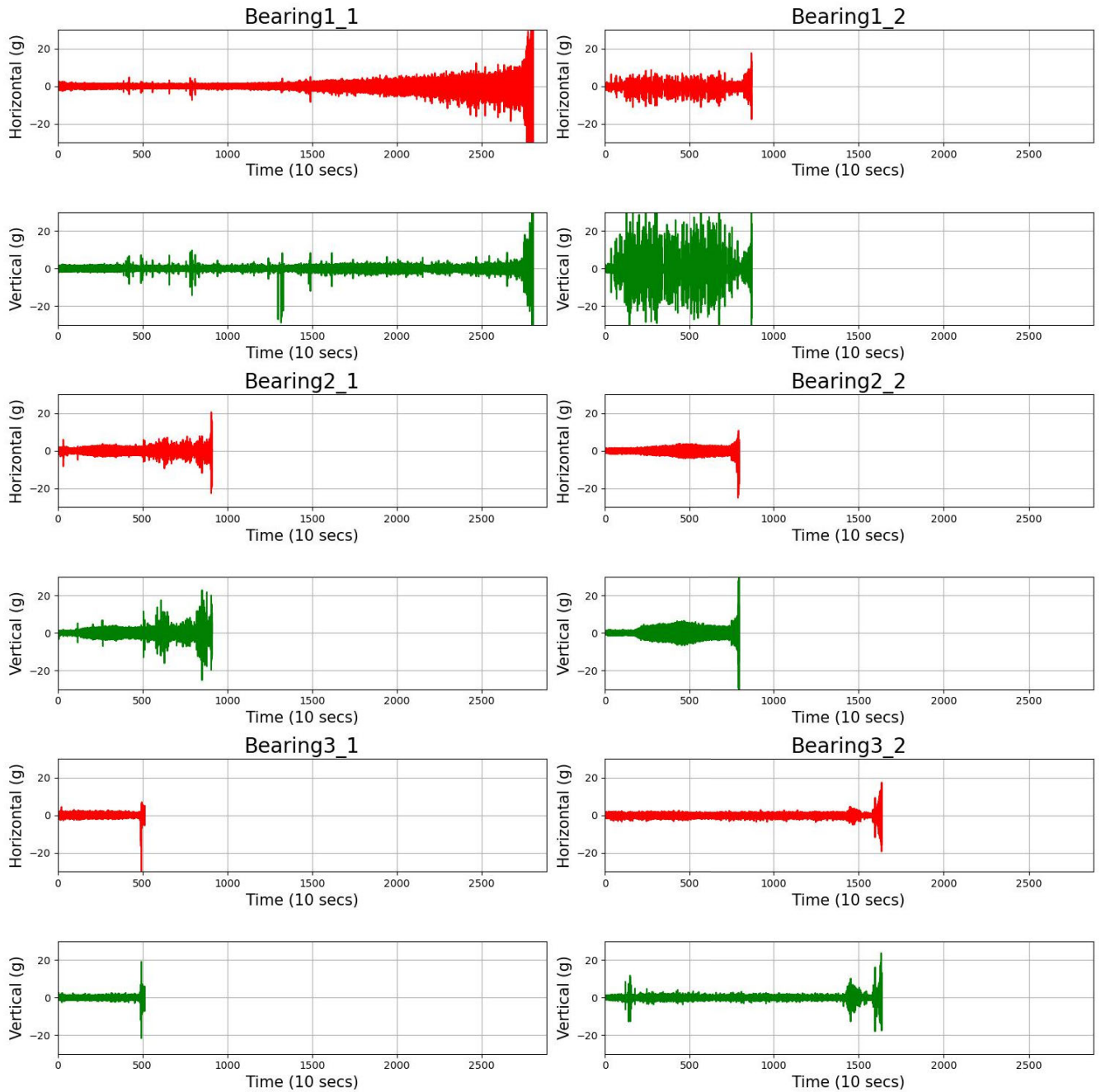


FIGURE 4. Raw vibration signal of six training sets, the red graph shows the vibration data of horizontal accelerometer, and the green plot shows the vibration data of vertical accelerometer.

multiple feature domains, enhancing the model's ability to capture complex patterns and dependencies in the data.

The third graph represents the rms of eight frequency subbands generated by three-level wavelet packet transform. This method integrates time and frequency information, offering a comprehensive view of how signal dynamics evolve and exhibit superior trend ability and monotonicity compared to time-based and frequency-based features. WPT-based features provide better RUL prediction accuracy

because they effectively capture both time-localized and frequency-specific information, enabling a more comprehensive analysis of signal dynamics compared to SOTA in the fifth column of Table 6. EMD features even offer richer information by capturing the evolution of faults than WPT, leading to improved prediction performance, as demonstrated in the sixth column of Table 6.

The bottom graph displays a series of 2D images representing vibration data that meets failure criteria, which

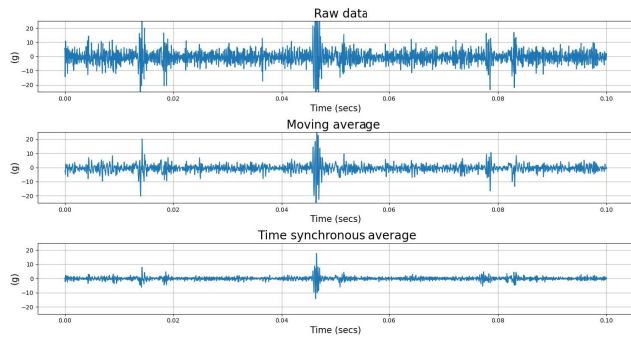


FIGURE 5. Raw vibration data, Moving average, and Time synchronized average of the horizontal accelerometer.

are used as input for CNN-based RUL prediction models. In our recent work, the pixel intensity-based image classification model demonstrated overall superior performance in fault identification, particularly in terms of computational resources and processing time [144]. For RUL predictions, GAF is better suited due to its ability to preserve spatial relationships and enhance feature extraction, whereas MTF's compact representation aligns well with the simpler structures and computational needs of conventional machine learning models [55].

The latter columns of Table 6. showed even higher prediction accuracy, as the probabilistic entropy-based HI is less sensitive to high probabilistic distribution data, making it suitable for use under varying load and speed conditions. HIs constructed using a residual hybrid network with a self-attention mechanism, combined with a fitting interval selection method to address abnormal fluctuations in the health curve, yield improved results. Additionally, the adaptive multi-scale feature extractor from the GCU-TCN network has achieved the highest accuracy to date because it can model intricate, non-linear relationships and capture multi-scale features, offering a more comprehensive representation of the bearing degradation process.

Visual inspection of the line plots for the 3-point moving average of RMS energy in the fifth sub-band wavelet feature across six training datasets reveals neither a clear trend nor smooth behavior, as shown in the top graph of Figure 7. This indicates that relying on a single feature is insufficient for practical health indicator construction due to significant variability. Multiple features must be considered to create a reliable HI with an irreversible degradation trend. The features should be monotonically correlated with the degradation process to discard irrelevant and redundant features that do not adequately capture the degradation mechanism. Therefore, a monotonicity-based feature selection criterion using Kendall's Tau coefficient was employed to create a bar chart of ten time-domain features across six training datasets, as shown in the bottom graph of Figure 7. Qin et al. [145] constructed the HI by dynamically adjusting the weights of metrics such as correlation, robustness, and monotonicity, based on the system's current state, data trends, and fault progression, as shown in the last column of Table 6.

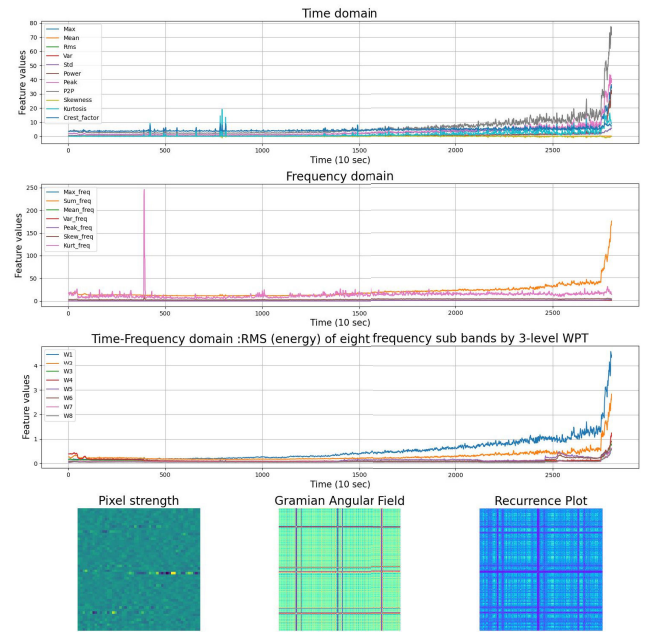


FIGURE 6. Raw vibration data, Moving average, and Time synchronized average of horizontal accelerometer.

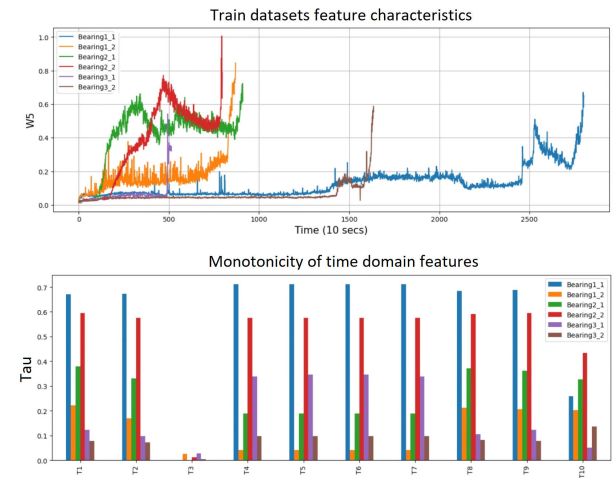


FIGURE 7. The characteristics of the three-point moving average wavelet-based feature were plotted across the entire lifetime of six training sets and the monotonicity metrics of ten time-domain features were evaluated across same six training datasets.

Correlation is more predictive during early degradation stages, while monotonicity becomes crucial in later stages when degradation trends stabilize. Additionally, robustness is prioritized during noisy or fluctuating conditions to ensure stable predictions, which contributes to the accuracy of the prediction.

C. RUL PREDICTION TIME (RPT)

The IEEE PHM 2012 Prognostic Challenge revealed the actual RUL for all 11 test experiments. Table 5 summarizes diverse criteria researchers use for determining RUL prediction times. RMS feature measures the energy content of the vibration signal and is less sensitive to fluctuations

and noise than other time domain statistical features. Therefore, it was used for choosing the RUL prediction time in [146]. A fault detection technique has been developed based on multivariate statistical process control, utilizing the wavelet packet method and considering only two stages: the normal operating stage and the irreversible degradation stage in [147]. The RUL prediction time was determined by detecting bearing anomalies using an appropriate threshold aided by a Box-Cox transformation in [148]. An Autoencoder was used for initial fault detection in high-dimensional data from multivariate industrial processes in [149]. Assessing bearing performance degradation and finding the RUL prediction time using a long short-term memory recurrent network was presented in [111]. A hidden Markov model (HMM) was employed to automatically detect state changes, allowing the RUL prediction time to be determined in [136]. The RUL prediction time of bearing was obtained using the logarithmic squared envelope spectrum-based diversity entropy method in [45]. In [53], the 3σ method, known for its simplicity and efficiency in detecting the RUL prediction point, is applied as per the literature. Information gain-based temporal segmentation was used to divide the degradation stages with different penalty factors for outlier removal, and the RUL prediction time was determined based on the identified transition point in [55].

Table 5 highlights variations in RUL prediction times based on classification models. Still, the most effective approach to finding RPT remains uncertain. The challenge lies in the diverse nature of systems, degradation patterns, and fault types, which means that the ideal RUL prediction time method may vary depending on the specific application and context, requiring further exploration and comparative studies. However, most research papers adhere to the RUL prediction times and failure criteria provided during the data release. Therefore, the following section will focus on achieving better accuracy scores based on time allocated during data release.

D. PROGNOSTIC MODELS

The prediction accuracy of an RUL prediction model can be assessed in two main ways. The first method follows the theme of the PHM 2012 competition, which involves evaluating the model based on specific competition criteria and benchmarks. The top graph on Figure 8 shows various time-domain features of vibration signals throughout the entire lifetime of the bearing, highlighting their behavior over time. The red line on the middle graph shows an example of a feature fusion approach to construct HI. The vibration data is given until the RUL prediction time, and the corresponding HI values in each time sample are constructed. Random fluctuations are typically present in an HI curve because of measurement noise, the stochastic nature of degradation processes, and variations in operational conditions. These fluctuations can undermine the stability of prediction results. Hence, we used a moving average smoothing technique

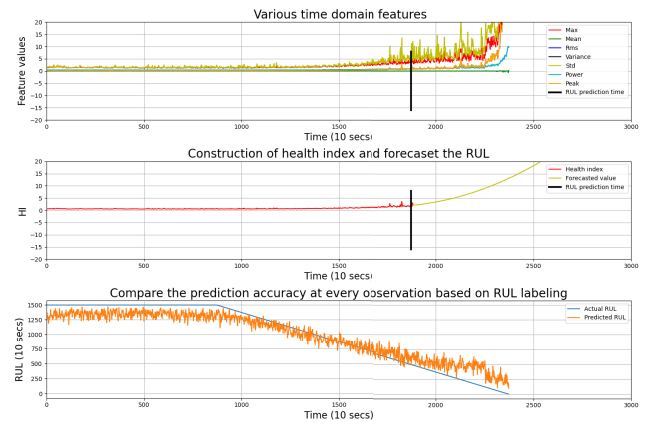


FIGURE 8. An example of feature fusion process and two technique for assessment of of RUL prediction models.

to remove such interferences. After constructing the HI, a forecasting model (curve fitting) is developed to project the HI beyond the black vertical line, with the time samples until the failure threshold is considered as the predicted RUL. The second method involves a straightforward evaluation of each sample's prediction accuracy using supervised ML models, where the accuracy is measured on a sample-by-sample basis, as shown in the third graph. The orange line shows the predicted RUL from CNN based DL model, and the blue line shows the actual RUL labeling by piecewise linear (PWL) function. The advantage of using PWL because of its ability to simplify and stabilize the prediction process, particularly in cases where very long RUL values could lead to less reliable or meaningful predictions. As in [77], the actual RUL is capped at a maximum value—set to 15,000 secs in all experiments—if it exceeds this limit.

Significant progress in RUL prediction is evident from the dataset releases, with Table 6 showing a steady increase in accuracy and a reduction in percentage error. This improvement reflects the model's enhanced performance and provides researchers with baseline values for further refinement. Some researchers demonstrate higher prediction accuracy using only a few experimental datasets from the 11 test sets. However, for the entire set of 11 tests, the most optimal RUL prediction model to date, to the best of our knowledge, is [145]. This model achieves an absolute mean percentage error of 5.71% and a score of 0.791, establishing a baseline for future incremental prediction improvements. It achieves the highest accuracy due to effective feature vector extraction and an accurate prediction regression model.

1) STATISTICAL AND CLASSICAL ML-BASED

Classical supervised ML model learns the relationship between input features (time domain, frequency domain, and time-frequency domain) and the labeled RUL, enabling them to predict the HI of similar, previously unseen data. After that HI construction before the RUL prediction time, they used either curve fitting (exponential, quadratic [78], double exponential [58] etc.) or some other dynamic regression model to get RUL time.

TABLE 5. Various RUL prediction time were suggested by different health stage classification methods, with time measured in 10-second units.

Experiments	End time	PHM challenge	RMS [146]	Wavelet [147]	Anomaly [148]	AE [149]	LSTM [111]	CNN-HS [150]	HMM [136]	TPDNN [45]	Robust-MBDL [53]	MLOF-TSC [55]
Bearing1_3	2375	1801	1676	891	1613	1494	1842	740	1684	1415	573	2251
Bearing1_4	1428	1138	1087	1083	1082	1085	1108	1880	1083	1090	339	1149
Bearing1_5	2463	2301	2410	1141	2306	2413	1653	1667	680	1206	161	2443
Bearing1_6	2448	2301	2407	1641	2035	2413	1656	2159	649	2019	146	2421
Bearing1_7	2259	1501	2201	885	2030	2199	2233	880	1026	1000	757	2214
Bearing2_3	1955	1201	1939	—	—	1939	1983	1995	—	1129	753	1947
Bearing2_4	751	611	740	—	—	735	763	743	—	560	139	737
Bearing2_5	2311	2001	2292	—	—	2311	2287	930	—	979	309	2300
Bearing2_6	701	571	684	—	—	685	712	687	—	519	129	682
Bearing2_7	230	171	221	—	—	223	—	225	—	179	58	224
Bearing3_3	434	351	318	—	—	—	451	—	—	313	82	421

TABLE 6. Data-driven AI-based RUL prediction results have evolved from 2012 to early 2024, with the most accurate methods of their time leading the field.

Testing datasets	Eol	RPT	RUL	LS-SVR [72]	EMD-SOM [64]	MLE [151]	CWT-CNN [48]	RVM [42]	Bi-GRU [108]	New-RVM [115]	State-space [60]	CABLSTM [116]	CAE-SDM [26]	SDAGA [67]	Res-HAS [152]	GCU-TCN [145]
Bearing1_3	23740	18010	5730	37	1.04	0.35	1.05	5.06	4.36	2.27	1.40	0.1	2.62	3.66	18.06	6.23
Bearing1_4	11719	11380	339	80	20.94	5.6	20.35	23.3	70.5	5.6	29.2	-3.4	17.4	2.65	10.26	8.1
Bearing1_5	24620	23010	1610	9	278.26	100	11.18	4.35	6.21	12.42	2.48	-1.94	5.59	4.35	12.46	5.88
Bearing1_6	24470	23010	1460	5	19.18	28.08	34.93	0.68	4.11	10.96	8.9	-18.05	3.42	0.68	7.27	6.55
Bearing1_7	22580	15010	7570	2	7.13	19.55	29.19	42.54	18.63	-22.46	25.36	1.72	1.06	10.57	20.82	0.79
Bearing2_3	19540	12010	7530	64	10.49	20.19	57.24	17.4	17.4	0.99	10.49	0.85	29.96	0.93	1.24	0.68
Bearing2_4	7500	6110	1390	10	51.8	8.63	1.44	12.23	1.44	5.76	6.47	-9.98	2.88	5.76	4.76	10.28
Bearing2_5	23100	20010	3090	440	28.8	23.3	-0.65	0.32	5.18	25.89	20.71	0.96	7.77	2.27	13.36	8.53
Bearing2_6	7000	5710	1290	49	20.93	58.91	-42.64	2.33	16.25	-10.85	4.65	-2.2	13.95	5.43	9.08	2.95
Bearing2_7	2290	1710	580	317	44.83	5.17	8.62	8.62	10.34	1.72	1.72	-20.5	8.62	56.9	9.76	5.54
Bearing3_3	4330	3510	820	90	3.66	40.24	1.22	3.66	6.1	-3.66	3.44	7.52	3.66	2.44	4.13	7.32
Mean				38.64	16.08	20.89	18.96	2.07	12.79	9.32	6.65	-4.08	6.24	5.33	9.88	4.71
Abs mean				100.27	44.28	28.18	25.59	10.95	14.59	12.57	10.44	6.11	8.81	8.69	10.11	5.71
Score				0.3066	0.355	0.4285	0.57	0.6682	0.619	0.64	0.6233	0.6508	0.686	0.6928	0.707	0.791

Linear regression is simple, but it often lacks the flexibility to capture complex patterns in degradation data. SVMs can handle non-linear relationships by mapping the input features into higher dimensions to minimize prediction error within a certain margin. Decision trees split the feature space into regions based on the input features and then assign a labeled RUL to each region. Random forests improve the labeling by building multiple trees and averaging their predictions, enhancing accuracy and robustness against overfitting [43]. Gradient Boosting is an ensemble method that builds decision trees sequentially, where each tree corrects the errors of its predecessor, offering better accuracy than single decision trees or random forests by focusing on difficult-to-predict instances [153]. Neural Networks can learn non-linear relationships between features and HI; although they lack temporal awareness, they were still effective for RUL prediction tasks [136].

In summary, statistical and classical ML models are well-understood, easy to implement, and interpretable and can be highly effective with well-engineered features. However, they often require extensive feature engineering and may struggle with capturing complex, non-linear relationships or temporal dependencies in degradation processes, limiting their accuracy in some cases.

2) DEEP LEARNING-BASED

Depending on the model type classified in Section III-D2, the DL architecture is tailored to capture relevant features and temporal dependencies. Here are some major works that achieve lower prediction errors and higher accuracy, which are summarized in Table 7. Using different features within the same prediction model results in two varying accuracy scores due to the distinct features used. This is because the type and quality of features directly influence the model's ability to learn and make accurate predictions. Some features may capture more relevant information or patterns from the data, leading to higher accuracy, while others may not be as informative, resulting in lower performance. Therefore, even when the underlying algorithm remains constant, the choice of features can significantly impact the model's predictive accuracy.

Besides the same features, the prediction model hyperparameter also determines model accuracy. For example, The third and fourth columns of Table 7 show disparate prediction outcomes of the CNN models due to differences in their hyperparameters and architectural configurations despite using the same input data. The model presented in the third column employs three convolutional layers with smaller kernel sizes (16×1 , 8×1 , and 4×1) and fewer filters

(4, 8, and 16), focusing on fine-grained feature extraction. In contrast, the model presented in the fourth column uses three layers with larger filters (64, 32, and 16) and deeper fully connected layers, enabling broader feature capture and richer representations. These variations in filter size, number, and layer depth significantly influence the models' feature extraction and predictive performance.

Similarly, the LSTM model used in the eighth and ninth columns used different numbers of network parameters can produce varying results, even when the same raw vibration data is used as input, as shown in Table. The hyperparameters for each DL model are listed just below Table 7. This variation occurs because different hyperparameters are used to build the model as presented in [105] and [106], respectively. A model with too few parameters may struggle to capture the complex patterns within the vibration data, leading to underfitting and poor performance. Conversely, a model with too many parameters might overfit the training data, performing well on training but poorly on new data. Therefore, finding the right balance in the number of parameters is crucial for achieving optimal results, even when the input data remains unchanged.

The tenth column of Table 7 (CLSTM) provides higher accuracy because it combines convolutional layers to capture spatial features with LSTM's ability to model temporal dependencies, effectively learning both spatial and sequential patterns in the data. The model CLSTM model used in the eleventh column shows significant accuracy compared to the model used in the tenth column because using two sensor data in the same ConvLSTM model improves accuracy because it provides complementary information, allowing the model to capture a broader range of system behaviors and anomalies. By combining the strengths of CNN and RNN, the TCN model, as shown in the twelfth column, achieves higher prediction accuracy. Stacking the soft threshold and attention mechanism in TCN further improves prediction accuracy, as shown in the thirteenth column. The third-to-last and second-to-last columns display the prediction results from the graph-based neural network. The model with the self-attention mechanism achieves higher accuracy compared to the others. By using advanced feature extraction with a dual-stage attention mechanism and an advanced prediction model like GRU, the authors achieve higher accuracy [110], as shown in the last column of Table 7. This is because probabilistic RUL prediction shows better than deterministic RUL prediction because it accounts for uncertainty and variability in the system's degradation, providing a more reliable and flexible forecast.

In summary, these DL techniques provide a broad overview. Still, the exact implementation of the prognostic model relies heavily on data preparation, including denoising, feature engineering, model configuration, and the computational resources available. Experimenting with various architectures and hyperparameters is typically required to discover the most effective model for a particular task. While we describe structural characteristics for comparison,

detailed model descriptions are necessary for specific contrasts.

3) TRANSFER LEARNING-BASED

Same-bearing RUL prediction involves using a model trained on one operating condition (specific load and RPM) and predicting RUL in another operating condition. In contrast, cross-bearing RUL prediction allows a model trained on one bearing to predict the RUL of a different bearing and different operating condition [121], [138], [154]. A rolling bearing accelerated life experimental dataset from Xi'an Jiaotong University (XJUT) is like PRONOSTIA datasets, and many researchers used that dataset as multisource TL to achieve higher accuracy [42]. Pre-trained models can be quickly adapted to new bearings or conditions, reducing the time and computational resources needed to train a model from scratch. However, some reservations exist; significant differences between the source and target domains can make negative transfer learning, leading to worse performance. When run-to-failure data is unavailable in the target domain, it happens. Careful selection of source data and fine-tuning strategies are required to avoid this [79].

Researchers used different experimental datasets and varying numbers of experiments for the source and target domains in TL studies. This inconsistency leads to biased comparisons, even when testing on the same experiments, as the differences in data and experimental setups can skew the results. As a result, providing a comparative table for TL outcomes across studies would be misleading. We have avoided the comparison Table in this Section to maintain fairness and accuracy.

In summary, TL learning provides a powerful approach to RUL prediction, enabling the development of robust models that can generalize across different machines, environments, and operating conditions. By transferring knowledge from source domains to target domains, TL models can achieve high prediction accuracy even in challenging scenarios with limited data. This makes them particularly valuable in industrial applications where data availability varies widely, and operational conditions change over time.

V. DISCUSSION

A. LESSON LEARNED FROM PROGNOSTIC MODELS

DL-based models generally outperform statistical and classical ML approaches in handling high-dimensional data and capturing long-term dependencies for RUL prediction. DL models, particularly AE, CNNs, LSTMs, and TCNs, are used as a powerful approach to RUL prediction in this dataset. DL models also automate feature extraction, capturing non-linear relationships in vibration data, and are scalable for large datasets.

In DNNs, raw data or time-frequency features can be used for training, allowing for the extraction of more abstract features. However, the training is time-consuming and prone to local optima or overfitting. In DBN, stacked RBMs followed by BP neural network learning fuse multi-feature

TABLE 7. RMSE values of various DL models in RUL prediction.

Testing datasets	DNN	CNN	CNN	MS-CNN	STFT-CNN	CNN-SA	LSTM	LSTM	CLSTM	CLSTM	TCN	TCN-SA	GCN	AGCN	SAGCN	P-DA-GRU
Bearing1_3	22.54	10.52	33.7	8.98	9	23.9	28.02	14.68	26.73	17.08	12.9	11.7	5.9	6.2	4.2	2.25
Bearing1_4	36.27	33.28	28.9	27.47	24.4	10.4	31.01	32.37	26.83	23.93	7.4	8.5	11.8	14.1	12.1	2.73
Bearing1_5	32.15	32.22	13	22.25	22.2	16.7	26.16	32.05	25.04	13.87	14.1	13	5.6	4.8	7.5	2.53
Bearing1_6	38.43	26.81	14.3	22.95	22.9	13.4	32.3	29.08	26.73	15.31	13	11.9	4.2	2.6	5.9	2.47
Bearing1_7	32.6	17.96	21.3	16.36	10.3	13.4	22.34	23.19	26.81	20.43	25.3	12.9	10.2	11.6	3.8	6.2
Bearing2_3	46.19	31.36	26	30.25	40.2	34.3	33.99	34.58	32.39	14.53	23.1	23	8.4	7.6	6.9	1.65
Bearing2_4	55.72	39.04	12.2	28.69	8.9	6.44	34.86	40.7	34.33	18.22	7.5	6.42	24.1	25.7	24.4	6.17
Bearing2_5	41.03	29.41	31.4	30.16	31.6	36.2	36.26	32.75	35.12	21.89	31.4	16.1	28.1	24.6	22.9	2.5
Bearing2_6	39.85	27.77	18.6	25.71	25.7	16	32.97	36.22	31.54	17.1	18	13.8	12.7	7.9	9.9	6.86
Bearing2_7	44.21	43.02	42	30.26	30.6	34.6	33.15	33.68	32.73	23.68	32.5	25	25.8	25.3	25.5	4.45
Bearing3_3	23.75	21.83		20.07	35.7		24.74	20.37	22.74	18.66			11.7	14.8	10.6	7.18
Mean	37.52	28.47	25.12	23.92	23.8	21.39	30.52	29.97	29.18	18.6	19.97	14.8	13.5	13.2	12.15	4.09

* The hyperparameters for each DL model are provided in the following references, listed sequentially from left to right [105], [105], [100], [105], [63], [100], [106], [105], [106], [106], [100], [100], [7], [74], [7], [110].

data but face long training times and overfitting. Genetic algorithms and Nesterov matrix optimization can address these issues. DBN outperforms BPNN in prediction and classification, learning directly from raw data without manual feature extraction. Preprocessing methods like FFT and integrated learning further improve DBN accuracy.

The LSTM method is less effective than CNN for feature extraction from large datasets and requires significantly more training time. LSTM effectively preserves valuable time information from vibration data but requires more extended training due to weight optimization. Heuristic algorithms like particle swarm optimization can enhance this process, and using methods like waveform entropy can improve learning efficiency. GRUs balance performance and efficiency, making them a preferred choice for RUL prediction when computational resources are limited or faster training is needed. Moreover, Bi-GRU often outperforms GRU by integrating information from both directions of a sequence, leading to better performance, especially in tasks that require a nuanced understanding of data patterns.

Attention-based DL models enhance feature extraction by focusing on the most relevant data, improving the capture of critical patterns and dependencies. Attention mechanisms refine feature extraction by integrating domain knowledge, leading to more precise and context-aware predictions.

Bearings are used in various industries with diverse purposes, such as ensuring safety, maintaining operational efficiency, and supporting financial sustainability. Recent advancements in prognostics have led to the development of numerous RUL prediction models. However, industrial adoption has faced challenges due to inherent assumptions, data requirements, and practical implementation issues. This work provides insights into selecting appropriate RUL prediction models by offering a balance between mathematical understanding and alignment with specific business objectives, paving the way for more effective and practical applications in industry.

B. CHALLENGES IN PROGNOSTICS

Accurately predicting the RUL of bearings from vibration data presents several significant challenges. First, the limited

availability of bearing degradation experimental data makes building robust models that can generalize well across different scenarios difficult. Second, determining the optimal time to make RUL predictions is crucial; predicting too early might result in inaccurate forecasts, while predicting too late may not allow sufficient time for maintenance interventions. Third, the interpretability of the prediction models is often a challenge, as complex AI-driven approaches like deep learning can act as black boxes, making it difficult to understand how the model arrives at its predictions. Finally, translating these RUL prediction models into practical industrial applications poses another challenge, as models must be reliable, prompt, scalable, and easy to integrate into existing systems to be helpful in real-world settings.

1) LIMITED NO OF BEARING DEGRADATION EXPERIMENT

Limited experimental data complicates RUL prediction due to uncertainties in degradation patterns. With fewer experiments, the model may struggle to capture the full range of wear and tear, leading to less accurate predictions and increased difficulty in generalizing to different operating conditions; as the percentage of training data increases, the model's prediction accuracy improves. Additionally, increasing the proportion of labeled data reduces the RMSE, enhancing prediction accuracy. This indicates that the model's ability to extract system health features and understand the degradation process improves with more comprehensive training data.

Noisy sensor readings and complex temporal dependencies among machine components further hinder accurate predictions. Addressing these challenges requires the effective use of sensor data to capture machines' intricate operational behaviors.

2) OPTIMAL TIME TO MAKE RUL PREDICTIONS

Currently, widely used methods for RUL prediction are based on ISO standards like the ISO/7919 and ISO/10816 series or industry-specific standards such as VDI/3834 for wind turbines. These standards only establish RUL prediction times for certain existing original HIs. However, there is no standard for determining the RUL prediction times for newly

developed HIs, particularly those without clear physical interpretations. The best time to predict RUL during the degradation process is when there is enough data to capture the initial signs of wear before the system experiences significant failure-related behavior. Ideally, this is when the degradation is steady and measurable, allowing the model to accurately forecast the remaining life of the system while still providing sufficient lead time for maintenance planning. Premature predictions may lack precision due to insufficient degradation data, while late predictions might not allow time for intervention.

3) INTERPRETABILITY OF MODEL

AI approaches were traditionally seen as difficult to explain due to their lack of transparency. However, Explainable AI aims to make AI systems understandable to humans [155], while Interpretable AI focuses on designing models that are inherently clear and transparent [153], [156].

An interpretable deep transfer learning model based on the transfer domain validity index was applied for RUL prediction, successfully enabling transfer prediction across different working conditions and different bearings datasets [157]. The weak interpretability of the DL models prevents their wide use in practical systems, and graph neural network (GNN) is a scalable approach for semi-supervised learning on graph-structured data [158]. Bearings' time series data was converted into a graph using regression shapelets, followed by a deep GNN combined with a GRU for RUL prediction in [159].

DL-based models remain difficult for humans to interpret and are further categorized into explainable and black-box models. Explainable ML clarifies why a model made a particular prediction, focusing on making the model's outputs understandable to humans. GNNs are one of the examples of an explainable ML model and operate within the graph domain to address the limitations of black-box models commonly found in many deep learning methods [160]. A comparative review of five explainable machine learning methods applied to RUL prediction and Grad-CAM showed the most robust model in image processing [161]. Recent work on multiplex aggregation heterogeneous GNN (MAHGNN) has shown more accurate prediction compared with GNN-based SOTA [162].

Explainable machine learning techniques can be classified into two groups: model agnostic and model specific. Model-agnostic techniques refer to methods that can be universally applied across different ML models, irrespective of their structure or type. Examples include LIME (Local Interpretable Model-agnostic Explanations) [163] and SHAP (SHapley Additive exPlanations) [164], which explain model predictions by approximating the contribution of individual features. These techniques are flexible and generalizable, making them ideal for interpreting diverse models such as decision trees, neural networks, and ensembles. On the other hand, model-specific techniques are tailored to specific

categories of ML models, leveraging their unique internal structures or properties. Examples such as Grad-CAM [165], saliency maps [166], and layer-wise relevance propagation (LRP) [167] are predominantly used for deep learning models, utilizing gradients, activations, or learned representations to interpret predictions in a manner specific to neural network architectures.

4) INDUSTRIAL IMPLEMENTATION

The fourth challenge is aligning the design of RUL prediction models with practical industrial implementation. While experimental datasets in laboratories generate valuable knowledge, the real value of RUL prediction models is realized only when they are applied in real industrial settings. This alignment is crucial because the ultimate goal is to save costs and improve efficiency in the industry. Therefore, it's essential to harmonize the model creation process (knowledge generation) with its industrial application (knowledge utilization) to ensure the technology delivers tangible benefits. Most RUL prediction models are developed in straightforward laboratory settings; extracting high-quality temporal features to minimize environmental noise is crucial. More robust feature extraction techniques are needed for real-world industrial applications.

In predicting bearings' RUL, computation time is just as critical as accuracy. Effective condition management requires swift decisions to prevent failures and minimize downtime. Quick computation allows for real-time monitoring and rapid response, enabling timely interventions that can extend the bearing's life and optimize operational efficiency. Therefore, balancing accuracy with fast computation is also challenging in industrial applications.

5) RECOMMENDATION

For accurate RUL prediction, improvements can be made across all three stages: data cleaning, effective feature vector extraction, and enhancing prediction model performance. Each stage showed ample cases for parallel exploration to optimize the overall prediction accuracy by refining data quality, extracting meaningful features, and fine-tuning the model's capabilities.

Researchers often use only horizontal accelerometer data because it provides more information, but prediction accuracy improves by including horizontal and vertical data. To achieve more accurate predictions, we recommend utilizing both types of accelerometer data despite the increased computational resource requirements.

Classical ML limited AI's role in RUL prediction. DL has revolutionized the field by enabling end-to-end prognostic methods that link growing monitoring data directly to machine health states. Future transfer learning approaches aim to apply knowledge across tasks, addressing application challenges.

DL methods require extensive historical data for training, which can be time-consuming and computationally intensive.

Predicting RUL using neural networks and deep learning offers high model accuracy and a longer prediction horizon. However, these methods come with significant challenges. They require extensive condition monitoring data for accurate predictions, involve high computation time, and can suffer from overfitting. Determining the optimal network structure is also challenging, and the neural network can be unstable, producing different results across multiple runs.

Most existing studies predict RUL across different working conditions within the same bearing dataset, making them unsuitable for data with significant distribution shifts. Few studies explore RUL prediction across different bearings and those that rely heavily on the availability of large amounts of labeled data. A multi-source adversarial transformer model is required to achieve greater generality, which can promise the current need for an industrial online accurate prediction model.

TL success depends on the assumption of related degradation behaviors across failures. If invalid, negative transfer can occur, reducing RUL prediction performance. This may happen if source domain degradation behaviors differ from the target domain, as seen with bearings of varying specifications. Metrics to assess cross-domain transferability are needed to select relevant source domains and improve model performance.

GANs can generate synthetic high-transferability data to enhance RUL prediction through transfer learning. However, these models sometimes fail to capture related degradation behaviors, and their transfer performance varies across scenarios. Therefore, it is recommended that transitive transfer learning be utilized to reuse source domain degradation behaviors in the target domain to mitigate negative transfer.

VI. CONCLUSION

This paper reviews state-of-the-art AI methods for bearing RUL prediction, including traditional machine learning, deep learning, and transfer learning on PRONOSTIA data. The datasets exhibit highly non-stationary characteristics, with end-of-life varying significantly even under the same operating conditions. When testing new RUL prediction algorithms, researchers are encouraged to compare their proposed methods against the benchmark results from this study. This could involve providing more comprehensive RUL predictions for datasets that were only partially or unsuccessfully predicted here or achieving comparable scores with methods that offer better computational efficiency or interpretability. Deep learning with transfer learning achieves superior accuracy in RUL prediction. However, each approach to signal denoising, feature engineering, and RUL prediction has strengths and weaknesses. Layer combination and parameter fine-tuning can further enhance prediction scores. The experimental results demonstrate that deep learning models with transfer learning capabilities achieve higher accuracy in RUL prediction.

REFERENCES

- [1] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 485–520, Nov. 2010.
- [2] J. Chen, R. Huang, Z. Chen, W. Mao, and W. Li, "Transfer learning algorithms for bearing remaining useful life prediction: A comprehensive review from an industrial application perspective," *Mech. Syst. Signal Process.*, vol. 193, Jun. 2023, Art. no. 110239.
- [3] H. Ghasemi, E. Shahrabi Farahani, M. Fotuhi-Firuzabad, P. Dehghanian, A. Ghasemi, and F. Wang, "Equipment failure rate in electric power distribution networks: An overview of concepts, estimation, and modeling methods," *Eng. Failure Anal.*, vol. 145, Mar. 2023, Art. no. 107034.
- [4] G. Klutke, P. C. Kiessler, and M. A. Wortman, "A critical look at the bathtub curve," *IEEE Trans. Rel.*, vol. 52, no. 1, pp. 125–129, Mar. 2003.
- [5] M. Rausand and A. Høyland, *System Reliability Theory: Models, Statistical Methods, and Applications*. Hoboken, NJ, USA: Wiley-Interscience, 2004.
- [6] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, Jun. 2012, pp. 1–8.
- [7] Y. Wei, D. Wu, and J. Terpenney, "Bearing remaining useful life prediction using self-adaptive graph convolutional networks with self-attention mechanism," *Mech. Syst. Signal Process.*, vol. 188, Apr. 2023, Art. no. 110010.
- [8] J. Zhuang, Y. Chen, X. Zhao, M. Jia, and K. Feng, "A graph-embedded subdomain adaptation approach for remaining useful life prediction of industrial IoT systems," *IEEE Internet Things J.*, vol. 11, no. 13, pp. 22903–22914, Jul. 2024.
- [9] J. Shang, D. Xu, M. Li, H. Qiu, C. Jiang, and L. Gao, "Remaining useful life prediction of rotating equipment under multiple operating conditions via multi-source adversarial distillation domain adaptation," *Rel. Eng. Syst. Saf.*, vol. 256, Apr. 2025, Art. no. 110769.
- [10] M. He, Z. Li, and F. Hu, "A novel RUL-centric data augmentation method for predicting the remaining useful life of bearings," *Machines*, vol. 12, no. 11, p. 766, Oct. 2024.
- [11] Z. Pan, Z. Meng, Z. Chen, W. Gao, and Y. Shi, "A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings," *Mech. Syst. Signal Process.*, vol. 144, Oct. 2020, Art. no. 106899.
- [12] J. Yang, Y. Peng, J. Xie, and P. Wang, "Remaining useful life prediction method for bearings based on LSTM with uncertainty quantification," *Sensors*, vol. 22, no. 12, p. 4549, Jun. 2022.
- [13] P. Ma, G. Li, H. Zhang, C. Wang, and X. Li, "Prediction of remaining useful life of rolling bearings based on multiscale efficient channel attention CNN and bidirectional GRU," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–13, 2024.
- [14] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.
- [15] I. Animah and M. Shafiee, "Condition assessment, remaining useful life prediction and life extension decision making for offshore oil and gas assets," *J. Loss Prevention Process Industries*, vol. 53, pp. 17–28, May 2018.
- [16] X. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, Nov. 2010.
- [17] Z. Luo, X.-B. Wang, and Z.-X. Yang, "An improved recursive ARIMA method with recurrent process for remaining useful life estimation of bearings," *Shock Vibrat.*, vol. 2022, pp. 1–16, Feb. 2022.
- [18] Z. Huang, Z. Xu, W. Wang, and Y. Sun, "Remaining useful life prediction for a nonlinear heterogeneous Wiener process model with an adaptive drift," *IEEE Trans. Rel.*, vol. 64, no. 2, pp. 687–700, Jun. 2015.
- [19] Z. Huang, Z. Xu, X. Ke, W. Wang, and Y. Sun, "Remaining useful life prediction for an adaptive skew-Wiener process model," *Mech. Syst. Signal Process.*, vol. 87, pp. 294–306, Mar. 2017.
- [20] H. Wang, H. Liao, X. Ma, and R. Bao, "Remaining useful life prediction and optimal maintenance time determination for a single unit using isotonic regression and gamma process model," *Rel. Eng. Syst. Saf.*, vol. 210, Jun. 2021, Art. no. 107504.

- [21] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A mixture of Gaussians hidden Markov model for failure diagnostic and prognostic," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, Aug. 2010, pp. 338–343.
- [22] F. Cartella, J. Lemeire, L. Dimiccoli, and H. Sahli, "Hidden semi-Markov models for predictive maintenance," *Math. Problems Eng.*, vol. 2015, pp. 1–23, Jan. 2015.
- [23] R. K. Singleton, E. G. Strangas, and S. Aviyente, "Extended Kalman filtering for remaining-useful-life estimation of bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1781–1790, Mar. 2015.
- [24] S. Dong, J. Xiao, X. Hu, N. Fang, L. Liu, and J. Yao, "Deep transfer learning based on bi-LSTM and attention for remaining useful life prediction of rolling bearing," *Rel. Eng. Syst. Saf.*, vol. 230, Feb. 2023, Art. no. 108914.
- [25] H. Cao, W. Xiao, J. Sun, M.-G. Gan, and G. Wang, "A hybrid data- and model-driven learning framework for remaining useful life prognostics," *Eng. Appl. Artif. Intell.*, vol. 135, Sep. 2024, Art. no. 108557.
- [26] W. Xu, Q. Jiang, Y. Shen, F. Xu, and Q. Zhu, "RUL prediction for rolling bearings based on convolutional autoencoder and status degradation model," *Appl. Soft Comput.*, vol. 130, Nov. 2022, Art. no. 109686.
- [27] F. Xu, N. Ding, N. Li, L. Liu, N. Hou, N. Xu, W. Guo, L. Tian, H. Xu, C.-M. Lawrence Wu, X. Wu, and X. Chen, "A review of bearing failure modes, mechanisms and causes," *Eng. Failure Anal.*, vol. 152, Oct. 2023, Art. no. 107518.
- [28] P. Dhungana, R. K. Singh, and H. Dhungana, "Machine learning model for fault detection in safety critical system," in *Proc. Int. Conf. Appl. Electron. Pervading Ind., Environ. Soc. Cham, Switzerland: Springer*, 2024, pp. 499–507.
- [29] M. Singh, K. Øvsthus, A.-L. Kampen, and H. Dhungana, "Initial fault identification for procedural decision making using biologically inspired condition management system," in *Proc. Int. Conf. Efficiency Perform. Eng. Neww. Cham, Switzerland: Springer*, Jan. 2024, pp. 641–657.
- [30] M. Singh, K. Øvsthus, A.-L. Kampen, and H. Dhungana, "Development of a human cognition inspired condition management system for equipment," *Int. J. Syst. Assurance Eng. Manage.*, vol. 1, pp. 1–10, Jun. 2024.
- [31] H. Dhungana, "Case based decision making in biologically inspired condition management system," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, Apr. 2024, pp. 335–339.
- [32] H. Dhungana, "Rule-based decision making in biologically inspired condition management system," in *Proc. 16th Int. Conf. Agents Artif. Intell.*, vol. 2, 2024, pp. 1245–1254.
- [33] H. Dhungana, "A machine learning approach for wind turbine power forecasting for maintenance planning," *Energy Informat.*, vol. 8, no. 1, pp. 1–25, Jan. 2025.
- [34] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.
- [35] Z. Zhao, T. Li, J. Wu, C. Sun, S. Wang, R. Yan, and X. Chen, "Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study," *ISA Trans.*, vol. 107, pp. 224–255, Dec. 2020.
- [36] H. Zhou, X. Huang, G. Wen, Z. Lei, S. Dong, P. Zhang, and X. Chen, "Construction of health indicators for condition monitoring of rotating machinery: A review of the research," *Exp. Syst. Appl.*, vol. 203, Oct. 2022, Art. no. 117297.
- [37] E. Zio, "Prognostics and health management (PHM): Where are we and where do we (need to) go in theory and practice," *Rel. Eng. Syst. Saf.*, vol. 218, Feb. 2022, Art. no. 108119.
- [38] C. Ferreira and G. Gonçalves, "Remaining useful life prediction and challenges: A literature review on the use of machine learning methods," *J. Manuf. Syst.*, vol. 63, pp. 550–562, Apr. 2022.
- [39] O. Matania, I. Dattner, J. Bortman, R. S. Kenett, and Y. Parnet, "A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery: Limitations and challenges," *J. Sound Vibrat.*, vol. 590, Nov. 2024, Art. no. 118562.
- [40] W. Li, R. Huang, J. Li, Y. Liao, Z. Chen, G. He, R. Yan, and K. Gryllias, "A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges," *Mech. Syst. Signal Process.*, vol. 167, Mar. 2022, Art. no. 108487.
- [41] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *J. Sound Vibrat.*, vol. 289, nos. 4–5, pp. 1066–1090, Feb. 2006.
- [42] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Trans. Rel.*, vol. 69, no. 1, pp. 401–412, Mar. 2020.
- [43] M. G. Alfarizi, B. Tajiani, J. Vatn, and S. Yin, "Optimized random forest model for remaining useful life prediction of experimental bearings," *IEEE Trans. Ind. Informat.*, vol. 19, no. 6, pp. 7771–7779, Jun. 2023.
- [44] P. Yu, M. Ping, J. Ma, and J. Cao, "Method to enhance time series rolling fault prediction by deep fast Fourier convolution," *Measurement*, vol. 228, Mar. 2024, Art. no. 114177.
- [45] R. Bai, K. Noman, K. Feng, Z. Peng, and Y. Li, "A two-phase-based deep neural network for simultaneous health monitoring and prediction of rolling bearings," *Rel. Eng. Syst. Saf.*, vol. 238, Oct. 2023, Art. no. 109428.
- [46] S. Yang, Y. Liu, Y. Liao, and K. Su, "A new method of bearing remaining useful life based on life evolution and SE-ConvLSTM neural network," *Machines*, vol. 10, no. 8, p. 639, Aug. 2022.
- [47] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9521–9530, Dec. 2019.
- [48] Y. Yoo and J.-G. Baek, "A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network," *Appl. Sci.*, vol. 8, no. 7, p. 1102, Jul. 2018.
- [49] Y. Zhu, J. Wu, X. Liu, J. Wu, K. Chai, G. Hao, and S. Liu, "Hybrid scheme through read-first-LSTM encoder-decoder and broad learning system for bearings degradation monitoring and remaining useful life estimation," *Adv. Eng. Informat.*, vol. 56, Apr. 2023, Art. no. 102014.
- [50] R. K. Jha and P. D. Swami, "Failure prognosis of rolling bearings using maximum variance wavelet subband selection and support vector regression," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 44, no. 1, pp. 1–13, Jan. 2022.
- [51] X. Liu, P. Song, C. Yang, C. Hao, and W. Peng, "Prognostics and health management of bearings based on logarithmic linear recursive least-squares and recursive maximum likelihood estimation," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1549–1558, Feb. 2018.
- [52] D. Chen, Y. Qin, Y. Wang, and J. Zhou, "Health indicator construction by quadratic function-based deep convolutional auto-encoder and its application into bearing RUL prediction," *ISA Trans.*, vol. 114, pp. 44–56, Aug. 2021.
- [53] K. Tran, H.-C. Vu, L. Pham, N. Boudaoud, and H.-S.-H. Nguyen, "Robust-MBDL: A robust multi-branch deep-learning-based model for remaining useful life prediction of rotating machines," *Mathematics*, vol. 12, no. 10, p. 1569, May 2024.
- [54] W. Jiang, Z. Li, A. Jiang, Y. Lei, and H. Wang, "Recurrence plot quantitative analysis-based fault recognition method of rolling bearing," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–8.
- [55] D. Hou, J. Chen, R. Cheng, X. Hu, and P. Shi, "A bearing remaining life prediction method under variable operating conditions based on cross-transformer fusioning segmented data cleaning," *Rel. Eng. Syst. Saf.*, vol. 245, May 2024, Art. no. 110021.
- [56] T. Yu, S. Li, and J. Lu, "Prediction model of bearing fault remaining useful life based on weighted variable loss degradation characteristics," *Meas. Sci. Technol.*, vol. 35, no. 9, Sep. 2024, Art. no. 096122.
- [57] G. Qiu, Y. Gu, and J. Chen, "Selective health indicator for bearings ensemble remaining useful life prediction with genetic algorithm and Weibull proportional hazards model," *Measurement*, vol. 150, Jan. 2020, Art. no. 107097.
- [58] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, May 2017.
- [59] Y. Ding and M. Jia, "Convolutional transformer: An enhanced attention mechanism architecture for remaining useful life estimation of bearings," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.
- [60] A. Kumar, C. Parkash, G. Vashishtha, H. Tang, P. Kundu, and J. Xiang, "State-space modeling and novel entropy-based health indicator for dynamic degradation monitoring of rolling element bearing," *Rel. Eng. Syst. Saf.*, vol. 221, May 2022, Art. no. 108356.
- [61] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.
- [62] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006.

- [63] X. Li, W. Zhang, and Q. Ding, "Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction," *Rel. Eng. Syst. Saf.*, vol. 182, pp. 208–218, Feb. 2019.
- [64] S. Hong, Z. Zhou, E. Zio, and K. Hong, "Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method," *Digit. Signal Process.*, vol. 27, pp. 159–166, Apr. 2014.
- [65] S. A. Aburakhia, R. Myers, and A. Shami, "A hybrid method for condition monitoring and fault diagnosis of rolling bearings with low system delay," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.
- [66] L. Ren, Y. Sun, H. Wang, and L. Zhang, "Prediction of bearing remaining useful life with deep convolution neural network," *IEEE Access*, vol. 6, pp. 13041–13049, 2018.
- [67] J. Zhuang, M. Jia, Y. Cao, and X. Zhao, "Semi-supervised double attention guided assessment approach for remaining useful life of rotating machinery," *Rel. Eng. Syst. Saf.*, vol. 226, Oct. 2022, Art. no. 108685.
- [68] S. Suh, P. Lukowicz, and Y. O. Lee, "Generalized multiscale feature extraction for remaining useful life prediction of bearings with generative adversarial networks," *Knowl.-Based Syst.*, vol. 237, Feb. 2022, Art. no. 107866.
- [69] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018.
- [70] A. Jovic, K. Brkic, and N. Bogunovic, "A review of feature selection methods with applications," in *Proc. 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2015, pp. 1200–1205.
- [71] L. Lu, Y. Tan, D. Oetomo, I. Mareels, and D. A. Clifton, "Weak monotonicity with trend analysis for unsupervised feature evaluation," *IEEE Trans. Cybern.*, vol. 53, no. 11, pp. 6883–6895, Nov. 2023.
- [72] E. Sutrisno, H. Oh, A. S. S. Vasan, and M. Pecht, "Estimation of remaining useful life of ball bearings using data driven methodologies," in *Proc. IEEE Conf. Prognostics Health Manage., Enhancing Saf., Efficiency, Availability, Effectiveness Syst. Through PHM Technol. Appl.*, Jun. 2012, pp. 1–7.
- [73] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *J. Biomed. Informat.*, vol. 85, pp. 189–203, Sep. 2018.
- [74] Y. Wei and D. Wu, "Remaining useful life prediction of bearings with attention-aware graph convolutional network," *Adv. Eng. Informat.*, vol. 58, Oct. 2023, Art. no. 102143.
- [75] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021.
- [76] L. Ren, J. Cui, Y. Sun, and X. Cheng, "Multi-bearing remaining useful life collaborative prediction: A deep learning approach," *J. Manuf. Syst.*, vol. 43, pp. 248–256, Apr. 2017.
- [77] J. Wu, C. Wu, S. Cao, S. W. Or, C. Deng, and X. Shao, "Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 529–539, Jan. 2019.
- [78] W. Ahmad, S. A. Khan, M. M. M. Islam, and J.-M. Kim, "A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models," *Rel. Eng. Syst. Saf.*, vol. 184, pp. 67–76, Apr. 2019.
- [79] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Data alignments in machinery remaining useful life prediction using deep adversarial neural networks," *Knowl.-Based Syst.*, vol. 197, Jun. 2020, Art. no. 105843.
- [80] T. H. Loutas, D. Roulias, and G. Georgoulas, "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression," *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 821–832, Dec. 2013.
- [81] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Rel. Eng. Syst. Saf.*, vol. 172, pp. 1–11, Apr. 2018.
- [82] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
- [83] L. Ren, Y. Sun, J. Cui, and L. Zhang, "Bearing remaining useful life prediction based on deep autoencoder and deep neural networks," *J. Manuf. Syst.*, vol. 48, pp. 71–77, Jul. 2018.
- [84] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, Apr. 2019.
- [85] W. Yang, Q. Yao, K. Ye, and C.-Z. Xu, "Empirical mode decomposition and temporal convolutional networks for remaining useful life estimation," *Int. J. Parallel Program.*, vol. 48, no. 1, pp. 61–79, Feb. 2020.
- [86] Y. Pang, L. Jia, and Z. Liu, "Discrete cosine transformation and temporal adjacent convolutional neural network-based remaining useful life estimation of bearings," *Shock Vibrat.*, vol. 2020, pp. 1–14, Jun. 2020.
- [87] B. Wang, Y. Lei, T. Yan, N. Li, and L. Guo, "Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery," *Neurocomputing*, vol. 379, pp. 117–129, Feb. 2020.
- [88] D. Verstraete, E. Drogue, and M. Modarres, "A deep adversarial approach based on multi-sensor fusion for semi-supervised remaining useful life prognostics," *Sensors*, vol. 20, no. 1, p. 176, Dec. 2019.
- [89] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*.
- [90] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
- [91] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Jun. 2014, pp. 1724–1734.
- [92] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, "Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2416–2425, Apr. 2019.
- [93] Y. Zou, S. Zhao, Y. Liu, Z. Li, X. Song, and G. Ding, "The transfer prediction method of bearing remaining use life based on dynamic benchmark," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
- [94] Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1–7.
- [95] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proc. 24th Int. Conf. Artif. Intell. (IJCAI)*, 2015, pp. 3939–3945.
- [96] C.-G. Huang, H.-Z. Huang, Y.-F. Li, and W. Peng, "A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing," *J. Manuf. Syst.*, vol. 61, pp. 757–772, Oct. 2021.
- [97] C. Wang, W. Jiang, X. Yang, and S. Zhang, "RUL prediction of rolling bearings based on a DCAE and CNN," *Appl. Sci.*, vol. 11, no. 23, p. 11516, Dec. 2021.
- [98] D. Chen, W. Hong, and X. Zhou, "Transformer network for remaining useful life prediction of lithium-ion batteries," *IEEE Access*, vol. 10, pp. 19621–19628, 2022.
- [99] C. H. Goay, N. S. Ahmad, and P. Goh, "Temporal convolutional networks for transient simulation of high-speed channels," *Alexandria Eng. J.*, vol. 74, pp. 643–663, Jul. 2023.
- [100] Y. Wang, L. Deng, L. Zheng, and R. X. Gao, "Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics," *J. Manuf. Syst.*, vol. 60, pp. 512–526, Jul. 2021.
- [101] H. Qiu, Y. Niu, J. Shang, L. Gao, and D. Xu, "A piecewise method for bearing remaining useful life estimation using temporal convolutional networks," *J. Manuf. Syst.*, vol. 68, pp. 227–241, Jun. 2023.
- [102] Y. Cao, Y. Ding, M. Jia, and R. Tian, "A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings," *Rel. Eng. Syst. Saf.*, vol. 215, Nov. 2021, Art. no. 107813.
- [103] W. Wang, G. Zhou, G. Ma, X. Yan, P. Zhou, Z. He, and T. Ma, "A novel competitive temporal convolutional network for remaining useful life prediction of rolling bearings," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
- [104] G. S. Chadha, U. Panara, A. Schwung, and S. X. Ding, "Generalized dilation convolutional neural networks for remaining useful lifetime estimation," *Neurocomputing*, vol. 452, pp. 182–199, Sep. 2021.
- [105] R. Wang, R. Shi, X. Hu, and C. Shen, "Remaining useful life prediction of rolling bearings based on multiscale convolutional neural network with integrated dilated convolution blocks," *Shock Vibrat.*, vol. 2021, no. 1, Jan. 2021, Art. no. 6616861.

- [106] S. Wan, X. Li, Y. Zhang, S. Liu, J. Hong, and D. Wang, "Bearing remaining useful life prediction with convolutional long short-term memory fusion networks," *Rel. Eng. Syst. Saf.*, vol. 224, Aug. 2022, Art. no. 108528.
- [107] Y. Li, H. Wang, J. Li, and J. Tan, "A 2-D long short-term memory fusion networks for bearing remaining useful life prediction," *IEEE Sensors J.*, vol. 22, no. 22, pp. 21806–21815, Nov. 2022.
- [108] D. She and M. Jia, "A BiGRU method for remaining useful life prediction of machinery," *Measurement*, vol. 167, Jan. 2021, Art. no. 108277.
- [109] J. Chen, H. Jing, Y. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process," *Rel. Eng. Syst. Saf.*, vol. 185, pp. 372–382, May 2019.
- [110] L. Cao, H. Zhang, Z. Meng, and X. Wang, "A parallel GRU with dual-stage attention mechanism model integrating uncertainty quantification for probabilistic RUL prediction of wind turbine bearings," *Rel. Eng. Syst. Saf.*, vol. 235, Jul. 2023, Art. no. 109197.
- [111] B. Zhang, S. Zhang, and W. Li, "Bearing performance degradation assessment using long short-term memory recurrent network," *Comput. Ind.*, vol. 106, pp. 14–29, Apr. 2019.
- [112] C.-G. Huang, H.-Z. Huang, and Y.-F. Li, "A bidirectional LSTM prognostics method under multiple operational conditions," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8792–8802, Nov. 2019.
- [113] A. Z. Hinchí and M. Tkouat, "Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network," *Proc. Comput. Sci.*, vol. 127, pp. 123–132, Jan. 2018.
- [114] Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu, "Remaining useful life estimation of engineered systems using vanilla LSTM neural networks," *Neurocomputing*, vol. 275, pp. 167–179, Jan. 2018.
- [115] G. Zhang, W. Liang, B. She, and F. Tian, "Rotating machinery remaining useful life prediction scheme using deep-learning-based health indicator and a new RVM," *Shock Vibrat.*, vol. 2021, no. 1, Jan. 2021, Art. no. 8815241.
- [116] J. Luo and X. Zhang, "Convolutional neural network based on attention mechanism and bi-LSTM for bearing remaining life prediction," *Appl. Intell.*, vol. 52, no. 1, pp. 1076–1091, Jan. 2022.
- [117] Y. Qin, S. Xiang, Y. Chai, and H. Chen, "Macroscopic–microscopic attention in LSTM networks based on fusion features for gear remaining life prediction," *IEEE Trans. Ind. Electron.*, vol. 67, no. 12, pp. 10865–10875, Dec. 2020.
- [118] Y. Lu, D. Tang, D. Zhu, Q. Gao, D. Zhao, and J. Lyu, "Remaining useful life prediction for bearing based on coupled diffusion process and temporal attention," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–10, 2024.
- [119] P. Lu, L. Ye, Y. Zhao, B. Dai, M. Pei, and Y. Tang, "Review of meta-heuristic algorithms for wind power prediction: Methodologies, applications and challenges," *Appl. Energy*, vol. 301, Nov. 2021, Art. no. 117446.
- [120] M. H. Sulaiman, Z. Mustaffa, M. M. Saari, and M. F. Abas, "Wind power forecasting with metaheuristic-based feature selection and neural networks," *Cleaner Energy Syst.*, vol. 9, Dec. 2024, Art. no. 100149.
- [121] C.-G. Huang, C. Men, M. Yazdi, Y. Han, and W. Peng, "Transfer fault prognostic for rolling bearings across different working conditions: A domain adversarial perspective," *Int. J. Adv. Manuf. Technol.*, pp. 1–19, Jun. 2022.
- [122] M. Miao, J. Yu, and Z. Zhao, "A sparse domain adaption network for remaining useful life prediction of rolling bearings under different working conditions," *Rel. Eng. Syst. Saf.*, vol. 219, Mar. 2022, Art. no. 108259.
- [123] G. Huang, Y. Zhang, and J. Ou, "Transfer remaining useful life estimation of bearing using depth-wise separable convolution recurrent network," *Measurement*, vol. 176, May 2021, Art. no. 109090.
- [124] Y. Cao, M. Jia, P. Ding, and Y. Ding, "Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network," *Measurement*, vol. 178, Jun. 2021, Art. no. 109287.
- [125] W. Mao, J. He, and M. J. Zuo, "Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1594–1608, Apr. 2020.
- [126] H. Cheng, X. Kong, G. Chen, Q. Wang, and R. Wang, "Transferable convolutional neural network based remaining useful life prediction of bearing under multiple failure behaviors," *Measurement*, vol. 168, Jan. 2021, Art. no. 108286.
- [127] M. Singh Rathore and S. P. Harsha, "Rolling bearing prognostic analysis for domain adaptation under different operating conditions," *Eng. Failure Anal.*, vol. 139, Sep. 2022, Art. no. 106414.
- [128] J. Zhuang, M. Jia, Y. Ding, and P. Ding, "Temporal convolution-based transferable cross-domain adaptation approach for remaining useful life estimation under variable failure behaviors," *Rel. Eng. Syst. Saf.*, vol. 216, Dec. 2021, Art. no. 107946.
- [129] Y. Ding, M. Jia, Q. Miao, and P. Huang, "Remaining useful life estimation using deep metric transfer learning for kernel regression," *Rel. Eng. Syst. Saf.*, vol. 212, Aug. 2021, Art. no. 107583.
- [130] M. Miao and J. Yu, "A deep domain adaptive network for remaining useful life prediction of machines under different working conditions and fault modes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021.
- [131] T. Hu, Y. Guo, L. Gu, Y. Zhou, Z. Zhang, and Z. Zhou, "Remaining useful life estimation of bearings under different working conditions via Wasserstein distance-based weighted domain adaptation," *Rel. Eng. Syst. Saf.*, vol. 224, Aug. 2022, Art. no. 108526.
- [132] J. Zhuang, M. Jia, and X. Zhao, "An adversarial transfer network with supervised metric for remaining useful life prediction of rolling bearing under multiple working conditions," *Rel. Eng. Syst. Saf.*, vol. 225, Sep. 2022, Art. no. 108599.
- [133] S. Fu, Y. Zhang, L. Lin, M. Zhao, and S.-S. Zhong, "Deep residual LSTM with domain-invariance for remaining useful life prediction across domains," *Rel. Eng. Syst. Saf.*, vol. 216, Dec. 2021, Art. no. 108012.
- [134] H. Cheng, X. Kong, Q. Wang, H. Ma, S. Yang, and G. Chen, "Deep transfer learning based on dynamic domain adaptation for remaining useful life prediction under different working conditions," *J. Intell. Manuf.*, vol. 34, no. 2, pp. 587–613, Feb. 2023.
- [135] Y. Xu, T. Xia, Y. Jiang, Y. Wang, D. Wang, E. Pan, and L. Xi, "A temporal partial domain adaptation network for transferable prognostics across working conditions with insufficient data," *Rel. Eng. Syst. Saf.*, vol. 250, Oct. 2024, Art. no. 110273.
- [136] J. Zhu, N. Chen, and C. Shen, "A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions," *Mech. Syst. Signal Process.*, vol. 139, May 2020, Art. no. 106602.
- [137] P. R. D. O. da Costa, A. Akçay, Y. Zhang, and U. Kaymak, "Remaining useful lifetime prediction via deep domain adaptation," *Rel. Eng. Syst. Saf.*, vol. 195, Mar. 2020, Art. no. 106682.
- [138] L. Shuang, X. Shen, J. Zhou, H. Miao, Y. Qiao, and G. Lei, "Bearings remaining useful life prediction across equipment-operating conditions based on multisource-multitarget domain adaptation," *Measurement*, vol. 236, Aug. 2024, Art. no. 115026.
- [139] Y. Ding, M. Jia, and Y. Cao, "Remaining useful life estimation under multiple operating conditions via deep subdomain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
- [140] C.-G. Huang, J. Zhu, Y. Han, and W. Peng, "A novel Bayesian deep dual network with unsupervised domain adaptation for transfer fault prognosis across different machines," *IEEE Sensors J.*, vol. 22, no. 8, pp. 7855–7867, Apr. 2022.
- [141] T. Hu, Y. Guo, L. Gu, Y. Zhou, Z. Zhang, and Z. Zhou, "Remaining useful life prediction of bearings under different working conditions using a deep feature disentanglement based transfer learning method," *Rel. Eng. Syst. Saf.*, vol. 219, Mar. 2022, Art. no. 108265.
- [142] W. Li, Z. Shang, M. Gao, S. Qian, and Z. Feng, "Remaining useful life prediction based on transfer multi-stage shrinkage attention temporal convolutional network under variable working conditions," *Rel. Eng. Syst. Saf.*, vol. 226, Oct. 2022, Art. no. 108722.
- [143] R. M. Mitrovic, Z. Z. Miskovic, M. B. Djukic, and G. M. Bakic, "Statistical correlation between vibration characteristics, surface temperatures and service life of rolling bearings—Artificially contaminated by open pit coal mine debris particles," *Proc. Struct. Integrity*, vol. 2, pp. 2338–2346, Jan. 2016.
- [144] H. Dhungana, S. K. Mukhiya, P. Dhungana, and B. Karic, "Deep learning-based fault identification in condition monitoring," 2024, *arXiv:2410.05889*.
- [145] Y. Qin, F. Gan, B. Xia, D. Mi, and L. Zhang, "Remaining useful life estimation of bearing via temporal convolutional networks enhanced by a gated convolutional unit," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108308.
- [146] W. Wang, "A model to predict the residual life of rolling element bearings given monitored condition information to date," *IMA J. Manage. Math.*, vol. 13, no. 1, pp. 3–16, Jan. 2002.

- [147] X. Jin, Y. Sun, J. Shan, Y. Wang, and Z. Xu, "Health monitoring and fault detection using wavelet packet technique and multivariate process control method," in *Proc. Prognostics Syst. Health Manage. Conf.*, Dec. 2014, pp. 257–260.
- [148] X. Jin, Y. Sun, Z. Que, Y. Wang, and T. W. S. Chow, "Anomaly detection and fault prognosis for bearings," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 9, pp. 2046–2054, Sep. 2016.
- [149] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, "Unsupervised electric motor fault detection by using deep autoencoders," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 2, pp. 441–451, Mar. 2019.
- [150] S. Suh, J. Jang, S. Won, M. S. Jha, and Y. O. Lee, "Supervised health stage prediction using convolutional neural networks for bearing wear," *Sensors*, vol. 20, no. 20, p. 5846, Oct. 2020.
- [151] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala, "A model-based method for remaining useful life prediction of machinery," *IEEE Trans. Rel.*, vol. 65, no. 3, pp. 1314–1326, Sep. 2016.
- [152] J. Zhu, Q. Jiang, Y. Shen, F. Xu, and Q. Zhu, "Res-HSA: Residual hybrid network with self-attention mechanism for RUL prediction of rotating machinery," *Eng. Appl. Artif. Intell.*, vol. 124, Sep. 2023, Art. no. 106491.
- [153] O. T. Bindingsbø, M. Singh, K. Øvsthus, and A. Keprate, "Fault detection of a wind turbine generator bearing using interpretable machine learning," *Frontiers Energy Res.*, vol. 11, Dec. 2023, Art. no. 1284676.
- [154] K. Liu and Y. Li, "Remaining useful life prediction across machines using multi-source adversarial online knowledge distillation," *Eng. Appl. Artif. Intell.*, vol. 130, Apr. 2024, Art. no. 107726.
- [155] A. Vassiliades, N. Bassiliades, and T. Patkos, "Argumentation and explainable artificial intelligence: A survey," *Knowl. Eng. Rev.*, vol. 36, p. 5, Jan. 2021.
- [156] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *Commun. ACM*, vol. 63, no. 1, pp. 68–77, Dec. 2019.
- [157] W. Mao, J. Liu, J. Chen, and X. Liang, "An interpretable deep transfer learning-based remaining useful life prediction approach for bearings with selective degradation knowledge fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.
- [158] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 31st Int. Conf. Learn. Represent.*, Sep. 2016.
- [159] X. Yang, Y. Zheng, Y. Zhang, D. S. Wong, and W. Yang, "Bearing remaining useful life prediction based on regression shaplet and graph neural network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
- [160] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.
- [161] D. Solís-Martín, J. Galán-Páez, and J. Borrego-Díaz, "On the soundness of XAI in prognostics and health management (PHM)," *Information*, vol. 14, no. 5, p. 256, Apr. 2023.
- [162] Y. Xiao, D. Liu, L. Cui, and H. Wang, "Heterogeneous graph representation-driven multiplex aggregation graph neural network for remaining useful life prediction of bearings," *Mech. Syst. Signal Process.*, vol. 220, Nov. 2024, Art. no. 111679.
- [163] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144.
- [164] S. M. Lundberg, P. G. Allen, and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [165] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
- [166] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *Proc. 2nd Int. Conf. Learn. Represent.*, Dec. 2013.
- [167] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLoS ONE*, vol. 10, no. 7, Jul. 2015, Art. no. e0130140.



HARIOM DHUNGANA (Member, IEEE) received the bachelor's degree in electronics and communication engineering and the Master of Science degree in information and communication engineering from Pulchowk Campus, Institute of Engineering, Tribhuvan University, Kathmandu, Nepal, in 2007 and 2011, respectively. He is currently pursuing the Ph.D. degree in computer science, specializing in software engineering, sensor networks, and engineering computing with the Department of Mechanical Engineering and Maritime Studies, Western Norway University of Applied Sciences. He has over a decade of industrial experience as a Senior Engineer at a telecom operator. From 2019 to 2020, he was a Research Scholar with the Department of Naval, Electrical, Electronic, and Telecommunications Engineering, University of Genova, Italy. His research interests include condition monitoring, cyber-physical systems, human cognition systems, fault diagnosis, prognostics, and health management of industrial equipment.



computing, and machine learning.

THORSTEIN RYKKJE (Member, IEEE) was born in Bergen, Norway, in 1988. He received the B.S. degree in mechanical engineering from Bergen University College, in 2016, and the master's degree in applied and computational mathematics from the University of Bergen, in 2018. He is currently pursuing the Ph.D. degree with the Western Norway University of Applied Sciences. His research interests include control systems, multi-body dynamics, rigid body dynamics, engineering



ALEXANDER S. LUNDERVOLD received the B.Sc. degree in mathematics, the M.Sc. degree in topology, and the Ph.D. degree in numerical analysis from the University of Bergen (UiB), Norway, in 2005, 2007, and 2011, respectively. He is currently a Professor in artificial intelligence with the Western Norway University of Applied Sciences and a Senior Data Scientist with Haukeland University Hospital, Bergen, Norway.

• • •