

Review

Not peer-reviewed version

A Review for Domain Adapted Continual Deep Learning Remaining Useful Life Estimation for Bearing Fault Prognosis Under Evolving Data Distributions

[Apeiranthitis Stamatis](#)*, [Christos Drosos](#)*, [Avraam Chatzopoulos](#)*, [Michail Papoutsidakis](#)*,
[Evangelos Pallis](#)*

Posted Date: 3 March 2026

doi: 10.20944/preprints202603.0102.v1

Keywords: remaining useful life; prognostics and health management; bearing fault prognosis; non-stationary data distributions; domain adaptation; continual learning; domain-adaptive continual learning; evolving data distributions; deep learning-based prognostics



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Review

A Review for Domain Adapted Continual Deep Learning Remaining Useful Life Estimation for Bearing Fault Prognosis Under Evolving Data Distributions

Apeiranthitis Stamatis*, Christos Drosos*, Avraam Chatzopoulos*, Michail Papoutsidakis* and Evangelos Pallis*

University of West Attica, Athens, 12243, Greece

* Correspondence: apiranthitiss@uniwa.gr (A.S.); drososx@uniwa.gr (C.D.); xatzopoulos@uniwa.gr (A.C.); mipapou@uniwa.gr (M.P.); epallis@uniwa.gr (E.P.)

Abstract

Estimating Remaining Useful Life (RUL) and predicting bearing faults based on data-driven models have become central components of modern Prognostics and Health Management (PHM) systems. Although deep learning models have demonstrated strong performance under controlled and stationary operating conditions, their reliability in real-world industrial and marine environments is limited. In practice, operating conditions, sensor properties, and degradation mechanisms evolve continuously over time, leading to non-stationary and shifting data distributions that violate the assumptions of conventional static learning approaches. To address these challenges, two research areas have gained increasing attention: Domain Adaptation (DA), which aims to mitigate distribution discrepancies across operating conditions or machines, and Continual Learning (CL), which enables models to learn sequentially while mitigating catastrophic forgetting. However, existing studies often examine these paradigms in isolation, limiting their effectiveness in long-term deployments, where domain shifts and temporal evolution coexist. This paper presents a comprehensive and systematic review of data-driven bearing fault prognosis and RUL prediction under evolving data distributions, adopting the framework of Domain-Adaptive Continual Learning (DAACL). By jointly examining the DA and CL methods, this review analyzes how these approaches have been individually and implicitly combined to cope with nonstationarity, knowledge retention, and limited label availability in practical PHM scenarios. We categorised existing methods, highlighted their underlying assumptions and limitations, and critically assessed their applicability to long-term, real-world monitoring systems. Furthermore, key open challenges, including scalability, robustness under sequential domain shifts, uncertainty handling, and plasticity–stability trade-offs, are identified, and research directions are outlined based on the identified limitations and practical deployment requirements of the proposed method. This review aims to establish a structured and critical reference framework for understanding the role of domain-adaptive CL in data-driven prognostics, clarifying current research trends, limitations, and open challenges in evolving data distributions. *Index Terms*— Remaining useful life prediction, Prognostics and health management, Bearing fault prognosis, Non-stationary data distributions, Domain adaptation, Continual learning, Domain-adaptive continual learning, Evolving data distributions, Deep learning–based prognostics

Keywords: remaining useful life; prognostics and health management; bearing fault prognosis; non-stationary data distributions; domain adaptation; continual learning; domain-adaptive continual learning; evolving data distributions; deep learning–based prognostics

I. Introduction

ROTATING machinery is widely used in complex industrial systems, manufacturing plants, transportation, and merchant vessels [1], [2]. Bearings are the core components of all rotating

machinery, which support the rotating shaft and reduce friction during operation [3], [4]. Owing to continuous mechanical loading and long operational lifetimes, bearings are inherently susceptible to progressive deterioration and failure. Common causes of bearing faults include excessive operating loads, improper or incorrect installation, inadequate lubrication, material fatigue, manufacturing variability, and insufficient maintenance practices [2], [5].

Failures in machinery can have severe consequences, ranging from unplanned downtime and significant economic losses to catastrophic accidents, environmental disasters, and, in safety-critical applications, loss of human life [6], [7], [8]. These risks have motivated the widespread adoption of Prognostics and Health Management (PHM) strategies, which aim to continuously assess machinery health and detect early stage degradation before functional failure occurs. From a practical perspective, PHM is not only a maintenance strategy but also a risk mitigation mechanism for safety-critical assets [6], [8], [9].

Within the PHM framework, Remaining Useful Life (RUL) estimation has emerged as a key prognostic task [1], [2], [10]. The RUL represents the time interval during which a component can continue to operate safely before reaching an unacceptable degradation or failure threshold [11]. Accurate RUL prediction enables timely maintenance planning, improves asset utilisation, reduces unexpected downtime, and supports informed operational decision-making. This makes bearing fault prognosis and RUL estimation prerequisites for any scalable and reliable PHM framework [12], [13], [14].

1.1. The Core Problem: Evolving Data Distributions in Real-World PHM

In real-world operating environments, rotating machinery rarely operates under fixed or repeatable conditions. Operating regimes evolve over time as a result of changes in load, speed, production demand, environmental conditions, and maintenance actions [5], [15]. As machinery continues to operate, components age, sensors may drift or be replaced, and degradation mechanisms can gradually change. Consequently, the statistical characteristics of condition-monitoring data are not stationary but evolve throughout the operational lifetime of the system [9], [16].

From a data-driven perspective, this behaviour leads to discrepancies between the data used during model development and those encountered during deployment [17], [18], [19]. Prognostic models are typically trained using historical datasets or data extracted from laboratory experiments conducted under specific, often controlled, conditions. Such data are commonly treated as source domains [20], [21], [22]. In contrast, data collected during real operations constitute the target domain, where operating conditions and degradation patterns may differ substantially [23], [24]. When the statistical properties of the source and target data differ, the assumption that the training and deployment data follow the same distribution no longer holds [9], [25]. This phenomenon is commonly referred to as a distribution shift [11].

In real-world applications, distribution shifts are commonly observed through mechanisms such as changes in operating conditions, sensor-related effects, and evolving degradation behaviour. Variations in speed or load can alter the signal characteristics without necessarily changing the underlying failure modes, whereas sensor recalibration, replacement, or gradual drift introduces additional variability into the measured data. Moreover, component aging, wear accumulation, and maintenance interventions may modify the relationship between the observed signals and the actual health state of the system [11], [19], [26]. In bearing applications, even nominally identical components may exhibit distinct degradation trajectories and lifetimes owing to manufacturing variability and differences in usage history [1], [27].

These evolving data distributions pose practical challenges for data-driven RUL predictions. Models developed under stationary assumptions may perform well during initial deployment but progressively lose accuracy and reliability as the operating conditions and degradation dynamics change [5], [19]. In long-term monitoring scenarios, such performance degradation accumulates over time, increasing uncertainty and reducing confidence in prognostic outputs. Consequently, static learning approaches struggle to provide robust and reliable RUL estimates in realistic industrial and marine environments characterised by continuous data evolution [20], [25], [28].

1.2. Limitations of conventional Deep Learning Models for RUL Prediction

Deep learning (DL) models have demonstrated strong performance in data-driven RUL prediction under controlled and stationary conditions. When sufficient labelled data are available and the operating regimes remain consistent, deep architectures can learn complex nonlinear relationships between sensor measurements and component degradation. This has led to their widespread adoption in bearing fault diagnosis and prognosis tasks [19], [29], [30].

However, these assumptions rarely hold in the real world. Most deep learning models for RUL prediction are trained offline using fixed datasets and are subsequently deployed as static-predictors. This workflow implicitly assumes that the data observed during operation are statistically similar to those used for training [20], [24]. As discussed in the previous section, the actual operating environment violates this assumption. When operating conditions change or degradation dynamics evolve, the representations learned during training using conventional DL models may no longer be appropriate, and the prediction accuracy can deteriorate rapidly [5], [16], [26].

A straightforward response to performance degradation is to retrain the model using newly collected data. While this approach may restore accuracy in the short term, it raises several practical concerns [10], [16], [31]. Retraining from scratch is computationally demanding and often requires access to large volumes of historical data, which may not always be available [32] [10]. Incremental retraining offers a more flexible alternative; however, it introduces a different problem: as the model adapts to new data, previously learned degradation patterns can be partially or completely forgotten. This phenomenon, commonly referred to as catastrophic forgetting, is particularly problematic in real-world settings, where historical degradation information remains relevant long after it has been observed [33], [34], [35].

Another practical limitation arises from the availability of labelled data. Reliable RUL labels are difficult to obtain in real operating environments because they typically require run-to-failure data and detailed maintenance records. In many cases, the labels are sparse, delayed or uncertain. Under these conditions, conventional supervised deep learning approaches struggle to maintain robust performance, especially when combined with distribution shifts and temporal data evolution [36], [37], [38].

Taken together, these issues highlight a fundamental limitation of conventional deep learning models for RUL prediction. Static learning approaches lack mechanisms to continuously adapt to changing operating conditions while retaining previously acquired knowledge. Consequently, their applicability in long-term real-world PHM deployments remains limited, motivating the need for learning paradigms that explicitly address both data evolution and knowledge preservation [15], [16], [20], [37].

1.3. Motivation for Integrating Domain Adaptation and Continual Learning.

The limitations discussed in the previous sections indicate that addressing evolving data distributions in real-world health monitoring systems requires more than just improvements to conventional deep learning models. In particular, two research directions have emerged as promising but incomplete responses to this challenge: Domain Adaptation (DA) and Continual Learning (CL).

Domain Adaptation focuses on reducing the discrepancy between the data used for training and the target data distributions [21], [29]. By aligning feature representations or model outputs across different machines or operating conditions, these techniques aim to improve generalisation when labelled data from the target domain are scarce or unavailable [39]. In bearing condition monitoring and RUL prediction, DA has been successfully applied to mitigate variations caused by changes in load, speed, sensor replacement, and environmental conditions [17] [29]. However, most DA approaches implicitly assume that the target domain is either static or available in batch form. In practice, operating conditions often evolve over time, making these assumptions unrealistic, and adaptation performed at a single point in time quickly becomes outdated [8], [11].

CL aims to handle learning over time by enabling models to update their knowledge incrementally as new data become accessible. Rather than retraining from scratch, CL aims to preserve previously learned information while incorporating new patterns, thereby mitigating catastrophic forgetting [20], [40]. This is particularly important in long-term scenarios, where systems

operate continuously and data are collected sequentially. However, many existing CL approaches focus mainly on temporal progression or task sequences and do not explicitly address the differences in data distributions between past and newly acquired data [2], [22]. Consequently, their performance may degrade when incoming data originate from operating conditions that differ significantly from those encountered during the earlier training stages [20], [40].

These later observations suggest that neither DA nor CL alone is sufficient to address the full complexity of real-world RUL prediction under evolving conditions. DA improves robustness across operating conditions but lacks mechanisms for long-term knowledge accumulation [7], [8], whereas CL enables sequential adaptation but often overlooks explicit domain shifts [15], [41].

Combining DA with CL presents a promising solution to these challenges by enabling models to dynamically adjust to evolving operating conditions without requiring complete retraining from scratch. Motivated by this, this review adopts the perspective of **Domain Adaptive Continual Learning**, in which models are designed to adapt continuously to new data distributions while retaining knowledge acquired under previous conditions. Rather than treating domain shift and temporal evolution as separate challenges, this paradigm views them as inherently coupled aspects of long-term system operation. By examining how DA and CL have been jointly or implicitly integrated in existing studies, this review aims to clarify current research trends, highlight open challenges, and provide a structured foundation for the development of robust, lifelong prognostic models.

1.4. Scope and Objectives of this Review

This review aims to provide a structured and critical overview of data-driven approaches for rotating machinery, specifically bearing fault prognosis and RUL prediction for nonstationary and evolving data distributions. Rather than exploring RUL prediction techniques based on conventional DL models and static learning scenarios, this study specifically focuses on long-term monitoring methods that address distribution shifts and data evolution, which are inherent in real-world industrial and marine PHM applications.

To address these challenges, this review adopts the perspective of Domain-Adaptive Continual Learning, emphasising the joint role of DA and CL in achieving robust and scalable prognostic models. Existing studies, although currently limited in number, have been examined not only based on their predictive performance but also with respect to their ability to cope with distribution shifts, temporal data evolution, and knowledge retention during long-term deployments.

The scope of this review is limited to data-driven and deep learning-based methods applied to bearing fault diagnosis and RUL prediction. Physics-based and hybrid approaches are discussed only when relevant for a contextual comparison. Particular attention is given to methods that explicitly or implicitly integrate DA and CL mechanisms, even when such integration is not the primary focus of the original study.

Specifically, this review aims to:

- analyze the sources and characteristics of evolving data distributions in bearing prognostics,
- review DA techniques relevant to cross-condition and cross-domain RUL prediction,
- review CL strategies applicable to sequential and long-term prognostic scenarios,
- examine existing studies that combine or bridge domain adaptation and CL concepts, and
- identify open challenges and future research directions toward reliable lifelong prognostic systems.

The remainder of this paper is organised as follows. Section 1 introduces the characteristics of evolving data distributions in bearing monitoring. Section 3 describes the methodology adopted for this systematic literature review. Sections 4 and 5 review the DA and CL methods, respectively, with an emphasis on their relevance to RUL prediction. Section 6 examines the approaches that integrate these paradigms under a unified Domain-Adaptive Continual Learning framework. Finally, Section 7 discusses the open challenges and outlines potential directions for future research.

2. Background – Related Works

Over the past decade, numerous reviews have addressed PHM, bearing fault diagnosis, and RUL prediction using traditional deep learning models under stationary operating conditions. Several surveys have also examined transfer learning and DA techniques for cross-condition and cross-machine prognostics, highlighting their effectiveness in mitigating distribution discrepancies between laboratory and field data. In parallel, a growing body of literature has reviewed CL theory and catastrophic forgetting, mainly from a general machine learning perspective, with limited focus on PHM applications.

However, existing reviews typically treat DA and CL as independent research directions and do not explicitly investigate their integration within a unified, lifelong prognostic framework. To the best of our knowledge, there is currently no comprehensive review that systematically analyzes Domain Adaptive Continual Deep Learning (DA-CL) for bearing fault prognosis and RUL estimation under evolving data distributions. Therefore, this review aims to fill this gap by synthesising recent advances at the intersection of DA and CL and critically examining their convergence toward scalable, robust, and deployable lifelong PHM systems.

3. Methodology

In this study, we adopted a literature review methodology in accordance with the PRISMA 2020 guidelines [42], [43], [44]. The PRISMA framework was selected to ensure the transparency of the methodology, reproducibility of the results, and rigor of the review process. The selection process included identification and screening of studies, eligibility assessment and final inclusion and exclusion criteria. The review followed a multi-stage process consisting of database searching, duplicate removal, title and abstract screening, full-text screening, and qualitative synthesis of selected studies. The overall workflow is illustrated in the following diagram.

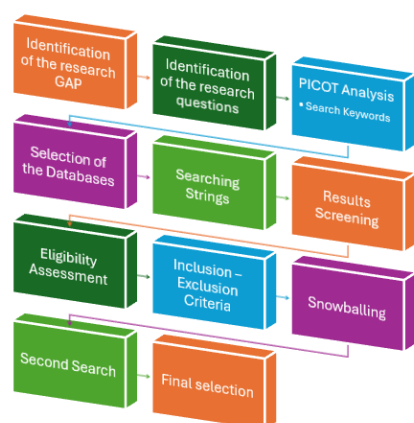


Figure 1. PRISMA-based workflow illustrating the identification, screening, eligibility assessment, and final selection of studies included in this review.

3.1. Research Objectives and Review Scope

Despite the significant advancements in DL models for RUL prediction, static models still have limitations in adapting to the dynamic and nonstationary nature of real-world industrial and marine environments. Although the existing literature addresses domain shifts through DA and TL and temporal evolution through CL in isolation, high-reliability applications increasingly require a synergistic approach to simultaneously mitigate distribution mismatches and adapt to incremental data streams without catastrophic forgetting. Furthermore, practical constraints, such as the scarcity of labelled data in target domains, privacy concerns, and the computational overhead of retraining, necessitate a rigorous examination of how these methodologies can be integrated to foster robust, lifelong prognostic frameworks. To systematically explore this emerging paradigm and evaluate its efficacy in advancing bearing fault prognosis under evolving conditions, this review critically analyzes and synthesises existing research on **“Adaptive Continual Deep Learning for Ball Bearing**

Fault Prognosis and Remaining Useful Life Estimation under Evolving Data Distributions". This review was guided by the following research questions:

RQ1: To what extent have Domain Adaptation techniques been successfully applied to bearing fault prognosis, and what are their underlying assumptions that limit their applicability in continually evolving environments?

RQ2: To what extent do continual learning methods address catastrophic forgetting in PHM, and how effective are they in enabling long-term adaptability for bearing RUL estimation?

RQ3: How can the integration of Domain Adaptation and Continual Learning principles into a unified deep-learning framework overcome the challenge of evolving data distributions for robust and adaptive Remaining Useful Life (RUL) estimation in bearing fault prognosis?

These research questions guided the review structure and classification of selected studies.

3.2. PICOT Framework

To ensure a structured and reproducible search strategy, the research questions were further operationalised using the PICOT framework, which is widely adopted in systematic reviews to define the core conceptual components of the investigation [45], [46].

3.3. P (Population / Problem):

Rotating machinery components, particularly focusing on rolling element bearings.

Bearings are critical mechanical elements subject to wear, fatigue, and degradation, often exhibiting nonstationary and evolving fault patterns that lead to unpredictable RUL and unplanned downtime.

3.4. I (Intervention):

Self-updating and adaptive deep learning models are capable of continuously learning from streaming condition-monitoring data, including online, continual, and adaptive learning.

3.5. C (Comparison):

Traditional deep learning models, domain-adaptive trained offline, and continual learning individual frameworks are deployed without adaptation to distributional changes or evolving operating conditions.

3.6. O (Outcomes):

Primary outcome:

Improved accuracy, robustness, and generalisation of remaining useful life (RUL) prediction under non-stationary and evolving data distributions.

Secondary outcomes:

- Reduced catastrophic forgetting in continual learning scenarios
- Real-time adaptability for online condition monitoring
- Improved reliability and safety of industrial predictive maintenance strategies

T (Time):

Real-time or near-real-time condition monitoring over the operational lifetime of machinery.

The search keywords listed in **TABLE I** were derived from the above analysis.

Table 1. PICOT-BASED KEYWORD GROUPS WERE USED FOR THE LITERATURE SEARCH.

Application Domain (P)	bearing rotating machinery
Prognostics Task (O)	RUL prediction Fault prognosis Health degradation

	Predictive maintenance
Learning Paradigm (I/C)	Deep learning Data driven models Domain adaption Incremental learning Continual learning

Table 2 presents the final search strings and the corresponding databases, while Columns 3 and 4 report the total number of retrieved records and their distributions over time.

Given the rapid evolution of the specific paradigm in RUL prediction, a literature search was conducted in two stages to ensure the temporal validity and currency of the review.

- Initial search: conducted in August 2025
- Update search: conducted five months later in January 2026, prior to manuscript submission.

The updated search used the same databases and eligibility criteria as those of the initial search. The search strings were slightly modified to focus on RQ3 and the integration of DA and CL into a unified framework. Only newly published records were considered during the update phase of the study.

Considering that only a limited number of studies addressing the integration of DA and CL within a unified framework were identified during the database search, and to reduce the risk of influential or emerging studies, a snowballing strategy was additionally employed. Snowballing was conducted by screening the reference lists of selected studies. All records identified through snowballing were subjected to the same eligibility criteria and screening processes as the database-derived records. The same process was applied to both the search stages.

3.7. Eligibility Criteria

The inclusion and exclusion criteria were defined prior to the search process to ensure consistency and objectivity in study selection. The selected papers must meet the following criteria:

3.8. First Stage – Initial Selection

- a. They should be published after 2020
- b. Duplicates n0=114
- c. Language English n1=0
- d. Available for download n2=5
- e. The articles should be published in a peer-reviewed journals and papers n3=0
- f. The paper should be an article n4=9
- g. Analysis by title, irrelevant to the research questions n5= 114

3.9. Second Stage - Analysis by Abstract, Introduction, Conclusion, Contributions, Keywords

- a. Empirical, theoretical, or methodological contributions
- b. Relevant to the research questions
- c. Key words like Adaptive Learning, Few Shots, Incremental (Continual) learning and Domain adaptation should be all included in the paper
- d. Contribution to the specific field of Domain adaptation, Continual Learning and their integration in a unified paradigm.
- e. Excluded n6= 100

3.10. Selection of Papers - Third stage - Complete Reading of the Papers. Extraction of Answers Related to Research Questions

- a. Relevant to the research questions
- b. The papers must include tables and graphs with experimental results and conclusions.

c. The research should utilise at least one well-known, publicly published university ball bearing dataset, such as NASA's bearing dataset, the Intelligent Maintenance Systems (IMS) dataset, or IEEE PHM2012.

d. If the researchers used a private dataset, details on the equipment, instrumentation, and data capturing methodology must be mentioned in the article.

e. Excluded n7= 26

3.11. Snowballing

Included n8= 5

Second search stage

Included n9= 5

Fig.2 illustrates the identification, screening, and selection of the included studies.

Table 2. Database search results and temporal distribution of the retrieved studies.

First Stage – Aug 2025			
Science Direct	("transfer learning" OR "domain shift" OR "domain adaptation" OR "adaptive learning" OR "Incremental learning" OR "Few-shot learning" OR "variable operating condition") AND "remaining useful life" AND "bearing"	60	2020:1 2021:6 2022:8 2023:10 2024:16 2025:18 2026:1
IEEE Xplore	(("All Metadata": transfer learning OR "All Metadata": domain shift OR "All Metadata": domain adaptation OR "All Metadata": adaptive learning OR "All Metadata": Incremental learning OR "All Metadata": few-shot learning OR "All Metadata": variable operating condition*) AND ("All Metadata": remaining useful life) AND ("All Metadata": bearing*))	192	2020:7 2021:23 2022:30 2023:40 2024:57 2025:35 2026:3
Scopus	TITLE-ABS-KEY (("transfer learning" OR "domain shift" OR "domain adaptation" OR "adaptive learning" OR "incremental learning" OR "Continual learning" OR "Few-shot learning" OR "variable operating condition") AND "bearing*" AND "remaining useful life")	163	2020:4 2021:19 2022:26 2023:31 2024:50 2025:33 2026:0
TOTAL		415	
Second Stage – Jan 2026			
Science Direct	("transfer learning" OR "domain shift" OR "domain adaptation") AND ("Incremental learning" OR "Continual Learning" AND "Catastrophic forgetting") AND "remaining useful life"	3	
IEEE	(("All Metadata": transfer learning OR "All Metadata": domain shift OR "All Metadata": domain adaptation) AND ("All Metadata": Incremental learning OR "All Metadata": Continual Learning OR "All Metadata": Catastrophic Forgetting) AND ("All Metadata": remaining useful life) AND ("All Metadata": bearing*))	2	
Total:		5	

4. Tackling Domain Shift: A Review of Domain Adaptation in PHM

Accurate RUL prediction is a cornerstone of PHM, as it facilitates the transition from conventional maintenance practices to condition-based and predictive maintenance strategies for rotating machinery. However, the efficacy of data-driven RUL models rests on a fundamental and often unrealistic assumption that the training (source) and field (target) data follow the same statistical distributions [23], [47]. In real-world industrial and marine environments, this assumption rarely holds owing to variations in operating conditions, sensor configurations, environmental factors, and component aging [19], [21], [29]. This discrepancy, known as a domain shift, leads to significant performance degradation in static deep learning models [8], [21]. In this context, DA has emerged as an effective approach for mitigating domain discrepancies and maintaining robust RUL prediction performance across various operational environments [19], [29].

Error! Reference source not found. provides an overview of the main DA paradigms considered in this review, highlighting their scope and inherent limitations under evolving operating conditions [21], [29], [47].

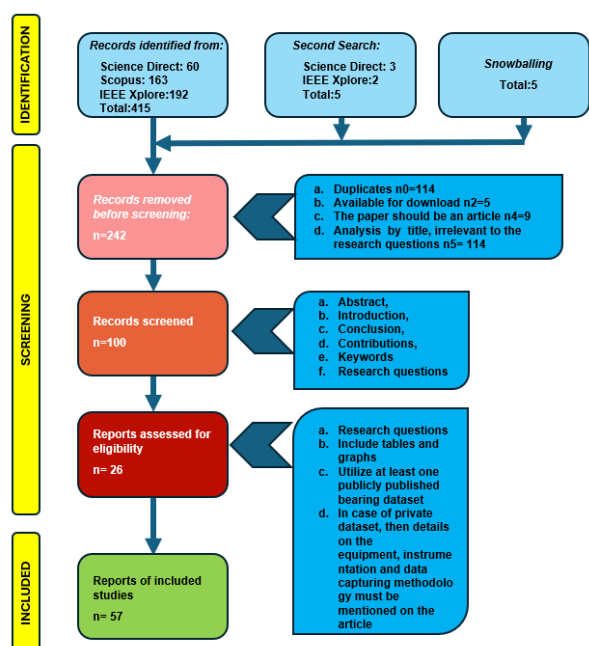


Figure 2. Flow diagram of the study selection process, including database search, screening, eligibility assessment, and inclusion of the studies.

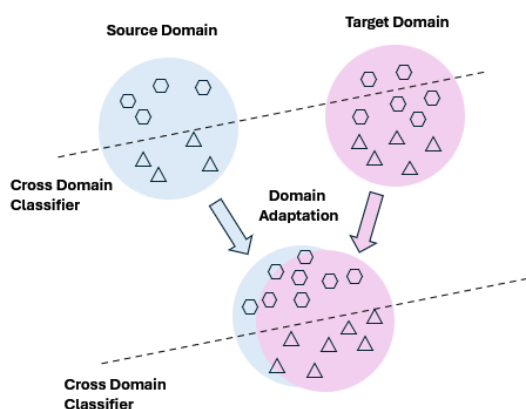


Figure 3. Overview of domain adaptation paradigms in PHM and their applicability under evolving operating conditions.

4.1. Domain Shift in PHM Applications

In PHM, a domain shift, also referred to as a distribution discrepancy, is defined as the discrepancy between the statistical distributions of data collected under different operating

conditions or environments [19], [23], [29]. For rotating machinery and bearings, such shifts commonly arise from the following reasons:

- Variations in the rotational speed, applied load, and duty cycles [8].
- Changes in ambient conditions such as temperature, humidity and noise [48]
- Sensor replacement, recalibration, reconfiguration or drift [21]
- Differences in machinery configurations or manufacturing tolerances [47].
- Progressive component degradation and maintenance interventions [8].
- Incomplete or imbalanced data [23].

These factors generate non-stationary data streams in which both the marginal input distribution $P(x)$ and conditional distribution $P(y | x)$ may change over time. Consequently, models trained on historical datasets or controlled laboratory conditions frequently fail to generalise to real-world applications, motivating the need for adaptive learning strategies [17], [19], [21].

4.2. Fundamentals of Domain Adaptation

DA can be viewed as a specific and constrained form of the broader Transfer Learning (TL) paradigm. TL aims to leverage the knowledge gained from one or more source tasks or domains to improve the learning performance in a related but different target task or domain. Unlike DA, which specifically focuses on mitigating distribution discrepancies between the source and target domains for the same task, transfer learning can involve transferring knowledge across different but related tasks, models, or domains. TL methods include fine-tuning pretrained models, feature extraction, and parameter sharing, and are widely used in PHM to address data scarcity, improve model generalisation, and accelerate training under new operating conditions [6], [8], [9], [29].

DA can be viewed as a specialised form of TL that addresses learning problems in which a source (S) and target (T) domain share the same prediction task (i.e. fault detection, classification, and RUL detection) but follow different data distributions, $P(S) \neq P(T)$ [19], [23]. Although the task is identical ($f: X \rightarrow Y$), the data between the two domains appear to be statistically different [8]. In PHM applications, the source domain typically consists of fully labelled running data and sometimes run-to-failure data collected under known conditions, whereas the target domain contains unlabelled or partially labelled operational data under unknown or evolving conditions [8], [21], [29].

The primary objective of DA is to help models “ignore” domain-specific variations and focus on domain-invariant representations or transferable model parameters, such that knowledge acquired from the source domain can be effectively applied to the target domain. This is particularly valuable in RUL prediction, where labelling target domain degradation data is often infeasible owing to cost, safety, and time constraints [11], [19], [47].

4.3. Categories of Domain Adaptation Methods in PHM

Instance-Based (Reweighting) Domain Adaptation

Instance-based DA methods attempt to mitigate distribution discrepancies by adapting source domain weights so that the reweighted product better matches the target domain. Although this strategy may be proven effective in some scenarios, it relies on the assumption that the source and target distributions have sufficient overlap. Moreover, the performance is sensitive to inaccurate weight estimation, limiting its robustness under severe domain shifts [9], [21].

4.4. Feature-Based Domain Adaptation

Feature-based DA methods aim to learn shared representations between source and target domains by extracting domain-invariant feature representations.

Common techniques include:

Distribution discrepancy minimisation, where statistical metrics such as Maximum Mean Discrepancy (MMD), Wasserstein distance, or Sinkhorn divergence are leveraged to quantify and subsequently minimise the distribution mismatch between labelled and target feature spaces [8], [29].

Adversarial learning involves a “min-max” game, where a domain discriminator is trained to distinguish source-labelled target features, whereas the feature extractor learns to confuse it. As a

result, the model learns domain-invariant feature representations that seem to be the same regardless of whether they belong to the training dataset (source) or the field (target) [19], [25], [49].

Subspace alignment and manifold learning are achieved by projecting the source and target data into a latent space common to both parties [9], [50].

The above approaches have been widely adopted in bearing fault prediction and RUL estimation, demonstrating improved cross-condition and cross-machine generalisation on benchmark datasets such as PHM2012 and XJTU-SY.

4.5. Parameter-Based Domain Adaptation

In Parameter-based DA techniques, features extracted from one or more pretrained source domain models are adapted to construct a target domain model via fine-tuning or regularisation. This approach is computationally efficient; however, it has the drawback that it requires at least a small amount of labelled target data and is prone to overfitting or negative transfer when domain discrepancies are substantial [11], [19].

4.6. Hybrid Domain Adaptation Approaches

Hybrid approaches combine feature-level alignment with parameter adaptation to exploit the complementary strengths of the two methods. These methods typically achieve improved robustness under moderate domain shifts but still rely on static alignment assumptions [11], [17], [48].

4.7. Representative DA Applications in Bearing RUL Prediction

Recent studies have demonstrated the effectiveness of DA in bearing fault prognosis under various operating conditions. Adversarial DA frameworks with weighted loss functions have been shown to enhance positive transfer while mitigating negative transfer effects. Multi-source DA approaches leverage knowledge from multiple related source domains to improve generalisation under previously unseen conditions. Additionally, self-supervised and entropy-minimisation-based DA methods aim to reduce the reliance on labelled data while enhancing feature transferability.

Zhang et al. (2025) introduced the Joint Domain-Adaptive Transformer (JDATransformer) model, which enhances cross-domain RUL prediction by aligning global feature distributions within the transformer framework. By leveraging attention mechanisms and joint maximum mean discrepancy, the JDATransformer addresses the challenges of distribution shift and degradation pattern learning [29]. *Ye et al. (2025)* introduced a novel Adaptive MAGNN-TCN. The framework combines multi-adaptive graph neural networks (MAGNNs) with temporal convolutional networks (TCNs) to directly model the interdependencies from raw sensor data. The MAGNN framework utilises a dynamic adjacency matrix that adjusts to reflect changing bearing operational states, thereby ensuring accurate predictions under varying conditions. The success of this approach lies in leveraging adaptive graph structures and multi-scale feature extraction to effectively capture the complex relationships inherent in the sensor data of mechanical systems [48]. *She et al. (2025)* proposed a transferable meta-learning with a task-adaptive loss function (TMeTAL) framework for RUL prediction. TMeTAL incorporates affine transformations to dynamically update the loss function in the inner loop, thereby enhancing adaptability to new tasks. The multi-kernel maximum mean difference (MK-MMD) is used to quantify distribution discrepancies across tasks and to improve parameter acquisition. The proposed method demonstrated superior performance in RUL prediction with small sample sizes and varying distributions [38]. *Shang et al. (2025)* introduced a multi-source adversarial distillation DA (MADDA) network. This method leverages a novel Source Aggregation Strategy (SAS) and Source Distillation Weighting Mechanism (SDWM) to automatically select relevant source domains and samples based on their similarity to the target domain. By emphasising the source information most pertinent to the target task, the model aims to enhance prediction accuracy [21].

Moreover, *Mao et al. (2025)* proposed a dynamic modelling-assisted tensor regression transfer learning method that integrates physics-based simulation models and self-supervised online data. The methodology involves pre-training a deep tensor domain adversarial model and using simulation libraries for various damage scenarios in conjunction with real-time data for online

prediction. By integrating a dynamic modelling mechanism and offline data, the proposed model demonstrated improved prediction accuracy for RUL prediction [11]. *Lu et al. 2025*, combines dynamic hybrid DA (DHDA) and attention contrastive learning (A-CL). The proposed model addresses the limitations of existing approaches by focusing on fine-grained information transfer between degradation features across domains and target domain specificity. The researchers also introduced an Enhanced Residual Convolutional Module (ERCM) as a feature extractor, which offers improved degradation feature extraction capabilities [17]. *J. Yang et al. (2025)* introduced a self-supervised dual-path meta-alignment (DPMA) network to address the challenge of RUL prediction under limited labelled data. The DPMA architecture incorporates meta-learning, which involves training a model to quickly learn from diverse prior tasks and adapt to new tasks within a limited sample space. Employing Bayesian inference, this study aims to enhance degradation feature learning, integrate domain space alignment techniques, and leverage lightweight self-attention networks. The lightweight self-attention network helps filter valuable information via nonlocal blocks, thereby improving the model's capacity to learn hidden variables, even with limited data [39]. *Han et al. (2025)* introduced a novel Sinkhorn Divergence-based Contrast Domain Adaptation (SD_CDA) method tailored for predicting the RUL of bearings. Leveraging an adversarial training framework, a temporal mixup strategy for data augmentation, and momentum contrast (MoCo) for mutual information extraction, the proposed approach effectively aligns the data distributions between domains. The incorporation of Sinkhorn divergence enhances model transferability, addressing the challenges faced by existing DA methods [50]. *Guo et al. (2025)* introduced a novel deep-learning dual-channel framework called CANN-GT-BDA, which leverages transfer learning and uncertainty quantification for an accurate RUL. By integrating AlexNet, GCU GCU-Transformer, and Bayesian optimisation techniques, the model effectively captured degradation patterns from time-series data while addressing domain shifts between datasets [47]. *Chen et al. (2025)* introduced a self-supervised knowledge distillation framework based on mutual learning. The proposed method utilises a teacher-student architecture that transfers knowledge via pseudo-labels and employs a mutual learning strategy to iteratively refine the teacher model and prevent source domain overfitting. Furthermore, the framework integrates feature-level domain adversarial training to decouple cross-domain features and utilises a sparse attention mechanism to efficiently extract degradation patterns [19].

Earlier studies have further highlighted the diversity of DA strategies for RUL prediction. *Zhang and Wang 2024*, introduces a Deep Subdomain Adaptation Time-Quantile Regression Network (DSATQRN) for cross-domain prediction of bearing RUL prediction. By combining deep subdomain adaptation for feature alignment with a temporal quantile regression network, the model effectively compresses the uncertainty intervals and captures the data-time correlations. The model utilises kernel density estimation for probabilistic distribution results [23].

Kumar et al. (2024) proposed an unsupervised DA strategy called Entropy-based Deep Domain Adaptation (EDDA), which aligns feature distributions by minimising the entropy of predictions in the target domain. The architecture utilises a Convolutional Neural Network (CNN) for feature extraction and optimises a combined loss function that minimises both the RUL prediction error on labelled source data and the entropy on unlabelled target data [9]. *Sun et al. 2023*, proposes a Domain-Adaptive Adversarial Network (DAAN) to predict RUL under various operating conditions. The method involves utilising Wavelet Packet Transform (WPT) for signal preprocessing and constructing two-dimensional data from the raw data. By leveraging transfer learning and the maximum mean discrepancy (MMD) algorithm, the network effectively maps vibration data from diverse operating conditions to a similar distribution, enabling optimal RUL prediction without requiring abundant labelled data from different operating conditions [6]. *Ding et al. 2023*, propose a multi-source domain generalization learning method that extracts generalized degradation feature representations from multiple available offline datasets collected under different known conditions. The model architecture uniquely combines Gated Recurrence Unit (GRU) and transformer structures to capture robust temporal features that are invariant across domains [14]. Finally, *Chou et al. (2023)* proposed a Generative Neural Network-Based Online Domain Adaptation (GNN-ODA) framework designed to solve RUL prediction tasks in which target domain data are incomplete or limited to

early stage degradation. The model employs a modified Auxiliary Classifier Generative Adversarial Network (ACGAN) to synthesise virtual run-to-failure paths, effectively reconstructing the missing late-stage failure data for the target domain. To address distribution shifts, an Online Domain Adaptation module continually aligns the feature distributions of the source domain with the augmented target data [8].

4.8. Open-Set and Unsupervised Domain Adaptation Challenges and Limitations of Domain Adaptation in Dynamic Environments

Traditional DA approaches usually operate under the premise of a closed set, which means that the label spaces of the source and target domains are identical. In real-world PHM applications, however, target domains may have fault modes that have never been observed before or incomplete degradation patterns. This makes open-set domain adaptation (OSDA) challenging (36). In addition, unsupervised DA (UDA) which is trained on large-scale annotated source datasets and applied to unlabelled target data, typically assumes a static target domain. Therefore, these models struggle in dynamically changing domains over time because adaptation is usually performed offline, and the models cannot continuously update themselves after deployment [49].

Although the DA effectively mitigates static distribution discrepancies, several critical limitations remain.

- A primary limitation of traditional DA methods is the assumption that the data in the source and target domains are balanced and complete, which means that all fault categories are present. However, in real-world online monitoring, the target data are often acquired sequentially. At the beginning of monitoring, only healthy data may be available, while fault data has not yet been generated [8]
 - Most DA methods rely on static alignment strategies, which are ineffective under continuous domain evolution [11], [19]
 - Adaptation is typically performed once per target domain, preventing lifelong learning. Most of the DA techniques assume that the target domain's data distribution is fixed and fully available during training, which prevents the model from engaging in lifelong or CL as the machinery's condition evolves over time [8], [11], [39]
 - Multitask learning and knowledge accumulation across sequential domains are neglected, leading to catastrophic memory losses. Moreover, in practical applications, machinery may operate under various new working conditions or equipment configurations that were not available during training [14], [16].
 - Most deep-learning-based DA methods focus solely on improving point prediction accuracy and do not account for the uncertainty inherent in the stochastic degradation processes of machinery components. This results in prediction results with unknown credibility, which affects decision-making for predictive maintenance [23]

Therefore, DA models often exhibit degraded performance when confronted with sequentially emerging operating conditions, highlighting the need for learning frameworks that support both distribution alignment and continual adaptation.

5. Learning Sequentially: A Review of Continual Learning in PHM

5.1. Motivation for Continual Learning in PHM

In real-world industrial and merchant vessel applications, machinery operates in highly dynamic environments characterised by diverse working conditions and equipment configurations, where machinery degradation evolves continuously over time [12], [20]. This operational variability introduces substantial challenges for data-driven prognostic models that are tasked with estimating the RUL prediction, as the statistical properties of the newly input sensor data streams often deviate from those of the models that were initially trained. Consequently, models trained on limited or static datasets struggle to generalise effectively to new, unseen domains, leading to a degradation in their predictive performance [22]. This phenomenon, as mentioned in the introduction, is known as a data or domain shift [31].

One way to mitigate domain drift is to retrain the model from scratch every time new data become available. However, this is impractical owing to the computational cost, limited available data, risk of losing valuable historical information, and, in some cases, data privacy or ownership constraints [20], [31], [37]. Ideally, PHM models should emulate human cognition learning, being capable of accumulating additional knowledge sequentially, reusing, and complementing what has already been learned in the past [20], [31], [37]. In this context, PHM models can leverage previously learned representations and efficiently adapt to novel degradation patterns without discarding prior experience, motivating the emergence of the CL paradigm, also known as lifelong or incremental learning [28]. CL aims to enable models to learn from a sequence of tasks, domains, or data distributions while retaining performance on previously encountered ones [28], [51]. However, despite extensive research efforts, a key challenge remains: the inherent plasticity–stability dilemma, which lies at the core of CL theory.

5.2. The Plasticity–Stability Dilemma and Catastrophic Forgetting

The plasticity-stability dilemma can be defined as the trade-off between a model's ability to learn new knowledge (plasticity) and its ability to retain previously acquired knowledge (stability) [20], [34]. When a model is updated to fit new data distributions, its weights and internal parameters are modified [22], [34]. However, even though plasticity is essential for adaptation, the performance of the models may drop dramatically when their parameters are excessively modified and previously learned representations are overwritten [37]. This phenomenon is known as catastrophic forgetting [1], [22]. Catastrophic forgetting refers to the inevitable and often irreversible decline in performance when a model overwrites or modifies the knowledge of previously observed domains (tasks or classes) while sequentially learning new domains [16], [52].

In PHM applications, this is particularly problematic because historical degradation patterns and failure modes are often critical for accurate RUL estimation. The loss of this knowledge can severely compromise prediction reliability and safety [31], [52], [53]. Depending on the application, it is crucial that the model either adapts quickly to newly acquired data (requiring high plasticity) or preserves most of the past experience (maintaining a high degree of stability) [28]. Therefore, the plasticity-stability dilemma is determined by the extent to which a model can update its internal parameters and weights to learn new distributions without forgetting previous experiences, thereby mitigating catastrophic forgetting [20], [22]. **Error! Reference source not found.** illustrates the fundamental trade-off between stability and plasticity [54]. The x-axis represents stability increasing from left to right, and the y-axis represents plasticity increasing from bottom to top. The Pareto Frontier depicts the sets of optimal trade-offs between stability and plasticity [54]. The curve represents the optimal plasticity-stability equilibrium, where improvement in one dimension necessarily compromises the other [54], [55]. In RUL prediction, the Pareto Frontier reflects the set of attainable compromises between retaining previously learned degradation knowledge and adapting to newly observed operating conditions. Its shape is not predefined but instead emerges empirically and is inherently dependent on the dataset, learning paradigm, and CL strategy applied to it.

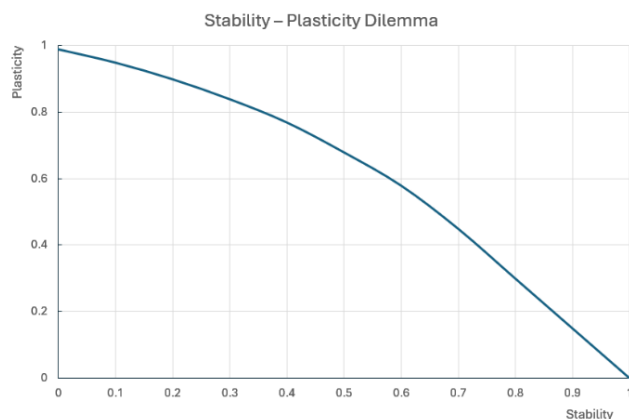


Figure 4. Illustration of the plasticity–stability trade-off in continual learning, highlighting the Pareto frontier for optimal adaptation and knowledge retention.

The severity of catastrophic forgetting is strongly influenced by the availability of data. In practice, access to historical data may be restricted because of storage limitations, privacy regulations, or data loss [22], [56]. Even when full data availability is theoretically possible, the computational cost of retraining models with all the accumulated data increases significantly over time [13], [20]. Consequently, CL methods in PHM must carefully consider strategies for storing, replaying, or approximating past data to maintain performance while learning sequentially.

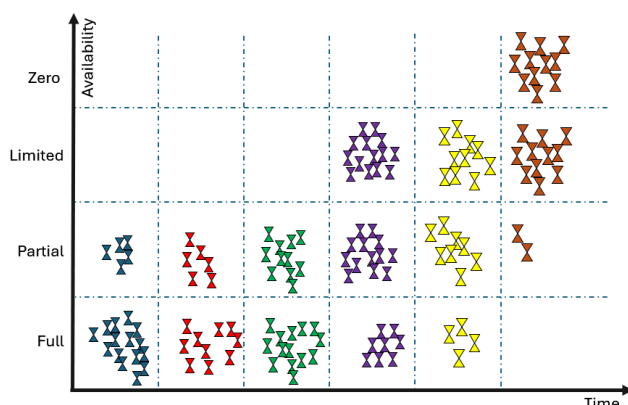


Figure 5. Data availability regimes in continual learning scenarios for PHM applications.

5.3. Core Continual Learning Strategies

CL has emerged as a promising paradigm that enables models to incrementally acquire and integrate new knowledge without retraining from scratch or requiring access to all historical data. This capability is particularly relevant in machinery prognostics, where operating conditions and degradation patterns evolve over time. However, to mitigate catastrophic forgetting in machinery prognostics, a variety of learning strategies have been proposed, which can be broadly categorised into three groups: regularisation-based, replay-based, and architecture-based approaches [16], [20], [57].

5.4. Regularization-Based Methods

Regularisation-based methods mitigate forgetting by constraining updates to important parameters to retain previously learned tasks. The core idea is to modify the loss function with an additional regularisation term that penalises changes to critical weights, thereby preserving past knowledge while allowing limited plasticity [20].

Seminal approaches include Elastic Weight Consolidation (EWC), memory-aware synapses (MAS), and Synaptic Intelligence (SI), which estimate parameter importance based on Fisher information or weight update trajectories, respectively. Other techniques, such as Learning without Forgetting (LwF), leverage knowledge distillation to preserve implicit representations of old tasks without storing raw data [20], [34], [40].

These methods are attractive for PHM applications because they do not require the explicit storage of historical datasets, thereby reducing memory and privacy issues. However, when there is a substantial difference between the previous and new tasks or operating conditions, their effectiveness decreases. In such cases, overly restrictive regularisation can hinder adaptation, leading to suboptimal RUL predictions for novel degradation patterns [13], [40].

5.5. Replay-Based Methods

Replay-based methods address catastrophic forgetting by reintroducing important task information from previous subsets during training on the new data. This is achieved by either storing representative samples (experience replay) or generating synthetic data that approximate the past distributions (generative replay) [40], [51].

Representative methods include Incremental Classifier and Representation Learning (iCaRL), Gradient Episodic Memory (GEM), and REMIND, which select or compress samples to maintain a memory buffer of previous experiences. Generative replay (GR) approaches employ generative models, such as GANs, to synthesise pseudo-samples that preserve historical knowledge without storing raw data [13], [56].

Replay-based methods generally achieve superior performance in mitigating forgetting and have demonstrated strong empirical results even in complex CL settings. However, their practical applicability in PHM is constrained by memory requirements, making them not highly scalable, computationally intensive, and challenging to accurately approximate complex degradation distributions with limited samples [16].

5.6. Architecture-Based and Parameter Isolation Methods

Architecture-based methods mitigate forgetting by allocating separate model components or parameter subsets (i.e. weights) to different tasks. By structurally isolating task-specific knowledge, these approaches preserve previously learned representations while allowing the model to retain plasticity when learning new tasks. Examples of representative networks include Progressive Neural Networks (PNNs), PackNet, and Dynamically Expandable Networks (DENs) [40], [56]. In general, these methods preserve old knowledge by either isolating parameters that are considered unimportant for past tasks, retaining spare connections for future learning, or expanding or sparing model capacity to accommodate new tasks.

Although these approaches effectively prevent catastrophic forgetting, they suffer from poor scalability because the model complexity often increases linearly with the number of tasks or operating conditions. This limitation can be partially mitigated by applying methods that add only subcomponents to the model or explicitly seek to reuse previously learned knowledge [20], [34]. However, in PHM systems with long operational lifetimes and continuously evolving conditions, such growth remains impractical because of the storage requirements, increased inference latency, and maintenance constraints. Additionally, many architecture-based methods often assume that the task identity is known during inference, which is rarely the case in real-world PHM scenarios [16], [22].

5.7. Continual Learning Scenarios in PHM Applications

CL approaches are often categorised based on the evolution of tasks, domains and classes over time. In the context of PHM, three CL scenarios are particularly relevant in PHM applications: domain-incremental learning, class-incremental learning, and task-incremental learning.[33], [34], [40], [58]:

Domain-Incremental Learning represents the most relevant scenario in which the prediction task remains unchanged (for example, RUL estimation); however, the data distribution shifts owing to varying operating conditions, environments, or sensor configurations. This setting is particularly common in rotating machinery and marine systems, where gradual or abrupt changes in the operating regime occur over long-term deployments.

Class-Incremental Learning arises when new types of faults or degradation modes appear over time, requiring the model to recognise previously unseen classes without access to data from earlier tasks. This is challenging in the context of PHM because of the scarcity of labelled fault data and the high costs associated with failure events.

Task-Incremental Learning involves sequential exposure to entirely new prediction tasks, such as transitioning from fault detection to RUL estimation or adapting a model to a different application domain. Although less common in operational PHM systems, this scenario highlights the need for modular and adaptable architectural design.

Recent research has explored various CL strategies to address these scenarios and mitigate catastrophic forgetting during the sequential data acquisition. *Shang et al. (2026)* introduced novel meta-contrastive learning tailored for semi-supervised learning. The proposed method is designed to ensure that the knowledge gained from the data contributes positively to the model performance with limited labelled data. To attenuate catastrophic forgetting, the authors introduced a meta-update

strategy that reduces the number of parameters involved in the meta-training and semi-supervised phases. Finally, to eliminate the need for manual tuning in the semisupervised phase, a weight optimisation strategy was proposed. This strategy treats the weights as learnable parameters and incorporates them into the outer loop of the meta-training phase [51]. *Lin et al. (2025)* in their research proposed a novel CL framework termed Diffusion-Integrated Dynamic Mixture Experts (DIMIX), designed to enhance fault diagnosis in dynamic and unknown environments. The model integrates a dynamic mixture-of-experts (MoE) architecture with a diffusion-based generative replay mechanism to mitigate catastrophic forgetting by synthesising pseudo-samples from known domains. It also includes a dynamic incremental clustering module (DICM) to autonomously identify and adapt to new and unknown operational conditions [28].

Regularisation-based strategies have also been widely adopted in PHM-oriented CL. *Zhou J and Qin Y. (2025)* proposes a Multistage Attention Convolutional Neural Network (MSACNN) designed to progressively refine feature representations using multiple attention mechanisms. To prevent catastrophic forgetting, the authors proposed a Knowledge Weight Constraint (KWC) mechanism, which acts as a regularisation term based on the importance of weight parameters and gradient information [13]. Similarly, *Zhou et al. (2025)* introduced a novel approach called the Multibranch Horizontal Augmentation Network (MBHAN). The core model, called the Time-Frequency Fusion Temporal Convolutional Network (TFFTCN), integrates a hierarchical self-attention (HSA) mechanism to capture multi-scale degradation features from both the time and frequency domains. To mitigate catastrophic forgetting, the method introduces a Memory Weight Constraint (MWC) regularisation term that penalises changes to important parameters from previous tasks [40]. *Benetia et al. (2025)* introduced Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN) architectures to effectively capture both global dependencies and localised feature patterns from sensor data. To mitigate catastrophic forgetting, the framework employs regularisation-based methods, specifically Elastic Weight Consolidation (EWC) and Synaptic Intelligence (SI), alongside a replay mechanism that retains crucial information from previous tasks [20]. *Wang et al. (2024)* introduced a Deep Residual Reservoir Computing (DeepESN) framework with a novel deep echo state network and residual block structure to enhance the CL model performance and reduce the model's training difficulties. To solve catastrophic forgetting, this study employs a regularisation-based Elastic Weight Consolidation (EWC) method [12].

Replay-based and hybrid approaches have further demonstrated their effectiveness in PHM scenarios. *Ren et al. (2024)* proposed a complementary CL framework for RUL prediction based on a Recurrent Convolutional Neural Network (RCCN) composed of stacked recurrent convolutional and fully connected layers. To mitigate catastrophic forgetting, the proposed framework employs a "long-term potentiation" mechanism that preserves pivotal memories by adjusting synaptic plasticity, ensuring that synapse weights crucial to previous learning remain stable. This is complemented by an "associative replay" technique, which consolidates old knowledge by selectively storing and replaying samples based on their informativeness, diversity, and novelty [30]. In complementary research, *Ren et al. (2024)* proposed a prognostic CL network built with "multi-kernel swarm convolution" (MKSC) blocks. The MKSC blocks allow the model to efficiently capture multi-scale degradation features and recalibrate the feature maps based on specific operating condition information. To mitigate catastrophic forgetting, the authors employed a "core space gradient projection" (CSGP) method, which does not strictly restrict weight change like conventional regularisation-based methods but rather guides the direction of gradient descent to avoid catastrophic forgetting [2]. *Ren et al. (2023)* proposed a novel perspective called the adaptive synapse constituted hybrid convolutional network (ASHCN). In the proposed network, each synapse (i.e. trainable parameter) is adaptively plastic; therefore, they can automatically adjust between more stable or more flexible, enabling the ASHCN to balance plasticity and stability during the CL of machinery RUL prediction. It also uses hybrid convolutional layers with multiscale paths and attention mechanisms to enhance feature extraction across varying operating conditions. To attenuate catastrophic forgetting, the authors introduced a hybrid strategy between regularisation-based and parameter importance CL [37]. *Que et al. (2024)* proposed a CL framework based on an Integrated Gated Recurrent Unit (iGRU) model to effectively capture temporal degradation information and

map it to RUL values. To mitigate catastrophic forgetting, the method employs the Orthogonal Weight Modification (OWM) algorithm, which utilises a projector to constrain weight updates to a direction orthogonal to the subspace of previously learned data [59]. *He et al. (2024)* introduced a Wasserstein generative adversarial network with a gradient penalty (WGAN-GP) and Knowledge Distillation. This approach uses a dual old-task retention strategy, where WGAN-GP generates representative old-task data to combat catastrophic forgetting, whereas Knowledge Distillation transfers knowledge from an old-task teacher model to a new-task student model [31]. Finally, *Cao et al. (2023)* proposed a lightweight flattened neural network, called the Temporal Cascade Broad Learning System (TCBLS), designed to capture temporal dependencies and extract both linear and nonlinear features for RUL prediction. To mitigate catastrophic forgetting, the authors employed a ridge regression strategy that calculates network weights analytically (via pseudoinverse) rather than using gradient-based retraining [10].

5.8. Challenges and Limitations of Continual Learning for PHM

CL provides a principled framework for addressing nonstationarity, data drift, and evolving degradation patterns in PHM systems. Effective machinery prognostics require models that can adapt sequentially to emerging operating conditions while retaining knowledge of previously encountered degradation patterns. Catastrophic forgetting remains a central challenge, necessitating incremental learning frameworks that balance plasticity and stability [34], [40], [51].

Despite the promising advances reported in the recent literature, several challenges remain in the application of CL for the PHM and RUL predictions. First, degradation processes often exhibit multiscale temporal characteristics, whereas many CL-based prognostic models rely on fixed scale feature extraction mechanisms. This mismatch limits their ability to capture complex degradation dynamics under variable operating conditions and long-term system evolution [1], [2].

Moreover, current approaches often lack forward knowledge transfer, failing to exploit previously learned representations to accelerate the learning of new degradation modes or operating regimes. This limits scalability and adaptability in long-term deployments, particularly when the models are exposed to frequent or abrupt changes in distribution shifts. In addition, knowledge redundancy and parameter inefficiency are rarely explicitly addressed, leading to constrained learnability as the number of tasks increases [2], [51].

Data-related challenges further exacerbate the plasticity–stability trade-off in PHM applications. Real-world PHM datasets are typically sparse, noisy, and highly imbalanced, with limited availability of labelled failure data. This further complicates the plasticity–non-stationarity trade-off. Under such conditions, achieving sufficient plasticity without sacrificing stability becomes particularly difficult, underscoring the need for CL strategies specifically tailored to prognostic scenarios [28], [30], [37].

From a methodological perspective, each major class of CL strategies exhibits inherent trade-offs [49]. Among the main strategies, regularisation-based methods are lightweight but struggle with significant domain shifts. Replay-based approaches are generally more effective in mitigating catastrophic forgetting but incur significant memory and computational overheads, limiting their scalability. Architecture-based methods preserve prior knowledge through the structural isolation of parameters; however, they often suffer from model growth, increased inference latency, and assumptions regarding task identity [2].

Overall, while no single approach fully resolves the plasticity–stability dilemma, CL remains a critical paradigm for enhancing the robustness, adaptability, and longevity of data-driven prognostic models [28], [30]. Addressing the identified limitations is essential for advancing practical and long-term PHM systems that can operate reliably under evolving conditions.

6. The convergence: Domain Adaptive Continual Deep Learning for Bearing Prognosis & RUL Estimation

6.1. Motivation for Integrating DA and CL in Real World PHM Applications

As discussed in the previous chapters, real-world industrial and marine rotating machinery systems operate in highly dynamic, non-stationary environments owing to variations in load, speed,

environmental factors, sensor configuration, and maintenance interventions. Such changes inevitably induce both distribution shifts across domains and temporal evolution of degradation patterns, which fundamentally challenge fault prognosis and RUL estimation when using conventional data-driven DL based on static learning frameworks [15], [16].

DA and CL have emerged as two complementary paradigms for mitigating these challenges. DA mitigates discrepancies between the source and target domains, enabling knowledge transfer across cross-machine and cross-condition scenarios [19], [29]. Conversely, CL addresses temporal shifts, enabling models to learn sequentially from streaming data while preserving historical knowledge and mitigating catastrophic forgetting [28], [51]. However, when applied independently, both approaches exhibit inherent limitations in their performance. DA methods typically assume a static or batched-accessible target domain and rely on offline adaptation, which limits their ability to handle sequential and incremental domain shifts and open set conditions [15]. In contrast, CL often neglects explicit domain discrepancies and struggles to generalise when new data originate from substantially different domains or distributions [25], [52].

In practical Prognostics and Health Management (PHM) scenarios, bearing monitoring systems must cope simultaneously with cross-domain variations (new operating conditions or different machines) and temporal evolution (which appears independently over time). The synergy of DA and CL within a unified Domain Adaptive Continual Deep Learning (DACL) framework addresses the evolving data distribution problem by formulating it as a sequence of incremental DA tasks with non-stationary target domains [16], [36]. This formulation enables the models to maintain a robust performance across diverse and progressively changing operating conditions.

6.2. Problem Formulation

In the DACL paradigm, rotating machinery prognostic data are received sequentially under evolving operating conditions. Let us define a sequence of domain pairs [5], [36], [56]:

$$\{(D_S^{(t)}, D_T^{(t)})\}_{t=1}^T \quad (1)$$

Where:

T = the number of indexes

$D_S^{(t)}$ = Historical Labelled Data from the source domain (i.e vibration data)

$D_T^{(t)}$ = New Unlabelled Data from the target domain (i.e vibration data from a different machine, or under different load, speed etc.)

At each stage t the model observes source labelled samples $(x_S^{(t)}, y_S^{(t)}) \sim D_S^{(t)}(X_S, Y_S)$, and unlabelled target samples $x_T^{(t)} \sim D_T^{(t)}(X_T)$. While the samples are assumed to be independently and identically distributed within each stage, the underlying domain distributions may vary across stages, $D_T^{(t)} \neq D_T^{(t-1)}$. In this way, the samples reflect non-stationary operating environments and progressive changes in degradation behaviour.

The objective of DACL is to train a sequence of models $f_{\theta_t}: X \rightarrow Y$, which at each stage can align the source and target distributions through DA ($\min L_{DA}(D_S^t, D_T^t)$), and preserve knowledge acquired from previous stages through CL ($L_{CL}(f_{\theta_t}, f_{\theta_{t-1}}) \leq \epsilon$). This is achieved by jointly minimising the task-specific prediction loss, DA loss that reduces distribution discrepancies, and CL regularisation term that constrains parameter updates to mitigate catastrophic forgetting. The formulation enables incremental adaptation to newly observed operating conditions while maintaining a stable bearing RUL prediction performance over previously encountered domains. The overall optimization of the model can be described as $\theta_t = \arg \min_{\theta} (L_{task} + \lambda_{DA} L_{DA} + \lambda_{CL} L_{CL})$

where the factors λ_{DA} and λ_{CL} weigh the relative importance of DA and CL within the DACL model. Error! Reference source not found. illustrates the Task Loss, Domain Adaptation Loss, and Continual Learning Loss interactions.

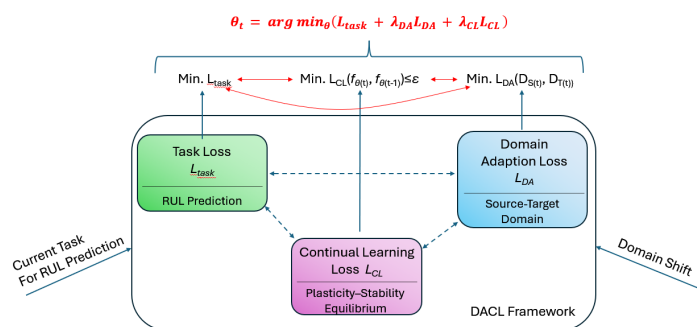


Figure 6. Interaction between task loss, domain adaptation loss, and continual learning regularisation within the DACL optimisation framework.

6.3. Integrated Learning Frameworks

Although there is still no abundance of studies on RUL prediction that use DA and CL unified within a single model, existing studies typically incorporate multiple, complementary mechanisms.

6.4. Inter–Intra Task Feature Alignment and Attention Mechanisms [7], [41], [52], [60]

These integrated models employ inter- and intra-task attention or cross-task attention mechanisms to align the latent representations across sequential tasks and evolving domains. The objective is to identify relationships and shared degradation patterns between tasks and domains while suppressing domain-specific representations. Thus, the model improves the stability of the learned representation under evolving operating conditions.

6.5. Pseudo-Labeling and Center-Aware Adaptation Strategies [25], [36], [41], [52]

In numerous PHM applications, target domain data remain unlabelled owing to the high cost and practical limitations of capturing run-to-failure data. Pseudo-labeling techniques are commonly adopted to enable unsupervised or semi-supervised adaptation. Combining pseudo-labeling with intra-tasks makes pseudo-labeling more robust, while mitigating the influence of noisy predictions and enhancing fine-grained feature alignment. Such strategies are particularly important for RUL prediction, where inaccurate pseudo-labels may accumulate over time, resulting in significant error propagation.

6.6. Rehearsal-Based Memory Mechanisms [5], [16], [35], [41], [52], [53]

To mitigate catastrophic forgetting, integrated DACL frameworks frequently employ rehearsal mechanisms that retain representative samples, latent features, and output logits from the previously observed domains. By replaying this information during adaptation, the model maintains continuity of knowledge while incorporating new domain information. Rehearsal-based approaches are particularly effective for prognostics owing to the recurring and evolving nature of degradation trajectories.

Error! Reference source not found. presents a conceptual flowchart of the domain-adaptive continual learning framework.

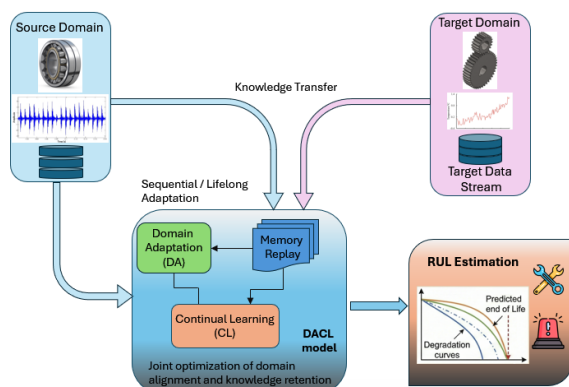


Figure 7. Conceptual architecture of a domain-adaptive continual learning framework for bearing fault prognosis and RUL estimation.

6.7. A Review of Domain-Adaptive Continual Learning Approaches

An overview of the DACL approaches for RUL prediction is presented below, based on the studies examined by the authors at the time of composition. However, as previously mentioned, the incorporation of DACL in RUL prediction and fault finding is still under development, and there are no abundant studies available yet. To provide a more comprehensive perspective, we investigated the use of DACL has been utilized in other fields, such as computer vision and disease diagnosis. Relevant insights from these domains are presented in rows 12-16. We selected some of the more interesting ones, also dated at the beginning of the DACL introduction and incorporation, for further analysis.

Table 3. Overview of Domain-Adaptive Continual Learning approaches for bearing prognosis and RUL prediction.

Author	Application/ Dataset	Model and Method	Integrated Learning Framework	Key Contributions	Limitations
Mao et al (2025) [36]	Online RUL prediction for rolling element bearings under varying and unknown operating conditions. (XJTY-SY dataset and IEEE PHM2012 PRONOSTIA platform dataset. UNSW Bearing Dataset)	Deep Incremental Regression TL Denoising Autoencoder (DAE) for vibration signal preprocessing Multi-layer LSTM feature extractor Fully-connected regression head for RUL prediction Wiener process-assisted incremental update mechanism for online DTL	Built around the Wiener process-assisted pseudo-labelling of online data to mitigate the drift of operating conditions.	Data-Model Interactive Prognostics Framework: Essential degradation information extracted from offline data and online dynamic tendency information provided by the Wiener process. No need for full-life data of target machine Robust under unknown and varying operating conditions	The proposed model initially requires historical full-life degradation data from similar types of machines. The Wiener process formulation assumes consistent degradation evolution. LSTM pre-training and feature adaptation increase the offline training costs. The authors did not test their model on real industrial feed data.
Xuan et al (2025) [61]	Fault prognostics and RUL prediction for predictive maintenance under evolving operating contexts for aircraft turbofan engines.	Bayesian Neural Network (BNN) with variational inference for probabilistic RUL regression.	Regularization-based lifelong learning with variance-based importance weighting, stability loss, variance boosting, and decoupled	Introduces a Bayesian Lifelong Learning Framework for Fault Prognostics. The model is trained with variance boosting and	Higher computational cost because the BNN is trained more slowly than other deterministic networks. The models were updated in sequential batches

	(NASA C-MAPSS dataset)		optimization (no memory, no replay)	decoupled optimization that enables a better stability-plasticity trade-off. Privacy-Preserving Continual Prognostics since previous datasets are not stored	but not in online streaming data. However, it has not been validated using real industrial data.
Guo & Sun (2025) [26]	Gas-path fault diagnosis of aero-engines under dynamic and non-stationary operating conditions. (High-fidelity physics-based turbofan simulation dataset from Xi'an Jiaotong University)	Continual domain adaptation framework (RCDAF) 1D CNN-based fault classifier with continuous unsupervised domain adaptation. Controllable Batch Normalization (CBN) mechanism	The proposed framework utilises memory-assisted continual domain adaptation with pseudo-labelling, supervised and pseudo-supervised contrastive learning (SCL+PSC), dual knowledge distillation, rehearsal buffer, controlled batch normalisation, and restraining-past regularisation.	Robust continual domain adaptation framework for aeroengine PHM. Introduces SCL plus PSC for robust semantic consistency under a domain shift. It demonstrates forgetting-resistant knowledge retention. The model is robust under Noise & Sensor Drift	The model relies on pseudo-labels, which may degrade when they become unreliable. Multi-objective optimisation and replay introduce additional training costs. Large, sudden data shifts can temporarily destabilise adaptation.
Zhao et al (2025) [5]	Rotating machinery fault diagnosis (HUST, WT gearbox, experimental, rolling mill experimental)	Dual shared space-driven domain continuous learning network (DSS-DCL). The model integrates: Shared Parameter Space Constraint and Parameter Distribution Uniformity, and Shared Feature Space Constraint - Multi-Perspective Knowledge	Rehearsal is utilised through experience replay with compact memory.	This study proposes a Dual shared space-driven domain continuous learning network (DSS-DCL), a unified model for domain-incremental fault diagnosis in dynamic environments. Introduces the parameter	Memory storing limitations Upper bound achieved by joint training to evolve the working conditions. Although compact, the model is directly dependent on LPD/LMPKD/focal multi-hyperparameter fine-tuning.

		Distillation (LMPKD)		<p>distribution uniformity loss (LPD) to expand the shared feasible parameter region and improve the stability-plasticity trade-off.</p> <p>It designs multiperspective knowledge distillation (LMPKD) to preserve multiscale representations and logits across domains.</p> <p>It employs a focal loss to mitigate the imbalance between the new domain data and replayed samples.</p>	
Kim et al (2025) [16]	RUL prediction under varying and time varying operating conditions (real-world machine – Robot Milling Dataset)	Fisher-informed continual learning (FICL) Sharpness Aware Minimization (SAM)	Rehearsal - A regularisation-based- continual learning that preserves old knowledge without storing old data explicitly is utilised, which is not the classic rehearsal model.	<p>Single predictive model for RUL prediction under varying operating conditions without the need to save the previously trained models.</p> <p>RUL prediction performance for previous operating conditions</p>	<p>This research is currently validated only for single-tool RUL prediction. This study assumes that the RUL decreases linearly with time.</p> <p>The theoretical relationship between FICL's loss landscape analysis and "mode connectivity" in deep learning optimization has not yet been fully explored</p>
Li et al (2025) [41]	RUL prediction of bearings specifically addressing challenges in wind turbines	Trend-Constrained Incremental Transfer Prognosis (TCITP).	Rehearsal: This study adopted experience replay-based incremental learning.	A novel feature fusion health indicator was proposed to predict bearing aging. A trend-	The proposed method cannot handle abrupt degradation. The generalisation to unlabelled target

	where working conditions are harsh and variable (XJTY-SY dataset, IEEE PHM2012 PRONOSTIA platform dataset, Bearings from on-site wind turbine high-speed shafts)	Demodulation Feature Fusion (DFF)-based Health Indicator Construction. Trend-Constrained Transfer Learning. Trend-Constrained Transfer Learning	Previously learned knowledge is retained and replayed during incremental updates.	constrained transfer learning method is presented in this paper for state estimation. An incremental prognosis method is developed to update the prognostic model for predicting the remaining useful life. It can update the prognostic model with newly acquired transfer pairing and review the previous knowledge with experience replay method	domains is limited. The approach assumes that the degradation trends across domains are comparable and monotonic
Guo et al (2025) [15]	RUL prediction for rolling element bearings under varying operating conditions and degradation trends. (PHM2012 PRONOSTIA platform dataset)	Stage-related Online Incremental Adversarial Domain Adaptation (SR-OIADA) transfer learning algorithm. Based on DANN: Feature extractor: 1D CNN (TCNN), Domain discriminator (with Gradient Reversal Layer), RUL regression predictor, Unsupervised stacked autoencoder (SAE) for online degradation stage detection	Buffer Rehearsal. Historical online samples are retained and reused at each checkpoint via incremental learning to prevent the forgetting of information. Moreover, Stage-related adversarial domain adaptation aligns source vs. target features (inter-domain) and stage-to-stage degradation representations (intra-process).	A novel online degradation stage division algorithm is proposed to adaptively detects the health status of the bearings monitored online. A Stage-related Online Incremental Adversarial Domain Adaptation (SR-OIADA) transfer learning algorithm is proposed that integrates incremental learning, transfer learning, and adversarial	The update times were defined empirically. However, these studies did not propose an adaptive checkpoint-triggering mechanism. The HI stage division was fixed at three stages.

				domain adaptation.	
Zeng et al (2024) [24]	Online self-driven RUL prediction for rolling bearings under cross-condition, cross-machine scenarios, and large distribution divergence. (PHM2012 PRONOSTIA platform dataset, XJTU-SY bearing dataset)	Bayesian Domain-Adversarial Regression Adaptation (BDRA), Deep Autoencoder (DAE) feature extractor, Tensorized domain discriminator (DANN-style), Bayesian LSTM regressor with Variational Inference (VI), Tucker tensor decomposition for degradation representation	Pseudo- Labelling, Online target data chunks receive pseudo-labels generated by the Bayesian pre-training network and are used for incremental regression updates. Moreover, Domain-adversarial training aligns source vs target (inter-domain) features using a tensorized discriminator; monotonicity-aware self-supervised learning aligns intra-degradation trends between successive online chunks	This study proposes a bidirectional information transfer mechanism for regression transfer learning in an online scenario. Utilizes core tensor for self-supervised information and Bayesian VI for pseudo-label information. Introduces an innovative framework for self-driven RUL transfer prediction in open environments. This enables lightweight prediction and reduces error accumulation without run-to-failure data from the target machines. Provides a confidence interval for prediction results, offering clear practical value	High model complexity. The current method may not provide a sufficiently rational interval for the prediction results because it uses a Gaussian distribution to estimate the regressor output. The authors noted that Weibull modelling better represents the failure behaviour. The model performance depends on the selection of an appropriate online chunk size for the training data.
Zeng et al (2024) [62]	Online RUL prediction of rotating machinery (rolling bearings) under unknown and drifting working conditions. (PHM2012	Tensor Domain-Adversarial Regression Adaptation with Interpretability (TDARA), Stacked Denoising Autoencoder (SDAE) for	The research assigns pseudo RUL labels generated by a pre-trained model to the online target data blocks. The pseudo-labels were then used	This study integrates a DANN with Tucker tensor decomposition to preserve the degradation structure, while aligning the domains. Tensor	The proposed method utilises an incremental continual learning method which is not evaluated under multitask CL benchmarks. The proposed method imposes a risk of

	PRONOSTIA platform dataset, XJTU-SY bearing dataset)	features extraction, Tensor Representation Module that captures degradation structure and temporal correlations, Tensor Domain Discriminator that Aligns source and target feature distributions using adversarial loss, LSTM regressor predictor trained on source RUL labels	to update the incremental regressions. Tensor domain adversarial training aligns source vs. target (inter-domain) features, and trend regularisation aligns the intra-degradation temporal structure across the blocks.	Tucker decomposition was utilised to identify key features and analyse interpretability at the geometric level. Extractes degradation trend information based on core tensor and established multi-scale evaluation criteria for transferability	gradual drift because it does not explicitly retain knowledge with replay or memory mechanisms. Trend modelling assumes monotonic degradation which may not generalise to all assets.
Xie et al (2024) [60]	Online RUL prediction with uncertainty quantification for continuously running machinery. (XJTU-SY bearing dataset, NCWP journal-bearing dataset, C-MAPSS turbofan engine dataset)	Incremental Contrast Hybrid Model (ICHM), and Contrastive Learning Transformer (CLformer), which predict the degradation trend and increments online. Enhanced Generalized Wiener Process (EGWP), calculates the RUL probability density function	The authors used an incremental contrastive learning method that aligns future latent features with actual future latent features (positive pairs), enforcing temporal/intra-trajectory consistency under distribution drift.	The proposed incremental contrast hybrid model (ICHM) is updated in real time to mitigate the prediction offset and align the past and future latent representations. Moreover, the model can adapt online in real time without the need for run-to-failure retraining.	This study uses RMS as the HI and monotonic trend for RUL thresholding, which may be less suitable for non-monotonic degradation applications. The proposed model appears to be sensitive to the selection of the IL and sliding window hyperparameters. The authors note that incremental data from very different domains may harm final prediction if deployed cross-domain
Liu et al (2024) [27]	Online industrial fault prognosis and RUL prediction under dynamic operating conditions using task-free continual learning. It	Online Task-Free Continual Fault Prognosis (OTFCFP) framework based on a Continual Neural Dirichlet Process Mixture (CN-DPM) model. Bayesian	The proposed work uses an Expansion-based task-free continual learning via Bayesian Mixture of Experts (CN-DPM)	A novel online task-free continual fault prognosis paradigm for practical industrial scenarios is proposed. Introduces a	The proposed model assumes the same sensor space; hence, distribution shifts are handled via expert expansion only. The authors did not use domain adaptation or feature

	focuses on real-time predictive maintenance using non-stationary sensor data streams. (CMAPSS, N-CMAPSS)	nonparametric Dirichlet Process Mixture, Attention-based Temporal Convolutional Network (TCN) for RUL regression, GRU-based Variational Autoencoder (VAE) for task identification and density estimation		CN-DPM-based continual learning framework that automatically identifies task boundaries and allocates new experts to them. It integrates a Mixture of Experts with Bayesian nonparametric learning to avoid catastrophic forgetting.	alignments. The model requires labelled RUL data for training.
Li et al (2024) [57]	Digital Twin (DT) for rolling bearings with online dynamic evolution and RUL prediction under varying working conditions. (PHM2012 PRONOSTIA)	End-Edge-Cloud Digital Twin architecture with a Condition-Adaptive Dynamic Continual Learning Digital Twin Model (CADCL-DTM)	The model utilises regularisation-based continual learning using EWC with condition-adaptive penalty.	This study proposes a realistic end-edge-cloud Digital Twin architecture for rolling bearings. It introduces CADCL-DTM, a condition-adaptive continual learning DT model. Designs different regularisation penalties for intra-condition and inter-condition task transitions. This enables the DT to continuously learn new working conditions while mitigating catastrophic forgetting.	In this study, the authors did not perform a standard domain adaptation. Condition changes are handled via regularisation rather than explicit distribution alignment. This requires a manual hyperparameter tuning. Different α values were manually selected for the intra-condition ($\alpha=500$) and inter-condition ($\alpha=150$) scenarios.
Li et al (2023) [32]	Industrial fault diagnosis of rotating machinery using industrial	Deep Continual Transfer Learning with Dynamic Weight	Continual transfer learning with rehearsal memory, DWA stability–	A novel DCTL-DWA framework is proposed for processing the	The model was tested only on private test rig data. The model does not incorporate real-

	streaming data and varying working conditions. (Private industrial dataset collected from a transmission test rig)	Aggregation (DCTL-DWA). Adversarial Continual Domain Adaptation (ACDA).	plasticity control, DANN-based domain adaptation, and triplet metric alignment	Industrial streaming data. The model learns the optimal stability–plasticity trade-off in each phase using the DWA. It maintains the diagnostic performance across various speeds and loads owing to the proposed ACDA.	time online learning; instead, it employs a session-based approach.
Zhuang et al (2023) [7]	Online Remaining Useful Life (RUL) prediction of rolling bearings under online unknown operating conditions. (PHM2012 PRONOSTIA, Auxiliary ABLT platform)	Multi-source Adversarial Online Regression (MAOR) with a three-stage pipeline Pseudo-domain extension, via encoder–decoder that adaptively generates multiple pseudo domains from a single source. Two-stage Multi-source Adversarial Domain Adaptation (MADA). Offline–online prediction framework	The model predicts pseudo-labelling to generate pseudo-domains and is used to guide adaptive weights and regression training. In addition, it performs domain-level adaptation and feature-level adaptations embedded into MADA to reduce the marginal divergence. Finally, adversarial multi-source alignment is incorporated for domain-invariant features.	A multi-source adversarial online regression (MAOR) framework is proposed for the RUL under unknown online conditions. Introduces a pseudo-label information-guided pseudo-domain extension with adaptive weighting and a domain-level adaptation. A two-stage MADA was designed for this study. It embeds feature-level adaptation to mitigate marginal feature divergence, which is often ignored in global adversarial alignment. An offline–online	The model performance degrades, and the training becomes unstable as the number of pseudo domains increases. Multiple hyperparameters require careful tuning to achieve optimal performance. The model is quite complex, incorporating pseudo-domain generation, two-stage adversarial training, and online updating, which increases the training and inference cost.

				prediction scheme with dynamic adaptive weighting for real-time updates was developed.	
Mao et al (2023) [25]	Online RUL prediction across machines (cross-machine and cross-condition) with unlabelled streaming target data and significant degradation characteristic divergence. (PHM2012 PRONOSTIA, XJTU-SY, In-house roller-bearing test rig)	Self-Supervised Deep Tensor Domain-Adversarial Regression Adaptation (SD-TDA-RA) with two stages: 1. Pretrained Tensor Domain-Adversarial Network, including a feature extractor, tensorized domain discriminator, Tucker decomposition, source regression predictor, and tendency regularizer. 2. Online Self-Supervised Fine-Tuning, incorporating Pseudo-labels and Self-supervised monotonicity loss computation	The model incorporates pseudo-labels for each online block generated by the pretrained network and is used for fine-tuning. Also features alignment through domain-adversarial alignment (DANN) on tensor core representations and tendency regularization via MMD	Proposes a novel tensor-based regression domain-adversarial adaptation for online RUL across machines. A tensorized domain discriminator was introduced to capture temporal degradation. It designs bidirectional information transfer, horizontal (pseudo-supervised across machines), and vertical (self-supervised from target monotonicity). A tendency regularizer is added to maintain the degradation trends while aligning the domains. <ul style="list-style-type: none"> • Develops a lightweight online fine-tuning strategy 	The proposed model exhibits complexity because it incorporates tensor decomposition (MDT + Tucker) and alternating minimisation. Multiple hyperparameters require careful fine-tuning. Knowledge retention relies on alignment and self-supervision rather than memory replay.
Li and Jha (2024) [56]	Smart healthcare disease detection using wearable medical sensors (WMS) under	Past-Agnostic Generative Replay (PAGE) is based on the following:	The model utilises generative replay using synthetic	The author proposed PAGE, a past-agnostic generative	The performance of the model relies on the ability of the GMM to capture the new domain

	domain-incremental adaptation. (CovidDeep, DiabDeep, MHDDeep)	Synthetic Data Generation (SDG) module, Pseudo labelling from new real data domain, Multi-dimensional Gaussian Mixture Model (GMM) is adopted for probability density estimation in the SDG, Extended Inductive Conformal Prediction (EICP) method is incorporated to generate confidence scores and credibility values for disease detection results	samples generated from the GMM and pseudo-labelling to create synthetic samples from the new real data.	replay framework for domain-incremental adaptation. This enables continual learning without storing past domain data. Introduces SDG with GMM density estimation for high-quality synthetic tabular data generation. Integrates Extended Inductive Conformal Prediction (EICP) for confidence and credibility estimation.	distribution. The proposed method targets only domain-incremental adaptation, not class- and task-incremental scenarios.
De Carvallo et al (2024) [52]	Vision-based unsupervised cross-domain task-incremental learning. (MNIST ↔ USPS, VisDA-2017, Office-31, Office-Home, DomainNet)	Cross-Domain Continual Learning (CDCL) is a transformer-based framework that unifies continual learning and UDA. Inter-Intra Task Cross-Attention that aligns category-level features between source and target and retains past alignment, intra task Center-Aware Pseudo- Labelling, Rehearsal with Logit Replay: stores selected source, target pairs, Backbone, a Compact convolutional tokenizer +	It utilises rehearsal through logit replay to maintain task boundaries. In addition, pseudo-labelling is performed for unlabelled target data to form confident cross-domain pairs. Finally, inter/intra-feature alignment consolidates the prior alignment by freezing the core parameters and learning new tasks.	It introduces CDCL, a unified framework for unsupervised cross-domain task-incremental learning. It proposes inter-intra-task cross-attention to maintain and extend category-level domain alignment over time, reducing feature-alignment catastrophic forgetting. An intra-task centre-aware pseudo-labelling pipeline was designed to select accurate cross-domain	The performance of the model relies on using a fixed rehearsal buffer. This research focused on vision UDA and did not evaluate non-visual applications. The model generates task-specific hyperparameters that are accumulated over time.

		transformer encoder		pairs and suppress noise. <ul style="list-style-type: none"> • Incorporates sample rehearsal and logit replay to consolidate past knowledge. 	
Rakshit et al (2024) [35]	Lifelong visual recognition under domain shift with unknown classes (Office-Home, DomainNet, UPR N-RSDA (remote sensing) dataset)	Incremental Open-Set Domain Adaptation (IOSDA), Multi-Domain & Class-guided GAN (MDCGAN), Multi-output Ensemble Open-set DA (MEOSDA)	The proposed model reconstructs previous domains using synthetic data (rehearsal), and pseudo-labelled data are used to train the next iteration of the MDCGAN.	Introduces a novel problem formulation of Incremental Open-set Domain Adaptation (IOSDA). It integrates replay, pseudo-labelling, and open-set adversarial DA in a unified lifelong framework. Releases UPRN-RSDA, a new IOSDA benchmark for remote sensing	The model performance relies on MDCGAN's ability of the MDCGAN to accurately model the prior domains; poor generation may propagate errors. Feature alignment is implicit via adversarial training rather than explicit inter/intra-class metric learning. Multi-head classifiers grow linearly with the number of domains, which can eventually lead to memory and inference overheads.
Nguyen M. et al (2022) [53]	Incremental multi-target unsupervised domain adaptation (MTDA) for object detection is motivated by real-world scenarios such as multi-camera video surveillance and adverse weather conditions. (PascalVOC, Cityscape, Wildtrack multi-camera dataset)	Multi-target Domain Adaptation with Domain Transfer Module (MTDA-DTM). Instead of storing past target data or duplicating detectors, the model incorporates MTDA-DTM, a lightweight Domain Transfer Module (DTM) that transfers source images into a joint representation space of all	Generative replay is utilised via the DTM. The past target distributions were approximated by transforming source images into pseudo-target images.	Introduces MTDA-DTM, a cost-efficient incremental MTDA framework for object detection. A novel Domain Transfer Module (DTM) that maps source images into a joint representation of all previous target domains is proposed. This enables incremental adaptation without storing	The proposed model is not applicable to time-series or prognostic domains. This requires careful tuning of the replay weight α for different domain shifts. Image-space pseudo-samples are not visually realistic.

		previously learned target domains.		previous target data or duplicating detectors. Prevention of catastrophic forgetting via pseudo-target replay.	
Truong et al (2024) [49]	Continual unsupervised domain adaptation (UDA) for semantic scene segmentation in self-driving cars. (GTA5 (synthetic, labelled source, Cityscapes (real, unlabelled target), IDD (Indian driving dataset), Mapillary Vistas)	Continual Unsupervised Domain Adaptation (CONDA), Bijective Maximum Likelihood (BML) loss, Bijective Network for Structured Output Modelling	This study utilised pseudo-labels generated via the EMA teacher model for unsupervised segmentation loss. Moreover, Alignment is performed in prediction distribution space using Bijective Maximum Likelihood	This study presents a novel Continual Unsupervised Domain Adaptation (CONDA) approach for semantic scene segmentation. It formulates continual domain adaptation by regularizing the distribution shift of predictions between source and target domains to avoid catastrophic forgetting introducing the Bijective Maximum Likelihood. The bijective network captures global structural information and represents segmentation distributions	The specific model was validated only on semantic segmentation data sets. It requires high computational power to train invertible networks. Its performance depends on the pseudo-labels generated by the EMA model. No explicit feature space alignment was performed; instead, alignment was conducted solely in the output distribution space.

6.8. Advantages of Integrated DA–CL Approaches for Bearing RUL Prediction

Through a comparative analysis of the unified integration of DA–CL frameworks versus the separate implementation of DA and incremental learning, we concluded that the unified integration offers several key advantages.

- **Improved Adaptability:** Integrated frameworks enable models to respond effectively to both abrupt and gradual changes in operating conditions while preserving previously learned degradation knowledge. This is extremely critical for machinery operating under diverse and evolving conditions [16] [15].

- **Enhanced generalisation:** By aligning feature distributions across domains and incrementally updating models, the approach generalises effectively to unseen operating states and machinery configurations [25], [53], [60].
- **Mitigation of Catastrophic Forgetting:** Unified domain adaptive continual learning maintains a balance between plasticity and stability, mitigating catastrophic forgetting [5], [15], [16].
- **Reduced dependence on labelled data:** DA minimises the need for extensive labelled target domain data, whereas CL efficiently incorporates limited new data, addressing practical data scarcity and labelling cost constraints [36], [53].
- **Enhanced Parameter Efficiency:** Modular adaptation strategies significantly reduce the number of trainable parameters, facilitating deployment in resource-constrained industrial systems [5], [16], [60].
- **Robustness under Uncertainty:** Continuous feature alignment and knowledge retention mechanisms support stable RUL predictions, reducing false alarms and enhancing reliability in safety-critical applications [5], [15], [24].
- **Computational Efficiency:** Incremental updates avoid costly retraining from scratch, reducing computational overhead and enabling real-time or near-real-time prognostics [5], [53], [60].

6.9. Limitations and Practical Implementation Challenges

Despite their promise, integrated DA-CL frameworks still face several unresolved challenges.

- **Scalability and Memory Constraints** [52], [53], [56]:
 - **Replay-based methods** improve forgetting mitigation but face limitations owing to memory and computational resource demands. This becomes more prevalent over long operational lifetimes with many sequential domains
 - Moreover, task-incremental learning approaches suffer from a theoretical limitation of infinitely growing parameters as new tasks are added
- **Sensitivity to noisy pseudo-label error.** Many methods rely on pseudo-labels generated by pre-trained networks. These labels can be imprecise because domain adversarial training cannot completely eliminate distribution differences, potentially leading to error accumulation [5], [24], [25].
- **Data Quality and Labelling Challenges.** Online frameworks must operate with imperfect data and a lack of ground truth labels for the target domain. Moreover, current models primarily focus on progressive run-to-failure degradation and cannot effectively address cases of abrupt degradation [5], [25], [36], [62].
- **Limited interpretability of adaptive and memory-based mechanisms.** In general, previously presented models are adaptive, and their memory-based strategies significantly improve continual and domain-adaptive learning performances. However, they often operate as black-box mechanisms, offering limited insight into how past knowledge is preserved and how new domain information is integrated [41], [62].
- **Dynamic Domain Identification** remains a major challenge. Real-world applications frequently lack explicit domain or task labels, complicating the application of parameter isolation or task-specific adaptation strategies [16], [27], [35], [36].
- Finally, there is a lack of standardised benchmarks that jointly evaluate DA and CL performance in the PHM. Although previous studies have explored the integration of DA and CL for PHM, there is currently no standardised benchmark that jointly evaluates both capabilities under a unified experimental protocol.
- Addressing these challenges is essential for translating integrated DA-CL approaches from academic studies to industrial applications.

7. Research Gaps - Challenges and Future Research Directions

Despite the rapid progress in DA, CL, and their emerging integration into the unified α DA-CL framework, several fundamental research gaps remain before such approaches can be considered mature and industrially deployable for bearing fault prognosis and RUL estimation under evolving data distributions.

7.1. Assumptions of Monotonicity and progressive Degradation Remain Restrictiv

Most, if not all, online and incremental prognostic models assume that equipment degradation follows a smooth and monotonic trajectory, which is often enforced through trend regularisation or Wiener-process formulations. In practice, however, industrial assets frequently exhibit non-monotonic degradation patterns owing to abrupt performance deterioration, intermittent fault occurrences, maintenance-induced state resets, and regime-switching operational modes. Current models struggle to handle non-monotonic and discontinuous degradation patterns. This may potentially reduce their reliability in safety-critical applications, including marine propulsion, power generation, and aerospace systems.

7.2. Robust Learning with Severely Limited or Noisy Supervision Is Challenging

The majority of DA and DACL frameworks rely on pseudo-labelling mechanisms to facilitate unsupervised or semi-supervised adaptation. Although pseudo-labelling allows incremental learning in unlabelled target domains, it also poses a considerable risk of error accumulation when the initial predictions are inaccurate. As domain adversarial training cannot fully eliminate distribution discrepancies, pseudo-label noise may propagate across the learning stages and lead to irreversible model drift. Therefore, future research should emphasise uncertainty-aware pseudo-labelling, confidence-weighted updates, and Bayesian or evidential learning frameworks that explicitly model prediction credibility during adaptation.

7.3. Scalability and Memory Efficiency Remain Major Bottlenecks for Lifelong Deployment

Replay-based CL methods demonstrate strong forgetting mitigation but suffer from increasing memory requirements as operational lifetimes and domain diversity increase. Architecture-expansion methods alleviate forgetting through structural isolation but introduce unbounded model growth, increased inference latency and maintenance complexity. Although recent studies have explored parameter-efficient modular adaptation and low-rank subspace learning, a principled solution for long-horizon deployment with bounded memory, computation, and model size is still lacking. This challenge is particularly relevant for embedded and edge PHM systems deployed in vessels, wind turbines, and industrial production lines.

7.4. Dynamic Domain Discovery and Task-Free Adaptation Remain Largely Unsolved Problems

Most existing DACL methods assume that domain boundaries or task transitions are known a priori or can be heuristically defined using time-windowing strategies. In real-world PHM systems, domain transitions occur gradually owing to load variations, environmental changes, sensor aging or drifting, and maintenance interventions. Without clear domain labels, the use of parameter isolation, task-specific regularisation, and rehearsal methods is challenging. Future research should focus on models that can autonomously recognise when the domain changes and adapt their learning mechanisms accordingly without manual intervention, thereby enabling task-free CL paradigms to be developed.

7.5. Model Interpretability Remains Limited

Although research on DA, CL, and combined DACL paradigms has gained ground, and a plethora of methods have been proposed to address domain shifts and catastrophic forgetting, these models largely operate as black boxes. However, for optimum industrial decision support, operators require transparent justifications for predictions, degradation drivers, and uncertainty margins.

7.6. The Lack of Standardized Benchmarks and Evaluation Protocols Hinders Reproducible Assessment

Existing studies use heterogeneous datasets drawn from public university testbeds and proprietary experimental rigs. Moreover, the studies utilise different evaluation metrics and adaptation protocols, making it difficult to assess the true progress of a specific DACL paradigm. There is a clear need for open, large-scale, multi-domain PHM benchmarks that support rigorous and reproducible assessments of DA capability, CL stability, catastrophic forgetting resistance, uncertainty quantification, and long-horizon robustness under realistic operating conditions.

8. Conclusions

This review presents a comprehensive and systematic analysis of data-driven models for rotating machinery fault prognosis and RUL estimation under evolving data distributions, with a particular focus on the emerging paradigm of Domain Adaptive Continual Learning (DACL). Modern Prognostic and Health Management (PHM) systems operate mainly in dynamic, non-stationary environments, where traditional deep learning models trained under fixed data distributions struggle to deliver robust, long-term predictive performance.

Through an extensive review of the recent literature, this study presents that DA and CL represent two complementary learning paradigms that tackle distinct yet interconnected challenges in real-world PHM applications. DA techniques focus on mitigating cross-domain discrepancies, whereas CL aims to alleviate catastrophic forgetting and enable incremental learning during sequential data evolution. However, literature research revealed that when these paradigms are applied in isolation, they exhibit intrinsic limitations that constrain their long-term applicability and performance in real-world industrial deployments.

The convergence of DA and CL into a unified DACL framework offers a principled solution to the evolving data distribution problem by enabling models to continuously align feature representations across domains while preserving the historical degradation knowledge of the data. Although the existing body of DACL-related studies remains relatively limited and at an early stage of development, the reviewed works indicate that such integration can enhance generalisation across machines and operating conditions, reduce reliance on labelled target domain data, mitigate catastrophic forgetting, and improve robustness under uncertainty. These features are imperative for making lifelong predictive maintenance possible and trustworthy in critical applications, such as marine propulsion systems, aerospace engines, and industrial production assets.

Nevertheless, this review highlights that DACL-based prognostics remains a developing research area. Significant challenges persist in terms of scalability, memory efficiency, robustness to noisy pseudo-labels, dynamic domain discovery, interpretability and benchmark standardisation. In addition, most existing studies have been validated using laboratory datasets or controlled test rigs. This underscores the need for large-scale validation in real-world industrial settings.

In conclusion, Domain Adaptive Continual Learning signifies a strategic and essential progression of data-driven prognostics toward autonomous, self-updating, and lifelong PHM systems. By bridging the gap between static model training and real-world operational dynamics, DACL frameworks establish the groundwork for next-generation intelligent and predictive maintenance solutions. Ongoing collaborative research at the intersection of deep learning, reliability engineering, and industrial and marine systems is essential to transform these academic advances into robust industry-deployable solutions that can eventually improve safety, reliability, efficiency, and operational sustainability.

References

1. T. Wang, D. Guo, and X.-M. Sun, 'Contrastive Generative Replay Method of Remaining Useful Life Prediction for Rolling Bearings', *IEEE Sensors Journal*, vol. 23, no. 19, pp. 23893–23902, 2023, doi: 10.1109/JSEN.2023.3308092.
2. X. Ren, Y. Qin, B. Li, B. Wang, X. Yi, and L. Jia, 'A core space gradient projection-based continual learning framework for remaining useful life prediction of machinery under variable operating conditions', *Reliability Engineering and System Safety*, vol. 252, 2024, doi: 10.1016/j.res.2024.110428.
3. S. Apeiranthitis, P. Zacharia, A. Chatzopoulos, and M. Papoutsidakis, 'Predictive Maintenance of Machinery with Rotating Parts Using Convolutional Neural Networks', *Electronics*, vol. 13, no. 2, p. 460, Jan. 2024, doi: 10.3390/electronics13020460.
4. S. Apeiranthitis, C. Drosos, M. Papoutsidakis, and A. Chatzopoulos, 'THE IMPORTANCE OF PROPER PLANNING AND MACHINERY MAINTENANCE OF MERCHANT VESSELS', *Hellenic Institute of Marine Technology*, no. 19th, Dec. 2025.

5. S. Zhao, Y. Bai, and C. Hou, 'A dual shared space-driven domain continuous learning network for rotating machinery fault diagnosis in dynamic environments', *Meas. Sci. Technol.*, vol. 36, no. 9, p. 096140, Sep. 2025, doi: 10.1088/1361-6501/ae09cb.
6. W. Sun, H. Wang, Z. Liu, and R. Qu, 'Method for Predicting RUL of Rolling Bearings under Different Operating Conditions Based on Transfer Learning and Few Labeled Data', *Sensors*, vol. 23, no. 1, 2023, doi: 10.3390/s23010227.
7. J. Zhuang, Y. Cao, M. Jia, X. Zhao, and Q. Peng, 'Remaining useful life prediction of bearings using multi-source adversarial online regression under online unknown conditions', *Expert Systems with Applications*, vol. 227, 2023, doi: 10.1016/j.eswa.2023.120276.
8. C. -B. Chou and C. -H. Lee, 'Generative Neural Network-Based Online Domain Adaptation (GNN-ODA) Approach for Incomplete Target Domain Data', *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–10, 2023, doi: 10.1109/TIM.2023.3246495.
9. A. Kumar *et al.*, 'Entropy-based domain adaption strategy for predicting remaining useful life of rolling element bearing', *Engineering Applications of Artificial Intelligence*, vol. 133, 2024, doi: 10.1016/j.engappai.2024.108575.
10. Y. Cao, M. Jia, P. Ding, X. Zhao, and Y. Ding, 'Incremental Learning for Remaining Useful Life Prediction via Temporal Cascade Broad Learning System With Newly Acquired Data', *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 6234–6245, Apr. 2023, doi: 10.1109/TII.2022.3201977.
11. W. Mao, J. Wang, K. Feng, Z. Zhong, and M. Zuo, 'Dynamic modeling-assisted tensor regression transfer learning for online remaining useful life prediction under open environment', *Reliability Engineering and System Safety*, vol. 263, 2025, doi: 10.1016/j.res.2025.111210.
12. T. Wang, H. Liu, D. Guo, and X.-M. Sun, 'Continual Residual Reservoir Computing for Remaining Useful Life Prediction', *IEEE Trans. Ind. Inf.*, vol. 20, no. 1, pp. 931–940, Jan. 2024, doi: 10.1109/TII.2023.3271661.
13. J. Zhou and Y. Qin, 'A Continuous Remaining Useful Life Prediction Method With Multistage Attention Convolutional Neural Network and Knowledge Weight Constraint', *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 7, pp. 11847–11860, Jul. 2025, doi: 10.1109/TNNLS.2024.3462723.
14. N. Ding, H. Li, Q. Xin, B. Wu, and D. Jiang, 'Multi-source domain generalization for degradation monitoring of journal bearings under unseen conditions', *Reliability Engineering and System Safety*, vol. 230, 2023, doi: 10.1016/j.res.2022.108966.
15. W. Guo, F. Li, P. Zhang, and L. Luo, 'A stage-related online incremental transfer learning-based remaining useful life prediction method of bearings', *Applied Soft Computing*, vol. 169, 2025, doi: 10.1016/j.asoc.2024.112491.
16. G. Kim *et al.*, 'Fisher-informed continual learning for remaining useful life prediction of machining tools under varying operating conditions', *Reliability Engineering & System Safety*, vol. 253, p. 110549, Jan. 2025, doi: 10.1016/j.res.2024.110549.
17. X. Lu, X. Yao, Q. Jiang, Y. Shen, F. Xu, and Q. Zhu, 'Remaining useful life prediction model of cross-domain rolling bearing via dynamic hybrid domain adaptation and attention contrastive learning', *Computers in Industry*, vol. 164, 2025, doi: 10.1016/j.compind.2024.104172.
18. '1-s2.0-S0031320322002527-main.pdf'.
19. Z. Chen, J. Chen, Z. Liu, and Y. Liu, 'Mutual-learning based self-supervised knowledge distillation framework for remaining useful life prediction under variable working condition-induced domain shift scenarios', *Reliability Engineering and System Safety*, vol. 264, 2025, doi: 10.1016/j.res.2025.111359.
20. M. A. Benatia, M. Hafsi, and S. Ben Ayed, 'A continual learning approach for failure prediction under non-stationary conditions: Application to condition monitoring data streams', *Computers & Industrial Engineering*, vol. 204, p. 111049, Jun. 2025, doi: 10.1016/j.cie.2025.111049.
21. 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', *Reliability Engineering and System Safety*, vol. 256, 2025, doi: 10.1016/j.res.2024.110769.
22. J. Hurtado, D. Salvati, R. Semola, M. Bosio, and V. Lomonaco, 'Continual Learning for Predictive Maintenance: Overview and Challenges', *Intelligent Systems with Applications*, vol. 19, p. 200251, Sep. 2023, doi: 10.1016/j.iswa.2023.200251.

23. T. Zhang and H. Wang, 'Quantile regression network-based cross-domain prediction model for rolling bearing remaining useful life', *Applied Soft Computing*, vol. 159, 2024, doi: 10.1016/j.asoc.2024.111649.
24. P. Zeng, W. Mao, Y. Li, N. Wang, and Z. Zhong, 'Bayesian Domain-Adversarial Regression Adaptation: A New Self-Driven Remaining Useful Life Prediction Method Across Different Machines With Uncertainty Quantification', *IEEE Sensors Journal*, vol. 24, no. 20, pp. 32673–32683, 2024, doi: 10.1109/JSEN.2024.3450470.
25. W. Mao, K. Liu, Y. Zhang, X. Liang, and Z. Wang, 'Self-Supervised Deep Tensor Domain-Adversarial Regression Adaptation for Online Remaining Useful Life Prediction Across Machines', *IEEE Transactions on Instrumentation and Measurement*, vol. 72, 2023, doi: 10.1109/TIM.2023.3265109.
26. C. Guo and Y. Sun, 'A Robust Continual Domain Adaptation Framework for Gas Path Fault Diagnosis of Aero-Engine under Dynamic Operating Conditions', *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1–14, 2025, doi: 10.1109/TAES.2025.3644280.
27. C. Liu, L. Zhang, Y. Zheng, Z. Jiang, J. Zheng, and C. Wu, 'Online industrial fault prognosis in dynamic environments via task-free continual learning', *Neurocomputing*, vol. 598, p. 127930, Sep. 2024, doi: 10.1016/j.neucom.2024.127930.
28. Tianjiao Lin, Liuyang Song, L. Cui, and H. Wang, 'Continual learning for unknown domain fault diagnosis in rotating machinery via Diffusion-Integrated Dynamic Mixture Experts', *Engineering Applications of Artificial Intelligence*, vol. 156, p. 111056, Sep. 2025, doi: 10.1016/j.engappai.2025.111056.
29. X. Zhang, Z. Li, and J. Wang, 'Joint domain-adaptive transformer model for bearing remaining useful life prediction across different domains', *Engineering Applications of Artificial Intelligence*, vol. 159, 2025, doi: 10.1016/j.engappai.2025.111615.
30. X. Ren, Y. Qin, B. Wang, X. Cheng, and L. Jia, 'A Complementary Continual Learning Framework Using Incremental Samples for Remaining Useful Life Prediction of Machinery', *IEEE Transactions on Industrial Informatics*, vol. 20, no. 12, pp. 14330–14340, 2024, doi: 10.1109/TII.2024.3450077.
31. X. He, C. Ding, F. Qiao, and J. Shi, 'An Incremental Remaining Useful Life Prediction Method Based on Wasserstein GAN and Knowledge Distillation', presented at the Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 2024, pp. 3857–3862. doi: 10.1109/SMC54092.2024.10831026.
32. J. Li, R. Huang, Z. Chen, G. He, K. C. Gryllias, and W. Li, 'Deep continual transfer learning with dynamic weight aggregation for fault diagnosis of industrial streaming data under varying working conditions', *Advanced Engineering Informatics*, vol. 55, p. 101883, Jan. 2023, doi: 10.1016/j.aei.2023.101883.
33. S. A. Bidaki *et al.*, 'Online Continual Learning: A Systematic Literature Review of Approaches, Challenges, and Benchmarks', Jan. 09, 2025, *arXiv*: arXiv:2501.04897. doi: 10.48550/arXiv.2501.04897.
34. L. Wang, X. Zhang, H. Su, and J. Zhu, 'A Comprehensive Survey of Continual Learning: Theory, Method and Application', Feb. 06, 2024, *arXiv*: arXiv:2302.00487. doi: 10.48550/arXiv.2302.00487.
35. S. Rakshit, H. Bandyopadhyay, N. Das, and B. Banerjee, 'Incremental Open-set Domain Adaptation', Aug. 31, 2024, *arXiv*: arXiv:2409.00530. doi: 10.48550/arXiv.2409.00530.
36. W. Mao, R. Guo, J. Wang, M. Zuo, and Z. Zhong, 'Wiener process-assisted online remaining useful life prediction with deep incremental regression transfer learning', *Reliability Engineering & System Safety*, vol. 267, p. 111867, Mar. 2026, doi: 10.1016/j.res.2025.111867.
37. X. Ren, H. Qiu, D. Chen, C. Peng, Y. Qin, and B. Wang, 'A Continual Learning Framework with Adaptive Synapses for Remaining Useful Life Prediction', presented at the 2023 Global Reliability and Prognostics and Health Management Conference, PHM-Hangzhou 2023, 2023. doi: 10.1109/PHM-HANGZHOU58797.2023.10482788.
38. D. She, Y. Luo, Y. Wang, S. Gan, X. Yan, and M. G. Pecht, 'A meta-transfer-driven method for predicting the remaining useful life of rolling bearing with few shot data', *Measurement: Journal of the International Measurement Confederation*, vol. 254, 2025, doi: 10.1016/j.measurement.2025.117882.
39. J. Yang, D. Sun, L. Wang, W. Zhang, and X. Wang, 'DPMA: Self-Supervised Dual-Path Meta Alignment Network for Remaining Useful Life Prediction with Limited Data and Unknown Working Conditions', *IEEE Transactions on Instrumentation and Measurement*, pp. 1–1, 2025, doi: 10.1109/TIM.2025.3595222.

40. J. Zhou, J. Luo, H. Pu, and Y. Qin, 'Multibranch Horizontal Augmentation Network for Continuous Remaining Useful Life Prediction', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 55, no. 3, pp. 2237–2249, Mar. 2025, doi: 10.1109/TSMC.2024.3519347.
41. X. Li *et al.*, 'Trend-constrained pairing based incremental transfer learning for remaining useful life prediction of bearings in wind turbines', *Expert Systems with Applications*, vol. 263, 2025, doi: 10.1016/j.eswa.2024.125731.
42. M. J. Page *et al.*, 'PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews', *BMJ*, p. n160, Mar. 2021, doi: 10.1136/bmj.n160.
43. M. J. Page *et al.*, 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
44. M. L. Rethlefsen *et al.*, 'PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews', *Syst Rev*, vol. 10, no. 1, p. 39, Jan. 2021, doi: 10.1186/s13643-020-01542-z.
45. K. A. Milner, D. Hays, S. Farus-Brown, M. C. Zonsius, E. Saska, and E. Fineout-Overholt, 'National evaluation of DNP students' use of the PICOT method for formulating clinical questions', *Worldviews Ev Based Nurs*, vol. 21, no. 2, pp. 216–222, Apr. 2024, doi: 10.1111/wvn.12709.
46. J. J. Riva, K. M. P. Malik, S. J. Burnie, A. R. Endicott, and J. W. Busse, 'What is your research question? An introduction to the PICOT format for clinicians', *J Can Chiropr Assoc*, vol. 56, no. 3, pp. 167–171, Sep. 2012.
47. J. Guo, Y. Song, Z. Wang, and Q. Chen, 'A dual-channel transferable model for cross-domain remaining useful life prediction of rolling bearings under uncertainty', *Measurement Science and Technology*, vol. 36, no. 3, 2025, doi: 10.1088/1361-6501/adbb0c.
48. Y. Ye, J. Wang, J. Yang, D. Yao, and T. Zhou, 'Adaptive MAGNN-TCN: An Innovative Approach for Bearings Remaining Useful Life Prediction', *IEEE Sensors Journal*, vol. 25, no. 4, pp. 7467–7481, 2025, doi: 10.1109/JSEN.2024.3506154.
49. T.-D. Truong, P. Helton, A. Moustafa, J. D. Cothren, and K. Luu, 'CONDA: Continual Unsupervised Domain Adaptation Learning in Visual Perception for Self-Driving Cars', in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA: IEEE, Jun. 2024, pp. 5642–5650. doi: 10.1109/CVPRW63382.2024.00573.
50. Y. Han, A. Hu, Q. Huang, Y. Zhang, Z. Lin, and J. Ma, 'Sinkhorn divergence-based contrast domain adaptation for remaining useful life prediction of rolling bearings under multiple operating conditions', *Reliability Engineering and System Safety*, vol. 253, 2025, doi: 10.1016/j.res.2024.110557.
51. X. Shang, H. Qiu, C. Jiang, P. Liang, S. Ding, and L. Gao, 'A remaining useful life estimation method based on meta contrastive learning', *Reliability Engineering & System Safety*, vol. 268, p. 111972, Apr. 2026, doi: 10.1016/j.res.2025.111972.
52. M. De Carvalho, M. Pratama, J. Zhang, C. Haoyan, and E. Yapp, 'Towards Cross-Domain Continual Learning', in *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, Utrecht, Netherlands: IEEE, May 2024, pp. 1131–1142. doi: 10.1109/ICDE60146.2024.00092.
53. L. T. Nguyen-Meidine, M. Kiran, M. Pedersoli, J. Dolz, L.-A. Blais-Morin, and E. Granger, 'Incremental Multi-Target Domain Adaptation for Object Detection with Efficient Domain Transfer', May 11, 2022, *arXiv*: arXiv:2104.06476. doi: 10.48550/arXiv.2104.06476.
54. S. Lai, Z. Zhao, F. Zhu, X. Lin, Q. Zhang, and G. Meng, 'Pareto Continual Learning: Preference-Conditioned Learning and Adaption for Dynamic Stability-Plasticity Trade-off', 2025, *arXiv*. doi: 10.48550/ARXIV.2503.23390.
55. T. Rudroff, O. Rainio, and R. Klén, 'Neuroplasticity Meets Artificial Intelligence: A Hippocampus-Inspired Approach to the Stability–Plasticity Dilemma', *Brain Sciences*, vol. 14, no. 11, p. 1111, Oct. 2024, doi: 10.3390/brainsci14111111.
56. C.-H. Li and N. K. Jha, 'PAGE: Domain-Incremental Adaptation with Past-Agnostic Generative Replay for Smart Healthcare', Mar. 13, 2024, *arXiv*: arXiv:2403.08197. doi: 10.48550/arXiv.2403.08197.
57. X. Li, X. Ma, H. Zhang, and D. Yuan, 'Condition-Adaptive Dynamic Evolution Method for Digital Twin of Rolling Bearings Based on Continual Learning', in *2024 China Automation Congress (CAC)*, Nov. 2024, pp. 5666–5671. doi: 10.1109/CAC63892.2024.10865632.

58. G. M. Van De Ven, T. Tuytelaars, and A. S. Tolias, 'Three types of incremental learning', *Nat Mach Intell*, vol. 4, no. 12, pp. 1185–1197, Dec. 2022, doi: 10.1038/s42256-022-00568-3.
59. Z. Que, X. Jin, Z. Xu, and C. Hu, 'Remaining Useful Life Prediction Based on Incremental Learning', *IEEE Transactions on Reliability*, vol. 73, no. 2, pp. 876–884, 2024, doi: 10.1109/TR.2023.3294939.
60. S. Xie *et al.*, 'Incremental Contrast Hybrid Model for Online Remaining Useful Life Prediction With Uncertainty Quantification in Machines', *IEEE Transactions on Industrial Informatics*, vol. 20, no. 12, pp. 14308–14320, 2024, doi: 10.1109/TII.2024.3450003.
61. Q. L. Xuan, M. Munderloh, and J. Ostermann, 'Lifelong Learning for Fault Prognostics in Predictive Maintenance with Bayesian Neural Networks', in *2025 25th International Conference on Software Quality, Reliability, and Security Companion (QRS-C)*, Hangzhou, China: IEEE, Jul. 2025, pp. 483–491. doi: 10.1109/QRS-C65679.2025.00066.
62. P. Zeng, W. Mao, and W. Zhang, 'Interpretability Analysis and Transferability Evaluation of Domain-Adversarial Regression Adaptation Model based on Tensor Representation', presented at the Proceedings of the 36th Chinese Control and Decision Conference, CCDC 2024, 2024, pp. 1919–1925. doi: 10.1109/CCDC62350.2024.10587731.

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