Yan et al.
Chinese Journal of Mechanical Engineering (2025) 38:5
https://doi.org/10.1186/s10033-024-01173-8

Chinese Journal of Mechanical Engineering

REVIEW

Open Access

Knowledge Driven Machine Learning Towards Interpretable Intelligent Prognostics and Health Management: Review and Case Study

Ruqiang Yan[1*], Zheng Zhou[1], Zuogang Shang[1], Zhiying Wang[1], Chenye Hu[1], Yasong Li[1], Yuangui Yang[1], Xuefeng Chen[1] and Robert X. Gao[2]

Abstract

Despite significant progress in the Prognostics and Health Management (PHM) domain using pattern learning systems from data, machine learning (ML) still faces challenges related to limited generalization and weak interpretability. A promising approach to overcoming these challenges is to embed domain knowledge into the ML pipeline, enhancing the model with additional pattern information. In this paper, we review the latest developments in PHM, encapsulated under the concept of Knowledge Driven Machine Learning (KDML). We propose a hierarchical framework to define KDML in PHM, which includes scientific paradigms, knowledge sources, knowledge representations, and knowledge embedding methods. Using this framework, we examine current research to demonstrate how various forms of knowledge can be integrated into the ML pipeline and provide roadmap to specific usage. Furthermore, we present several case studies that illustrate specific implementations of KDML in the PHM domain, including inductive experience, physical model, and signal processing. We analyze the improvements in generalization capability and interpretability that KDML can achieve. Finally, we discuss the challenges, potential applications, and usage recommendations of KDML in PHM, with a particular focus on the critical need for interpretability to ensure trustworthy deployment of artificial intelligence in PHM.

Keywords PHM, Knowledge driven machine learning, Signal processing, Physics informed, Interpretability

1 Introduction

Prognostics and health management (PHM) is an engineering discipline to extend life cycle of physical systems in service, including anomaly detection to identify binary health state, fault diagnosis to isolate the fault location, fault prognosis to predict remaining useful life, and condition-based maintenance to optimize maintenance schedule. A general workflow is composed of sensor data acquisition, feature extraction, and decision making, where feature extraction is a cornerstone to convert high dimensional sensor reading to low dimensional states and the changes in these states can be detected using pattern recognition approaches. To understand the degradation process of physical systems, there are mainly two distinguishable kinds of approaches to extract feature for PHM, i.e., physics-based and data-driven. For the former, degradation process is represented by concrete physical variables, like crack length [1] or stiffness [2]. The development of physics-based methods has led to various advances in PHM knowledge, like degradation patterns

*Correspondence:
Ruqiang Yan
yanruqiang@xjtu.edu.cn
[1] State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH 44106, USA

in life-cycle period [3] or performance model between damage variables with responsible variables [4]. For the latter, measurement data is expected to contain information about the degradation process of the physical systems, like changes in frequency components [5]. Signal processing [5] and machine learning are two mainly approaches to extract such pattern from data. In the last decades, signal processing has developed a variety of transformation methods to analyze data for PHM, like Fourier transformation or wavelet transformation. These transformation methods map data from original representation space to another, making patterns discriminative. Compared with implicit mathematical modeling evolution of system state variables through physics-based approaches or signal processing methods, machine learning (ML) provides a powerful learning system for feature extraction in PHM [6]. This powerful learning system shows strong flexibility and generalization ability in PHM with support of various types of models, like classical multilayer perception (MLP) or currently popular Transformer. These success scenarios are grounded in the different inductive bias of different ML models. For example, (1) MLP is a universal approximation function and can approximate any desired functions or operators in pattern recognition. (2) Convolution neural network has symmetry property of translation, and this inductive bias can improve its generalization ability with respect to period feature, like impulses in vibration signal. (3) Recurrent neural network is relative to wave equation or differential equation and it can capture temporal relation from time series data. (4) Graph neural network is a natural choice to model unstructured relation within multiple sensors or multi-modal data. Essentially, these general inductive biases are merely the intrinsic characters of the models. If there is no further specification, they still require sufficient data samples to train an efficient learning system, while data scarcity and imbalance are persistent problems in PHM. Due to the conservative maintenance strategy, variable conditions, and unknown fault types in PHM domain, sensor data is usually represented in a long tail distribution. Such data with insufficient variation will result in overfitting for ML models. In addition, PHM domains have another requirement for ML models to be interpretable and trustworthy. As PHM domain is risk sensitive, a wrong prediction can lead to heavy loss of life and property. From this perspective, an unexplained model does not allow users to participate in the machine's decision loop, nor does it allow users to make trustworthy decision.

To solve the abovementioned difficulties, current researches attempt to embed specific domain knowledges into ML models to improve generalization ability and interpretability, ranging from expert knowledges, physics knowledges to signal processing knowledges. Expert knowledges, like causal graphs [7] or probability equations [8], have been integrated into ML models to achieve some desired properties such as variable dependence and smoothness in states prediction for PHM. Physics knowledges, like physical equations [9] or simulation [10], also show great success to enhance physical interpretability of neural representation extracting from sensor readings. As long-term development research in PHM, signal processing knowledges provide rich data analysis techniques to specialize the ML pipeline for PHM tasks [5]. Knowledge has broad definition with understanding of empirical facts. In science or engineering, knowledge generally comes from the observation of experimental phenomenon, theory modeling of the first principle, and information system of computational science. As knowledge is expected to have invariant applicability for entire domain, generalization can be seamlessly expected to be improved. More importantly, knowledge integration can constrain the function of ML models into a specific space, and then the prediction logic can conform to the specific knowledge constraint and be more interpretable. Based on the classification of interpretable machine learning approaches (i.e., post-hoc and ad-hoc), most of knowledge-driven methods should fall into the ad-hoc category, as these methods typically involve active intervention in the modeling process, like feature transformation or optimization regularization. This is why more and more attention has been paid in this topic, including physics-informed machine learning [11], theory-guided machine learning [12], or science-guided machine learning [13]. Especially in PHM domain, various knowledge, like inductive experience, physics model, signal processing, have been developed for over decades. Recent research tendency shows the scalability of ML models to be combined with various knowledges in PHM. Through the integration of domain knowledges with data-driven learning systems, such hybrid approaches promote ML pipeline to capture more interpretable patterns from sensor readings and show potential to trustworthy artificial intelligence in engineering applications. Therefore, a systematic review is needed to investigate the common ground and diversity in literatures and then to capture its main research direction.

To summarize existing research in knowledge driven machine learning (KDML) in PHM and identify trends and gaps, we provide a hierarchical structure to organize the advancement of KDML research. Rather than the current literature search methodology to group KDML mainly based on embedding approaches (including three categories: data-centric, model structure-centric, and optimization centric) [14–16], our hierarchical structure refers to a recent survey on informed machine learning

[17] and will provide a roadmap from the knowledge sources, knowledge representations and the knowledge embedding approaches. Especially for PHM task, this roadmap will emphasize the currently mainstream trends of KDML in terms of knowledge source and representation, where signal processing and physical model are the two themes that dominate. Compared to general KDML in other engineering fields [11, 18], KDML in PHM will concern more about the specific challenges in lifecycle management and dynamic operational conditions in predictive maintenance to pursue a reliable and trustworthy solution. In addition, there also exists some reviews of interpretable machine learning for PHM [19, 20], which also paid attention to knowledge-driven machine learning approaches, such as physics-informed neural networks and signal processing-informed neural networks. The difference between our review with existing literatures is that we will establish a hierarchical framework of KDML to provide a practical roadmap, including knowledge sources, knowledge representations, and knowledge embedding approaches.

Our goal is to investigate the main research directions and approaches of knowledge driven machine learning in PHM and then provide basis references for potential users in PHM domain towards interpretable and trustworthy ML applications. We start from the knowledge in PHM by where we can conclude knowledge and how to represent it. The difference between classical ML approaches and knowledge driven machine learning is illustrated to distinguish how to integrate knowledge into ML pipeline. Through hierarchical review in knowledge sources and ML pipeline, we introduce the main research approaches to implement KDML in PHM. Our contributions are listed as follows:

(1) We utilize an inclusive concept to ordinate different approaches like physics-informed or theory-guided to integrate knowledge and data-driven models in PHM domain, named as knowledge driven machine learning. This clarification states the common ground of recent researches which integrate knowledge independent of data acquisition into ML pipeline to improve generalization ability and interpretability.
(2) We provide a hierarchical overview on KDML implementation in PHM by knowledge sources and ML pipeline. This literature analysis methodology emphasizes the classification of different knowledge sources, and then potential users in PHM domain can identify where they can be used.
(3) We present extensive case studies to show the results of KDML in PHM for various knowledges, including signal processing, physics model, and inductive experience. Although the knowledge sources of these case studies are diverse, they all follow a similar hierarchical framework of KDML about knowledge representation and knowledge embedding.

The structure of this paper is as follows: In Section 2, following a brief description of knowledge in PHM domain, the concept of knowledge driven machine learning is introduced. In Section 3, the development of KDML in PHM is summarized and classified depending on different knowledge sources. For each knowledge source, we present the knowledge embedding approaches with respect to ML pipeline. In Section 4, we give several case studies of different knowledge sources to describe generalization and interpretability performance of KDML in PHM. We discuss the future development direction and challenges in Section 5, and conclude in Section 6.

2 Concepts of KDML for PHM

In this section, we will provide the details of hierarchical structure roadmap for KDML in PHM, generally including knowledge sources, knowledge representations and knowledge embedding approaches. In Subsection 2.1, three main knowledge sources in PHM are summarized from advancement of scientific paradigms in PHM, including inductive experience, physics model, and signal processing. In Subsection 2.2, concrete forms of knowledge representations in PHM are described. In Subsection 2.3, we describe the main purpose of KDML in PHM and show the knowledge embedding approaches.

2.1 Knowledge Sources in PHM

As knowledge has broad definition in different fields, we restrict the discussion of knowledge in terms of engineering domain, especially PHM. In general, knowledge is a result from empirical observation or scientific experiment. Therefore, knowledge source usually changes with the evolution of scientific paradigm, ranging from empirical science, theoretical science, computational science, and to current data science. In the early stage of PHM research, a general way is to do statistical analysis to build relationship between external physical phenomenon and inherent degradation factor, where knowledge can be represented in a fault tree or fault causal graph. With deepening the understanding of systems, the first principle theory is used to construct theoretical model, such as lumped parameter models or state space model, where knowledge can be represented in an equation or formula. If further employing advanced computing technique, we can utilize system theory, like finite element method (FEM) or computational fluid dynamic (CFD), to approximate the complex physical system, where knowledge can be represented in such computational system. If big data of the system is available, advanced data processing techniques can be used to recognize fault pattern from data, like using signal processing or ML, where knowledge implicitly exists in such data driven models.

Here, to make knowledge and data more distinguishable in the scientific paradigm of data science, we exclude the knowledge implicit in ML models, while signal processing will still be considered as it is normally considered to be parallel to ML models in PHM field. In addition, we will group theoretical science and computational science into one category of knowledge source, physics model, as the two paradigms are both originated from the first principle. For empirical science, we specialize it as inductive experience in PHM field and split it to two subset of knowledge source, that is, empirical model with explicit formula in terms of experience modeling (like tool wear model) and without explicit formula in terms of inductive knowledge (like fault tree). Figure 1 illustrates the relationship from scientific paradigms to PHM knowledge sources.

2.2 Knowledge Representations in PHM

After defining the three main knowledge sources in PHM, the next thing is to extract specific knowledge representations from different knowledge sources. This category further subdivides the knowledge sources and provides an interface for the following knowledge embedding. Here, we give a brief overview of these knowledge representations in PHM.

Causal graph: Causal graph represents the relation between different variables, where relation is denoted by edge and variable is denoted by node. A classic example is fault tree, that all the system states of interest are organized in a causal graph and fault reason can be traced in this graph.

Logic rule: Logic rules refer to a symbolic language that demonstrates a phenomenon with a set of Boolean expressions and logical operators (e.g., $\wedge, \vee, \neg, \ldots$). For example, the rule "if condition A and B exists, then event C occurs" can be described as "$A \wedge B \Rightarrow C$". The simple representation is easy for users to understand and apply.

Physics equation: Physics equation is derived from the first principle including the basic governing equations, aiming to describe the working principle of certain physics process or phenomenon. The equations are generally unique for different physics entities. For vibration analysis in mechanical systems, dynamic models are engaged based on the Newton's law.

Simulation: Simulation data refer to the high-fidelity data generated by computer software based on the detailed dynamic models derived from the first principle. The most used simulation method is the finite element method that solves real engineering problems and obtains high-fidelity data as prior knowledge.

Signal equation: Signal equation represents the perspective to analyze the signal, where the perspective can be time domain, frequency domain, and time-frequency

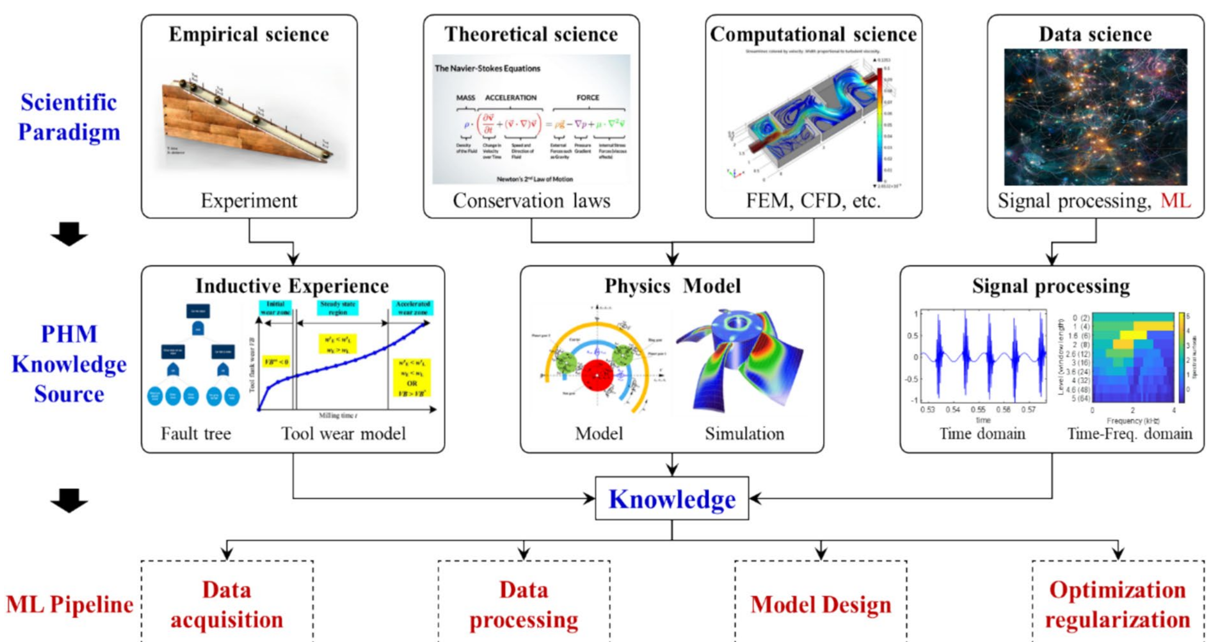

Figure 1 Knowledge sources in PHM domain

domain. For example, with Fourier transform, signal can be view as a linear combination of sinusoidal and cosinusoidal functions (complex exponentials).

Probability equation: Probability equation is capable of describing relationships among random variables. This relationship can be derived from observations and statistical analysis of likelihood. On this basic, probability equations guide observers to make predictions about the likelihood of future events.

Statistics property: Statistical property characterizes the description of behavior from a population, which can usually be achieved through mathematical tools. As a form of empirical knowledge, Statistical property reflects the understanding of the data from observers. This understanding can help observers infer the distribution, trends, and relationships of similar data population to make better predictions and decisions.

Others: Despite the above knowledge representations can cover mainstream knowledge in PHM domain, there still exists some other knowledge for specific application or expert knowledge, like attribute, hybrid model. Therefore, we group these specific methods to an additional category.

2.3 Knowledge Driven Machine Learning Pipeline

The significant difference between traditional machine learning and knowledge driven machine learning is information source for them to train a learning system. For traditional machine learning, data is the starting point of the whole pipeline and then a learning algorithm is designed to approximate the underlying function behind data. In addition, traditional approaches will consider conciseness of algorithm following the Occam's razor principle to reach a simplest hypothesis. The purpose of traditional ML pipeline can be formulated as follows:

$$\text{goal} \propto \text{task performance} + \text{algorithm conciseness}.$$

For knowledge driven machine learning, data and knowledge are two parallel information sources for the learning pipeline. As shown in Figure 1, knowledge can be integrated into each step of pipeline. Even for data acquisition, knowledge can help to optimize sensor placement or acquisition strategy. Therefore, the purpose of knowledge driven machine learning can be formulated as follows:

$$\begin{aligned}\text{goal} \propto\ & \text{task performance} + \text{algorithm conciseness} \\ & + \text{knowledge conformit}.\end{aligned}$$

Through integrating knowledge into ML pipeline, generalization ability and interpretability of the learning system are expected to be improved. For generalization, knowledge itself is an invariant representation with respect to distribution shift in data. For interpretability, knowledge integration will actively modify the modeling process of ML pipeline or passively test the prediction of ML models. Such active and passive interpretability is the key to achieve trustworthy AI in PHM.

We utilize a Sankey diagram to visualize the roadmap from scientific paradigms to knowledge driven ML pipeline in Figure 2. The paths illustrate the direction on how to start from a specific knowledge source to embed it into a specific module in ML pipeline. For example, the scientific paradigm of data science can derive the knowledge source in terms of signal processing, and then this kind knowledge source can be described in two kinds of knowledge representations, that is, signal equation and statistic property. Finally, signal equation or statistic property can be integrated in to ML pipeline, including data processing, model design and optimization regularization. Despite each knowledge representation has the potential to adjust each part in ML pipeline, the practical integration approach usually depends on the reality of research, and the literature review shows that model design and optimization regularization are the two main embedding parts in ML pipeline.

3 Description of Knowledge Driven Machine Learning in PHM

In this section, we give a detailed description of KDML approaches in PHM. We organize this description in a hierarchical framework, in which we first group KDML approaches by three knowledge sources, inductive experience, physical model, and signal processing and then for each knowledge source, we group the literatures by their embedding methods in ML pipeline.

3.1 Inductive Experience

Engineers would gain rules and draw conclusions from historical observation or experimentation of mechanical equipment, which is so called "expert knowledge" or "empirical knowledge" in the field of empirical science. In contrast to theoretical science, the justification of empirical science depends on a large number of experiments. Fortunately, empirical knowledges have been obtained from massive experiments with the development of PHM. These knowledges are valuable to guide the learning process and provide interpretability for the trained models. According to whether the empirical knowledge can be represented with algebraic expression, it can be categorized into experience models and inductive knowledge, as shown in Figure 3.

As the approaches to integrating inductive experience with the learning process are mainly concentrated

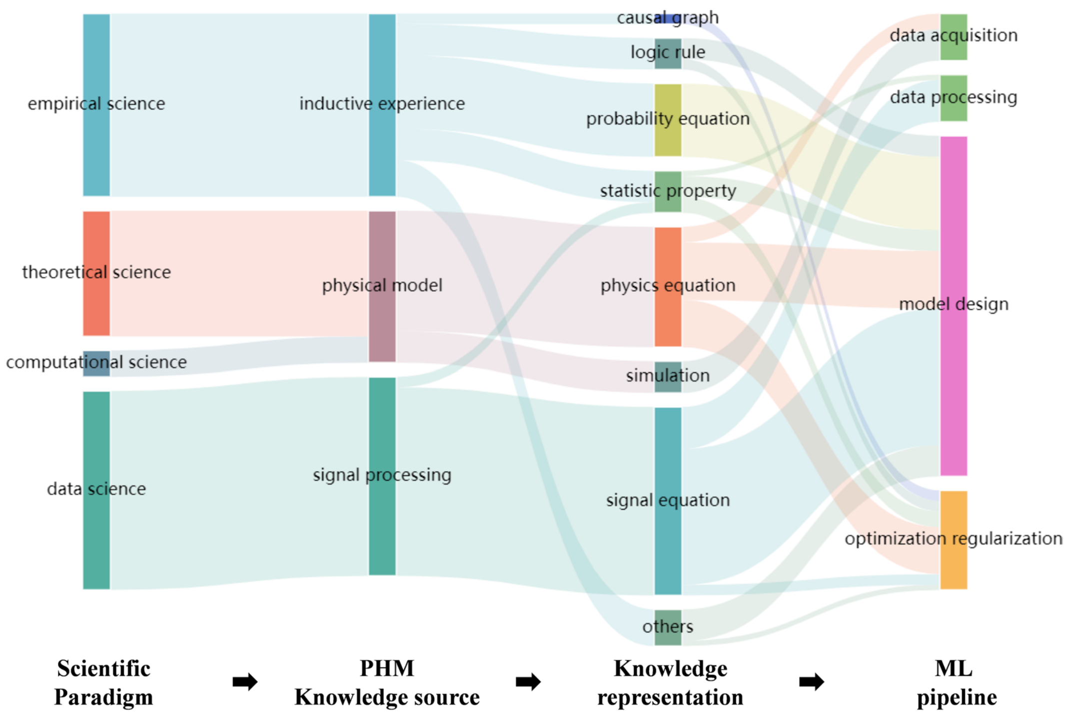

Figure 2 Overview of Knowledge driven machine learning in PHM domain

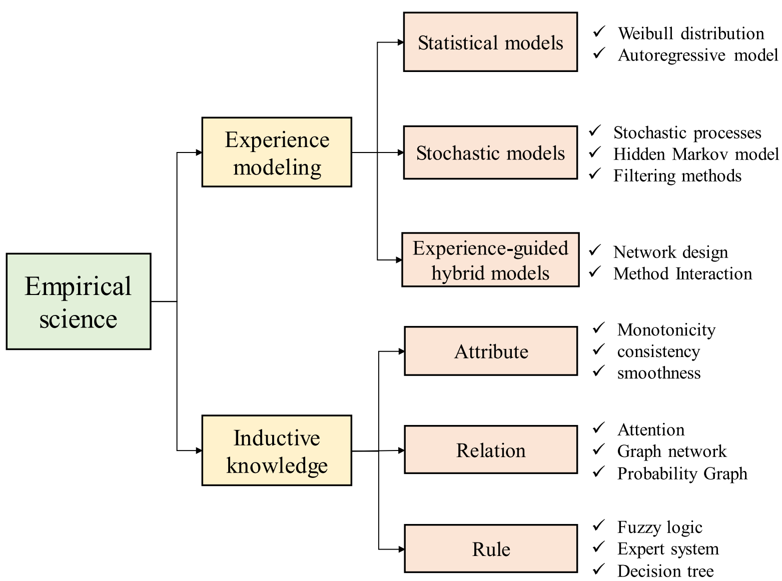

Figure 3 Classification of empirical science

on model design and optimization regularization, we do not classify inductive experience-based methods according to different stages of machine learning pipeline in this section. Instead, we divide them based on different forms of knowledge representation, as described below.

3.1.1 Empirical Model Informed Learning

Empirical modeling strives to establish a correspondence between the model and the modeled object, which represents a process of summarizing from a large amount of experimental data. Empirical models are capable of reflecting the understanding of the system being analyzed from manufacturers, which can usually be proxied in the form of mathematical models. Typically, they do not consider the underlying principles and physical mechanisms of the modeled objects. Therefore, the agent process reflects more on the dependencies of observed behaviors rather than on the inherent mechanisms by which behaviors occur. For instance, the capacity fading of lithium-ion batteries usually follows the form of exponential degradation, which is a principle derived from massive experimental observations [21]. In fact, this failure behavior is physically caused by microscopic changes in the material. This empirical knowledge can be inherited to achieve reasonable modeling of unknown lithium-ion battery. For empirical model informed learning, it integrates inductive empirical knowledge with machine learning methods to achieve good interpretability.

Empirical model informed learning can be categorized into statistical property, probabilistic equations and experience-guided hybrid models. It should be emphasized that the first two are constructed based on the empirical knowledge from the monitoring objects and are inherently interpretable [22]. That is, additional interpretable tools or modules are generally not needed to interpret the model. While for the latter, hybridization with other methods often requires unique interpretable designs. The research in this area has obtained some advancements, which will be summarized and discussed below.

(1) Statistical property

Statistical property is knowledge mined from data and formed into experience, which is expressed by mathematical form. The autoregressive (AR) model is a classic statistical model with good mathematical expression and interpretability. Many researchers have attempted to apply it in degradation modeling. This method extracts association relationships from monitoring data and extends them to the future for predicting degradation trends. For instance, Qian et al. [23] employed multi-dimensional AR model to track the extension of bearing defects, thereby achieving real-time RUL prediction. Ordóñez et al. [24] employed auto-regressive integrated moving average (ARIMA) to predict the next state of the system. Subsequently, the predicted results are fed into the support vector machine (SVM) for RUL prediction of aircraft engines. Weibull distribution is also a classic statistical model, which is usually utilized to describe the RUL probability distribution. Kundu et al. [25] utilized General Log-Linear Weibull distribution to estimate the impact of some external factors (humidity, temperature, pressure, etc.) on degradation progression. On this basis, multiple models under different failure behaviors were constructed to estimate RUL. Kundu et al. [26] further considered the influence of working conditions and degradation degree in Weibull distribution according to the engineering experience, and established a single prediction model for the entire life cycle. Experimental results revealed that modeling these two factors obtained higher accuracy.

(2) Probabilistic equations

Probabilistic equations can attempt to agent the physical evolution process from complex systems through probabilistic modeling. Among them, stochastic processes and dynamic Bayesian networks are the most common models. For example, Zhang et al. [8] utilized the Arrhenius model which is an exponential model to represent the temporal relation of physics degradation process, and induced it as the mean function of Gaussian process (GP) for prognosis of Heating, ventilation, and air-conditioning system. Jin et al. [27] used a gyroscope drift model to replace the global model in GP for prognosis, where the drift model described the fault feature of ball bearings to better capture the trajectory of degradation. Zhang et al. [28] considered uncertainty caused by external factors in degradation modeling based on the Wiener process. To be specific, unobservable factors are modeled by Brownian motion, and measurable covariates are modeled with an Ornstein-Uhlenbeck process, which is thereby linked to degradation rate. Dynamic Bayesian networks model the dependence of variables from adjacent time steps, which are generally utilized to estimate or predict the degradation state of the system. Li et al. [29] developed an improved hidden Markov model (HMM) for tool state prediction, in which changes in cutting conditions are modeled through conditional adaptive state transitions. Lyu et al. [30] believed that the capacity regeneration phenomenon is not conducive to RUL prediction of lithium battery, so variational mode decomposition is utilized to decompose the signal into trend signal and capacity regeneration signal. Then they employ particle filtering to process the trend signal for RUL prediction. Cui et al. [31] proposed a time-varying Kalman filter for bearing RUL prediction, where different degradation stages correspond to different degradation model to achieve accurate RUL prediction. Zheng et al. [32] exploited relevance vector machine to extract to predict the future system state of Lithium-ion battery and an improved Sage-Husa adaptive Kalman filtering is constructed to enhance the filtering effect.

(3) Experience-guided hybrid models

Experience-guided hybrid models usually utilize empirically extracted knowledge as the logic to interpret the model. As a typical black box model, many scholars try to enhance interpretability by embedding empirical knowledge into the network. Pei et al. [33] combined the superiority of deep learning and stochastic process to develop a hybrid method for bearing prognosis. A deep neural network was used to extract low-dimensional feature from high-dimensional data, and then a diffusion process was used to fit the temporal low-dimensional feature and estimate the uncertainty of degradation process. Inspired by state space models (SSMs), Li et al. [34] proposed a life cycle modeling method for mechanical systems. In the proposed method, emission function and state transition in SSMs are parameterized using DNN. Moreover, the coupling competition degradation prior is embedded into the state transition network as knowledge, and the attention heat map interprets the competition relationship among the three mechanisms. Deng et al. [35] employed the Wiener process to connect the RUL prediction task with DNN and the uncertainty quantification task. This surrogate modeling method approach improves near-failure prediction accuracy in an interactive manner and provides interpretability to operators. Hu et al. [36] designed deep belief networks (DBNs) to extract hidden features and select features with high tendency to be fed into local linear embedding to construct health indicators. Diffusion process was utilized to model degradation evolution according to the health indicators and the uncertainty from prediction results could be quantified. Sun et al. [37] combined Winner process with back propagation neural network to construct a cutting tool condition degradation model, which obtained excellent RUL prediction results. Dai et al. [38] proposed an interpretable wavelet kernel network as RUL prediction model, and the Wiener process was used to evaluate reliability and provide prediction uncertainty. Experiment results conducted on a bearing dataset verified the advantages of the fusion of the two methods. Chen et al. [39] designed a hybrid prognostic approach for bearings including DNN, Winner process and Kalman filtering. More specifically, gated recurrent unit network is utilized to extract degradation representation, Wiener process is employed to adaptively update the degradation state, and Kalman filtering is designed to estimate model parameters and infer real-time RUL distributions.

Empirical model informed learning can well model or describe the behavior of monitored objects through mathematical relationships, which are completely white-box and inherently interpretable. However, empirical knowledge cannot have the high fidelity of physical models, which makes it easy to fail under unknown conditions, causing the established model difficult to generalize. Furthermore, it can be concluded that hybrid methods based on empirical models have more advantages, and this is the future research direction.

3.1.2 Inductive Knowledge Informed Learning

Induction is a reasoning process from individuality to generality. The prerequisite for reasoning and knowledge discovery is the large amount of data collected from historical experiments or daily observations. By analyzing the data and discovering their common attributes, the general rules can be obtained, which we call the "inductive knowledge". It is a similar concept to the empirical model, since the empirical models are sometimes constructed on the observed data. For ease of understanding, we distinguish them by the manifestation of knowledge in this study. That is, empirical models can be described with concrete algebraic expression, while inductive knowledge is a more abstract and broader concept that covers intuitive ideas, perspectives or assumptions concluded from rich engineering practice. It is hard to express them with relational algebraic equations. However, by integrating inductive knowledge into the learning process, the learned models can avoid overfitting in specific scenarios and provide extra interpretability for the calculating mechanisms.

There are massive attempts to embed these knowledges into the intelligent models in existing literatures. The critical difficulty of such type of work is to summarize correct, generalized and optimization-promoting knowledge, thereby improving the interpretability and meanwhile maintaining the performance. According to the different forms of inductive knowledge, inductive knowledge informed learning can be classified as attribute informed learning, relation informed learning, and rule informed learning.

(1) Attribute informed learning

In attribute informed learning, knowledge comes from some key attributes or characters that conform to the observed phenomena, e.g., the monotonicity, consistency, and smoothness of the state indicators during the degradation process of equipment, the separability between normal and abnormal frequency spectrum, and the singularity of faulty signals. These attributes are summarized from data and describe physics rules. Therefore, researchers integrate them with learning process by modifying the model structure or the optimizing objective, which guides the model to learn some interpretable attributes. Zhou et al. [40] stated the differences between stochastic degradation process and proxy regression labels, and proposed a dynamic governing network to model the degrading trajectory of machines, where discretized ordinary differential equation was parameterized to ensure the monotonicity of the degrading trajectory. Compared with traditional methods, the predicted RUL curve is monotonically decreasing, providing more interpretability for the proposed model. Yan et al. [41] proposed an interpretable weight learning framework with two-stage convex optimization, where the first optimizing objective is to enlarge the separability of health indicators, and the second objective combines the monotonicity and fitness properties of the degrading states to generate piecewise health index. With the property constraints, the model weights reveal informative frequency bands. Motivated by the effectiveness of statistical complexity in quantifying potential dynamic changes, Yan et al. [42] proposed a weight-based sparse degradation model and introduced a set of entropy-based health indicators to quantify the interpretable model weights. Experimental results confirm that the learned weights magnify the weak fault features in online incipient fault detection. Besides, Zhou et al. [43] leveraged the advantages of singular values in revealing weak fault information, and designed graph-modeled singular values that combines graph theory and singular value decomposition (SVD). By constructing graph with singular values, the proposed method realizes a balance between sensitivity to early fault and robustness to noise.

(2) Casual graph informed learning

Although the operating mechanisms of complex systems are hard to obtain, the dependence, causality or logical relationship between variables can sometimes be induced according to observations or physical structures. This form of knowledge can be viewed as an auxiliary constraint to regulate model training, which also endows the model with partial interpretability. The approaches to knowledge integration mainly focus on designs of model structures, such as using attention mechanism, selecting graph neural networks, constructing probabilistic graph model, etc. For example, Ma et al. [7] proposed a graph-based abnormal sensor localization and fault detection method for lithium battery packs, which constructs a graph autoencoder based on the layout of voltage sensors. The reconstruction of both signals and the graph structures are required, which greatly reduces the time delay for fault detection. Li et al. [44] focused on the impulse characteristic of vibration signals and argued that the attention weights to different segments should follow sparse distribution. Sparse constraints are then added in the Transformer encoder. Post-hoc attention visualization reveals that attention weights concentrate on the fault related impulse. Bi and Zhao [45] proposed an orthogonal self-attentive variational autoencoder model, where the casual relationship between process variables and the temporal dependency can be extracted via spatial and temporal self-attention layer. Liao et al. [46] proposed a self-attention assisted physics-informed neural network (PINN) for aeroengine life prediction, where self-attention mechanism is embedded into PINN to learn more accurate physical relationship. Martel et al. [47] mapped the monitored data and time into the latent variable by partial differential equation. The latent features are then used as health approximation to explicitly deduce the temporal behavior of data. When dealing with complex problems like health management in varying working conditions, the causality between health states, working conditions and signals can serve as prior knowledge to assist model training. Li et al. [48] proposed a casual disentanglement network to realize cross-machine knowledge generalization, which combines domain separation loss, classification loss and disentangled loss to capture invariant fault information. Hu et al. [49] utilized information theory to deduce the equivalent optimization objective to extracting condition-independent features, and designed an adversarial training manner by minimizing variational upper bound and self-supervised learning. The learned model can reduce false or missing fault detection and adapt to time-varying scenarios.

(3) Logic rule informed learning

Logic rule informed learning expect to embed rule-based knowledge into model learning. Various expression forms of the extracted rules (e.g., fuzzy logic, decision tree, expert systems) can be inserted as prior knowledge into the deep networks, thereby endowing the network with interpretability. For example, Wong et al. [50] proposed a fuzzy extreme learning machine for industrial fault diagnosis, where fuzzy membership functions, rule-combination matrix are embedded into the extreme learning machine. Without loss of accuracy, the output weights are utilized to form the class and confidence for any rules, providing explicit knowledge in an interpretable manner. Yu and Liu [51] inserted the inductive symbolized rules and confidences into a deep belief network, which enables the model to determine the network adaptively. Wu et al. [52] proposed a cluster-based hidden Markov model to learn the mapping between critical performance index and RUL. Then a semantic rule-based inference module is attached to recognize the root factor for performance degradation. Steenwinckel et al.

[53] proposed a knowledge and data joint driven anomaly detection model to improve the model's representation ability. According to abnormal feedback, knowledge driven methods can generate new rules to realize the adaptive update. Zhou et al. [54] argued that the reasons for decreasing interpretability are parameter over-optimization and the deviations to expert judgements. They evaluated model interpretability by measuring the consistency index of rules, consistency index of rule set, and over-optimization index, achieving both high accuracy and good interpretability. Furthermore, Ming et al. [55] proposed an interpretable diagnosis model integrating both rule base and probability table. Interpretability constraints are added to the adaption evolution strategy of covariance matrix to ensure the model interpretability.

3.1.3 Summary for Inductive Experience in KDML

By contrast to signal processing techniques and physics models that explicitly describes the properties of signal and physical entities, inductive experiences are observed and concluded from historical phenomena. Their interpretability is less explicitly revealed compared to the former two sorts of knowledge. However, there are massive regular patterns implicitly hidden in engineering practice. By integrating them into the optimization process, the model would avoid overfitting and learn the core relationship between input and output. The interpretability and performance of model would be improved simultaneously.

3.2 Physical Model

Physics informed machine learning (PIML) incorporates physical knowledge and data-driven machine learning model. It starts from the observation of working principles for the physical entity, establishing physical models that describe certain physics phenomena or process. Physical models are usually governed by differential equations, and these equations could be derived into different forms for different orientations, i.e., state space models, discrete difference forms, finite element analysis, etc. Physics informed machine learning constructs physical models from the first principle, aiming to figure out the basic governing equations for physical phenomena or process as prior information for machine learning.

Data-driven models establish the relation between inputs and outputs through violent mapping, leading to the lack of interpretability. However, in PIML process, physical models reveal the causality between system excitation and response by governing equations, which assigns physical meaning to machine learning results. Physics knowledge observed from the corresponding physical models is usually unique, and preserves within one or a group of physical entities. For the battery RUL prediction problem, electrochemical models and equivalent-circuit models based on Kirchhoff's law are established to relate the state of health and discharge time. Besides, for most mechanical systems, conservation laws are governing the major parts, including mass, momentum, and energy conservations. As for the degradation process, there are certain phenomenological models for different scenes, such as Paris law for crack propagation. The following works show the discovery mechanism for physics knowledge starting from the first principle. Luo et al. [56] conducted a dynamic model of a mechanical system and assumed that concerned stiffness is coupled parameter of the degradation variable. By using a polynomial function to approximate this relation, this method can reveal the degradation process. Mojallal and Lotfifard [57] utilized Hybrid Bond-Graph theory to construct a multi-physics graphical model of a mechanical system to capture its causal relation, and this method was developed for fault detection and isolation in wind turbines. The mechanism shares the same for different physics entities. Thus, through the engaging methods of physical knowledge and learning process, PIML methods can be divided into physics informed data augmentation, physics informed network architecture design and physics informed loss function construction, as is shown in Figure 4.

3.2.1 Physics Informed Data Augmentation

Physics informed data augmentation aims to incorporate physics knowledge for class imbalance problems in the field of interpretable intelligent diagnostics. High-fidelity physical models are essential to generate time series data, and these models typically composed of the simulation model-based approach which utilizes computers to obtain time series data by solving physical models. To ensure the fidelity, simulation models pay more attention to the details in physical systems, apart from the general conservation equations applied in other subsections. Thus, these models generally require large computational costs and subject to computer science paradigm. The simulation approach focuses more on understanding and interpreting the physical system, and needs to consider the physical meaning and practical application requirements, so as to ensure the interpretability and validity of the model. It is worth noting that while there are also generative models based on GAN and other generative models to generate virtual data, we do not consider this a simulation model when considering that it has no actual physical significance. However, if it is possible to further add physical constraints to the GAN to improve the results, in which case the GAN itself is driven by knowledge, we support such a paradigm.

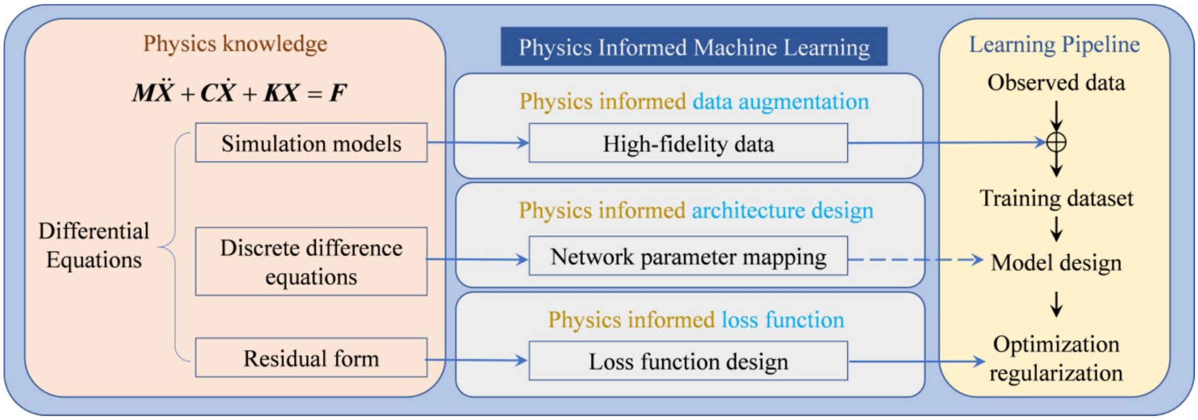

Figure 4 Classification of physics informed machine learning

Researchers have conducted extensive studies on simulation model-based KDMLs. Some scholars use the centralized parameter method to model physical systems and solve simulation models to obtain data to guide the network to learn the knowledge that has physical meaning. Cui et al. [10] developed a comprehensive dynamic model by dividing the degradation process of rolling bearings into four different stages while considering the coupled excitation of time-varying morphology and stiffness. The response of the simulation model was solved to obtain a large amount of performance degradation data for RUL prediction. Yu et al. [58] established a dynamics model of the rotor-bearing system, constructed a source domain dataset, and utilized the diagnostic knowledge obtained from the simulation data to realize domain adaptive transfer learning for mechanical equipment. Ni et al. [59] utilized modal-property-dominant features to learn the underlying physical knowledge embedded in the training and test data, and proposed a physical-informed residual network that considers the physical information in the data.

Since the lumped parameter method simplifies the physical system too much, in order to obtain higher fidelity data, other studies use the finite element method to construct the simulation model and extract a priori information about the physics. Kohtz et al. [60] developed a physically based finite element model for lithium batteries and fused the finite element results with experimental data to construct a multi-fidelity model for battery state of health estimation. He et al. [61] used numerical simulation data of fluid-structure interaction in place of actual monitoring data to realize the RUL prediction of centrifugal pumps based on a ladder network. Agarwal et al. [62] developed a finite element analysis model of the electric motor in an electric vehicle, which introduced faulty defect conditions, obtained surrogate current data from the simulation model, and finally classified the faults using a support vector machine. Zhao et al. [63] developed a finite element model of gears to further obtain a crack propagation model so that the distribution of failure time and RUL can be predicted. Then, the collected condition monitoring data are fused by Bayesian method to estimate the crack length and realize more accurate RUL prediction.

In addition, there are other domain-specific simulation model modeling approaches that also allow access to high-fidelity data. The best known of these is the C-MAPSS simulation model, which provides degradation monitoring data for turbine engines and is widely used for learning and training in the field of remaining life prediction. López de Calle-Etxabe et al. [64] used Matlab Simcape to build a physical model of a linear actuator for generating data in health and failure scenarios, and fused the simulated data with actual measurements to establish a diagnostic model. Djeziri et al. [65] used Bond Graph to physically model the wind turbine to obtain data on normal operation and failures, and used this data to predict the RUL based on geolocation principal.

Physics informed data augmentation approach typically contributes very little to explain the final output of the model or the structure of the network, but they learn by applying the data generated from modeling scientific knowledge in the domain. This introduces prior knowledge of data distribution into the model, corrects the initial iterations of the model, and guides the model to use the priors to discover interpretable knowledge in the actual data. However, the difficulty of this method is that a reliable physical model of the system needs to be established for the simulation data to increase its similarity to the observed data, further ensuring the accuracy of the model iteration direction.

3.2.2 Physics Informed Network Architecture Design

Physics informed network architecture design combines the mathematical calculations of the network model with the solving process of the physical model, assigning specific physical meanings to network parameters, thereby enhancing interpretability. Chen et al. [66] introduced an ordinary differential equation solver based on the residual connection in residual networks or the recursive calculation in recurrent neural networks. This method employs the discrete form or the recursive form of differential equations to neural networks, mapping network parameters to the physical quantities in the equation. Based on this approach, researchers started to study the equations governing health state of certain equipment and incorporated physics equations and neural networks to fault diagnosis and health assessment. Dourado and Viana [1] developed a physics-informed neural network for bias estimation in corrosion-fatigue prognosis. This method utilized the Paris Law to describe the phenomenon of corrosion-fatigue crack propagation process, and used a neural network to model the corrosion part. In this work, the physics knowledge provides the relation of variables in prognosis model while the neural network provides a parameter update method in physics model. Nascimento et al. [67] provided a tutorial on integrating ordinary differential equations into RNN and used two case studies to illustrate it, that is, a fatigue crack growth model and a dynamic two-degree-of-freedom system. The proposed method worked as the classical system parameter identification. Firstly, an evolution model of physical variables is derived based on physics knowledge, such as conservation laws or constitutive relation. Then, a neural network is utilized to fit some variables in the evolution model which are difficult to model or observable. Nascimento and Viana [68] integrated cumulative damage model into RNN for crack length prediction of a fleet of aircraft, and they developed several RNN cell styles to combine physics with data. Yucesan and Viana [69] buried degradation model for bearing fatigue in RNN cells and calibrated the proposed model by visual grease inspection. Besides, Yucesan and Viana [70] improved their bearing fatigue prognosis model by considering the uncertainty caused by grease quality variation. Yucesan and Viana [71] further proposed a hybrid method for wind turbine main bearing fatigue prediction, where a physics model was used to model the L10 fatigue life and a neural network was used to model the grease degradation process. Tipireddy and Tartakovsky [72] presented a physics-informed Gaussian process for monitoring and forecasting of power grid dynamics. The proposed method utilized a random process to model the evolution of unknown physical variables and then compute the covariance matrix from it. Samundsson et al. [73] proposed a variational integrator network to model dynamic system, in which the structure of the neural network matches the discrete-time equation of motion. Chao et al. [74] proposed a hybrid prognosis method by fusing physics features with data features. This fusion first used unscented Kalman filters to estimate the unobservable model parameters, and then merged and fed these parameters with measurements into a deep neural network for prediction. Nascimento et al. [75] developed a hybrid method integrating a reduced-order physics model with RNN for lithium-ion battery modeling and prognosis. This method utilized the Nernst and Butler-Volmer equations to represent the battery discharge process and encoded this reduced-order model into the RNN cell structure.

The design of physics informed network architecture makes the combination of physical knowledge and neural networks more closely. The mapping of physical models and the mathematical models of neural networks assigns the internal parameters of the network with physical meanings and further brings physical interpretability.

3.2.3 Physics Informed Loss Function

Physics informed loss functions are derived from governing equations based on the first principle, constricting the optimization direction of neural networks, and further ensuring the interpretability of learning results. Zhang and Liu [9] established a Parsimony-enhanced sparse Bayesian learning method to discover the governing Partial Differential Equations (PDE) of nonlinear dynamic systems and validated its application in anomaly detection. This method utilized Bayesian inference method to reduce the information loss before the sparse regression procedure. Lutter et al. [76] developed a Deep Lagrangian Network to learn the equations of mechanical motion (i.e., robot tracking control) while ensuring physical plausibility. This method derives the loss function of the neural network from the Lagrangian equation and applies symmetric and positive constraints to the parameter matrix based on the properties of physical quantities. Zhang et al. [77] incorporated the physics knowledge into neural network from the loss function aspect. This method leveraged available yet incomplete physics information, such as governing equation or states relation, to match network parameters and physical ones, and then encoded a physics loss to constrain the solution space of the neural network. Further, Zhang et al. [78] embedded physics knowledge into a CNN for building structure response prediction. This method established a dynamic model to derive a physics loss term, so as to alleviate overfitting problem of CNN. Unfortunately, due to the agnostic of partial variables, this physics loss can only be verified on mathematical simulation but cannot be used on the real engineering data. Chen et al. [79] proposed a degenerate consistent recursive network to study the physical characteristics of bearing degradation process, and trained the network through loss function constraints to improve the accuracy and interpretability of bearing fault prediction results. Freeman et al. [80] raised a new physically guided rotor blade imbalance fault detection framework, which combines non-invasive fault features obtained from turbine power signals with environmental condition data to customize loss functions for neural networks to enhance fault detection capabilities. Shen et al. [81] established a fault threshold model based on physical knowledge and combined it with a neural network model. By designing a loss function, the influence of physical knowledge is selectively amplified to achieve bearing fault detection. Xu et al. [82] proposed a physically constrained variational neural network for evaluating the wear status of external gear pumps. In this method, spectral methods are involved to establish a pressure pulsation model for gear pumps which is further transformed into a physical loss term, constraining the learning process of neural networks while enhancing physical meanings of learning features. Wang et al. [83] designed physics informed loss function based on the fluid mass conservation law in machine learning process, connecting the learning results to piston wears of axial piston pumps, and realizing the interpretable health assessment.

The loss function determines the direction of network optimization, and the addition of physical loss functions directly constrains the network optimization process with physical knowledge. Designing loss functions based on the different tasks and target systems and balancing the weights of each item in the loss function according to the order of magnitude of the described physical quantities makes the combination of physical knowledge and neural networks clearer and enhances the interpretability of the network.

3.2.4 Summary for Physical Model in KDML

Physics informed data augmentation utilizes the data generated by the simulation model to train the deep network model, allowing the model to obtain a reasonable initial value prior. At the same time, this method relies heavily on the quality of the data from the simulation model, which makes it difficult to obtain high-fidelity simulation data when dealing with extremely complex physical systems. The physics informed loss function, on the other hand, introduces the physics loss as a regularization term to guide the optimization direction of the network. While controlling the direction of the network output to make it more consistent with physical laws, the regularization term also reduces the search space of the network parameters and increases the interpretability of the model's predicted output. However, the physical models they build are generally low-fidelity, and it is extremely difficult to derive a high-fidelity physical model into the formula form of the regularization term. In addition, the internal structure of the network of the above two methods is still a black-box model. In contrast, the physics informed network architecture design happens to enhance the interpretability of some parameters in the network structure. But it usually requires different network structures in different scenarios, and the network structure is less generalizable. The above three methods increase the interpretability of the network in terms of initial value selection, optimization direction and model structure, respectively. Nevertheless, for highly complex physical models, how to further embed the knowledge into the network according to the above three methods needs further research.

3.3 Signal Processing

Signal processing informed neural networks (SPINN) is an interpretable network that combines the prior information of signal processing technology and the data-driven capabilities of deep learning. The prior information comes from the signal analysis and feature extraction methods that have been extensively developed and widely used in the field of health management, such as time domain, frequency domain, and time-frequency domain analysis techniques. Since signal processing prior has been validated in industrial applications, SPINN can often obtain better performance with a reasonable prior. In addition, SPINN can be easily interpreted from the signal processing perspective and thus better understood by users.

From the constituent element of network training, there are three kinds of methods to embed the signal processing prior into the network and construct SPINN, i.e., data, model, and optimization, as is shown in Figure 5.

1) Data: signal processing methods can be easily used for data filtering, data enhancement, and feature extraction. The processed data are then directly used by the network for health management.
2) Model: signal processing methods are used as a component in a network with fixed/learnable parameters or guiding theory to design new network structures.
3) Optimization: signal characteristic priors obtained by signal processing methods are used to generate regularization on the optimization target.

Recent work on SPINN is introduced from the three aspects.

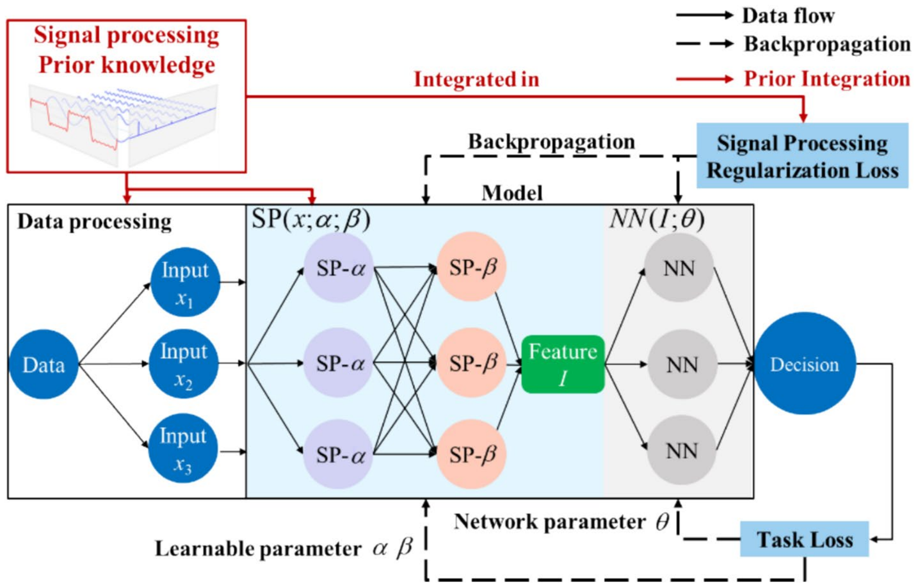

Figure 5 Signal processing informed neural networks

3.3.1 Signal Processing Informed Data Preprocessing

As the simplest way to realize SPINN, data processed by signal processing method can directly introduce prior information without influencing the model and optimization.

1) Data transform or filtering realized by Fourier transform, short-time Fourier transform, wavelet transform, and so on can provide another view of the signal which is better for health management tasks. Lahcen et al. [84] utilized stationary wavelet transform to extract features from raw ECG signals and then input the decomposed wavelet sub-bands into a 1D-CNN for heart disease diagnosis. Wen et al. [85] applied S transform, a time-frequency technique, to transfer the vibration signals to images, and then input into a CNN to classify these images for fault diagnosis. Li et al. [86] utilized short-time Fourier transform to get time-frequency representation of the vibration signal. Then a CNN is used to extract multiscale features and realize RUL prediction.
2) The adjustable parameters of the signal processing method and the procedure of signal analysis and synthesis provide a chance for data augmentation. To solve class-imbalance problem in fault diagnosis, Fu et al. [87] proposed local wavelet similarity fusion to augment high-quality fault data. Wavelet packet transform is used to decompose signal into different frequency bands and the amplitudes of wavelet coefficients in selected bands are distorted by a similarity-based weighting strategy with reference samples. Signals reconstructed from the distorted coefficients serve as the augmented samples. Due to the scarcity of labeled failure data, Kulevome et al. [88] utilized analytic wavelets to obtain augmented scalograms. By successively adjusting the predefined decay parameter of the generalized Morse wavelet, scalograms with different energy concentrations are obtained. Then augmented scalograms and the original scalograms are combined as training data for model optimization.
3) The last way to embed signal processing priors in data is feature extraction. Since a large number of feature extraction and indices design methods based on expert experience have been developed, effective utilization of these methods can embed priors in input and simplify the model's feature extraction module. Meng et al. [89] proposed an ensemble learning method for online joint strength prediction in ultrasonic metal welding. The variability in joint strength is decomposed into a large-scale term characterizing the influence of physics conditions and a small-scale residual term characterized by online sensing data of power, displacement, microphone, and acoustic emission. Discrete wavelet transform is used to extract features from sensing data, which is used by multilayer perceptron (MLP) for tool condition prediction and gradient boosting machines for prediction. Zhao et al. [90] proposed a local feature-based gated recurrent unit network for machine health monitoring. Classical time-frequency methods are used to extract typical features, which are used by a

GRU network and fully connected layers for prediction and diagnosis. Zhu et al. [91] used wavelet transform to get the time-frequency representation of vibration and then input it into a multiscale CNN for bearing remaining useful life (RUL) prediction. Ren et al. [92] utilized several time-frequency techniques to extract features from vibration signals and then input them into a deep neural network for bearing RUL prediction.

3.3.2 Signal Processing Informed Model Design

Currently, the primary approach to implementing SPINN is embedding the signal processing method in the design of models, which deeply integrates the priors of signal processing methods and the data-driven capabilities of deep learning.

1) A simple approach to designing a prior informed model is to directly use signal processing methods as processing or activation layers in the network, with fixed parameters and clear physical meanings. Sadoughi et al. [93] utilized spectral kurtosis and envelope analysis methods to extract sidebands from raw signals and minimize non-transient components. Then a fixed convolution layer designed based on shaft speed and characteristic frequency is utilized to extract fault features from the demodulated signal, which are input into a CNN model for bearing fault diagnosis. Based on the multi-resolution analysis, Jiang et al. [94] proposed a 4D wavelet convolution layer to process infrared thermography, characterizing complex spatiotemporal degradation-related features. Then the features are denoised by a deep image stream denoiser layer and utilized for RUL prediction. Wang et al. [95] embedded discrete wavelet transform (DWT) as a frequency mapping layer in the network. After obtaining wavelet representation, data-driven convolution and frequency attention are utilized for feature extraction, enabling noise-robust fault diagnosis. Liu et al. [96] introduced Harr wavelet transform as a signal preprocessor of the discriminator part in the structure of a generative adversarial network (GAN), enabling higher data generation quality. Ren et al. [97] integrated dual tree complex wavelet into transformer to obtain shift invariance and more discriminative features.
2) Making key parameters of signal processing methods trainable or partially networked is an important approach to implementing end-to-end data-driven SPINN.

Due to the equivalence of convolution and filtering, a series of works have emerged that parameterize the convolutional kernels using signal transformation theory. Ganguly et al. [98] utilized wavelet kernels to initialize the CNN kernels and construct a specific feature learning by neural architecture design, to detect and discriminate signal or multiple partial discharge locations in high voltage power apparatus. Li et al. [99] utilized a continuous wavelet kernel to replace the first convolutional layer of CNN to design meaningful filters in an end-to-end manner and then realized machine fault diagnosis. Chen et al. [100] embedded different time-frequency transforms into convolutional layers as different trainable kernels, realizing a first-layer physically interpretable model for fault diagnosis.

Signal denoising often requires parameters that are adapted to the signals. Data-driven approaches are well-suited to address this issue effectively. Shang et al. [5] proposed the concept of SPINN and designed a wavelet denoising network for fault diagnosis. Zhao et al. [101] analyzed the regularization constraints of denoising problems in the reproducing kernel Hilbert space (RKHS). The parameters controlling the bandwidth and signal smoothness are selected as trainable parameters and designed as an interpretable denoising layer. Zhao et al. [102] implemented denoising based on the soft thresholding method, where the threshold is adaptively determined by a network branch. Shao et al. [103] improved the autoencoder by replacing the activation function with Morlet wavelet function, enabling better feature extraction for fault diagnosis. The wavelet parameters controlling frequency bandwidth and central frequency are optimized by the fruit fly optimization algorithm.

3) Signal processing algorithms and the feature extraction procedure can be reformulated as network structures. Based on the similarity of the convolutional network and wavelet transform, scattering transform is proposed by Mallat [104, 105]. Furthermore, Liu et al. [106] developed a normalized wavelet scattering convolutional network for fault diagnosis. Informed by signal time-scale representation theory, Kim et al. [107] designed a health-adaptive time-scale representation layer and embedded it into CNN for gearbox fault diagnosis. Based on the lifting wavelet transform theory, Pan et al. [108] designed a LiftingNet for adaptative feature learning from noisy data, which is used for bearing fault diagnosis. The predictor and updater are realized by the convolutional layer. Yuan et al. [109] designed smart lifting wavelet kernels with strict theoretical constraints from signal processing theory, enabling impact fault feature extrac-

tion for machine fault diagnosis. To track mechanical degradation, Jiang et al. [110] proposed learnable lifting scheme with spatiotemporal dynamic convolution layer. The infrared thermography is proposed by the lifting scheme and the fused subbands energy features are used for degradation prediction. Informed by DWT, Fink et al. [111] developed a learnable deep discrete wavelet transform for high-frequency time series analysis in machine monitoring. This method integrated the fast discrete wavelet transform into an unsupervised autoencoder framework, which makes both the wavelet bases and denoising thresholds fully learnable. Inspired by the similarity between DWT and autoencoder, Shang et al. [112, 113] proposed learnable M-band wavelet network to obtain discriminative reconstruction error between normal and abnormal signal. Zhao et al. [114] developed a model-driven deep unrolling method for fault diagnosis, which unrolled a corresponding optimization algorithm into a neural network with the characteristics of interpretability and noise robustness. An et al. [115] unrolled a sparse coding model and developed an adversarial algorithm unrolling network for anomaly detection. The encoder and decoder in the generator are designed based on the sparse coding algorithm. Based on morphological analysis, Ye et al. [116] developed a deep morphological convolutional network with learnable structure elements for gearbox fault diagnosis.

Feature extraction priors can also be realized in a network structure. Wang et al. [117] incorporated wavelet transform, square envelope, and Fourier transform into the input layer of extreme learning machine (ELM). Sparsity measures are induced into the hidden nodes of ELM to establish an interpretable neural network for machine condition monitoring. Borghesani et al. [118] embedded the signal processing method of gear diagnostics into the network and achieved adaptive spectrum editing. A fault index related to the average log-ratio is obtained through the network. Based on signal processing priors on the bearing, Lu et al. [119] developed a weighting layer to assign higher weights for features located closer to the bearing fault characteristic frequency. Xie et al. [120] designed frequency learning branches and corresponding loss functions for different fault signals. This allows the network to adaptively extract frequency priors and combine them with deep features for fault diagnosis.

To discover mappings between two infinite-dimensional function spaces, Rani et al. [121] proposed wavelet neural operator, a data-driven framework consisting of uplifting transformation, wavelet integral block, and downlifting transformation. Then generative adversarial wavelet neural operator is constructed to obtain the distribution of multivariate time series data and the reconstruction error is used as a fault indicator.

3.3.3 Signal Processing Informed Optimization Constraint

By leveraging signal processing methods, signals can be described from different perspectives, allowing for additional constraints to be imposed on network training.

With extra frequency information reconstruction loss, Russell et al. [122] enhanced the data compression performance of deep autoencoder for industrial condition monitoring. Yao et al. [123] utilized the statistical property of signal in the wavelet domain to weight different scale coefficients, applying a regularization constraint on the temporal reconstructor. This regularization enables the model to capture both the temporal and frequency patterns of the signal. Dai et al. [124] proposed an acceleration-guided acoustic signal denoising framework based on a learnable wavelet transform for slab track condition monitoring. The acceleration-guided wavelet feature alignment constraint introduces clean signal information and improves the robustness of condition monitoring.

3.3.4 Summary for Signal Processing in KDML

In conclusion, although embedding priors into data can be easily realized, the selection of priors relies on expert experience and cannot be modified in a data-driven manner. Once the selection of the prior is incorrect, it can easily impact the performance of the model. When embedding knowledge into the model design, the most popular way is to parameterize signal processing methods. As for optimization loss design, priors generally come from the distribution characteristics of the signal in a particular representation domain. Moreover, this often overlaps with the previous two methods because introducing priors from a signal processing perspective often requires the object being optimized to have been processed by signal processing methods.

4 Case Study

In this section, we provide four case studies to show the result KDML can bring in PHM, including inductive experience, physical model, and signal processing. In each case study, we describe its knowledge source and representation, knowledge embedding approach, and experimental result.

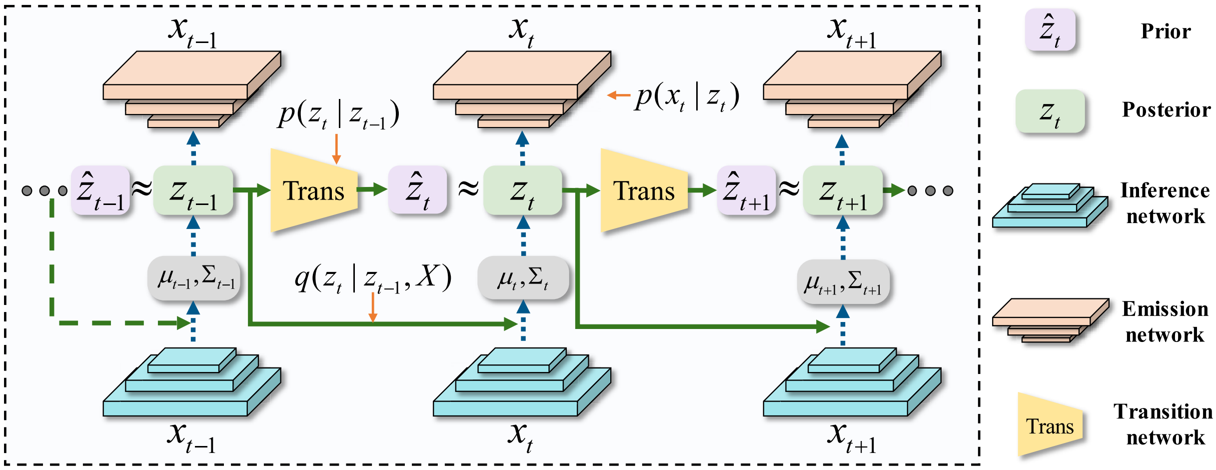

Figure 6 Framework of C^{2}D^{2}M^2 [34]

4.1 Life-Cycle Modeling Driven by Coupling Competition Degradation

4.1.1 Knowledge Source and Representation

In this subsection, an empirical model informed learning method constructed by Li et al. [34] is described in detail, whose name is Coupling Competition Degradation based Deep Markov Model (C^{2}D^{2}M^2). The overall network framework is inspired by the probabilistic graph of SSMs. As illustrated in Figure 6, the representation ability of C^{2}D^{2}M^2 can be improved by parameterizing the emission function and transition function in SSM. Variational inference is employed to estimate true degradation states, which is realized through the inference network in Figure 6. The specific objective loss function can be found in Ref. [34]. The source of knowledge is the prior assumption of SSM on the degradation process, which is the two independent assumptions of HMM, the homogeneous Markov assumption and the observation independence assumption. To be specific, the observed variable at a certain moment depends only on the latent state variable at that moment, while the latent state variable at a certain moment depends only on the latent state variable from the previous moment. This dependence is represented by designing a unique neural network, which can infer the hidden degradation state from noisy observation signals. Moreover, the knowledge of coupling competition degradation mechanism (CCDM) induced from experimental phenomena is embedded into the transition network as prior. The empirical knowledge of CCDM is that the degradation of entities generally does not obey a single mechanism, but a coupling of multiple mechanisms, among which there is a competitive relationship.

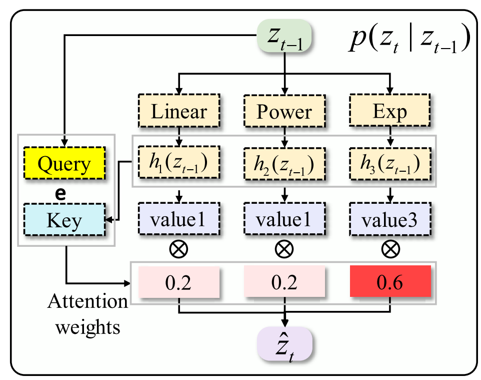

Figure 7 Transition network structure [34]

4.1.2 Knowledge Embedding Approach

The difference equations from three degradation mechanisms are encoded into the transition network, including linear degradation, exponential degradation and power rate degradation. Figure 7 illustrates the specific embedding approach. The differential equations of the latent degradation states at adjacent moments are derived based on these three degradation modes. In the transition network, the latent state z_{t-1} at the previous moment is fed into the differential equation to estimate the latent state at the next moment. The query matrices are obtained by linearly projecting z_{t-1}, and the outputs of the three degradation modes are linearly projected to obtain the key matrices. The dot product operation of query and key matrices is utilized to calculate the similarity and obtain the attention weights. The attention weights are then employed to perform a weighted summation of the outputs from three degradation modes to obtain prior estimation $\hat{z}_t$ at the next moment. The competition relationships can be reflected in the optimization of attention weights.

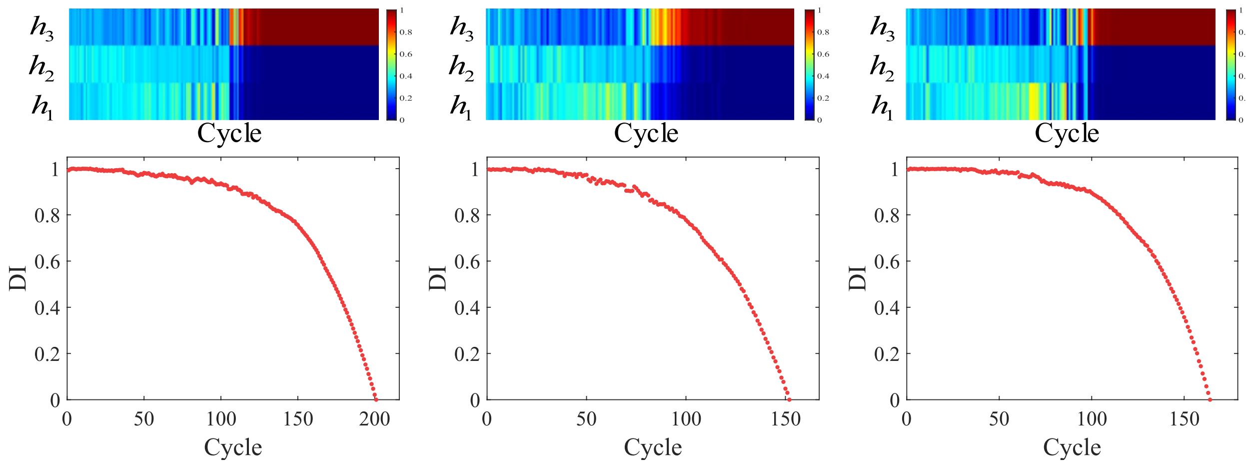

Figure 8 DI curves from three instances and the corresponding attention weights [34]

Table 1 RMSE comparison with existing methods

Algorithms	FD001	FD002	FD003	FD004
AE	14.40	23.16	17.42	24.05
VAE	14.70	23.25	18.12	24.17
DMM	14.00	22.26	16.22	23.16
BiGRU-AS [126]	13.68	20.81	15.53	27.31
SUR-TSMAE [127]	14.46	21.10	17.16	22.61
BiLSTM-ED [128]	14.74	22.07	17.48	23.49
$C^2D^2M^2$	**12.32**	**20.81**	**15.32**	**22.43**

4.1.3 Experimental Results

In the experiment, Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset is employed for validation and the detailed dataset information can be found in Ref. [125]. Figure 8 shows one-dimensional degradation curves generated from several engines, which is inferred from latent features. The attention weights learned by the network are visualized accordingly, where linear, power rate and exponential degradation are denoted as $h1$, $h2$ and $h3$, respectively. For the early stage, $h1$ and $h2$ obtains higher weights, indicating a tendency towards linear and power rate degradation. As for the later stage, the attention weights from exponential degradation are the highest, which reveals that it dominates the competitive process. The network architecture itself is developed based on SSM and is therefore interpretable. Post-hoc attention heat map analysis indicates the evolution of engine operating conditions during the degradation process, which can also provide interpretability for monitoring personnel. To illustrate the superiority of embedded knowledge, the experiment also designed a RUL prediction framework based on similarity matching. Comparison approaches include autoencoders (AE), variational autoencoders (VAE), deep Markov models (DMM), and three existing methods. Table 1 shows the comparison results of the root mean square error (RMSE) from seven methods. It can be observed that the RUL prediction results of $C^2D^2M^2$ on the four sub-datasets are the best (bold values). This shows that embedding the degradation pattern prior can encourage the model to learn degradation indicators that reflect the true latent degradation state.

4.2 Physics Informed Neural Networks for Fault Severity Identification

Physics knowledge is derived from the first principle of target systems. To introduce the detailed incorporating process of physics knowledge and machine learning, a PINN model for axial piston pump fault severity identification by Wang et al. [83] is chosen as an example. The overall framework of fault severity identification is shown in Figure 9.

4.2.1 Physics Knowledge Derivation

This work concentrates on the fluid mass conservation law within a control volume as the first principle, aiming to analyze instantaneous pressure change and figure out the basic governing pressure build-up equation as prior physics knowledge. The objective control volume is chosen as a section of fluid pipeline at pump outlet.

Pressure analysis is based on the Euler method and lamped parameter approach, which declare the assumption of uniformly pressure distribution within the control volume. The physics knowledge in this case is derived as a differential equation of pressure p, time t and the physical parameters to be determined.

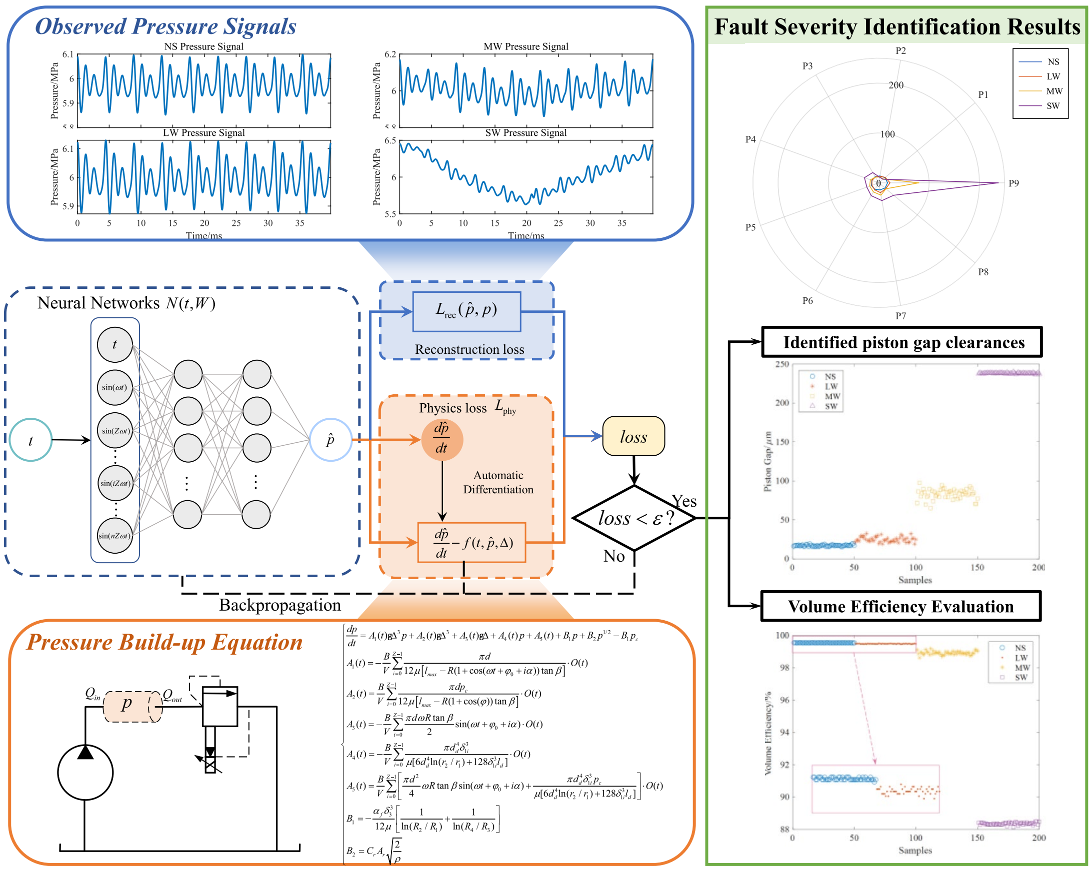

Figure 9 The overall framework of fault severity identification based on PINN [83]

4.2.2 Physics Knowledge Embedding Approach

In this case, physics knowledge is derived to certain loss functions and refines the learning process of neural networks, which is classified into the physics informed loss function case as reviewed previously.

The detailed approach is based on a PINN model which is composed of neural network model for data science and physics equations for theoretical science. The neural network model is involved to estimate the natural functional mapping from time coordinates to pressure amplitude. As for the physics knowledge, it is embedded in the optimization process of neural networks by design corresponding loss functions. In order to obtain the loss function, automatic differentiation method is firstly applied to the neural works, evaluating the differentiations of estimated pressure. Then, the loss function can be parameterized by the network outputs, which bridges the physics equations and neural networks. This function is formulated as the residual of the former mentioned physics equation, which contains fault severity information, describing the gap clearance of piston/cylinder block interface.

4.2.3 Experimental Results

To analysis the results of physics knowledge embedded learning approach, a fault simulation experiment is performed on an axial piston pump test bench. The severity of piston wear is classified into four degrees, including normal state, slight wear, medium wear, and severe wear. In this experiment, the pressure signal is collected at pump outlet by a high frequency pressure sensor with sampling frequency of 10240 Hz.

After optimizing the proposed model, the undetermined physics parameters inside physics informed loss function is identified. There are 50 signal samples fed into PINN model under each fault mode. The fault severity identification results are depicted in the results part of Figure 9, and the identified gap clearances are illustrated by a polar coordinate system, with the angular domain P1–P9 representing the 9 pistons inside the axial piston pump, and the radial axis representing the magnitude of gap clearance. Besides, the maximum identified clearances are scattered in a new figure. The proposed method is capability of distinguish the position of wearing pistons and the fault severities. Moreover, the identified results are substituted to the physics equations and further used to calculate the physical performance degradation indicator, i.e., volume efficiency. The proposed model reveals the relation between the degradation stage of pump performance and piston wears, offering an interpretable approach for the fault severity identification of axial piston pumps.

4.3 Adversarial Algorithm Unrolling Network

4.3.1 Knowledge Source and Representation

Data collected in industrial scenarios often cover massive noise. Intelligent anomaly detection methods can obtain good performance in accuracy, but fail to ensure the credibility of the detection results. Sparse coding is a representation learning algorithm with explicit probabilistic inference formula. The form of its loss function is relevant to the noise distribution, which makes the denoising features well-explainable. Therefore, An et al. [115] construct such a signal processing informed learning framework, where knowledge source is the explainable sparse coding theory, and the knowledge can be represented as the solving algorithm of the multilayer sparse coding model.

4.3.2 Knowledge Embedding Approach

The informed learning framework is built upon adversarial autoencoder (AAE) and algorithm unrolling techniques. It consists of four steps. First, a sparse coding model is built based on data priors for encoding and decoding process. Second, the iterative solving algorithm is derived to determine the basic encoding/decoding units. After that, unrolling technique is applied to basic encoding/decoding units. The encoder, decoder and discriminator of AAE are modified to form the adversarial algorithm unrolling network. In such way, the knowledge contained in interpretable sparse coding approach is embedded into the reconstruction network. Finally, post-hoc interpretability analysis can be realized by visualizing the reconstructed and component features.

Table 2 Detection indicators on the SQI dataset

Algorithms	TPR	FPR	AUC	F1 score	ACC
VAE	71.77	**0.03**	99.71	83.52	92.26
AAE	96.73	1.86	98.75	95.91	97.74
GANomaly	31.62	0.05	97.56	48.00	81.29
AAU-Net	**96.82**	0.25	**99.81**	**98.04**	**98.94**

4.3.3 Experimental Results

To evaluate the detection performance and interpretability of the unrolling network, we carry out a fault experiment on the SQI dynamics simulator [115]. Four health states, i.e., normal, root crack, wear and missing teeth, are preset on the parallel shaft gearbox. The rotating frequency of input shaft is 30 Hz. The sampling frequency is 20480 Hz. Samples are truncated with a window length of 1024 and an overlap ratio of 0.8. Finally, 5780 normal samples are selected for training, and 5780 normal samples and 2173 abnormal samples are saved for testing. The proposed method is compared with other anomaly detection algorithms including VAE [129], AAE [130], and GANomaly [131], and their detection performances are listed in Table 2. The bold values in Table 2 represent the best result in the column.

As can be observed, AAU-Net obtained almost the best performance on five detection indicators, indicating that embedding correct knowledge into learning process can boost the model performance. With a decision threshold of 0.5, the rolling network achieves the accuracy of 98.94 and the F1 score of 98.04, greatly widening the gap between other intelligent detection

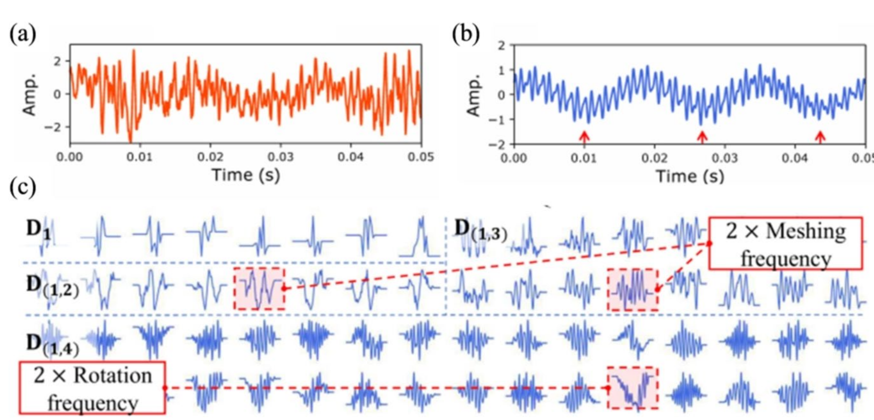

Figure 10 Visualization on the raw signal, reconstructed signals and component features [115]: (**a**) Raw signal, (**b**) The decoded features by adversarial algorithm rolling network, (**c**) Component features of the learned dictionary

methods. More importantly, the visualization on the reconstructed features and component features can indicate related frequency information of the system, as shown in Figure 10. According to the rotating frequency and the structural parameters of the gearbox, the meshing frequency between the planetary gear and the ring gear is computed to be 500 Hz. In Figure 10(a), system-related features are submerged by noises and hard to be detected, while in Figure 10(b), the decoded signal shows clear 2× rotating frequency and 2× meshing frequency, where the former is pointed with red arrows. Besides, the atoms ($\mathbf{D_1}$ to $\mathbf{D_{(1,4)}}$) of the learned dictionaries are listed in Figure 10(c), showing the component features at different scales. Among them, the components of rotating frequency and meshing frequency can be found. It indicates that the proposed model can extract system dynamics-related features from normal signals, helping users to better understand the model and improving the credibility of the results.

4.4 SPINN: Denoising Fault-Aware Wavelet Network

Here, a SPINN model informed by wavelet denoising and kurtosis-based feature selection is provided [5].

4.4.1 Signal Processing Priors

The signal processing priors are derived from a universal fault diagnosis procedure of signal processing: filter-feature-decision. For the specifical representation, firstly, wavelet transform is used to provide nonstationary signal representation from the time-frequency domain. Then, in the feature extraction stage, the feature should be noise-robust and fault-related. Thus, wavelet hard threshold denoising is used, which can effectively remove noise from the signal. Finally, to obtain discriminative features for fault diagnosis, index-based feature selection is important. For wavelet coefficients, the energy of the coefficients can be used as the index. In addition, spectral kurtosis (SK) is also used to search the optimal band-pass filter for fault information extraction.

4.4.2 Signal Processing Informed Model Design

In this section, signal processing priors mentioned above are used as the design principles of the network module. The network design also follows the flow of filter-feature-decision.

In the stage of filter, since wavelet transform is similar to the definition of convolutional layers, embedding wavelet transform priors can be achieved by replacing the convolutional kernels with wavelet basis functions. The scale parameter is set to learnable, and the translation parameter is replaced by the stride parameter. Additionally, multiple wavelet bases are used in the fused wavelet convolutional layer to obtain different fault features. In the feature extraction stage, since the coefficients of the noise are zeroed in the hard threshold denoising function, data-driven denoising is realized by constructing a noise classification network. The coefficients classified as noise are zeroed. Then the energy of wavelet coefficients is used for generating a data-driven new index for feature selection. Furthermore, an SK-based optimization loss is used for better wavelet kernel optimization. Finally, in the decision stage, a decision layer composed of a general convolutional network is incorporated. The

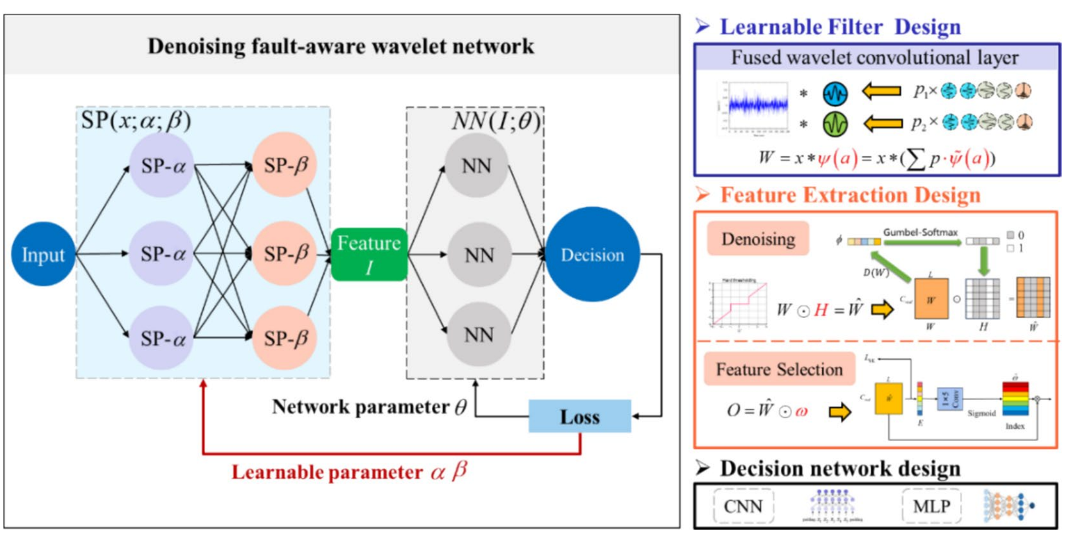

Figure 11 SPINN: denoising fault-aware wavelet network [5]

Table 3 Diagnosis performance of different models

Model	Max-acc (%)	Min-acc (%)	Avg-acc (%)
CNN	76.42	74.67	75.55 ± 0.55
WCNN	89.52	88.77	88.47 ± 0.62
MKCNN	87.77	86.03	86.72 ± 0.49
DSN	83.84	80.79	82.27 ± 0.79
SincNet	90.39	89.52	89.74 ± 0.29
DFAWNet	92.36	93.32	92.76 ± 0.49

overall design of signal processing informed denoising fault-aware wavelet network (DFAWNet) is shown in Figure 11.

4.4.3 Experimental Results

Firstly, different methods are validated on the Machinery Failure Prevention Technology dataset [132]. The results shown in Table 3 indicate that DFAWNet performs best among shallow convolution neural network (CNN) [133], an anti-noise model with a wide kernel size of 64 in the first layer (WCNN) [134], a multiscale kernel model with multi-resolution property (MKCNN) [135], a model with residual shrinkage module (RSNet) [102], an explainable model with a learnable sinc function as the filter (SincNet) [136].

For interpretability, this network is analyzed from a frequency perspective on the Machinery Failure Prevention Technology dataset. As shown in Figure 12, for the normal and fault signals, the frequencies of optimal filters are between 10 kHz and 14 kHz. With data-driven training, the frequency of wavelet kernels shifts from 6.10 kHz to 12.21 kHz, positioning it in the middle of 10 kHz and 14 kHz. This experiment verified the adaptive fault feature extraction ability of DFAWNet.

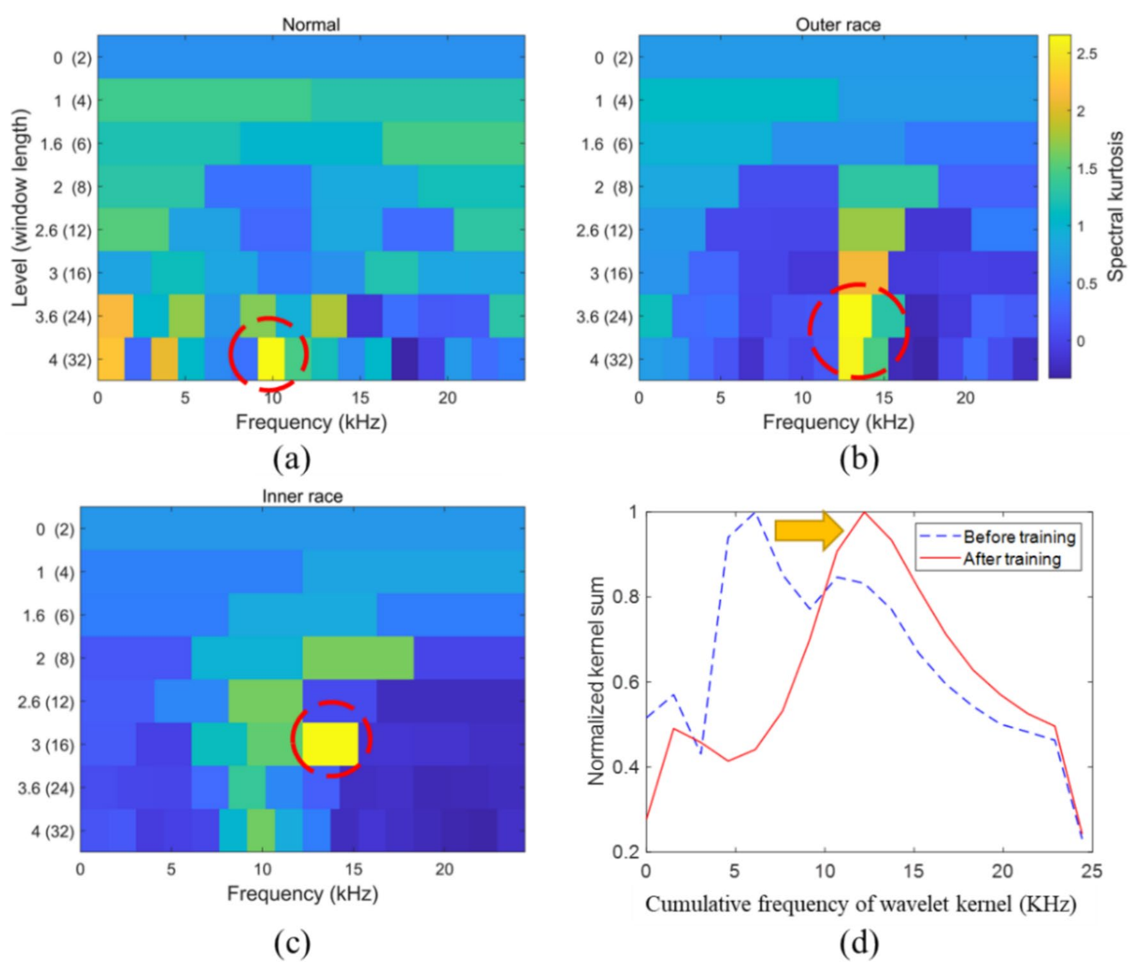

Figure 12 Fault characteristic frequency bands and frequency change of wavelet kernels [5]: (**a**) Normal, (**b**) Inner race fault, (**c**) Outer race fault, (**d**) Frequency change of wavelet kernels after training

4.5 Physics-Informed Machine Learning for Pressure Sensor Placement Optimization

In this subsection, a physics-informed machine learning approach for surrogate modeling of wind pressure and optimization of pressure sensor placement is illustrated [137], which shows potential in structural health monitoring for civil structures.

4.5.1 Knowledge Source and Representation

The knowledge prior of this work is derived from conservation laws and represented by partial differential equations. To optimize the pressure sensor placement in a civil structure subjected to different wind conditions, this work developed a surrogate model to predict the full-field pressure from scattered sensor measurements. To construct the surrogate model, physical knowledge informed by a turbulence model is introduced. Such knowledge prior makes the establishment of surrogate model without the requirement of a large amount of measurement data owing to the highly condensed knowledge embedded in the physical principles.

4.5.2 Knowledge Embedding Approach

To construct the surrogate model to predict the full-field wind pressure, a turbulence model represented by partial differential equations is used to design a loss function to regular the optimization process of neural network. Therefore, this knowledge driven approach can be classified into the physics informed optimization regularization.

The objective of this approach is two-fold, as shown in Figure 13(a). The first is to use a physics-informed neural network to construct the surrogate model for wind pressure prediction. In this step, the physical equations will be utilized to design a loss function to train neural network, so this surrogate model can recover the full-field pressure profile from scattered measurement data with the help of highly condensed physical knowledge. The second is to use the above physics-informed surrogate model as a fast evaluation predictor to learn the best placement for a given number of pressure sensors. In the second step, the problem is stated as an optimization problem to achieve the best predictive accuracy over a wide range of wind conditions.

4.5.3 Experimental Results

To evaluate performance of the proposed method, a finite element method (FEM) is used to generate synthetic data of a classical flat roof. As the optimization problem of sensor placement has to decide both the number and location of sensors, this work first investigates the effort of the number of sensors for one specific wind condition and then optimize their locations for various conditions. The results can be found in Figure 13(b, c). It can be observed that the proposed method can predict a more accurate full-field pressure and get a moderate number of sensors.

4.6 Physics-Informed Machine Learning for Data Augmentation in Battery Prognostics

In this subsection, a physics-informed machine learning model for battery state of health prognostics [138] is introduced. In this case, a physical mode is used to

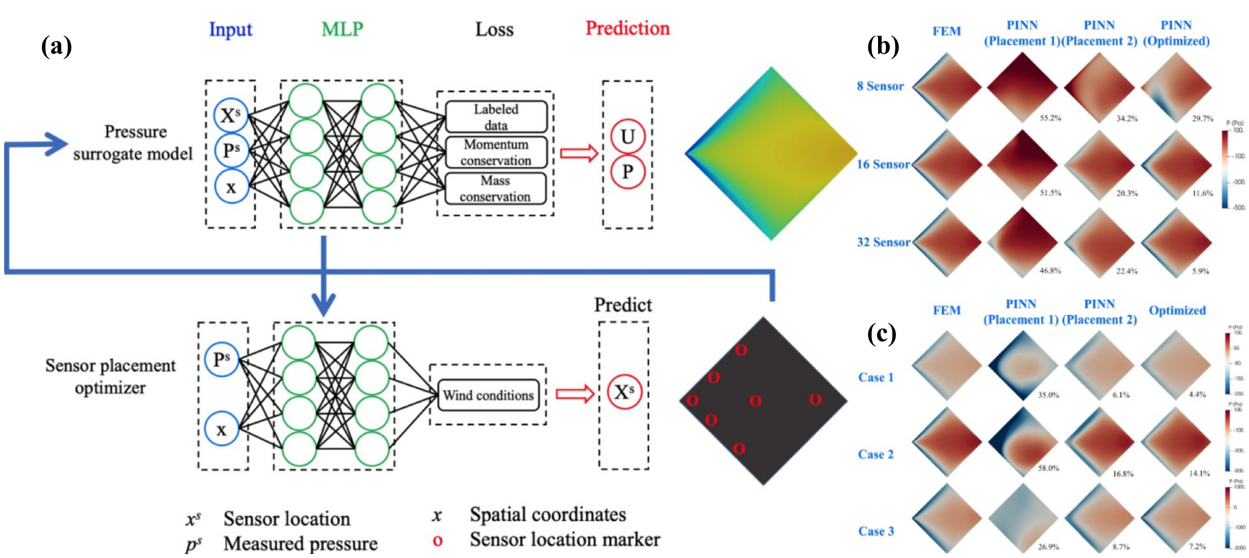

Figure 13 PINN for optimization of pressure sensor placement [137]: (**a**) Overview of PINN model, (**b**) Optimization results of different number of sensors, (**c**) Optimization results under different wind conditions

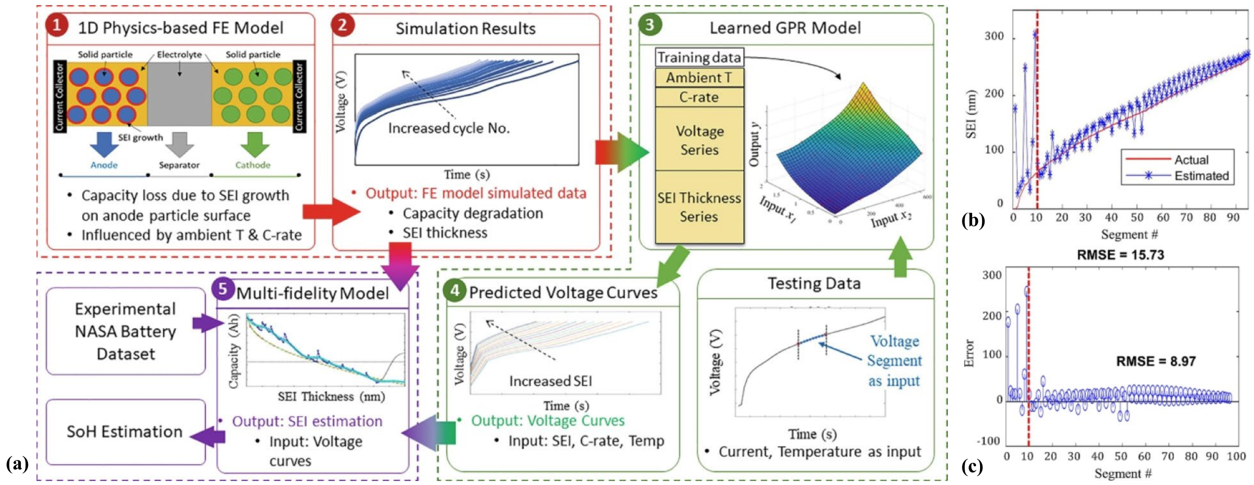

Figure 14 PIML for battery prognostics [138]: (**a**) Framework of PIML method for battery prognostics, (**b**) The estimated result of the health state of battery, (**c**) The estimated error with respect to segments

generate synthetic data and then experimental data will be fused to develop a multi-fidelity framework for battery prognostics.

4.6.1 Knowledge Source and Representation

Knowledge prior of this work depends on the physical model of the influences of dominating aging mode. More specifically, this knowledge is represented by a physics-based finite element model to describe the relation between solid electrolyte interface growth with anode particle surfaces and capacity loss of battery.

4.6.2 Knowledge Embedding Approach

To establish a multi-fidelity framework, the physics-based finite element model is used to generate simulated data. Then the simulated data is combined with experimental data to train a gaussian process regression model. In this case, physical knowledge is embedded into the data acquisition process in ML pipeline. The entire framework is shown in Figure 14(a).

4.6.3 Experimental Results

Numerical results are presented to evaluate the performance of the proposed method, as shown in Figure 14(b, c). It can be observed that the proposed model can perform well at later cycles of battery health, while remains a high error in the early cycles. Such physics-informed data augmentation method allows to train a predictive model with few-shot realistic samples.

5 Challenges, Potential Applications and Recommendations

The insight of knowledge driven machine learning is not a recent objective in PHM. A historical effort is to employ various L-norm regularization methods to constrain model to have some specific properties, like robustness to measurement noise. An additional and parallel way is hybrid methods in PHM, such as integrating physics model with data-driven model or directly combining multiple data-driven models. These approaches have gained a lot of attention to improve prediction accuracy and developed for a long time. On the other hand, with the quick development of Deep Learning since 2012, many researches shifted to novel neural network structures or learning algorithms. Less attention was focused on knowledge driven, especially for physics model. The recent PINN work in 2017 [139] revived awareness of additional benefits that such knowledge driven paradigm can bring, such as generalization ability and interpretability, and thus exploded hot spots of this topic in PHM. Moreover, the desire for trustworthy AI in PHM domain also requires such evolution. However, there still exists challenges in knowledge driven machine learning.

5.1 Challenges

5.1.1 The Tradeoff between Fidelity and Complexity in Physical Models

In modeling the same physical system, there exist multiple methods that can be used to construct models with different fidelities. As the fidelity of the model increases, the accuracy of the obtained solutions to the physical model increases accordingly, thus enabling better embedding of knowledge into the network. However, high fidelity models are often accompanied by increased solution complexity, which is not advantageous for network design and training for learning. On the other hand, physical systems in the real world are becoming more and more complex, causing difficulty to build high-fidelity physical models, and oversimplification of the physical systems can lead to model distortion. Therefore, how to balance the fidelity and complexity of the physical model is very important for the design of physical informed neural networks.

5.1.2 Quantification of Interpretability

Deep learning models have extremely complex structures and parameters. In order to explain the internal mechanisms of these models, metrics need to be developed to assess the amount and quality of interpretable information contained in the network structure and output. This further brings the issue of quantifying interpretability. The quantitative data provided by such metrics can also further help us to understand and optimize the models, to achieve trade-offs between accuracy and transparency, and to provide reliable explanations of the predicted results. It is through quantifying interpretability that we can strike a balance between conflicting metrics and further advance the application of deep learning in the field of intelligent diagnosis to achieve more reliable and understandable diagnostic results.

5.1.3 Iterative "Human in the Loop"

Interpretable neural network can be regarded as an offline human-in-the-loop strategy by embedding the existing expert experience and knowledge into the network, and the output of the network is fed back to the expert for final interpretation and decision making. However, how to design iterative "human-in-the-loop" intelligent systems so that they can most effectively collaborate with human experts and ensure that decisions and feedback from both sides can be accurately aligned is a major challenge. This further requires the development of intuitive and easy-to-use user interfaces so that human experts can easily interact with intelligent networks. In addition, consideration needs to be given in protecting sensitive data generated during human-computer interactions to prevent data leakage or misuse.

5.1.4 Modularization and Efficient Collaboration of Physical Knowledge

In the field of interpretable intelligent diagnosis, the applied expert experience and knowledge is complex and diverse. Effectively integrating this information into a learning system is a challenging task. One possible idea is to modularize the physical knowledge separately. The modularization approach allows different domains and levels of physics knowledge to be decomposed into smaller modules, each focusing on a specific task. This provides flexibility and scalability in embedding physical knowledge and enables knowledge sharing and reuse. Suitable knowledge modules can be appropriately selected for different tasks and problems. The reorganization and synergy of module knowledge increases accuracy while further enhancing the interpretability of the learning system and understanding the decision-making process of the system as each module is inherently understandable.

5.2 Potential Applications

5.2.1 Interpretable AI towards Trustworthy Decision

The ultimate goal of PHM is to ensure human security and reduce maintenance costs, so this domain is risk sensitive to wrong decision. This is why advanced data-driven models have been slow to spread in this conservative domain. Despite pure data-driven models can provide state-of-the-art performance, the loss caused by making wrong prediction even once is unacceptable in PHM domain. Therefore, how to enhance transparency of data-driven model is a capital step to gain the trust of decision makers. Knowledge driven machine learning provides an active strategy to intervene ML pipeline, and thus convert its prediction logic to an interpretable level that humans can understand.

5.2.2 Active Optimization for Machine Learning Pipeline

Through embedding knowledge into ML pipeline, KDML is expected to learn intrinsic pattern from data which can be generalized to different data domains. For example, the construction of physical models is based on the first principle, like energy conservation law for thermodynamics systems or momentum conservation law for mechanical systems. In this case, physical models can describe the system dynamics under different working conditions or initial conditions. Therefore, knowledge is expected to be broadly adaptable to these varying conditions. Moreover, this active embedding is a potential way to troubleshoot invalidation in ML pipeline and then to optimize it. With respect to the three main parts in ML pipeline, data, model, and optimization, knowledge can guide their construction before forming ML pipeline. For example, an important but seldom noticed issue is data acquisition. Structure knowledge of the physics system can help to optimize sensor layout to monitor principal components. Knowledge of specific fault types can help to optimize parameter setting of sensors, like suitable sampling frequency to capture fault features under limited data storage. Therefore, knowledge still has additional benefits to optimize ML pipeline, except for generalization ability improvement.

5.2.3 Knowledge Discovery to Feedback Smart Manufacturing

Machine learning methods, especially neural networks, are transforming from simple data processing tools to complete knowledge discovery frameworks. Outcome of the ML pipeline is an operational model, which can learn pattern and relation from finite data samples. Based on the learned pattern or relation, further investigation can be performed to find causal mechanism between physics variable, then to construct interpretable theory or hypothesis. In PHM domain, knowledge can be discovered from operation and maintenance data to feedback product design or manufacturing. For example, ML methods can build a causal graph from monitoring data to represent relation within system variables, and then vulnerable spot of a complex system can be located through inference of causal graph. This potential application also points out the importance of interpretability, that the learned pattern or relation should be totally comprehensible in theory.

5.3 Usage Recommendations for Different PHM Problems

In this subsection, we will discuss how to select the appropriate knowledge driven methods for specific industrial scenarios for PHM domain, including anomaly detection, fault diagnosis, and fault prognosis.

5.3.1 Anomaly Detection

Anomaly detection is understood to identify abnormal data whose dynamic behavior deviates significantly from healthy states. Through the literature analysis of current research as shown in Figure 15, it can be observed that there are mainly two roadmaps to implement KDML for anomaly detection, i.e., integrating statistic property for optimization regularization and integrating signal equation for model design, which can be a recommendation for potential usage.

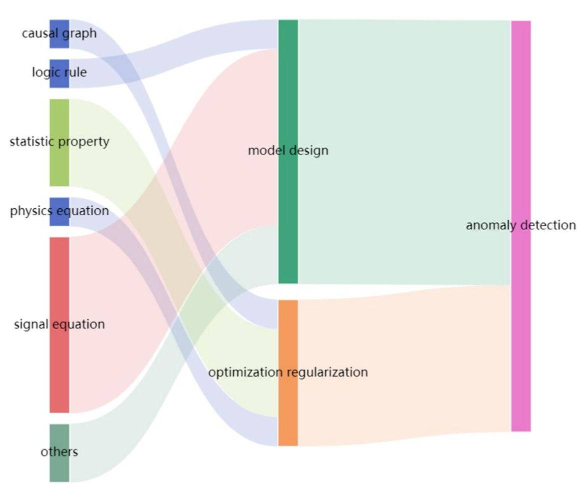

Figure 15 Sankey diagram for KDML in anomaly detection

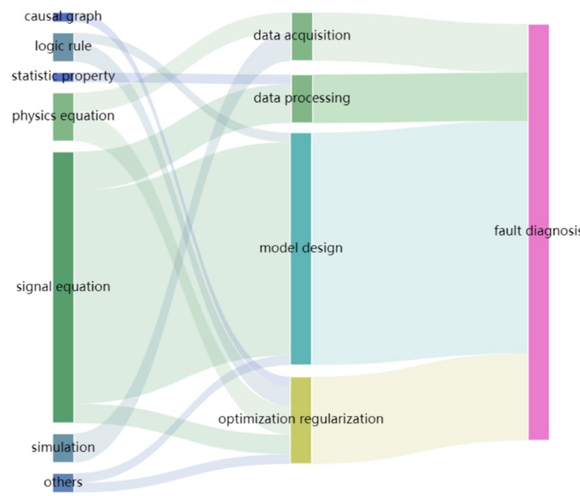

Figure 16 Sankey diagram for KDML in fault diagnosis

5.3.2 Fault Diagnosis

Fault diagnosis is used to identify fault types or fault locations with diverse dynamics behaviors. As shown in Figure 16, we can find that signal equation is the majority among knowledge representations and has been embedded in several parts of ML pipeline, including data processing, model design, and optimization regularization. As signal processing has been a well-researched technique for fault diagnosis over the past decades and has different representations for different physical systems,

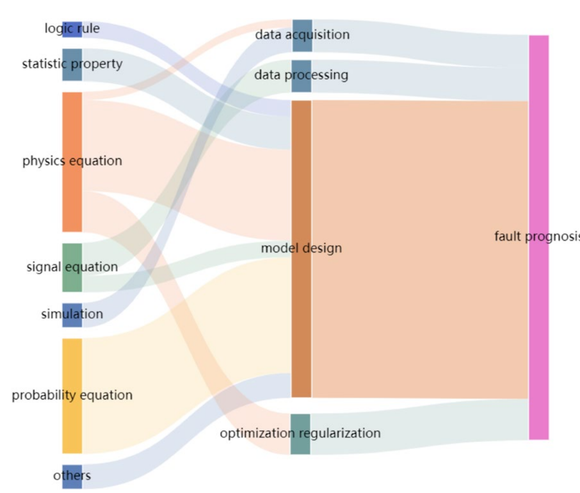

Figure 17 Sankey diagram for KDML in fault prognosis

a recommended roadmap is to utilize signal equation to design a specific ML model for fault diagnosis.

5.3.3 Fault Prognosis

Fault prognosis focuses on predicting the time at which a physical system will not perform its function. As shown in Figure 17, we can observe that physics equation and probability equation are the two main kinds of knowledge representation methods, such as cumulative fatigue models or stochastic degradation processes. For the knowledge embedding approaches, model design is the main part to be modified for KDML. It is recommended to use physical knowledge for KDML to provide interpretable solutions for fault prognosis.

6 Conclusions

In this paper, we proposed a universal concept, knowledge driven machine learning, for integrating diverse knowledge into machine learning pipeline in PHM domain. Our main contribution is to classify KDML in a hierarchical framework, ranging from knowledge sources, knowledge representation, to knowledge embedding in ML pipeline. In addition, we provided several case studies to illustrate usage of different KDML methods in PHM domain. The proposed hierarchical framework of KDML and extensive case studies can help PHM users to find suitable way for their applications.

In addition, popularity of knowledge driven machine learning indicates a new research paradigm in science and engineering. In the era of big data, development of data science is unstoppable. Some even refer to the rise of data science as "the end of theory" [140], as data driven model can easily generate an operable model from big data. But knowledge driven machine learning provides a new perspective to evaluate the effect brought by data science, that knowledge, even theory, can be extracted from data. It means KDML starts a new direction to find theory rather than its end.

Acknowledgements
Not applicable.

Authors' Contributions
RY: Review-editing & supervision; ZZ: Writing-review & editing; ZS: Writing-review & editing; ZW: Writing-review & editing; CH: Writing-review & editing; YL: Writing-review & editing; YY: Writing-review & editing; XC: Review-editing & supervision; RG: Review-editing & supervision. All authors read and approved the final manuscript.

Funding
Supported in part by Science Center for Gas Turbine Project (Project No. P2022-DC-I-003-001) and National Natural Science Foundation of China (Grant No. 52275130).

Availability of Data and Materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing Interests
The authors declare no competing financial interests.

Received: 7 June 2024 Revised: 10 October 2024 Accepted: 23 December 2024
Published online: 17 January 2025

References

[1] A Dourado, F A C Viana. Physics-informed neural networks for missing physics estimation in cumulative damage models: A case study in corrosion fatigue. *J Comput Inf Sci Eng*, 2020, 20:061007.
[2] J Luo, M Namburu, K Pattipati, et al. Model-based prognostic techniques. *Proceedings AUTOTESTCON 2003, IEEE Systems Readiness Technology Conference*, IEEE, 2003:330-340.
[3] B Li. A review of tool wear estimation using theoretical analysis and numerical simulation technologies. *International Journal of Refractory Metals and Hard Materials*, 2012, 35:143-151.
[4] M A Chao. *Combining deep learning and physics-based performance models for diagnostics and prognostics*. ETH Zurich, 2021.
[5] Z Shang, Z Zhao, R Yan. Denoising fault-aware wavelet network: A signal processing informed neural network for fault diagnosis. *Chin J Mech Eng*, 2023, 36:9.
[6] Z Wei, Z Zhao, Z Zhou, et al. Collaborative-sequential optimization for aero-engine maintenance based on multi-agent reinforcement learning. *Expert Syst Appl*, 2024, 247:123358.
[7] G Ma, S Xu, C Cheng. Fault detection of lithium-ion battery packs with a graph-based method. *J Energy Storage*, 2021, 43:103209.
[8] J Zhang, C Liu, R X Gao. Physics-guided Gaussian process for HVAC system performance prognosis. *Mech Syst Signal Process*, 2022, 179:109336.
[9] Z Zhang, Y Liu. Parsimony-enhanced sparse Bayesian learning for robust discovery of partial differential equations. *Mech Syst Signal Process*, 2022, 171:108833.
[10] L Cui, X Wang, H Wang, et al. Remaining useful life prediction of rolling element bearings based on simulated performance degradation dictionary. *Mech Mach Theory*, 2020, 153:103967.
[11] G E Karniadakis, I G Kevrekidis, L Lu, et al. Physics-informed machine learning. *Nat Rev Phys*, 2021, 3: 422-440.
[12] A Karpatne, G Atluri, J H Faghmous, et al. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans Knowl Data Eng*, 2017, 29: 2318-2331.
[13] N Sharma, Y A Liu. A hybrid science-guided machine learning approach for modeling chemical processes: A review. *AIChE J*, 2022, 68:e17609.
[14] W Deng, K T P Nguyen, K Medjaher, et al. Physics-informed machine learning in prognostics and health management: State of the art and challenges. *Appl Math Model*, 2023, 124: 325-352.
[15] Y Xu, S Kohtz, J Boakye J, et al. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. *Reliab Eng Syst Saf*, 2023, 230:108900.
[16] S P Zhu, L Wang, C Luo, et al. Physics-informed machine learning and its structural integrity applications: state of the art. *Philos Trans R Soc Math Phys Eng Sci*, 2023, 381:20220406.
[17] L Rueden, S Mayer, K Beckh, et al. Informed machine learning – A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Trans Knowl Data Eng*, 2023, 35: 614-633.
[18] D Li, K Deng, M Zhao, et al. Knowledge-driven machine learning: Concept, model and case study on channel estimation. arXiv preprint arXiv:2012.11178, 2020.
[19] S Vollert, M Atzmueller, A Theissler. Interpretable machine learning: A brief survey from the predictive maintenance perspective. *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation* (ETFA), IEEE, 2021: 01-08.
[20] G Chen, J Yuan, Y Zhang, et al. Enhancing reliability through interpretability: A comprehensive survey of interpretable intelligent fault diagnosis in rotating machinery. *IEEE Access*, 2024.

[21] B Saha, K Goebel. Modeling li-ion battery capacity depletion in a particle filtering framework. *Proceedings of the Annual Conference of the PHM Society*, 2009.
[22] S Vollert, M Atzmueller, A Theissler. Interpretable machine learning: A brief survey from the predictive maintenance perspective. *2021 26th IEEE Int. Conf. Emerg. Technol. Fact. Autom*. ETFA, Vasteras, Sweden: IEEE, 2021: 1-8.
[23] Y Qian, R Yan, R X Gao. A multi-time scale approach to remaining useful life prediction in rolling bearing. *Mech Syst Signal Process*, 2017, 83: 549-567.
[24] C Ordóñez, F S Lasheras, J Roca-Pardiñas, et al. A hybrid ARIMA–SVM model for the study of the remaining useful life of aircraft engines. *J Comput Appl Math*, 2019, 346.
[25] P Kundu, S Chopra, B K Lad. Multiple failure behaviors identification and remaining useful life prediction of ball bearings. *J Intell Manuf*, 2019, 30: 1795-1807.
[26] P Kundu, A K Darpe, M S Kulkarni. Weibull accelerated failure time regression model for remaining useful life prediction of bearing working under multiple operating conditions. *Mech Syst Signal Process*, 2019, 134:106302.
[27] S M M Hassani N, J Jin, J Ni. Physics-based Gaussian process for the health monitoring for a rolling bearing. *Acta Astronaut*, 2019, 154: 133-139.
[28] S Zhang, Q Zhai, Y Li. Degradation modeling and RUL prediction with Wiener process considering measurable and unobservable external impacts. *Reliab Eng Syst Saf*, 2023, 231:109021.
[29] W Li, T Liu. Time varying and condition adaptive hidden Markov model for tool wear state estimation and remaining useful life prediction in micro-milling. *Mech Syst Signal Process*, 2019, 131: 689-702.
[30] G Lyu, H Zhang, Y Zhang, et al. An interpretable remaining useful life prediction scheme of lithium-ion battery considering capacity regeneration. *Microelectron Reliab*, 2022, 138:114625.
[31] L Cui, X Wang, H Wang, et al. Research on remaining useful life prediction of rolling element bearings based on time-varying Kalman filter. *IEEE Trans Instrum Meas*, 2020, 69: 2858-2867.
[32] X Zheng, H Wu, Y Chen. Remaining useful life prediction of lithium-ion battery using a hybrid model-based filtering and data-driven approach. *2017 11th Asian Control Conf.* ASCC, Gold Coast, QLD: IEEE, 2017: 2698-2703.
[33] H Pei, X S Si, C H Hu, et al. An adaptive prognostics method for fusing CDBN and diffusion process: Application to bearing data. *Neurocomputing*, 2021, 421: 303-315.
[34] Y Li, Z Zhou, C Sun, et al. Life-cycle modeling driven by coupling competition degradation for remaining useful life prediction. *Reliab Eng Syst Saf*, 2023, 238:109480.
[35] Y Deng, A D Bucchianico, M Pechenizkiy. Controlling the accuracy and uncertainty trade-off in RUL prediction with a surrogate Wiener propagation model. *Reliab Eng Syst Saf*, 2020, 196:106727.
[36] C H Hu, H Pei, X S Si, et al. A prognostic model based on DBN and diffusion process for degrading bearing. *IEEE Trans Ind Electron*, 2020, 67: 8767-8777.
[37] H Sun, D Cao, Z Zhao, et al. A hybrid approach to cutting tool remaining useful life prediction based on the Wiener process. *IEEE Trans Reliab*, 2018, 67: 1294-1303.
[38] L Dai, J Guo, J L Wan, et al. A reliability evaluation model of rolling bearings based on WKN-BiGRU and Wiener process. *Reliab Eng Syst Saf*, 2022, 225:108646.
[39] Z Chen, T Xia, Y Li, et al. A hybrid prognostic method based on gated recurrent unit network and an adaptive Wiener process model considering measurement errors. *Mech Syst Signal Process*, 2021, 158:107785.
[40] Z Zhou, T Li, Z Zhao, et al. Time-varying trajectory modeling via dynamic governing network for remaining useful life prediction. *Mech Syst Signal Process*, 2023, 182:109610.
[41] T Yan, D Wang, T Xia, et al. Online piecewise convex-optimization interpretable weight learning for machine life cycle performance assessment. *IEEE Trans Neural Netw Learn Syst*, 2024: 1-13.
[42] T Yan, D Wang, M Zheng, et al. Interpretable sparse learned weights and their entropy based quantification for online machine health monitoring. *Mech Syst Signal Process*, 2023, 199:110493.
[43] T Zhou, E L Droguett, A Mosleh. Physics-informed deep learning: A promising technique for system reliability assessment. *Appl Soft Comput*, 2022, 126:109217.
[44] Y Li, Z Zhou, C Sun, et al. Variational attention-based interpretable transformer network for rotary machine fault diagnosis. *IEEE Trans Neural Netw Learn Syst*, 2022: 1-14.
[45] X Bi, J Zhao. A novel orthogonal self-attentive variational autoencoder method for interpretable chemical process fault detection and identification. *Process Saf Environ Prot*, 2021, 156: 581-597.
[46] X Liao, S Chen, P Wen, et al. Remaining useful life with self-attention assisted physics-informed neural network. *Adv Eng Inform*, 2023, 58:102195.
[47] S Cofre-Martel, E Lopez Droguett, M Modarres. Remaining useful life estimation through deep learning partial differential equation models: A framework for degradation dynamics interpretation using latent variables. *Shock Vib*, 2021, 2021: 1-15.
[48] J Li, Y Wang, Y Zi, et al. Causal disentanglement: A generalized bearing fault diagnostic framework in continuous degradation mode. *IEEE Trans Neural Netw Learn Syst*, 2023, 34: 6250-6262.
[49] C Hu, J Wu, C Sun, et al. Mutual information-based feature disentangled network for anomaly detection under variable working conditions. *Mech Syst Signal Process*, 2023, 204:110804.
[50] S Y Wong, K S Yap, H J Yap, et al. On equivalence of FIS and ELM for interpretable rule-based knowledge representation. *IEEE Trans Neural Netw Learn Syst,* 2015, 26: 1417-1430.
[51] J Yu, G Liu. Knowledge extraction and insertion to deep belief network for gearbox fault diagnosis. *Knowl-Based Syst*, 2020, 197:105883.
[52] Z Wu, H Luo, Y Yang, et al. K-PdM: KPI-oriented machinery deterioration estimation framework for predictive maintenance using cluster-based hidden Markov model. *IEEE Access*, 2018, 6: 41676-87.
[53] B Steenwinckel, D D Paepe, S Vanden Hautte, et al. FLAGS: A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing expert knowledge with machine learning. *Future Gener Comput Syst*, 2021, 116: 30-48.
[54] Z Zhou, Z Ming, J Wang, et al. A novel belief rule-based fault diagnosis method with interpretability. *Comput Model Eng Sci*, 2023, 136: 1165-1185.
[55] Z Ming, Z Zhou, Y Cao, et al. A new interpretable fault diagnosis method based on belief rule base and probability table. *Chin J Aeronaut*, 2023, 36: 184-201.
[56] J Luo, M Namburu, K Pattipati, et al. Model-based prognostic techniques [maintenance applications]. *Proc. AUTOTESTCON 2003 IEEE Syst. Readiness Technol. Conf.*, 2003: 330-340.
[57] A Mojallal, S Lotfifard. Multi-physics graphical model-based fault detection and isolation in wind turbines. *IEEE Trans Smart Grid*, 2018, 9: 5599-5612.
[58] K Yu, Q Fu, H Ma, et al. Simulation data driven weakly supervised adversarial domain adaptation approach for intelligent cross-machine fault diagnosis. *Struct Health Monit*, 2021, 20: 2182-2198.
[59] Q Ni, J C Ji, B Halkon, et al. Physics-informed residual network (PIResNet) for rolling element bearing fault diagnostics. *Mech Syst Signal Process*, 2023, 200:110544.
[60] S Kohtz, Y Xu, Z Zheng, et al. Physics-informed machine learning model for battery state of health prognostics using partial charging segments. *Mech Syst Signal Process*, 2022, 172:109002.
[61] R He, Y Dai, J Lu, et al. Developing ladder network for intelligent evaluation system: Case of remaining useful life prediction for centrifugal pumps. *Reliab Eng Syst Saf*, 2018, 180: 385-393.
[62] D Agarwal, M Kumari, B Srinivasan, et al. Fault diagnosis and degradation analysis of PMDC motors using FEA based models. *2020 IEEE Int. Conf. Power Electron. Smart Grid Renew. Energy PESGRE2020*, Cochin, India: IEEE, 2020: 1-6.
[63] F Zhao, Z Tian, Y Zeng. Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method. *IEEE Trans Reliab*, 2013, 62: 146-159.
[64] K López De Calle-Etxabe, C Ruiz-Cárcel, S Starr, et al. Hybrid modelling for linear actuator diagnosis in absence of faulty data records. *Comput Ind*, 2020, 123:103339.
[65] M A Djeziri, S Benmoussa, R Sanchez. Hybrid method for remaining useful life prediction in wind turbine systems. *Renew Energy*, 2018, 116: 173-187.

[66] R T Q Chen, Y Rubanova, J Bettencourt, et al. Neural ordinary differential equations. *Adv. Neural Inf. Process. Syst.*, 2018, 31.
[67] R G Nascimento, K Fricke, F A C Viana. A tutorial on solving ordinary differential equations using Python and hybrid physics-informed neural network. *Eng Appl Artif Intell*, 2020, 96:103996.
[68] R G Nascimento, F A C Viana. Cumulative damage modeling with recurrent neural networks. *AIAA J*, 2020, 58: 5459-5471.
[69] Y A Yucesan, F A C Viana. Hybrid physics-informed neural networks for main bearing fatigue prognosis with visual grease inspection. *Comput Ind*, 2021, 125:103386.
[70] Y A Yucesan, F A C Viana. A hybrid physics-informed neural network for main bearing fatigue prognosis under grease quality variation. *Mech Syst Signal Process*, 2022, 171:108875.
[71] Y A Yucesan, F A C Viana. A physics-informed neural network for wind turbine main bearing fatigue. *Int J Progn Health Manag*, 2023, 11.
[72] R Tipireddy, A Tartakovsky. Physics-informed machine learning method for forecasting and uncertainty quantification of partially observed and unobserved states in power grids. arXiv preprint arXiv:1806.10990, 2018.
[73] S Saemundsson, A Terenin, K Hofmann, et al. Variational integrator networks for physically structured embeddings. *Proc. Twenty Third Int. Conf. Artif. Intell. Stat.*, PMLR, 2020: 3078-3087.
[74] M A Chao, C Kulkarni, K Goebel, et al. Fusing physics-based and deep learning models for prognostics. *Reliab Eng Syst Saf*, 2022, 217:107961.
[75] R G Nascimento, M Corbetta, C S Kulkarni, et al. Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis. *J Power Sources*, 2021, 513:230526.
[76] M Lutter, C Ritter, J Peters. Deep Lagrangian networks: Using physics as model prior for deep learning. arXiv preprint arXiv:1907.04490, 2019.
[77] R Zhang, Y Liu, H Sun. Physics-informed multi-LSTM networks for metamodeling of nonlinear structures. *Comput Methods Appl Mech Eng*, 2020, 369:113226.
[78] R Zhang, Y Liu, H Sun. Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling. *Eng Struct*, 2020, 215:110704.
[79] X Chen, M Ma, Z Zhao, et al. Physics-informed deep neural network for bearing prognosis with multisensory signals. *J Dyn Monit Diagn*, 2022: 200-207.
[80] B Freeman, Y Tang, Y Huang, et al. Physics-informed turbulence intensity infusion: A new hybrid approach for marine current turbine rotor blade fault detection. *Ocean Eng*, 2022, 254:111299.
[81] S Shen, H Lu, M Sadoughi, et al. A physics-informed deep learning approach for bearing fault detection. *Eng Appl Artif Intell*, 2021, 103:104295.
[82] W Xu, Z Zhou, T Li, et al. Physics-constraint variational neural network for wear state assessment of external gear pump. *IEEE Trans Neural Netw Learn Syst*, 2022: 1–11.
[83] Z Wang, Z Zhou, W Xu, et al. Physics informed neural networks for fault severity identification of axial piston pumps. *J Manuf Syst*, 2023, 71: 421-437.
[84] L E Bouny, M Khalil, A Adib. An end-to-end multi-level wavelet convolutional neural networks for heart diseases diagnosis. *Neurocomputing*, 2020, 417: 187-201.
[85] L Wen, L Gao, X Li, et al. A jointed signal analysis and convolutional neural network method for fault diagnosis. *Procedia CIRP*, 2018, 72: 1084-1087.
[86] X Li, W Zhang, Q Ding. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliab Eng Syst Saf*, 2019, 182: 208-218.
[87] S Fu, L Lin, Y Wang, et al. High imbalance fault diagnosis of aviation hydraulic pump based on data augmentation via local wavelet similarity fusion. *Mech Syst Signal Process*, 2024, 209:111115.
[88] D K B Kulevome, H Wang, B M Cobbinah, et al. Effective time-series data augmentation with analytic wavelets for bearing fault diagnosis. *Expert Syst Appl*, 2024, 249:123536.
[89] Y Meng, C Shao. Physics-informed ensemble learning for online joint strength prediction in ultrasonic metal welding. *Mech Syst Signal Process*, 2022, 181:109473.
[90] R Zhao, D Wang, R Yan, et al. Machine health monitoring using local feature-based gated recurrent unit networks. *IEEE Trans Ind Electron*, 2018, 65: 1539-1548.
[91] J Zhu, N Chen, W Peng. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Trans Ind Electron*, 2019, 66: 3208-3216.
[92] L Ren, J Cui, Y Sun, et al. Multi-bearing remaining useful life collaborative prediction: A deep learning approach. *J Manuf Syst*, 2017, 43: 248-256.
[93] M Sadoughi, C Hu. Physics-based convolutional neural network for fault diagnosis of rolling element bearings. *IEEE Sens J*, 2019, 19: 4181-4192.
[94] Y Jiang, T Xia, D Wang, et al. Spatiotemporal denoising wavelet network for infrared thermography-based machine prognostics integrating ensemble uncertainty. *Mech Syst Signal Process*, 2022, 173:109014.
[95] H Wang, Z Liu, D Peng, et al. Interpretable convolutional neural network with multilayer wavelet for noise-robust machinery fault diagnosis. *Mech Syst Signal Process*, 2023, 195:110314.
[96] Y Liu, H Jiang, C Liu, et al. Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis. *Knowl-Based Syst*, 2022, 252:109439.
[97] J Ren, C Hu, Z Shang, et al. WavFormer: An interpretable wavelet-constrained transformer for industrial acoustics diagnosis. *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, IEEE, 2024: 1-6.
[98] B Ganguly, S Chaudhury, S Biswas, et al. Wavelet Kernel based convolutional neural network for localization of partial discharge sources within a power apparatus. *IEEE Trans Ind Inform*, 2020: 1–1.
[99] T Li, Z Zhao, C Sun, et al. WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis. *IEEE Trans Syst Man Cybern Syst*, 2021, 52: 2302-2312.
[100] Q Chen, X Dong, G Tu, et al. TFN: An interpretable neural network with time-frequency transform embedded for intelligent fault diagnosis. *Mech Syst Signal Process*, 2024, 207:110952.
[101] B Zhao, C Cheng, G Tu, et al. An interpretable denoising layer for neural networks based on reproducing Kernel Hilbert space and its application in machine fault diagnosis. *Chin J Mech Eng*, 2021, 34:44.
[102] M Zhao, S Zhong, X Fu, et al. Deep residual shrinkage networks for fault diagnosis. *IEEE Trans Ind Inform*, 2020, 16: 4681-4690.
[103] H Shao, M Xia, J Wan, et al. Modified stacked autoencoder using adaptive Morlet wavelet for intelligent fault diagnosis of rotating machinery. *IEEE-ASME Trans Mechatron*, 2022, 27: 24-33.
[104] J Bruna, S Mallat. Invariant scattering convolution networks. *IEEE Trans Pattern Anal Mach Intell*, 2013, 35: 1872-1886.
[105] J Andén, S Mallat. Deep scattering spectrum. *IEEE Trans Signal Process*, 2014, 62: 4114-4128.
[106] C Liu, X Ma, T Han, et al. NTScatNet: An interpretable convolutional neural network for domain generalization diagnosis across different transmission paths. *Measurement*, 2022:112041.
[107] Y Kim, K Na, B D Youn. A health-adaptive time-scale representation (HTSR) embedded convolutional neural network for gearbox fault diagnostics. *Mech Syst Signal Process*, 2022, 167:108575.
[108] J Pan, Y Zi, J Chen, et al. LiftingNet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification. *IEEE Trans Ind Electron*, 2018, 65: 4973-4982.
[109] J Yuan, S Cao, G Ren, et al. LW-Net: an interpretable network with smart lifting wavelet kernel for mechanical feature extraction and fault diagnosis. *Neural Comput Appl*, 2022.
[110] Y Jiang, T Xia, D Wang, et al. A spatiotemporal dynamic wavelet network for infrared thermography-based machine prognostic. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 54: 1658-1665.
[111] G Michau, G Frusque, O Fink. Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series. *Proc Natl Acad Sci*, 2022, 119:e2106598119.
[112] Z Shang, Z Zhao, R Yan, et al. M-band wavelet network for machine anomaly detection from a frequency perspective. *Mechanical Systems and Signal Processing*, 2024, 216: 111489.
[113] Z Shang, Z Zhao, S Wang, et al. Anomaly detection from a frequency perspective: M-band wavelet packet anomaly detection network. *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2024: 5550-5554.
[114] Z Zhao, T Li, B An, et al. Model-driven deep unrolling: Towards interpretable deep learning against noise attacks for intelligent fault diagnosis. *ISA Trans*, 2022, 129: 644-662.

[115] B An, S Wang, F Qin, et al. Adversarial algorithm unrolling network for interpretable mechanical anomaly detection. *IEEE Trans Neural Netw Learn Syst*, 2023: 1-14.
[116] Z Ye, J Yu. Deep morphological convolutional network for feature learning of vibration signals and its applications to gearbox fault diagnosis. *Mech Syst Signal Process*, 2021, 161:107984.
[117] D Wang, Y Chen, C Shen, et al. Fully interpretable neural network for locating resonance frequency bands for machine condition monitoring. *Mech Syst Signal Process*, 2022, 168:108673.
[118] P Borghesani, N Herwig, W Wang, et al. Embedding signal processing knowledge in neural networks-an application to gear diagnostics. *AIAC 2023 20th Aust. Int. Aerosp. Congr. 20th Aust. Int. Aerosp. Congr., Engineers Australia Melbourne*, 2023: 669-677.
[119] H Lu, V P Nemani, V Barzegar, et al. A physics-informed feature weighting method for bearing fault diagnostics. *Mech Syst Signal Process*, 2023, 191:110171.
[120] T Xie, Q Xu, C Jiang, et al. The fault frequency priors fusion deep learning framework with application to fault diagnosis of offshore wind turbines. *Renew Energy*, 2023, 202:143-153.
[121] J Rani, T Tripura, H Kodamana, et al. Generative adversarial wavelet neural operator: Application to fault detection and isolation of multivariate time series data. arXiv preprint arXiv:2401.04004, 2024.
[122] M Russell, P Wang. Physics-informed deep learning for signal compression and reconstruction of big data in industrial condition monitoring. *Mech Syst Signal Process*, 2022, 168:108709.
[123] Y Yao, J Ma, Y Ye. Regularizing autoencoders with wavelet transform for sequence anomaly detection. *Pattern Recognit*, 2023, 134:109084.
[124] B Dai, G Frusque, Q Li, et al. Acceleration-guided acoustic signal denoising framework based on learnable wavelet transform applied to slab track condition monitoring. *IEEE Sens J*, 2022: 1–1.
[125] A Saxena, K Goebel, D Simon, et al. Damage propagation modeling for aircraft engine run-to-failure simulation. *2008 Int. Conf. Progn. Health Manag.*, 2008: 1-9.
[126] Y Duan, H Li, M He, et al. A BiGRU autoencoder remaining useful life prediction scheme with attention mechanism and skip connection. *IEEE Sensors Journal*, 2021, 21(9): 10905-14.
[127] S Fu, S Zhong, L Lin, et al. A novel time-series memory auto-encoder with sequentially updated reconstructions for remaining useful life prediction. *IEEE Trans. Neural Netw. Learn. Syst.*, 2022, 8.
[128] W Yu, I Y Kim, C Mechefske. Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. *Mech. Syst. Signal Proc.*, 2019, 129: 764-780.
[129] L Li, J Yan, H Wang, et al. Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder. *IEEE Trans. Neural Netw. Learn. Syst.*, 2021, 32(3): 1177-1191.
[130] D Li, Q Tao, J Liu, et al. Center-aware adversarial autoencoder for anomaly detection. *IEEE Trans. Neural Netw. Learn. Syst.*, 2022, 33(6): 2480-2493.
[131] S Akcay, A Atapour-Abarghouei, T P Breckon. GANomaly: Semi-supervised anomaly detection via adversarial training. *Proc. Asian Conf. Comput. Vis. Cham*, Switzerland: Springer, 2018: 622-637.
[132] E Bechhoefer. Condition based maintenance fault database for testing diagnostics and prognostic algorithms. *MFPT Data*, 2013.
[133] Y LeCun, L Bottou, Y Bengio, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
[134] W Zhang, G Peng, C Li, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, 2017, 17(2): 425.
[135] R Liu, F Wang, B Yang, et al. Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions. *IEEE Transactions on Industrial Informatics*, 2019, 16(6): 3797-3806.
[136] M Ravanelli, Y Bengio. Speaker recognition from raw waveform with sincnet. *2018 IEEE Spoken Language Technology Workshop (SLT)*, IEEE, 2018: 1021-1028.
[137] Q Zhu, Z Zhao, J Yan. Physics-informed machine learning for surrogate modeling of wind pressure and optimization of pressure sensor placement. *Computational Mechanics*, 2023, 71(3): 481-491.
[138] S Kohtz, Y Xu, Z Zheng, et al. Physics-informed machine learning model for battery state of health prognostics using partial charging segments. *Mechanical Systems and Signal Processing*, 2022, 172: 109002.
[139] M Raissi, P Perdikaris, G E Karniadakis. Physics informed deep learning (Part I): Data-driven solutions of nonlinear partial differential equations. arXiv preprint arXiv:1711.10561, 2017.
[140] C Anderson. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *The Wired*, 2008-06-23.

Ruqiang Yan received the Ph.D. degree in mechanical engineering from the *University of Massachusetts at Amherst, MA, USA*, in 2007. He is currently a Professor of mechanical engineering with *Xi'an Jiaotong University, China*. His research interests include data analytics, AI, and energy-efficient sensing for health diagnosis of large-scale, complex, dynamical systems. Dr. Yan is a Fellow of *ASME* (2019) and *IEEE* (2022). He is also the Editor-in-Chief of the *IEEE Transactions on Instrumentation and Measurement*, an Associate Editor of the *IEEE Sensors Journal*, and Associated Editor-in-Chief of *Chinese Journal of Mechanical Engineering* and Editorial Board Member of *Journal of University of Science and Technology of China*.

Zheng Zhou received the B.S. degree in mechanical engineering from *Xi'an Jiaotong University*, China, in 2019. He is currently working toward his Ph.D degree in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University*. From 2023, he is jointly working in *Intelligent Maintenance and Operations Systems lab* at EPFL. His current research is focused on physics-informed machine learning, automatic machine learning, and fault prognosis.

Zuogang Shang received the B.S. degree in mechanical engineering from *Xi'an Jiaotong University*, China, in 2020. He is currently working toward his Ph.D degree in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University*. His current research is focused on explainable deep learning, mechanical fault diagnosis, and anomaly detection.

Zhiying Wang received the B.S. degree in mechanical engineering from *Dalian University of Technology,* China, in 2020. He is currently working toward his Ph.D degree in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University*, China. His current research interests include physics-informed machine learning, dynamic modeling of hydraulic pimps, and digital twins.

Chenye Hu received the B.S. degree in mechanical engineering from *Xi'an Jiaotong University*, China, in 2020. He is currently working toward his Ph.D degree in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University*. His current research interests include deep learning and mechanical fault diagnosis.

Yasong Li received the B.S. degree in mechanical engineering from *Jiangsu University*, China, in 2020. He is currently working toward his Ph.D degree in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University*, China. His current research interests include interpretable deep learning, attention mechanism, remaining useful life prediction.

Yuangui Yang received the B.S. degree in mechanical engineering from *Xi'an Jiaotong University*, China, in 2020. He is currently working toward his Ph.D degree in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University*. His current research interests include explainable deep learning, graph neural network, and gearbox fault diagnosis.

Xuefeng Chen received the Ph.D. degree in mechanical engineering from *Xi'an Jiaotong University,* China, in 2004. He is currently a Full Professor with *School of Mechanical Engineering, Xi'an Jiaotong*

University. He has authored more than 100 SCI publications in areas of composite structure, aeroengine, wind power equipment, and so on. His research interests include intelligent maintenance, advanced sensing technology, and deep learning. Dr. Chen was the recipient of the National Excellent Doctoral Thesis Award in 2007, the First Technological Invention Award of the Ministry of Education in 2008, the Second National Technological Invention Award in 2009, the First Provincial Teaching Achievement Award in 2013, and the First Technological Invention Award of the Ministry of Education in 2015. He was awarded the Science and Technology Award for Chinese Youth in 2013. He is also the Executive Director of the Fault Diagnosis Branch in Chinese Mechanical Engineering Society.

Robert X. Gao received the Ph.D. degree in mechanical engineering from the *Technical University of Berlin,* Berlin, Germany, in 1991. He is currently the Cady Staley Professor of engineering and the Chair of the *Department of Mechanical and Aerospace Engineering, Case Western Reserve University,* Cleveland, OH, USA. He holds 13 patents, authored or coauthored three books and more than 400 technical papers, including more than 200 journal papers. His research interests include physics-based sensing, multiresolution data analysis, stochastic modeling, and machine learning for improving the observability of cyber physical systems, toward improved process and product quality control. Dr. Gao is a Senior Editor for *IEEE/ASME Transactions on Mechatronics*. He was the Lead Guest Editor of the Special Issue on Data Science-Enhanced Manufacturing of the *ASME Journal of Manufacturing Science and Engineering*, and was an Associate Editor for several journals of the IEEE ASME and IFAC. He is a Fellow of ASME, SME, and CIRP, and a Distinguished Fellow of IIAV, the International Institute of Acoustics and Vibration. He was the recipient of several professional awards, including the ASME Milton C. Shaw Manufacturing Research Medal, IEEE Best Application in Instrumentation and Measurement Award, SME Eli Whitney Productivity Award, ASME Blackall Machine Tool and Gage Award, IEEE Instrumentation and Measurement Society Technical Award, NSF Early CAREER Award, and several best paper awards. He was a Distinguished Lecturer of the IEEE Instrumentation and Measurement Society and IEEE Electron Devices Society.