Journal of Manufacturing Systems 77 (2024) 652–661

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Technical paper

Deep expert network: A unified method toward knowledge-informed fault diagnosis via fully interpretable neuro-symbolic AI

Qi Li, Yuekai Liu, Shilin Sun, Zhaoye Qin *, Fulei Chu

State Key Laboratory of Tribology, Department of Mechanical Engineering, Tsinghua University, Beijing, 100084, PR China

ARTICLE INFO

Keywords:
Deep expert network
Fault diagnosis
Neuro-symbolic AI
Rotating machinery
Fully interpretability

ABSTRACT

In recent years, intelligent fault diagnosis (IFD) based on Artificial Intelligence (AI) has gained significant attention and achieved remarkable breakthroughs. However, the black-box property of AI-enabled IFD may render it non-interpretable, which is essential for safety-critical industrial assets. In this paper, we propose a fully interpretable IFD approach that incorporates expert knowledge using neuro-symbolic AI. The proposed approach, named Deep Expert Network, defines neuro-symbolic node, including signal processing operators, statistical operators, and logical operators to establish a clear semantic space for the network. All operators are connected with trainable weights that decide the connections. End-to-end and gradient-based learning are utilized to optimize both the model structure weights and parameters to fit the fault signal and obtain a fully interpretable decision route. The transparency of model, generalization ability toward unseen working conditions, and robustness to noise attack are demonstrated through case study of rotating machinery, paving the way for future industrial applications.

1. Introduction

The current era is witnessing the emergence of the fourth industrial revolution, which has a profound impact on the manufacturing system through the utilization of artificial intelligence [1]. Given the complexity of safety-critical assets, such as aero-engines and wind turbines [2,3], their sustainability and security have become pressing concerns. For rotating machinery, the vibration signal containing fault information usually passes through multiple transmission paths to reach the sensor measuring point. This complexity makes it difficult to directly establish the relationship between signals and fault modes [4]. Therefore, intelligent fault diagnosis (IFD) approaches have gained attention over the past decade with the development of deep learning (DL) theories.

DL enables complex computational models consisting of multiple layers to acquire hierarchical representations of data, incorporating various levels of abstraction [5]. In essence, DL not only captures relationships between variables but also involves understanding the underlying principles that govern these relationships and deriving meaningful insights from them [6]. Consequently, it becomes apparent that DL can extract relevant information from raw sensor data.

The IFD approach relies on a data-driven paradigm that learns a mapping from the signals measured from machines such as vibration time series to health state. This approach eliminates the necessity for manual feature engineering and selection and reduces the reliance on human labor [7,8]. By integrating insights from several models, each model functions as a unique expert, leading to a more reliable and precise assessment of the individual's health status [9], which can be applied to real-time monitoring of unseen signals [10].

Despite its success, recent literature on IFD raises concerns about safety, reliability, and interpretability, as well as the limitations of pure DL in safety-critical assets [11]. In particular, interpretability has become increasingly crucial in safety-critical applications to facilitate a better understanding of the mechanism of the learned function and to establish a trustworthy and AI-enabled prognostics and health management system [12,13]. The interpretable AI methods can be divided into two categories, that is, transparent models and opaque models that require post-hoc explanation [14]. Transparent models, also referred as "white-box" models, are inherently interpretable, allowing humans to easily understand and trace their internal processes and decision-making mechanisms. Examples include linear regression, decision trees, and rule-based systems. On the other hand, opaque models, also known as "black-box" models, such as deep neural networks, necessitate the use of post-hoc explanation techniques to decipher their decisions. These techniques include methods like SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and

* Corresponding author.
E-mail addresses: liq22@tsinghua.org.cn (Q. Li), lykai@mail.tsinghua.edu.cn (Y. Liu), sunsl0115@163.com (S. Sun), qinzy@mail.tsinghua.edu.cn (Z. Qin), chufl@mail.tsinghua.edu.cn (F. Chu).

https://doi.org/10.1016/j.jmsy.2024.10.007
Received 6 February 2024; Received in revised form 12 July 2024; Accepted 8 October 2024
Available online 25 October 2024

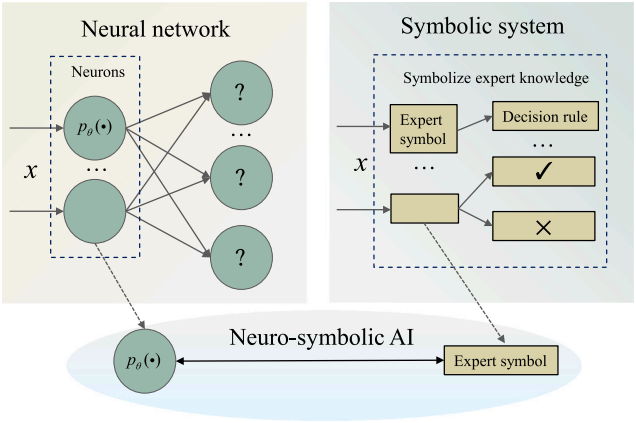

Fig. 1. Neuro-symbolic AI.[17].

feature importance analysis, which provide insights into the model's behavior after it has made predictions [15].

However, rather than post-hoc explaining the black-box model, Ref. [16] advocates for the learning of inherently interpretable and transparent models to avoid making high-stake decisions solely based on black-box models.

To bridge the advantages of IFD and the interpretability of expert knowledge, the concept of Neuro-symbolic AI is promising to give a choice [18]. Neuro-symbolic AI, as shown in Fig. 1, is a type of AI that combines symbolic programming and neural networks [17]. The term *neural* typically refers to artificial neural networks, which have made significant advances in the last decade, guided by the philosophy of connectionism [19]. Using DL techniques can learn from raw data and are robust against errors or outliers in the data, but are often difficult to interpret and cannot easily incorporate expert knowledge. In contrast, the term *symbolic* enables the use of expert knowledge and is explicit to understand and interpret through the explicit representation of knowledge, such as logic or symbol. However, symbolic programming is brittle and cannot be easily trained. In particular, symbolic programming represents knowledge in a logical, symbolic form and uses rules and logical reasoning to make decisions [20].

Given these strengths and limitations, researchers have endeavored to unify these two approaches in order to overcome their individual weaknesses and create more powerful AI systems. Particularly in the context of safety-critical assets, it is imperative to enhance the performance and transparency of the model through the utilization of neuro-symbolic AI.

Under the idea of neuro-symbolic AI, this paper proposes a novel IFD framework with transparent neuro-symbolic node (NSN), which are easily understood to maintenance engineers. To the best of our knowledge, this is the first attempt to introduce neuro-symbolic AI for fully interpretable IFD. The key contributions are as follows.

- To fulfill a fully interpretable IFD, a unified method namely Deep Expert Network (DEN) is proposed, utilizing expert knowledge via neuro-symbolic AI. Through sparsification, it enables the identification of transparent decision routes with a sparse white-box structure.
- In DEN, rather than the black-box fashion like conventional neural nodes, we introduce NSN incorporating signal processing operators, statistical operators, and logic operators with expert knowledge to build a fully interpretable learning space within the network.
- The NSN is connected by gate functions that determine the weights in the model, forming signal processing, statistical and logical layers that are trainable in an end-to-end manner.
- Case studies of rotating machinery are conducted to evaluate the proposed DEN framework. The results demonstrate promising diagnostic performance, characterized by transparency, robustness against noise attacks, and generalizability to unseen working conditions.

The remainder of this paper is arranged as follows. Section 2 reviews related work on IFD and symbolic regression. Section 3 provides the details of the proposed DEN. Section 4 presents the experimental results and analysis. Section 5 draws conclusions.

2. Preliminary

2.1. Intelligent fault diagnosis with expert knowledge

In recent years, significant progress has been made in the development of IFD models. These models, based on statistical learning principles, can be represented as follows: $h_\theta : \mathcal{X} \to \mathcal{Y}$, where h_θ denotes the mapping function with learnable parameters, and $\mathcal{X}, \mathcal{Y}$ represent the spaces of collected signals and monitoring indices or fault types, respectively [21]. The models have hypothesis space, which has all of the decision functions:

$$\mathcal{H} = \{h | Y = h_\theta(X), \theta \in \mathcal{R}^n\} \tag{1}$$

where the parameter θ can be learned by minimize the empirical risk:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}(\theta) = \underset{\theta}{\operatorname{argmin}}\, \frac{1}{N} \sum_{n=1}^{N} \ell(y_n, h(x_n; \theta)). \tag{2}$$

To improve the performance of fault diagnosis, researchers have attempted to incorporate expert knowledge, such as permutation entropy, wavelet packet transform, envelope analysis, and neuro-fuzzy system, into DL structures. For example, Rajabi et al. [22] built a multi-output adaptive neuro-fuzzy inference system and permutation entropy (MANFIS-PE) model to detect the frequency band associated with fault resonance. To establish a DL structure embedded with time-synchronous resampling mechanism, He et al. [23] incorporated discrete Fourier transform (DFT) techniques into an autoencoder (AE) named as DFT-IDFT AE. The parameters of DFT-IDFT AE are initialized as the DFT coefficients. After model training, the DFT-IDFT AE has the capability to reconstruct the vibration signals using the computed spectral average. Zhu et al. [24] proposed multi-sensor fusion deep belief network (MSF-DBN) to diagnose the continuously changing, uncertain, and unknown thruster faults. Chen et al. [25] emphasize the importance of prior knowledge of the system for fault diagnosis and presented a model GCN-SA combining structural analysis (SA) and graph convolutional network (GCN). Han et al. [26] combined short-time wavelet entropy (STWE), long short-term memory network (LSTM) and support vector machine (SVM). Mazzoleni et al. [27] combined Hidden Markov Model and SVM with envelope analysis to build a fuzzy Inference System (FIS) for condition monitoring. Huang et al. [28] utilized wavelet packet decomposition (WPD) and convolutional neural network (CNN) to extract multiple features.

The aforementioned IFD achieve remark progress due to the neural networks with expert knowledge, but the lack of deep understanding of the complex black-box model remains an issue. Consequently, researchers attempt to explore the physical interpretation of model or learned features. There is no doubt that incorporating expert knowledge has the potential to promote interpretability. Li et al. [29] proposed a WaveletKernelNet (WKN) toward interpretable IFD by replacing the first layer of CNN with a trainable continuous wavelet convolutional Layer. Wang et al. [11] extended extreme learning machine (ELM) as an interpretable neural network with signal processing technique including wavelet transform, square envelope and four sparsity measures of normalized squared envelope spectrum.

Apparently, the approaches discussed in this study can be considered as hybrid, combining elements of DL and expert knowledge. The

model benefits from the utilization of numerous parameters in the learning function Eq. (1). However, understanding these parameters can be challenging in most cases. On the other hand, the model that relies on expert experience results in a learning function that is predominantly configured manually. In this paper, we reveal the unified formulation for these methods in Section 3. The target of this work is to develop a model that exhibits both high performance and interpretability.

2.2. Symbolic regression

Symbolic regression [30] (SR) is a machine learning technique that aims to find mathematical expressions accurately predicting the output of a given input dataset. Unlike traditional statistical learning used by IFD, SR learns a mapping $\phi_\theta : \mathcal{X} \to \mathcal{Y}$ from the dataset to target label, assuming the existence of a simplified equation that may be easy to interpret, where ϕ are the equation structure. In this fashion, the model has a hypothesis space that encompasses all of the decision functions:

$$\boldsymbol{\Phi} = \{\phi | Y = \tilde{\phi}_\theta(X), \theta \in \mathcal{R}^n, \tilde{\phi} \in \mathcal{E}^z\} \tag{3}$$

where the $\mathcal{E}^z$ is the symbolic space requiring pre-defined operator such as the basic operations (+, ×) and functions (exp, sin, etc.). To address the requirement for interpretable, several SR methods have been proposed. For example, Cranmer et al. [31] utilized pySR framework to distill symbolic representation form the sparse latent representation learned by a graph neural network.

It is known that genetic programming-based SR is notably slow when facing vast predefined function. Therefore, SR has also incorporated various deep learning techniques such as symbolic embedding, reinforcement learning, pre-training model and gradient-based optimization. Kamienny et al. [32] proposed an end-to-end fashion to simultaneously learn the mathematical equation and its parameter. Petersen et al. [33] used a recurrent neural network to generate a distribution then can be mapped to mathematical expression. The whole task can be trained via reinforcement learning based on risk-seeking policy gradient. Biggio et al. [34] leveraged large-scale pre-training Transformer to predict the symbolic expression. Kim et al. [35] proposed a neural network-based architecture equation learner (EQL) network to carry out symbolic regression with compact constraint, which shows the capability to learn a transparent symbolic expression and to extrapolate outside of the distribution of training data.

However, the above SR method only focuses on simple basic operations such as +, × [32], and is limited to handling a maximum of 10 input features [33]. As a result, traditional SR is unable to effectively address complex variables such as vibration signals. In this paper, we propose an extension of SR by incorporating expert knowledge from signal processing, statistical features, and logic symbols. Our approach aims to enhance the accuracy and interpretability of predictions in the context of IFD. By using the proposed DEN, we anticipate the potential to enable health management in a more comprehensible manner.

3. Deep expert network

To construct the fully interpretable IFD that has comprehensible learning space by expert knowledge, the DEN is proposed, offering a unified neuro-symbolic AI perspective on IFD. Firstly, the IFD in Section 2.1 can be unified in the following framework:

$$\phi : \mathcal{X} \to \mathcal{S}, \phi\left(x, \theta_\phi\right) = s, \tag{4}$$

$$\tau : \mathcal{S} \to \mathcal{F}, \tau\left(s, \theta_\tau\right) = f, \tag{5}$$

$$\sigma : \mathcal{F} \to \mathcal{Y}, \sigma\left(f, \theta_\sigma\right) = y, \tag{6}$$

where ϕ, τ, σ represent the mapping form input space to enhanced signal space, feature space, and label space, respectively; $\theta_\phi, \theta_\tau, \theta_\sigma$ represent associated parameters. However, the black-black property of neural network such as the mapping ϕ above renders the fully interpretable of the decision route.

To achieve the fully interpretable decision route, we use NSN rather than traditional neural node to represent expert knowledge as symbols, giving the interpretable sematic of learning space using the expert knowledge base.

Specifically, as illustrated in Fig. 2, three modules are built by NSN as:

$$\phi : \mathcal{X} \to \mathcal{S}, \phi\left(x, w_\phi, \theta_\phi\right) = s, \tag{7}$$

$$\tau : \mathcal{S} \to \mathcal{F}, \tau\left(s, w_\tau, \theta_\tau\right) = f, \tag{8}$$

$$\sigma : \mathcal{F} \to \mathcal{Y}, \sigma\left(f, w_\sigma, \theta_\sigma\right) = y, \tag{9}$$

where ϕ, τ, σ are set from expert knowledge base representing signal processing module, statistical module and logical module, respectively.

In each module, we have connected weight w_ϕ, w_τ, w_σ to perform gate functions that determine the usage of the symbol in the model. For gradient-based optimization, both w and θ should be guaranteed to be differentiable. In the following subsections, a example of DEN used in this paper is demonstrated.

3.1. Signal processing layer

The expert knowledge needs firstly to be abstracted into symbolic operators in the signal processing layer (abbreviated to processing layer). The detailed symbols used in signal processing layer are listed in Table 1. It is noted that these symbolic operators are common in the signal processing field and the arguments and parameters involved in NSN are different in terms of specific symbols. For example, move average (MA) filter [36] is a simple filter used to smooth out short-term fluctuations in a signal and highlight longer-term trends. Additionally, other symbolic operators are detailed in the referenced literature. After obtaining these symbolic operators, we introduce the details of our module in a forward propagation way. Initially, a multidimensional input signal $\mathbf{x} \in \mathbb{R}^{B\times C\times L}$ is inputted to the module where L is the sample length, C is the number of channels. For illustrative purposes, we consider a single $x \in \mathbb{R}^{1\times C_{in}\times L}$ as example. To eliminate the influence of different signal processing unit and normalize the amplitude, the instance normalization to channel signal x_c is applied.

$$\tilde{x}_c = \frac{x_c - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}}, \tag{10}$$

where $\mu_c = \frac{1}{L}\sum_{l=1}^{L} x_{c,l}$ represents the mean of c channel, $\mu_c = \frac{1}{L}\sum_{l=1}^{L}(x_{c,l} - \mu_c)$ denotes the standard deviation of each channel and L is the number of sample points of each channel. By doing so, the module can effectively handle the input signal while preserving its statistical characteristics, thereby improving its efficacy. Then the output $\tilde{x}$ is weighted for the following symbolic calculation as $w_g\tilde{x} \in \mathbb{R}^{1\times C_{out}\times L}$ where w_g represent channel-wise vector specifying the connective of each channel of the signal as:

$$\mathbf{w_g}\tilde{\mathbf{x}} = \begin{pmatrix} w_{11}\tilde{x}_1 & +\cdots & w_{1C_{in}}\tilde{x}_{C_{in}} \\ \vdots & \ddots & \vdots \\ w_{C_{out}1}\tilde{x}_1 & +\cdots & w_{C_{out}C_{in}}\tilde{x}_{C_{in}} \end{pmatrix}. \tag{11}$$

Following Eq. (7) with signal processing symbols ϕ set from Table 1, we have:

$$\phi(x) = [\phi_1(x_{n-arity}; \theta_1), \ldots, \phi_H(x_{n-arity}; \theta_H)] \tag{12}$$

where $\phi(x) \in \mathbb{R}^{1\times H\times L}$, H represent the number of the symbols ϕ and the n-arity is according to the arity of the symbol. For instance, Hadamard product is binary operator where MA filter is unitary operator. Then, we use skip connection to avoid gradient vanishing [37] by:

$$\tilde{\phi}(x) = \phi(x) + w_s\tilde{x} \tag{13}$$

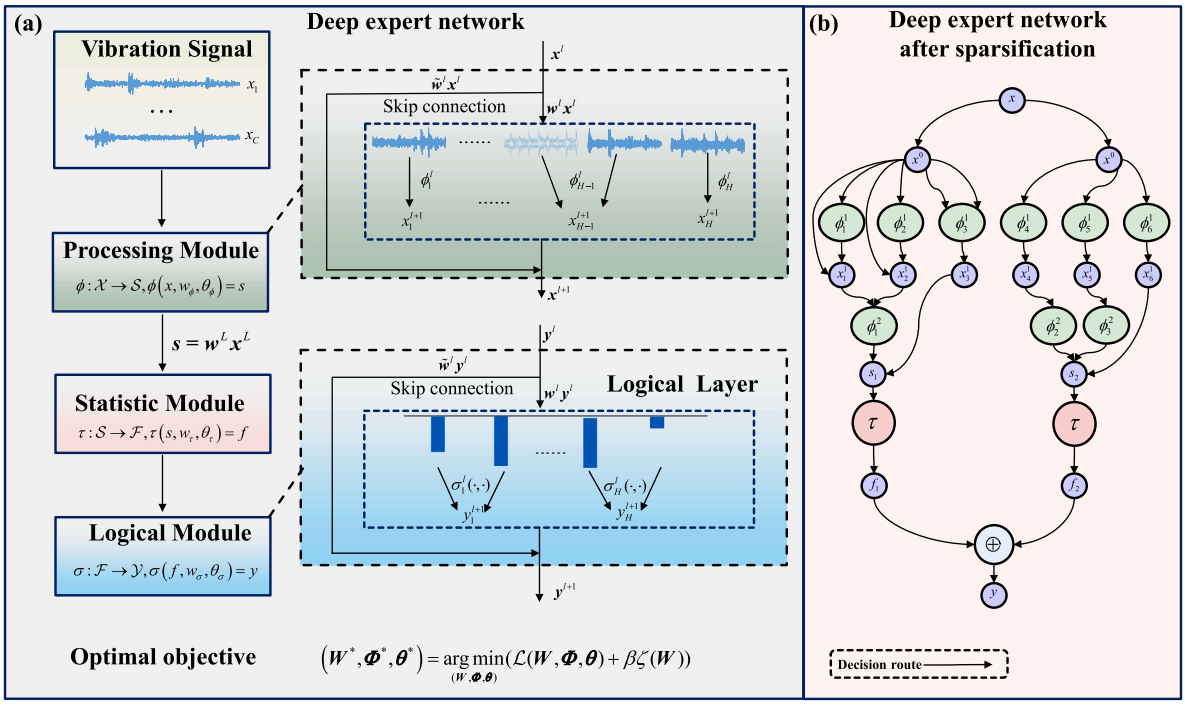

Fig. 2. Deep expert network: (a) the whole architecture; (b) DEN after sparsification.

Table 1
Signal processing symbol and their semantics.

Signal processing symbol	Equation
Haddam product [38]	$\odot(x_1, x_2) = x_1 \odot x_2$
Fourier transform [39]	$\mathcal{F}f(x) = F(\omega) = \int_{-\infty}^{+\infty} f(x)\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}x$
Square	$Squ(x) = x^2$
Morlet filter [40]	$*_{Morlet}(x; s) = \frac{A}{\sqrt{s}} e^{-\beta\frac{(t/s)^2}{2}} e^{j\omega t/s} * x$
Laplace filter [41]	$*_{Laplace}(x; s) = \frac{A}{\sqrt{s}} e^{-\frac{\zeta}{\sqrt{1-\zeta^2}\omega t/s}} e^{j\omega t/s} * x$
MA filter [36]	$*_{MA1}(x) = [1, 1] * x$
MA2 filter	$*_{MA2}(x) = [1, 1, 1] * x$
Differential filter [42]	$*_{D}(x) = [-1, 2, -1] * x$

Table 2
Logic rule symbols and their semantics.

Logic rule symbol	Equation of formal semantics
Implication	$M(x \to y) = \min\{1, 1 - x + y\}$
Biconditional	$M(x \leftrightarrow y) = 1 - \lvert x - y\rvert$
Negation	$M(\neg x) = 1 - x$
Conjunction	$M(x \wedge y) = \min\{x, y\}$
Disjunction	$M(x \vee y) = \max\{x, y\}$
Strong conjunction	$M(x \otimes y) = \max\{0, x + y - 1\}$
Strong disjunction	$M(x \oplus y) = \min\{1, x + y\}$

where w_s is the connected weight and $\tilde{\phi}(x)$ represent the input of the next signal processing layer. After passing through several signal processing layers, we obtain signal representation $s = x^L$ with transparent structure, where L is the number of layers. The representation enhances the health state-related information and can be used for later monitoring purposes.

3.2. Statistical layer

After signal processing layers, the signals from various channels are further weighted to measure the importance of different statistical symbols operator as $\tau\left(s, w_\tau, \theta_\tau\right) = f$. This connection is represented by the equation:

$$\mathbf{f} = \tau(\mathbf{w}_\tau \mathbf{s}). \tag{14}$$

Here, the input signal $\mathbf{s} \in \mathbb{R}^{N\times C\times L}$ is transformed into the output feature vector $\mathbf{f} \in \mathbb{R}^{N\times(n\cdot C)\times 1}$ through the feature mapping function. The term $n \cdot C$ represents n statistical features across C channels, resulting in a total of $n \cdot C$ features. For simplicity, this paper only use a typical feature kurtosis without inner parameter θ, as it has been shown to be an effective tool to measure the magnitude of transients in vibration signals [43]. The kurtosis can be fomulated as:

$$\mathrm{Kurtosis}(x) = \frac{\frac{1}{L}\sum_{l=1}^{L}(x_l - \mu)^4}{(\frac{1}{L}\sum_{l=1}^{L}(x_l - \mu)^2)^2}. \tag{15}$$

3.3. Logical layer

Prior to fault identification, the weighted feature is passed into logical module and the forward procedure is similar to processing layer and statistical layer as

$$\mathbf{y} = \sigma(\mathbf{w}_\sigma \mathbf{f}). \tag{16}$$

Here, σ refers to the continuous fuzzy logics that expand Boolean logic by allowing truth degrees from the closed interval [0, 1] instead of discrete truth-values in {0, 1}, and replacing Boolean connectives with differentiable real-valued operators [44]. This module employs real-valued Lukasiewicz logic [45] for logical inference. The symbols of logic rule and their formal semantics are listed in Table 2. Moreover, it should be noted that the max operation in logic rules is not continuous. Consequently, the concept of smooth maximum of two arguments is introduced as:

$$\max(x, y) \approx \frac{x \cdot e^{x/T} + y \cdot e^{y/T}}{e^{x/T} + e^{y/T}} \tag{17}$$

where T determines the level of differentiable approximation, and the right-hand side approaches the max function as T approaches zero. By employing the smooth maximum technique, all logic rules can be approximated smoothly.

3.4. The whole structure of DEN

Without losing the generality, DEN in this paper utilize three layer of signal processing layer, one layer of feature layer and one layer of logical layer. The data flow of the network can be formulated as:

$$x^0 \xrightarrow[w_s^0]{\phi_{1:c*m}^0} x^1 \xrightarrow[w_s^1]{\phi_{1:c*m}^1} x^2 \xrightarrow[w_s^2]{\phi_{1:c*m}^2} s \xrightarrow{\tau_{1:c}} f \xrightarrow[w_s]{\sigma_{1:c*m}} \tilde{y} \tag{18}$$

where $\phi_{1:c*m}^0$ is the first signal processing layer with c signal processing operator and each operator use m time, i.e. the $H = c * m$ in Eq. (12). w_s^0 is the skip connection weight in Eq. (13). After the logical module, the logical results with the truth degree are obtained.

To identify the health state, a linear combination layer is applied to assign weights to the logical results. This enable the use of the results to calculate loss function to conduct health monitoring and fault diagnosis.

$$\tilde{\mathbf{y}}' = \mathbf{w}_s \tilde{\mathbf{y}}. \tag{19}$$

The loss function employed in this study is the classical cross-entropy (CE), defined as follows:

$$\mathcal{L}(P(y), P_{W,\Phi,\theta}(\tilde{y}|x)) = -\frac{1}{N}\left(\sum_{i=1}^{N} P(y_i) \log P_{W,\Phi,\theta}(\tilde{y}_i|x_i)\right) \tag{20}$$

where W, Φ, θ are the weight of the connection, all of the symbolic operator and the parameters gathered from different symbols respectively.

3.5. Training and prune for sparcification

In DEN, we also adhere Occam's Razor principle [46] that demands the extraction of the most compact representation. Therefore, the architecture constrain is applied by employing L1 regularization on connection weights as:

$$\zeta(W) = \sum_i \sum_j \sum_l |w_{i,j,l}| \tag{21}$$

where i, j represent the row and column of each layer weight and l represent the l layer. Therefore, the final optimization goal optimized by stochastic gradient decent [47] is:

$$\left(W^*, \Phi^*, \theta^*\right) = \arg\min_{(W,\Phi,\theta)} (\mathcal{L}(W, \Phi, \theta) + \beta\zeta(W)) \tag{22}$$

where the β is the tradeoff between CE and L1 regularization.

In consideration of the varying tradeoffs between performance and sparsity for different applications, we propose the use of an iterative prune strategy (IPS) to achieve compact expression by removing of the lowest connected weight and its associated symbolic operator. A detailed description of the algorithm of IPS is presented in Algorithm 1 (see [48]). After training, we use D_{test} to test the performance of the DEN_{Pt}.

4. Cases study

4.1. Experiment setting

Rotating machinery plays a crucial role as a key category of safety-critical assets [49]. For instance, aero engines can lead to numerous mechanical faults, including rotor/stator rub-impact, high-cycle fatigue of blades, shaft cracks, unbalance-misalignment coupling faults, and other mechanical failures [50]. This study specifically centers on examining the primary bearing fault types associated with rotating machinery.

The case study is carried out with two datasets. The first one is Ottawa dataset under time-varying conditions [51]. The experiment was conducted to collect vibration signal on a mechanical-failure simulator (MFS-PK5M) with 200 kHz sampling rate as illustrated in Fig. 3. The Ottawa dataset comprises vibration data collected from bearings

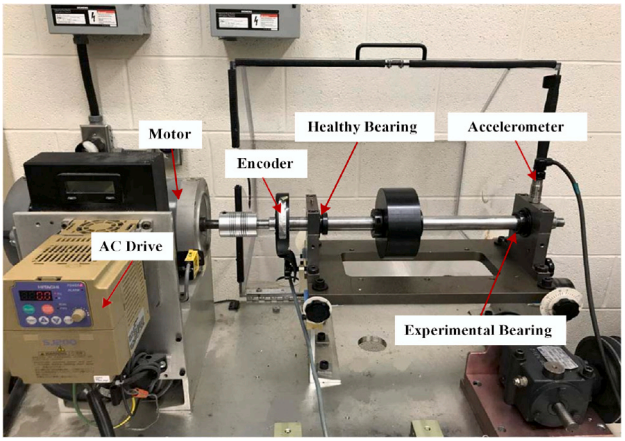

Fig. 3. Fault diagnosis test bench of the Ottawa dataset under time-varying speeds.

Table 3
Different speed conditions of the ottawa dataset. SI,SD,SID,SDI represent Speed Increase, Speed Decrease, Speed Increase then Decrease, Speed Decrease then Increase, respectively.

Bearing health condition	SI	SD	SID	SDI
Health	14.1–23.8	28.9–13.7	14.7–25.3–21.0	24.2–14.8–20.6
Inner race fault	12.5–27.8	24.3–9.9	15.1–24.4–18.7	25.3–14.8–19.4
Outer race fault	14.8–27.1	24.9–9.8	14.0–21.7–14.5	26.0–18.9–24.5

with different health conditions under various time-varying speed conditions. The health conditions included in the dataset are normal (Nor), faulty with an inner race (IF) defect and faulty with an outer race (OF) defect. Each bearing types is associated with a set of four rotational speeds: increasing speed, decreasing speed, increasing then decreasing speed and finally decreasing then increasing speed, as shown in Table 3. The speeds of different working conditions are denoted SI, SD, SID, SDI, respectively. It should be noted that the concepts of domain and working condition are also infer to speed in this case. Then in each working condition, the training dataset, validation dataset and testing dataset are construct. For single domain diagnosis, i.e., the training set and the testing set follow the same distribution for Section 4.3, we use SI domain. For the analysis of robustness and generalization in Section 4.5, we follow the domain generalization (DG) setting [37]. For example, we set DG task T0 for SD, which means we use SI, SDI, SID to minimize the empirical risk of the diagnosis model when we testing the model performance in the SD domain. Similarly, DG task T1 and T2 means we test the model in SDI and SID domain, respectively, using the model trained from the other domains.

For the second case study, we utilized our self-powered condition monitoring dataset acquired through a piezoelectric energy harvester [52]. In Fig. 4, the test rig incorporates an arc-shaped piezoelectric sheet positioned between the outer race of a rolling bearing and the bearing pedestal. Under compressive load resulting from rotational motion, the piezoelectric sheet produces electricity. The voltage signals captured by the energy harvester provide detailed frequency information that serves as a sensor for identifying faults in the rolling bearing. The dataset comprises voltage signals representing four health conditions: Nor, IF, ball fault (BF), and OF. These signals, sampled at 40960 Hz, are segmented at various intervals to create 100 samples each for the training, validation, and test sets. Each sample consists of 4096 data points. The objective here is a diagnostic task wherein the model is trained at speeds of 1 Hz and 15 Hz, and subsequently tested at a speed of 10 Hz.

All the diagnostic models are implemented in python 3.8, sympy 1.10.1 and PyTorch 1.12. The model training are realized based on the Linux-5.10.102.1 and GeForce RTX 3090 GPU.

Algorithm 1 IPS for sparsity

Require: Signal processing symbols ϕ ,statistical symbols τ,logic rule symbols σ
Require: Initialization of connective weight W ,Initialization of parameters θ
Require: Training and validation dataset D_{train} , $D_{validation}$
Require: Initialization optimization
Require: Prune time T_P ,Epoch ϵ in each prune procedure
Require: Other hyperparameters such as Batch B
Connect all symbols by connective weight
$t \leftarrow 0$ ▷ Model training start
$e \leftarrow 0$
$b \leftarrow 0$
while $t < T_P$ **do**
  while $e < \epsilon$ **do**
    while b<B **do**
      Get batch data x, y from D_{train}
      Forward propagation according to Eq. (18) and Eq. (19)
      Optimize the weights W and parameters θ according to Eq. (22)
      $b \leftarrow b + 1$
    end while
    Validate the performance of the DEN by $D_{validation}$
    $e \leftarrow e + 1$
  end while
  Prune the connective weight W by setting the ones with the lowest L1-norm to zero according to Ref. [48]
  Fine-tuning the connective weight W and parameters θ according to Eq. (22)
  Save the DEN_{P_t}
  $t \leftarrow t + 1$
end while

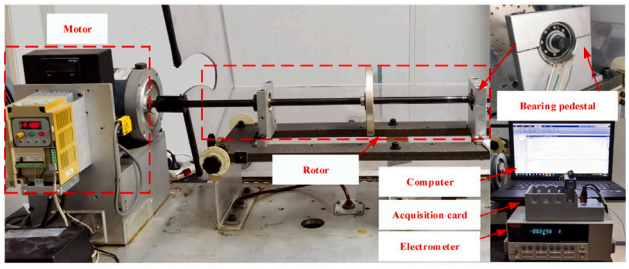

Fig. 4. Test rig of self-powered fault diagnosis.

4.2. Compared methods

To the best of our knowledge, the DEN is the first IFD using a neuro-symbolic structure. DEN not only perform fault identification, but also generate fully interpretable decision route rather than implicit black-box representations. To evaluate the effectiveness of our method, we compared it to several IFD methods with different parameters. One of these methods is ResNet fashion method [37], a well-known convolutional neural network in multiple domains including IFD [53]. As a neural fashion backbone, we choose the ResNet-18 version to compare. Additionally, we also evaluated the WKN model [29], which uses continue Morlet wavelet convolution as the expert knowledge in the ResNet.

Using 4 time of each symbolic operator, we train a DEN as our basic model and carry out multiple prunes with training and pruning algorithm to obtain the sub-model DEN_{P1-P5}. We also set up the model without parameter T (DEN w/o T) in Eq. (17) to evaluate the smooth approximation and a model without the statistical feature Kurtosis (DEN w/o K) replaced by mean, to evaluate the statistical features.

4.3. Diagnosis results of single domain

As shown in Table 4, the proposed transparent DEN_{P3} achieve relatively high accuracy (97.3% ± 1.6%), which is regarded as the default DEN model. ResNet has a larger standard deviation, and is considered a black-box model that are not fully interpretable for the industrial maintainers. And WKN obtain higher performance than ResNet but cannot compete with DEN. The DEN w/o T cannot be easily trained which indicates that the T smooth the model training. Interestingly, the DEN w/o Kurtosis obtains a lower performance than DEN with Kurtosis, indicating that Kurtosis is a more appropriate statistic feature for performing an aggregator role than mean in this case. In summary, these results show the effectiveness of the DEN structure compared related models and the ablation of DEN.

Table 4
Diagnosis accuracy of Ottawa dataset.

Methods	Acc (%)	Methods	Acc (%)	Methods	Acc (%)
ResNet	90.2 ± 12.7	DEN_{P2}	93.7 ± 3.4	DEN_{P5}	94.3 ± 3.6
WKN	90.5 ± 12.2	DEN_{P3}	97.3 ± 1.6	DEN w/o T	89.5 ± 1.7
DEN_{P1}	93.7 ± 3.4	DEN_{P4}	94.7 ± 1.1	DEN w/o K	59.1 ± 14.5

4.4. Learned subnetwork

The distinguishing characteristic of DEN is the ability to not only obtain the model predictions like the traditional black-box deep learning fashion but also to derive explicitly fully interpretable decision route in Fig. 5.

Overall, the behavior of signal processing in DEN constructs a series of filter banks to enhance the fault related frequency band supervised by fault label. This process is similar to modeling the human expert to identify the target frequency band. Fig. 6 depicts the related waveform of three health condition frequency responses. Notably, the filters in s_2 perform the joint effect of low-pass filter and band-pass filters where the filters in s_4 perform the joint effect of high-pass and band-pass filters. Furthermore, it is evident that the feature kurtosis can identify three different states of health, which are highlighted in the red and yellow circle by multiple weighted filter banks.

Notably, without the requirement of manually designing the filter banks combination, the DEN can automatically learn the signal

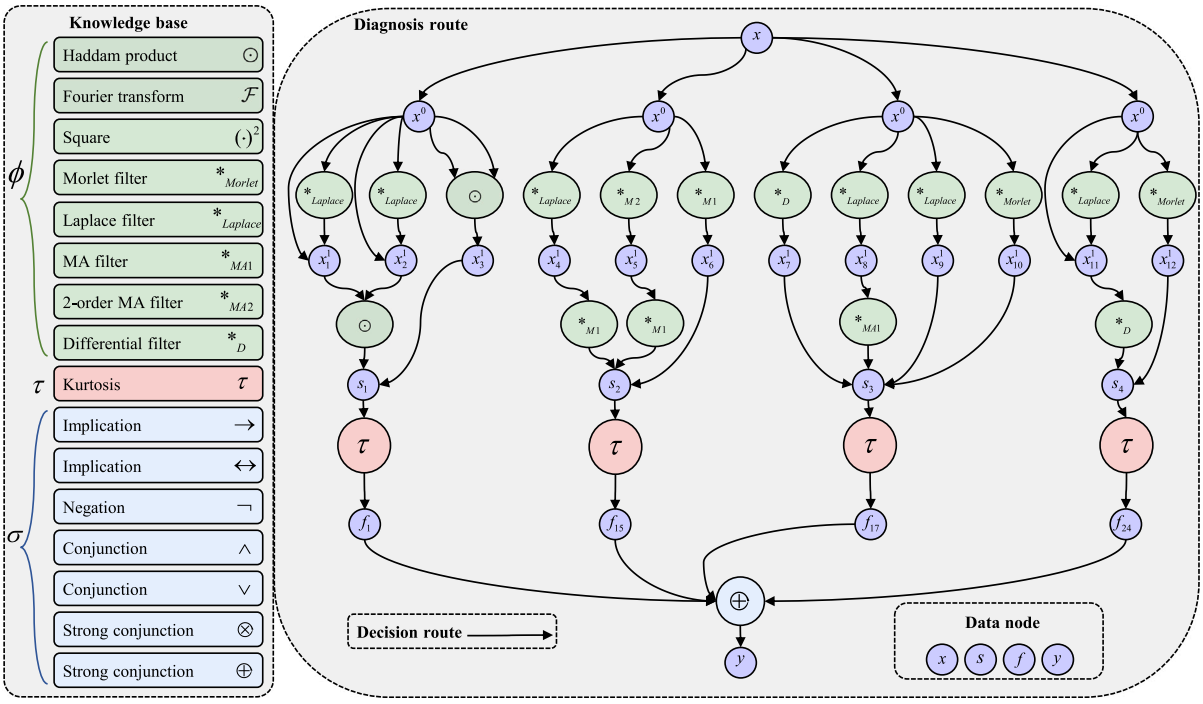

Fig. 5. The learned fully interpretable decision route.

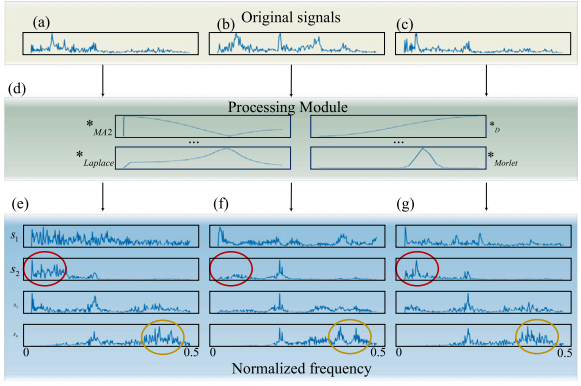

Fig. 6. Frequency spectrum of original signals, filters and enhanced signals: (a) original signal of Nor, (b) original signal of IF, (c) original signal of OF, (d) weighted filters in filter module, (e) enhanced signals of Nor, (f) enhanced signals of IF, (g) enhanced signals of OF.

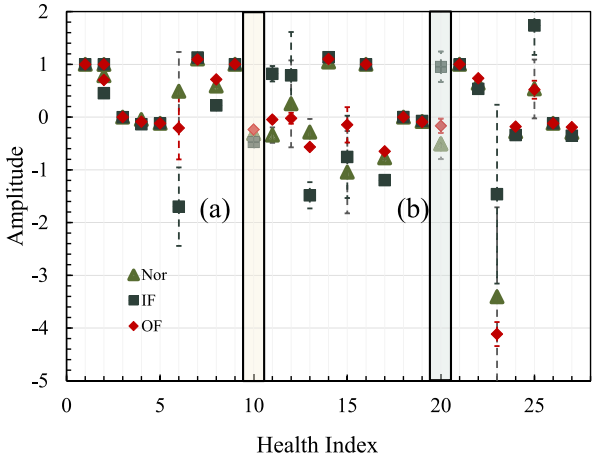

Fig. 7. Health index output after logical module: (a) not interested by DEN, (b) interested by DEN.

enhancements to extract the more distinguishable features. After the learned signal processing step, the statistic feature of the enhanced signals are extracted.

As demonstrated in Fig. 7, due to the weight constrain $\zeta(W)$, only a few health indexes are interested by DEN such as health index 10. Consequently, the features extracted by DEN effectively represent the underlying health condition in Fig. 7(b). Therefore, DEN can yield fully interpretable decision route that represent important health indexes of three different health condition after the logical module. There explicit decision route are fully interpretable and can be further research and understood the signal behavior.

4.5. Generalization analysis of multiple domains

Then, we investigated the basic generalization ability of the three models through empirical risk minimization without fancy techniques. In the generalization analysis, we defined three tasks as mentioned in Section 4.1: T0, T1, T2. For instance, the T0 task involves training from seen domains under domain of SI, SID, and SDI, thereby generalizing to the unseen target domain, i.e., the working condition of SD. The diagnosis performance of generalization tasks is shown in Fig. 8. Notably, the generalization task is more challenging than the task discussed in Section 4.3 because the model has no access to the target domain although we use multiple source domain. Then, it is observed in the performance that three model achieve similar accuracy on average, but the standard deviation of WKN and ResNet is larger than DEN.

Furthermore, the t-SNE [54] is used for dimension reduction of extracted features of T1. The visualization of the reduced two-dimension is shown in Fig. 9. S0, S1, and S2 represent the data in source domains while T indicates the test target domain. Therefore, S0IF indicates the feature from inner race fault signal in the first domain. It is observed that three models can cluster well with decision boundaries cross different working conditions. However, the feature of S0IF cannot be effectively clustered by WKN and ResNet.

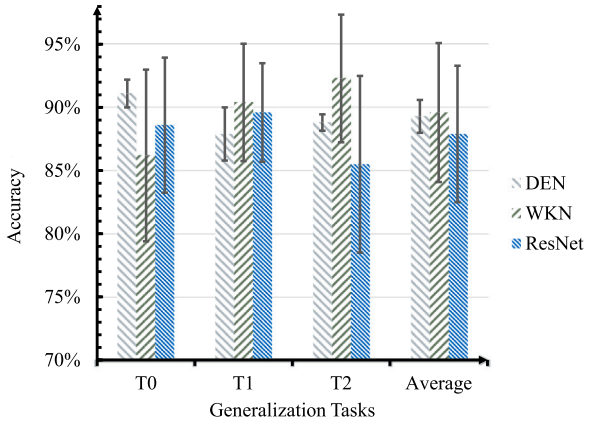

Fig. 8. Diagnosis performance of generalization tasks.

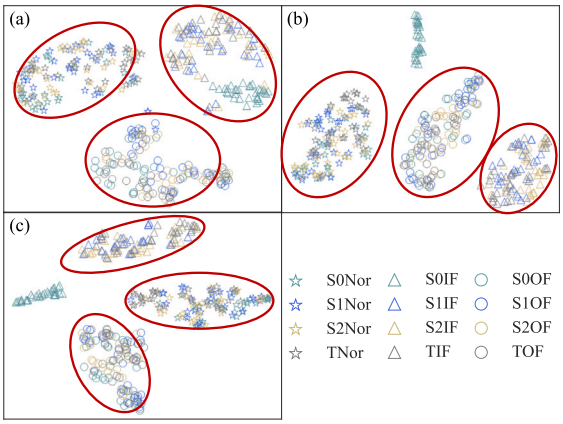

Fig. 9. Feature visualization of T1 via T-SNE: (a) DEN, (b) WKN, (c) ResNet.

Table 5
Diagnosis accuracy of self-powered fault diagnosis dataset.

Methods	Acc (%)	Methods	Acc (%)	Methods	Acc (%)
ResNet	52.4 ± 4.2	DEN_{P2}	98.8 ± 3.4	DEN_{P5}	95.4 ± 2.0
WKN	61.9 ± 19.4	DEN_{P3}	99.62 ± 0.2	DEN w/o T	57.0 ± 14.1
DEN_{P1}	98.2 ± 1.3	DEN_{P4}	94.7 ± 9.7	DEN w/o K	56.1 ± 22.2

Similarly, in the self-powered fault diagnosis dataset, Table 5 illustrates the diagnostic accuracy. Both ResNet and WKN performed unsatisfactory results in the generalization task. WKN improved its generalization to some extent by incorporating expert knowledge. For DEN, it was observed that after three iterations of the prune strategy, DEN achieved good performance. Furthermore, the ablation experiments on parameters T and Kurtosis also validated the effectiveness of their respective modules.

4.6. Robustness analysis

To assess the robustness of three models, we add Gaussian noise with different signal-to-ratio (SNR) [55] to the testing signal. The diagnosis accuracy of signal domain task in Section 4.3 is shown in Fig. 10(a).

Notably, the fully data-driven model is highly vulnerable to noise attacks. This may lead to mistrust of the black-box model by industrial maintainers. This highlights the importance of understanding the behavior of a model, whether it is merely fitting the dataset or learning remarkable knowledge. Although attacked by the noise, DEN exhibits greater robustness than the others to a certain extent. Furthermore, we also test the robustness of the generalization model of case 1 as shown

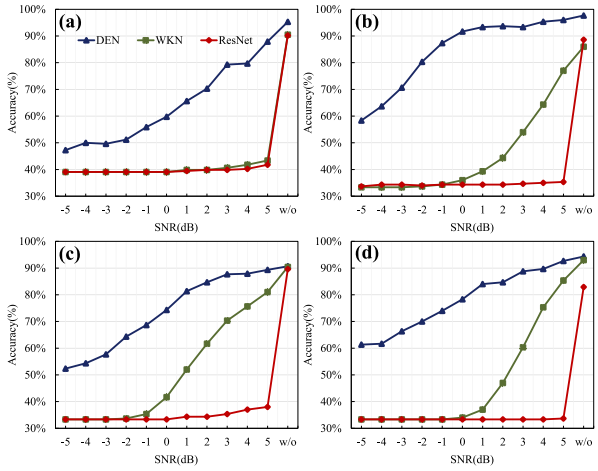

Fig. 10. Diagnosis accuracy with different SNRs: (a) single domain task of SI domain, (b) generalization task of T0, (c) (b) T1, (d) T2.

in Fig. 10(b–d). Surprisingly, WKN demonstrates enhanced resistance to noise attacks due to the training of multiple domain signal. Models with expert knowledge are better equipped to extract invariant features as the data diversity increase, i.e., more data from a variety of working conditions. DEN is evidently is the most robust backbone model under varying SNR noise conditions.

5. Conclusion and future work

Via neuro-symbolic AI, we present a unified perspective by a transparency DEN with expert knowledge for fault diagnosis. In the DEN, we first introduce neuro-symbolic operator to establish a clear learning space of the network, which narrows the gap between neuro-symbolic AI and IFD. All of these operators are connected with trainable weights to decide the connections. Then, through gradient-based sparse learning, both model structure weights and parameters can be optimized to fit the fault signal and obtain the compact decision route. In the case study, the model actually like the human expert to identify the target frequency band, feature extraction and decision making, which shows its properties on transparency. Through experiment on multi varying working and varying SNR noise attack, the DEN get better generalization ability and robustness even with fewer parameters than the competitive models. Furthermore, due to the clearly defined semantic space for the network, constructed through expert knowledge and trainable weights that determine the connections, a fully interpretable decision route can be established to inform the decision-maker. In the future, we will improve the interpretability of neuro-symbolic operator and also introduce other symbols into our DEN to adapt the variable speed condition.

CRediT authorship contribution statement

Qi Li: Conceptualization, Investigation, Methodology, Software, Visualization, Writing – original draft. **Yuekai Liu:** Data curation, Formal analysis, Investigation. **Shilin Sun:** Data curation, Investigation, Methodology, Writing – review & editing. **Zhaoye Qin:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Fulei Chu:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China (Grant No. 11972204) and the Science Center for Gas Turbine Project (Grant No. P2022-B-III-002–001)

References

[1] Liu Y, Jiang H, Yao R, Zhu H. Interpretable data-augmented adversarial variational autoencoder with sequential attention for imbalanced fault diagnosis. J Manuf Syst 2023;71:342–59. http://dx.doi.org/10.1016/j.jmsy.2023.09.019, URL https://linkinghub.elsevier.com/retrieve/pii/S0278612523002017.

[2] Chai Z, Zhao C, Huang B. Multisource-refined transfer network for industrial fault diagnosis under domain and category inconsistencies. IEEE Trans Cybern 2022;52(9):9784–96. http://dx.doi.org/10.1109/TCYB.2021.3067786.

[3] Qin Y, Qian Q, Luo J, Pu H. Deep joint distribution alignment: A novel enhanced-domain adaptation mechanism for fault transfer diagnosis. IEEE Trans Cybern 2022;1–11. http://dx.doi.org/10.1109/TCYB.2022.3162957.

[4] Li S, Li T, Sun C, Yan R, Chen X. Multilayer grad-CAM: An effective tool towards explainable deep neural networks for intelligent fault diagnosis. J Manuf Syst 2023;69:20–30. http://dx.doi.org/10.1016/j.jmsy.2023.05.027, URL https://linkinghub.elsevier.com/retrieve/pii/S0278612523001024.

[5] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015-05-28;521(7553):436–44. http://dx.doi.org/10.1038/nature14539, URL https://www.nature.com/articles/nature14539.

[6] Zhang W, Yang G, Lin Y, Ji C, Gupta MM. On definition of deep learning. In: 2018 world automation congress. WAC, IEEE; 2018-06, p. 1–5. http://dx.doi.org/10.23919/WAC.2018.8430387, URL https://ieeexplore.ieee.org/document/8430387/.

[7] Lei Y, Yang B, Jiang X, Jia F, Li N, Nandi AK. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mech Syst Signal Process 2020;138:106587. http://dx.doi.org/10.1016/j.ymssp.2019.106587, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327019308088.

[8] Modi S, Lin Y, Cheng L, Yang G, Liu L, Zhang WJ. A socially inspired framework for human state inference using expert opinion integration. IEEE/ASME Trans Mechatronics 2011;16(5):874–8. http://dx.doi.org/10.1109/TMECH.2011.2161094.

[9] Ji C, Lu X, Zhang W. A biobjective optimization model for expert opinions aggregation and its application in group decision making. IEEE Syst J 2021;15(2):2834–44. http://dx.doi.org/10.1109/JSYST.2020.3027716.

[10] Mo Z, Zhang Z, Miao Q, Tsui K-L. Sparsity-constrained invariant risk minimization for domain generalization with application to machinery fault diagnosis modeling. IEEE Trans Cybern 2022;1–13. http://dx.doi.org/10.1109/TCYB.2022.3223783, URL https://ieeexplore.ieee.org/document/9976035/.

[11] Wang D, Chen Y, Shen C, Zhong J, Peng Z, Li C. Fully interpretable neural network for locating resonance frequency bands for machine condition monitoring. Mech Syst Signal Process 2022;168:108673. http://dx.doi.org/10.1016/j.ymssp.2021.108673, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327021009961.

[12] Han T, Li Y-F. Out-of-distribution detection-assisted trustworthy machinery fault diagnosis approach with uncertainty-aware deep ensembles. In: EI, Reliab Eng Syst Saf In: EI, 2022;226:108648. http://dx.doi.org/10.1016/j.ress.2022.108648.URL https://linkinghub.elsevier.com/retrieve/pii/S0951832022002836,

[13] Liang W, Tadesse GA, Ho D, Li F-F, Zaharia M, Zhang C, Zou J. Advances, challenges and opportunities in creating data for trustworthy AI. Nat Mach Intell 2022;4(8):669–77. http://dx.doi.org/10.1038/s42256-022-00516-1, URL https://www.nature.com/articles/s42256-022-00516-1.

[14] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. In: EI, Inf Fusion In: EI, 2020;58:82–115. http://dx.doi.org/10.1016/j.inffus.2019.12.012.URL https://linkinghub.elsevier.com/retrieve/pii/S1566253519308103,

[15] Rong Y, Leemann T, Nguyen TT, Fiedler L, Qian P, Unhelkar V, Seidel T, Kasneci G, Kasneci E. Towards human-centered explainable AI: A survey of user studies for model explanations. IEEE Trans Pattern Anal Mach Intell 2024;46(4):2104–22. http://dx.doi.org/10.1109/TPAMI.2023.3331846, arXiv:2210.11584.

[16] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1(5):206–15. http://dx.doi.org/10.1038/s42256-019-0048-x, URL http://www.nature.com/articles/s42256-019-0048-x.

[17] Sun JJ, Tjandrasuwita M, Sehgal A, Solar-Lezama A, Chaudhuri S, Yue Y, Costilla-Reyes O. Neurosymbolic programming for science. In: NeurIPS 2022 AI for science workshop. 2022, URL http://arxiv.org/abs/2210.05050. arXiv:2210.05050 [cs].

[18] Garcez Ad, Lamb LC. Neurosymbolic AI: the 3rd wave. Artif Intell Rev 2023;56(11):12387–406. http://dx.doi.org/10.1007/s10462-023-10448-w, URL https://link.springer.com/10.1007/s10462-023-10448-w.

[19] Hitzler P, Eberhart A, Ebrahimi M, Sarker MK, Zhou L. Neuro-symbolic approaches in artificial intelligence. In: EI, Natl Sci Rev In: EI, 2022;9(6):nwac035. http://dx.doi.org/10.1093/nsr/nwac035.URL https://academic.oup.com/nsr/article/doi/10.1093/nsr/nwac035/6542460,

[20] Badreddine S, d'Avila Garcez A, Serafini L, Spranger M. Logic tensor networks. In: EI, Artificial Intelligence In: EI, 2022;303:103649. http://dx.doi.org/10.1016/j.artint.2021.103649.URL https://linkinghub.elsevier.com/retrieve/pii/S0004370221002009,

[21] Murphy KP. Probabilistic machine learning: an introduction. Adaptive computation and machine learning, Cambridge, Massachusetts London, England: The MIT Press; 2022.

[22] Rajabi S, Saman Azari M, Santini S, Flammini F. Fault diagnosis in industrial rotating equipment based on permutation entropy, signal processing and multi-output neuro-fuzzy classifier. Expert Syst Appl 2022;206:117754. http://dx.doi.org/10.1016/j.eswa.2022.117754, URL https://linkinghub.elsevier.com/retrieve/pii/S0957417422010326.

[23] He M, He D. A new hybrid deep signal processing approach for bearing fault diagnosis using vibration signals. Neurocomputing 2020;396:542–55. http://dx.doi.org/10.1016/j.neucom.2018.12.088, URL https://linkinghub.elsevier.com/retrieve/pii/S0925231219304904.

[24] Zhu D, Cheng X, Yang L, Chen Y, Yang SX. Information fusion fault diagnosis method for deep-sea human occupied vehicle thruster based on deep belief network. In: EI, IEEE Trans Cybern In: EI, 2022;52(9):9414–27. http://dx.doi.org/10.1109/TCYB.2021.3055770.Conference Name: IEEE Transactions on Cybernetics,

[25] Chen Z, Xu J, Peng T, Yang C. Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge. IEEE Trans Cybern 2022;52(9):9157–69. http://dx.doi.org/10.1109/TCYB.2021.3059002.

[26] Han Y, Qi W, Ding N, Geng Z. Short-time wavelet entropy integrating improved LSTM for fault diagnosis of modular multilevel converter. IEEE Trans Cybern 2022;52(8):7504–12. http://dx.doi.org/10.1109/TCYB.2020.3041850.

[27] Mazzoleni M, Sarda K, Acernese A, Russo L, Manfredi L, Glielmo L, Del Vecchio C. A fuzzy logic-based approach for fault diagnosis and condition monitoring of industry 4.0 manufacturing processes. In: EI, Eng Appl Artif Intell In: EI, 2022;115:105317. http://dx.doi.org/10.1016/j.engappai.2022.105317.URL https://linkinghub.elsevier.com/retrieve/pii/S0952197622003566,

[28] Huang D, Zhang W-A, Guo F, Liu W, Shi X. Wavelet packet decomposition-based multiscale CNN for fault diagnosis of wind turbine gearbox. In: EI, IEEE Trans Cybern In: EI, 2021;1–11. http://dx.doi.org/10.1109/TCYB.2021.3123667.Conference Name: IEEE Transactions on Cybernetics,

[29] Li T, Zhao Z, Sun C, Cheng L, Chen X, Yan R, Gao RX. WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis. In: EI, IEEE Trans Syst Man Cybern: Syst In: EI, 2020;52(4):2302–12. http://dx.doi.org/10.1109/TSMC.2020.3048950.Conference Name: IEEE Transactions on Systems, Man, and Cybernetics: Systems,

[30] La Cava W, Orzechowski P, Burlacu B, de França FO, Virgolin M, Jin Y, Kommenda M, Moore JH. Contemporary symbolic regression methods and their relative performance. In: 35th Conference on Neural Information Processing Systems. NeurIPS 2021, 2021, URL http://arxiv.org/abs/2107.14351.

[31] Cranmer M, Sanchez Gonzalez A, Battaglia P, Xu R, Cranmer K, Spergel D, Ho S. Discovering symbolic models from deep learning with inductive biases. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, editors. Advances in neural information processing systems. vol. 33, Curran Associates, Inc.; 2020, p. 17429–42, URL https://proceedings.neurips.cc/paper/2020/file/c9f2f917078bd2db12f23c3b413d9cba-Paper.pdf.

[32] Kamienny P-A, d'Ascoli S, Lample G, Charton F. End-to-end symbolic regression with transformers. In: 36th conference on neural information processing systems. NeurIPS 2022, 2022, URL http://arxiv.org/abs/2204.10532 arXiv:2204.10532 [cs].

[33] Petersen BK, Landajuela M, Mundhenk TN, Santiago CP, Kim SK, Kim JT. Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients. 2021, URL http://arxiv.org/abs/1912.04871 arXiv:1912.04871 [cs, stat].

[34] Biggio L, Bendinelli T, Neitz A, Lucchi A, Parascandolo G. Neural symbolic regression that scales. In: Proceedings of the 38 th international conference on machine learning. ICML, 2021, URL http://arxiv.org/abs/2106.06427 arXiv:2106.06427 [cs].

[35] Kim S. Integration of neural network-based symbolic regression in deep learning for scientific discovery. 2022, URL https://github.com/samuelkim314/DeepSymReg original-date: 2020-04-16T16:04:10Z.

[36] Bianchi FM, Grattarola D, Livi L, Alippi C. Graph neural networks with convolutional ARMA filters. IEEE Trans Pattern Anal Mach Intell 2021;1. http://dx.doi.org/10.1109/TPAMI.2021.3054830, URL https://ieeexplore.ieee.org/document/9336270/.

[37] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Computer vision and pattern recognition. 2015, URL http://arxiv.org/abs/1512.03385 arXiv:1512.03385 [cs].

[38] Li T, Zhou Z, Li S, Sun C, Yan R, Chen X. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. Mech Syst Signal Process 2022;168:108653. http://dx.doi.org/10.1016/j.ymssp.2021.108653, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327021009791.

[39] Lee-Thorp J, Ainslie J, Eckstein I, Ontanon S. Fnet: Mixing tokens with Fourier transforms. 2022, arxiv URL http://arxiv.org/abs/2105.03824 arXiv:2105.03824 [cs].

[40] Ashmead J. Morlet wavelets in quantum mechanics. Quanta 2012;1(1):58–70. http://dx.doi.org/10.12743/quanta.v1i1.5, URL http://quanta.ws/ojs/index.php/quanta/article/view/5.

[41] Chen Q, Dong X, Tu G, Wang D, Zhao B, Peng Z. TFN: An interpretable neural network with time-frequency transform embedded for intelligent fault diagnosis. 2022, arxiv URL http://arxiv.org/abs/2209.01992 arXiv:2209.01992 [cs, eess].

[42] Wang J, Han F, Li Y, Wang Z, Du W. First-order differential filtering spectrum division method and information fusion multi-scale network for fault diagnosis of bearings under different loads. Meas Sci Technol 2022;33(7):075014. http://dx.doi.org/10.1088/1361-6501/ac6661, URL https://iopscience.iop.org/article/10.1088/1361-6501/ac6661.

[43] Shang Z, Zhao Z, Yan R. Denoising fault-aware wavelet network: A signal processing informed neural network for fault diagnosis. Chin J Mech Eng 2023;36(1):9. http://dx.doi.org/10.1186/s10033-023-00838-0, URL https://cjme.springeropen.com/articles/10.1186/s10033-023-00838-0.

[44] Barbiero P, Ciravegna G, Giannini F, Zarlenga ME, Magister LC, Tonda A, Lio' P, Precioso F, Jamnik M, Marra G. Interpretable neural-symbolic concept reasoning. 2023, arXiv URL http://arxiv.org/abs/2304.14068 arXiv:2304.14068 [cs, stat].

[45] Marchioni E, Wooldridge M. Łukasiewicz logics for cooperative games. Artificial Intelligence 2019;275:252–78. http://dx.doi.org/10.1016/j.artint.2019.03.003, URL https://linkinghub.elsevier.com/retrieve/pii/S0004370219301328.

[46] Champion K, Lusch B, Kutz JN, Brunton SL. Data-driven discovery of coordinates and governing equations. Proc Natl Acad Sci 2019;116(45):22445–51. http://dx.doi.org/10.1073/pnas.1906995116, URL http://arxiv.org/abs/1904.02107 arXiv:1904.02107 [stat].

[47] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2017, arXiv URL http://arxiv.org/abs/1412.6980 arXiv:1412.6980 [cs] version: 8.

[48] Molchanov P, Tyree S, Karras T, Aila T, Kautz J. Pruning convolutional neural networks for resource efficient inference. 2017, arXiv URL http://arxiv.org/abs/1611.06440 arXiv:1611.06440 [cs, stat].

[49] Prabith K, Krishna IRP. The numerical modeling of rotor–stator rubbing in rotating machinery: a comprehensive review. Nonlinear Dynam 2020-07;101(2):1317–63. http://dx.doi.org/10.1007/s11071-020-05832-y, URL https://link.springer.com/10.1007/s11071-020-05832-y.

[50] Jin Y, Zhou X, Quan X, Zhang X, Lu K, Wang J. Topological structures of vibration responses for dual-rotor aeroengine. Mech Syst Signal Process 2024-02;208:111053. http://dx.doi.org/10.1016/j.ymssp.2023.111053, URL https://linkinghub.elsevier.com/retrieve/pii/S0888327023009615.

[51] Huang H, Baddour N. Bearing vibration data collected under time-varying rotational speed conditions. Data Brief 2018;21:1745–9. http://dx.doi.org/10.1016/j.dib.2018.11.019, URL https://linkinghub.elsevier.com/retrieve/pii/S2352340918314124.

[52] Zhang L, Zhang F, Qin Z, Han Q, Wang T, Chu F. Piezoelectric energy harvester for rolling bearings with capability of self-powered condition monitoring. In: EI, Energy In: EI, 2022;238:121770. http://dx.doi.org/10.1016/j.energy.2021.121770.URL https://linkinghub.elsevier.com/retrieve/pii/S0360544221020181,

[53] Wang X, Shen C, Xia M, Wang D, Zhu J, Zhu Z. Multi-scale deep intra-class transfer learning for bearing fault diagnosis. Reliab Eng Syst Saf 2020;202:107050. http://dx.doi.org/10.1016/j.ress.2020.107050, URL https://linkinghub.elsevier.com/retrieve/pii/S0951832020305512.

[54] Van der Maaten L, Hinton G. Visualizing data using t-SNE. In: EI, J Mach Learn Res In: EI, 2008;9.(11).

[55] Huang J, Cui L, Zhang J. Tracking the location of bearing outer raceway defects using multidimensional synchronous signal fusion and tensor rank-1 decomposition. In: EI, Measurement In: EI, 2022;198:111137. http://dx.doi.org/10.1016/j.measurement.2022.111137.URL https://linkinghub.elsevier.com/retrieve/pii/S0263224122003980,