

Reliable and Intelligent Fault Diagnosis with Evidential VGG Neural Networks

Hanting Zhou, Wenhe Chen, Longsheng Cheng, Darren Williams, Clarence W. de Silva, *Fellow, IEEE*, and Min Xia, *Senior Member, IEEE*

Abstract—With the emergence of Internet-of-Things and big data technologies, data-driven fault diagnosis approaches, notably deep learning-based methods have shown promising capabilities in achieving high accuracy through end-to-end learning. However, these deterministic neural networks cannot incorporate the prediction uncertainty, which is critical in practical applications with possible out-of-distribution data. The present paper develops a reliable and intelligent fault diagnosis framework based on evidence theory and improved VGG neural networks, which can achieve accurate and reliable diagnosis results by incorporating additional estimation of the prediction uncertainty. Specifically, the paper treats the predictions of the VGG as subjective opinions by placing a Dirichlet distribution on the category probabilities and collecting the evidence from data during the training process. A specific loss function assisted by evidence theory is adopted for the VGG to obtain improved uncertainty estimations. The proposed method, which incorporates Evidential VGG (EVGG) neural networks, as termed here, is verified by a case study of the fault diagnosis of rolling bearings, in the presence of sensing noise and sensor failure. The experimental results illustrate that the proposed method can estimate the prediction uncertainty and avoid overconfidence in fault diagnosis with out-of-distribution (OOD) data. Also, the developed approach is shown to perform robustly under various levels of noise, which indicates a high potential for use in practical applications.

Index Terms—Trustworthy AI, fault diagnosis, evidence theory, VGG neural networks, uncertainty estimation.

I. INTRODUCTION

Fault detection and diagnosis play a significant role in reducing system downtime and avoiding unexpected failure of industrial machines, in a wide range of sectors including manufacturing, construction, transportation, and healthcare. With the emergence of Internet-of-Things (IoT) and big-data technologies, data-driven approaches using

machine learning (ML) and deep learning (DL) have been increasingly implemented in fault diagnosis for various industrial systems [1]. DL-based methods have demonstrated unprecedented prediction accuracies and remarkable generalization capabilities in many pattern recognition tasks. However, the black-box nature of deep learning still limits its adoption in safety-critical applications in the engineering sector, making it difficult to trust the results [2]. Also, most of the current deep learning approaches are unable to distinguish between samples in-domain and out-of-domain, as well as the sensitivity to domain shifts [3]. Due to the lack of reliability and uncertainty estimates of artificial intelligence (AI) systems, these deep learning models can produce overconfident predictions. These main factors related to uncertainty in the deep learning process are data noise and model inference errors [4]. Thus, uncertainty is an important indicator of the model reliability of deep learning and the possible distribution shift of the data. Proper estimation of uncertainty can support the decision-making in recognizing when models are likely to fail or to what extent we can trust the diagnosis results, making the AI system more trustworthy.

Predictive uncertainty in deep neural networks (DNN) generally can result from two sources: model uncertainty and data uncertainty [5]. Model uncertainty, or epistemic uncertainty, measures the uncertainty in estimating the model parameters given the training data. It is generated by an imbalance in the training data distribution and may be reducible as the size of the training data increases [6]. Data uncertainty, or aleatoric uncertainty, is irreducible uncertainty, and it arises from the noise inherent in the observations or class overlap, usually caused by a defect in the sensor [7]. Unlike model uncertainty, data uncertainty cannot be reduced by observing more data samples due to the inherent property of the data distribution. When the training and testing data distributions are mismatched due to noise interference or new class situations, data uncertainty occurs. Accurate uncertainty predictions help to interpret the confidence levels, capture domain shifts in out-of-distribution (OOD) samples, and realize trustworthy prediction [8]. However, uncertainty estimation in practice is still challenging and to be further investigated.

In practical engineering scenarios, sensing noise and sensor failure have a significant impact on the collected signal, which will make inaccurate interferences leading to unacceptable fault diagnosis results [9]. Unfortunately, the existing studies on DL-based fault diagnosis have focused more on the accuracy of models without assessing the uncertainty of the diagnostic results. This can lead to overconfident predictive

Manuscript received 4 August 2022; accepted 3 February 2023. This work was supported in part by the China Scholarship Council (CSC) under Grant No.202106840063, the Royal Society under Grant IECNSFC211012. (Corresponding authors: Min Xia; Longsheng Cheng)

Hanting Zhou, Wenhe Chen, and Longsheng Cheng are with the School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China. (e-mail: zhouhanting@njust.edu.cn; chenwenhe@njust.edu.cn; cheng_longsheng@njust.edu.cn)

Clarence W. de Silva is with the Department of Mechanical Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. (e-mail: desilva@mech.ubc.ca)

Darren Williams and Min Xia are with the School of Engineering, Lancaster University, Lancaster LA1 4YW, U.K. (e-mail: d.williams8@lancaster.ac.uk, m.xia3@lancaster.ac.uk)

models for practical scenarios, as they make wrong decisions, largely affecting operational reliability [14]. DL algorithms can only be fully integrated for intelligent fault diagnosis if prediction uncertainty is incorporated. Quantifying uncertainty in a reliable manner is a key feature of trustworthy AI used in IFD to avoid false trust and its serious consequences, including increased safety risks, system downtime, and maintenance costs [10]. Reliable estimation of uncertainty will provide valuable and reliable decisions in the decision-making process.

Traditional uncertainty estimation methods estimate uncertainty by sampling methods, such as dropout [11]. Bayesian-based methods provide a theoretical framework for uncertainty estimation by casting dropout training in conventional deep networks as a Bayesian approximation of a Gaussian Process [12]. Then the uncertainty estimates can be obtained by calculating the variance of multiple predictions for different dropout masks. It can be applied to any pre-trained network using a dropout layer in achieving a lower estimation risk [13]. However, it still has some problems, including the intractability of directly inferring the posterior distribution of the weights given the data and how to choose a weight prior [4]. Even though sampling techniques can be used to better assess model uncertainty, data uncertainty has not been explicitly modeled yet with overconfident predictions [4]. In addition, sampling methods usually ignore the relationship between data and model uncertainty which increases the risk of underestimating the uncertainty.

To address these problems, researchers have attempted to quantify data uncertainty by modeling network activation and weights through parameter probability distributions. Recently, approaches of Bayesian Neural Nets (BNNs) have been proposed to estimate the prediction uncertainty by approximating the moments of the posterior forecast distribution [13]. However, its posterior predictive distribution cannot be computed in closed form and relies heavily on Monte Carlo integration to approximate the posterior predictive density. Thus, another direction for predicting data uncertainty has been explored, such as evidential deep learning and lightweight probabilistic deep networks [14, 15]. These methods can assess data uncertainty in DNN by directly estimating the predictive posterior parameters as their output, and this effectively reduces the risk of underestimation of uncertainty. Also, they require no sampling and minimal changes to the standard neural network structures [10]. This approach is promising for data uncertainty estimation in practical settings, which can provide robust and trustworthy prediction results. It has been used in a robotic system, leading to good performance. However, to our knowledge, it has not been utilized in practical fault diagnosis with data uncertainty.

The present paper develops a novel and original framework for trustworthy fault diagnosis with uncertainty estimation of deep neural network predictions. Inspired by the very deep convolutional neural network, which was originally proposed by the Visual Geometry Group (VGG network) [16], this paper improves VGG architecture by adding the Batch

Normalization (BN) layers. Then, incorporating the theory of evidence, our improved VGG-architecture network realizes the reliable fault diagnosis with uncertainty estimation, termed evidential VGG (EVGG). Also, this paper adopts an evidential loss function with Kullback-Leibler (KL) divergence as a penalty term to generate more evidence for the correct labels and decrease the misleading evidence from misclassified samples. The improved method can estimate the uncertainty of the prediction which is vital for a trustworthy AI system and can detect possible OOD data that includes sensing noise and sensor failure. The performance of the proposed EVGG is verified through two case studies in different practice scenarios. The main contributions of the paper are as follows:

1) A novel framework of intelligent fault diagnosis that incorporates uncertainty estimation is proposed. It can capture prediction uncertainties of DL-based approaches for general neural network structures. This framework requires minimal changes in the learning process, which has significant value in practical applications.

2) Based on evidence theory, the proposed method achieves uncertainty estimation for the prediction results. The neural network structure replaces the *softmax* parameter sets with the parameters of a Dirichlet density and then achieves the predictions as a distribution over possible *softmax*. In addition, the weights of the standard backpropagation neural network are optimized by the specific evidential loss function with KL divergence as the penalty terms.

3) By developing an improved VGG model with batch normalization added after each convolutional layer can prevent the problem of gradient disappearance and reduce the model uncertainty. Integration of evidence theory, the present work achieves end-to-end trustworthy and intelligent fault diagnosis. The proposed method achieves reasonable uncertainty estimation with high diagnostic performance under various levels of sensing noise and sensor failure. The evidence for trust in the predicted results of the DL models leads to trustworthy predictions, with the capability of detecting OOD observations.

The remainder of the present paper is organized as follows. Section II introduces the evidence theory. Section III presents the details of the proposed method of EVGG with evidential loss function. The details of experimental studies are presented in Section IV. Section V gives the conclusion and possible future work.

II. THE PRINCIPLE OF EVIDENCE THEORY

Dempster-Shafer evidence theory (DST) is a generalization of the Bayesian theory to include subjective probabilities [17]. There exists a discriminative framework that represents the set of unique possible states. On assigning belief mass to subsets of the discriminative framework, the belief truth can be any of the possible states. Then, subjective logic (SL) formalizes DST's notion of belief distribution over a discernment framework as a Dirichlet distribution [18]. This theoretical framework can quantify belief mass and uncertainty based on the principles of evidence theory. For each independent class

m from the M classes, SL provides a belief mass a_m and uncertainty w , satisfying the following equation:

$$w + \sum_{m=1}^M a_m = 1 \quad (1)$$

where $w \in [0,1]$, $a_m \geq 0$. The belief mass a_m and uncertainty w can be computed as

$$a_m = v_m / V, w = M / V \quad (2)$$

where $V = \sum_{i=1}^M (v_i + 1)$ is the total evidence, v_m means the evidence of the m^{th} class. Eq. (1) and Eq. (2) indicate that evidence and belief mass are positively correlated, while uncertainty is inversely proportional to them. For example, when there is no evidence to support a result, its belief mass is zero and the uncertainty value is one. The Dirichlet distribution parameter β_m can be calculated using:

$$\beta_m = v_m + 1 \quad (3)$$

Then the belief mass a_m can be derived from the parameter β_m , as follows:

$$a_m = (\beta_m - 1) / V \quad (4)$$

where the total evidence V refers to the Dirichlet strength, with the corresponding Dirichlet distribution given by:

$$V = \sum_{i=1}^M \beta_i \quad (5)$$

Standard neural network classifiers will output a probability assignment of possible classes for each sample. The Dirichlet distribution $D(\cdot)$ represents the probability density function for possible values of the probability mass function. It can be characterized by the Dirichlet distribution parameter $\beta = [\beta_1, \dots, \beta_M]$ as follows:

$$D(\mathbf{p} | \beta) = \begin{cases} \frac{1}{B(\beta)} \prod_{i=1}^M p_i^{\beta_i-1} & \text{for } \mathbf{p} \in V_M, \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

in which \mathbf{p} is a simplex representing class assignment probabilities, V_M is the K -dimensional unit simplex, given by

$$V_M = \left\{ \mathbf{p} \mid \sum_{i=1}^M p_i = 1, 0 \leq p_i, \dots, p_M \leq 1 \right\} \quad (7)$$

Given an opinion, the expected probability of the m^{th} class can be computed as follows:

$$\hat{p}_m = \beta_m / V \quad (8)$$

where \hat{p}_m corresponds to the mean of a Dirichlet distribution. The deep neural networks first estimate evidence v_m over each class. Then, the belief mass a_m is obtained which can be used to calculate the uncertainty w of a sample by Eq. (1). When the observed sample is biased toward a specific m^{th} class, the corresponding evidence v_m will update the Dirichlet parameter β_m in the Dirichlet distribution, which generates class probabilities \hat{p}_m by Eq. (8).

III. THE PROPOSED METHOD

This paper develops the EVGG architecture as the neural network, which can output classification results with the corresponding prediction uncertainty, by integrating evidence theory and the improved VGG-style architecture with the batch normalization layer. The EVGG model uses a specific evidential loss function which can be minimized subject to neural net weights using standard backpropagation. This evidential loss function can support the generation of more evidence for the correct classification samples as well as the removal of misleading evidence from misclassification samples, leading to more accurate and credible predictions.

A. The Evidential Loss Function with Evidence Theory

Given a sample i , x_i is the observation and y_i is the class label that is one-hot encoding. When the sample i belongs to the m^{th} class, y_{im} equals one otherwise to zero. The present paper adopts the evidential loss function using the technique of the maximum likelihood, treating as Dirichlet distribution a prior on the likelihood, then resulting in

$$\begin{aligned} \mathcal{L}_i(\Theta) &= -\log\left(\int \prod_{m=1}^M p_{im}^{y_{im}} \frac{1}{B(\beta_i)} \prod_{m=1}^M p_{im}^{\beta_{im}-1} d\mathbf{p}_i\right) \\ &= -\sum_{m=1}^M \log\left(\int p_{im}^{y_{im}} \frac{1}{B(\beta_i)} \prod_{m=1}^M p_{im}^{\beta_{im}-1} d\mathbf{p}_i\right) \\ &= -\sum_{m=1}^M y_{im} \log(\hat{p}_m) \\ &= -\sum_{m=1}^M y_{im} \log\left(\frac{\beta_{im}}{V_i}\right) \\ &= \sum_{m=1}^M y_{im} (\log(V_i) - \log(\beta_{im})) \end{aligned} \quad (9)$$

where Θ represents the network parameters and minimize $\mathcal{L}_i(\Theta)$ concerning the β_i parameters. To improve the loss function, a KL divergence has been added as a penalty term to regularize the predictive distribution [14, 19]. It can improve the assigning accuracy of evidence to correctly classified samples, decreasing the misleading evidence for misclassified samples, given as follows:

$$\begin{aligned} KL[D(\mathbf{p}_i | \tilde{\beta}_i) \| D(\mathbf{p}_i | \mathbf{I})] &= \log\left(\frac{\Gamma(\sum_{m=1}^M \tilde{\beta}_{im})}{\Gamma(M) \prod_{m=1}^M \Gamma(\tilde{\beta}_{im})}\right) \\ &+ \sum_{m=1}^M (\tilde{\beta}_{im} - 1) [\psi(\tilde{\beta}_{im}) - \psi(\sum_{m=1}^M \tilde{\beta}_{im})] \end{aligned} \quad (10)$$

where $D(\cdot)$ is the Dirichlet distribution. \mathbf{p}_i is a simplex representing class assignment probabilities. \mathbf{I} is the parameter vector of M ones, which can increase the effectiveness of the KL divergence. $\Gamma(\cdot)$ represents the gamma function, and $\psi(\cdot)$ is the digamma function. $\tilde{\beta}_i$ represents the Dirichlet parameter after removing the non-misleading evidence from the predicted parameter, given by:

$$\tilde{\beta}_i = \mathbf{y}_i + (1 - \mathbf{y}_i) \odot \beta_i \quad (11)$$

where \odot represents the Hadamard product. The evidence loss with the regularizing term can be calculated as

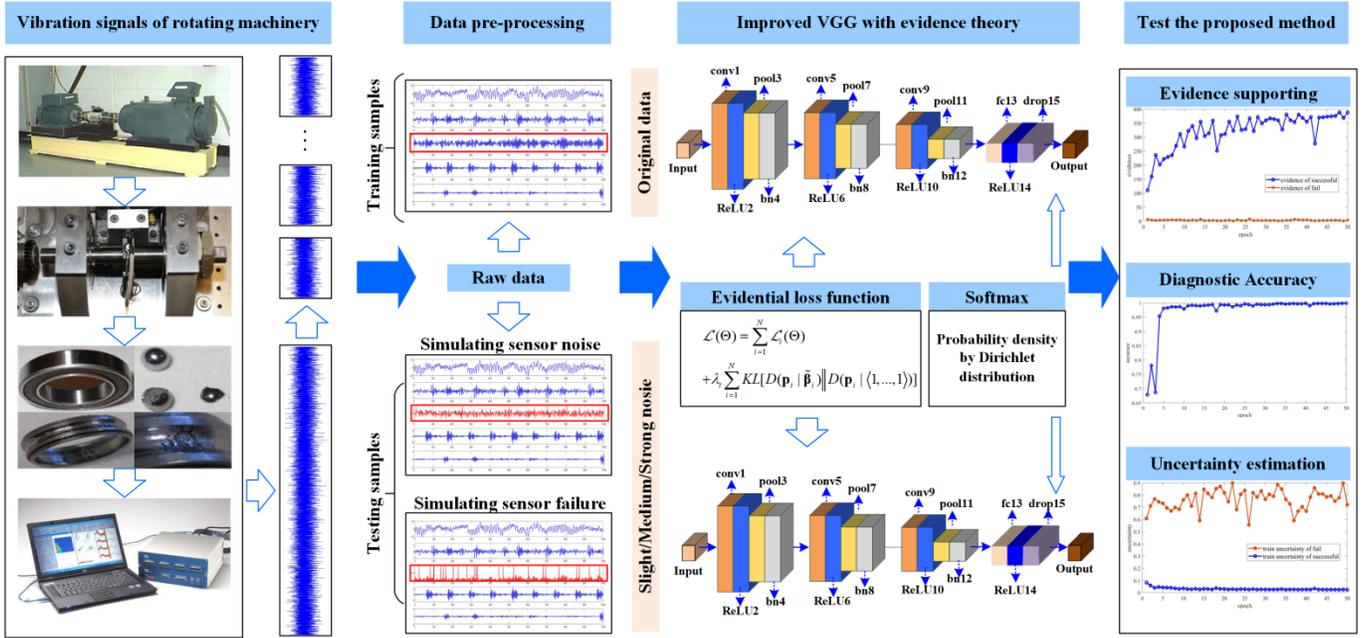


Fig. 1. The framework of the proposed EVGG-based reliable IFD with uncertainty estimation.

$$\mathcal{L}(\Theta) = \sum_{i=1}^N \mathcal{L}_i(\Theta) + \lambda_t \sum_{i=1}^N KL[D(\mathbf{p}_i | \tilde{\beta}_i) || D(\mathbf{p}_i | \langle 1, \dots, 1 \rangle)] \quad (12)$$

where $\lambda_t = \min(1, t/10) \in [0, 1]$ represents the annealing coefficient, t is an index for the current training epoch, and $D(\mathbf{p}_i | \langle 1, \dots, 1 \rangle)$ is the Dirichlet uniform distribution.

B. EVGG-based IFD with Uncertainty Estimation

The overall framework of the proposed trustworthy intelligent fault diagnosis approach by EVGG with uncertainty estimation is shown in **Fig. 1**. The original sensory data is used as the training set. After the model is trained and implemented, monitoring data with possible OOD are input into the model. Both sensor noise interference and sensor failure scenarios are considered to represent a typical practical situation. The proposed method adopts the specific evidential loss function and constructs the *softmax* probability with Dirichlet density. The EVGG model can be used for effective fault diagnosis with high accuracy. Meanwhile, it provides an estimation of the predictive uncertainty, which can indicate whether the prediction is trustworthy or not. Large uncertainty can be associated with possible OOD of the online monitoring data, in which case further inspection and analysis can be triggered. The entailed procedure is summarized as follows:

Step 1: Sense and collect vibration data of the monitored machine with different conditions. Preprocess the data and construct the two-dimensional feature maps of the samples to form the training set.

Step 2: Construct EVGG by integrating evidence theory and the improved VGG-like model. Adopt the evidential loss function with KL divergence as a penalty term. Calculate the density of each class probability by the Dirichlet distribution parametrized over evidence.

Step 3: Train the EVGG model with the training set. The EVGG model is shown to have excellent diagnostic accuracy

and a significant difference in the evidence for correct and incorrect classification results.

Step 4: Deploy the EVGG model for online fault diagnosis. The model will be able to detect OOD data that may include sensing noise, sensor failure, or new fault types. In the present study, sensing noise and sensor failure are considered by adding specific noise to the original data.

Step 5: If large predictive uncertainty is observed, further inspection and analysis of the machine will be triggered.

IV. DATA DESCRIPTION AND PRE-PROCESSING

In the present paper, the effectiveness of the proposed trustworthy intelligent fault diagnosis approach using EVGG is evaluated through the fault diagnosis of rolling bearings.

A. Experimental Setup and Data Description

(1) CASE 1 CWRU Dataset

In this experiment, the roller bearing dataset collected from a motor drive system by Case Western Reserve University (CWRU) is used [20]. The test stand is shown in **Fig. 2**. The vibration signals from CWRU include the drive-side bearing data at a sampling frequency of 12 kHz under a 1hp load (1772rpm). The monitored conditions included one normal condition and nine faulty conditions. Three types of faults were included, specifically, inner race fault (IRF), ball fault (BF), and outer race fault (ORF). Each failure type had three severity levels including 0.007, 0.014, and 0.021 inches that corresponded to slight, moderate, and severe faults respectively, as given in **TABLE I**.

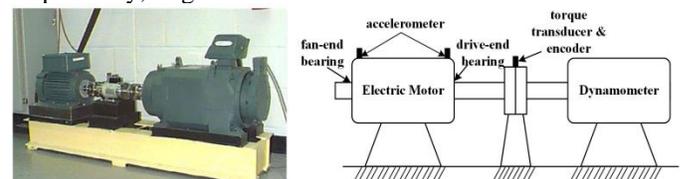


Fig. 2. Tested rolling bearing of CWRU.

TABLE I
DESCRIPTION OF CWRU DATASETS IN THE PRESENT PAPER

Fault level	Fault type	Load (hp)	Speed (rpm)	Samples	Class Label
normal	/	1	1772	1000	C1
slight fault (0.007")	IRF	1	1772	1000	C2
	BF	1	1772	1000	C3
	ORF	1	1772	1000	C4
medium fault (0.014")	IRF	1	1772	1000	C5
	BF	1	1772	1000	C6
	ORF	1	1772	1000	C7
severe fault (0.021")	IRF	1	1772	1000	C8
	BF	1	1772 </td <td>1000</td> <td>C9</td>	1000	C9
	ORF	1	1772	1000	C10

(2) CASE 2 XJTU-SY Dataset

The roller bearing dataset collected by Xi'an Jiaotong University and the Changxing Sumyoung Technology Co. (XJTU-SY) is used [21]. The experiment setup is shown in Fig.3. The vibration signals at a sampling frequency of 25.6 kHz were collected. Under three working conditions, the faults include inner race fault (IRF), cage fault (CF), outer race fault (ORF), inner race and outer race fault (IORF), and mix fault (MF, including inner race, ball, cage, and outer race fault), details of which are outlined in TABLE II. Since the XJTU-SY dataset contains the run-to-failure data, only the data in the faulty states of each fault type are selected as the samples.

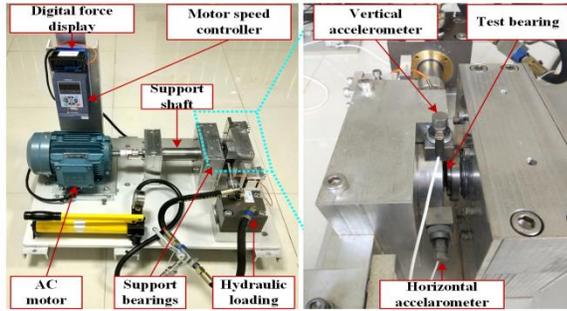


Fig. 3. Tested rolling bearing of XJTU-SY.

TABLE II
DESCRIPTION OF XJTU-SY DATASETS IN THE PRESENT PAPER

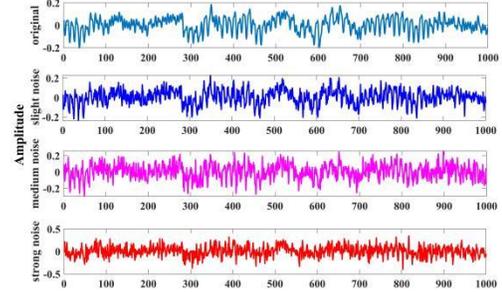
Fault type	Bearing	Operating	Fault level	Samples	Class Label
normal	Bearing 2_3	Condition 2	severe	1000	C1
ORF	Bearing 2_5	Condition 2	severe	1000	C2
IRF	Bearing 2_1	Condition 2	severe	1000	C3
CF	Bearing 2_3	Condition 2	severe	1000	C4
IORF	Bearing 1_5	Condition 1	severe	1000	C5
MF	Bearing 3_2	Condition 3	severe	1000	C6

B. Sensing Noise and Sensor Failure on the Testing Dataset

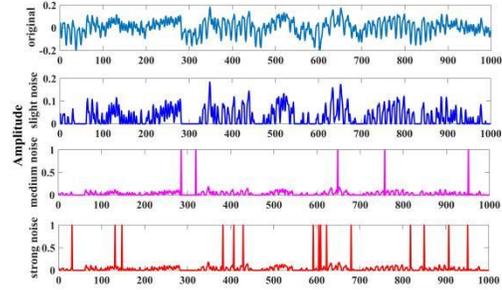
Both sensing noise and sensor failure are common issues in practice and could impact the prediction reliability using traditional DL-based fault diagnosis methods. The present study considers both sensor noise and sensor failure, in evaluating the uncertainty estimation capability of the proposed EVGG method. First, background noise during the operation of the equipment and communication noise through the transmission channels can be presented by additive white Gaussian noise (AWGN) [22]. Second, sensor failures may have various patterns. In the model, salt-and-pepper (SAP) noise of different levels is added to represent typical sensor

failures. SAP is the common type of impulse noise and takes maximum and minimum values in the dynamic range [23].

To simulate real conditions, various signal-to-noise ratios (SNR) are added to the test data samples. Based on the literature, the SNR is set to 4dB, 0dB, and -4dB for slight, medium, and strong noise conditions, respectively [18]. And the intensity of the SAP noise (d) is chosen to be 0, 0.005, and 0.02, respectively. The original dataset is subjected to different levels of additive noise using the *awgn* and *imnoise* functions in MATLAB 2020a. Subplot a) of Fig. 4 presents the different Gaussian noise disturbances in C1 of CWRU, while subplot b) of Fig. 4 plots the different levels of SAP noise disturbance in the CWRU normal conditions.



a). Different Gaussian noise disturbances in C1 of CWRU.



b). Different SPA noise disturbances in C1 of CWRU.

Fig. 4. Different types and levels of noise disturbance.

C. Data Conversion and Sampling

Given its excellent feature extraction capability, the improved VGG model is used in the present paper as the basic DL structure. An image transformation method is adopted to convert the vibration signal to two-dimensional feature maps [24]. In the data conversion process, this paper uses a feature map of size 784, where the vibration signal of 784 data points is converted to a 28*28 two-dimensional feature map, details given in Fig. 5.

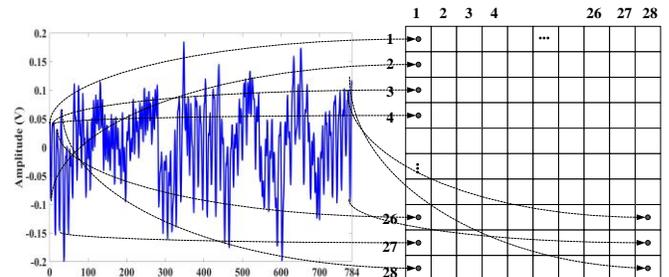


Fig. 5. Schematic diagram of data conversion.

To increase the sample size, the overlap sliding segmentation method [25] is used to expand the number of training and testing samples. Every segmentation signal

includes 784 sample points and the length of sample overlap for two neighbor segments is 684 sample points. After the segmentation, for each class, 1000 samples are obtained. Then dataset is split randomly into subsets of training data (70%) and testing data (30%). The performance of the fault diagnosis is quantified using the following four standard measures: accuracy, precision, recall, and F1 score which are widely used in the literature:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (16)$$

where TP, FP, FN, and TN represent the number of true positive, false positive, false negative, and true negative outcomes, respectively.

V. CASE VALIDATION

In this study, an improved EVGG model with batch normalization layers is composed and the hyperparameters are given in **TABLE III**. In the experiments, the Adam optimizer is used as the default setting for training.

TABLE III

PARAMETER SETTINGS OF THE EVGG MODEL

Model	Parameters	Number or Type
EVGG	input	28*28*1
	convolution layers 1	Kernel 5*5*1*20, stride [1 1]
	activation layers 2	ReLU
	pooling layers 3	Maximum pooling [2 2]
	batch_normalization 4	Batch_normalization
	convolution layers 5	Kernel 5*5*20*40, stride [1 1]
	activation layers 6	ReLU
	pooling layers 7	Maximum pooling [2 2]
	batch_normalization 8	Batch_normalization
	convolution layers 9	Kernel 5*5*40*60, stride [1 1]
	activation layers 10	ReLU
	pooling layers 11	Maximum pooling [2 2]
	batch_normalization 12	Batch_normalization
	full-connected layers 13	300 fully connected layer
	activation layers 14	ReLU
dropout 15	50% dropout	
output	10 classes	

CASE 1 CWRU Dataset

A. Diagnosis Performance on Original Data

The original vibration data for the ten conditions without noise are used first to evaluate the proposed EVGG. The diagnostic performance and uncertainty estimation ability are presented in **Fig. 6**. With the increase of the training epoch, the training accuracy of the EVGG model for original data gradually increases and finally stabilizes at 100% as shown in subplot (b) of **Fig.6**. Correspondingly, the diagnostic accuracy for the testing data set is maintained at 99.37% as shown in subplot (d). This indicates that the EVGG model has satisfactory diagnostic performance. For the uncertainty estimation, as shown in subplots (a) and (c), the estimated total evidence of the correct classifications is high and the

evidence of misclassifications is low, which represents the actual situation. The estimated uncertainty should be opposite to the estimated evidence. Therefore, the uncertainty for correctly classified samples is low and that for misclassified samples is high. The experimental result on the original data shows that the proposed EVGG can perform accurate fault diagnosis and provide valid uncertainty estimation.

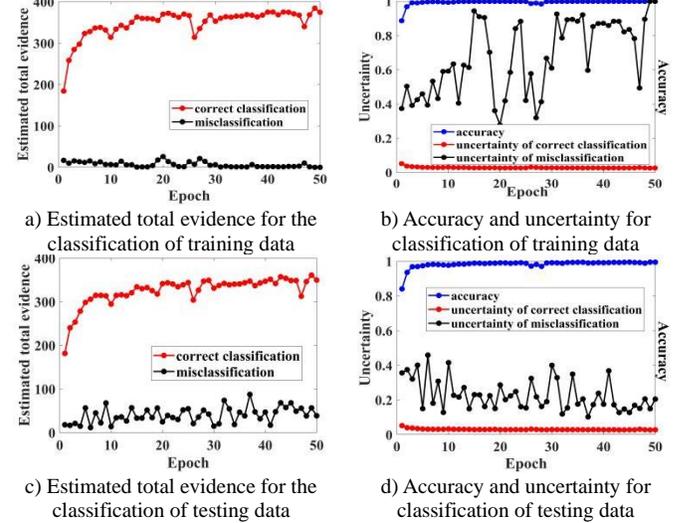


Fig. 6. Training and testing process diagram of the original data from CWRU.

B. Diagnosis Performance on Data with Sensing Noise

To verify the effectiveness of the proposed EVGG on fault diagnosis with sensing noise, Gaussian white noise is added to the testing samples of the original vibration data. For each experiment, only the testing samples of one class are corrupted by noise. **TABLE IV** presents the experimental results of the medium ball fault (C6) under the medium noise disturbance. The diagnosis accuracy of C6 is 95.13% is still high, but slightly lower than for the other classes without noise. As the only class with medium noise interference, the developed EVGG predicts a large degree of uncertainty (61.64%) in the diagnostic results of C6. This new indicator of uncertainty tells the degree of the untrustworthy diagnostic results from the DL model. It indicates possible OOD data observation, and further investigation should be triggered. This is of significant value for practical applications.

TABLE IV
FAULT DIAGNOSIS AND UNCERTAINTY ESTIMATION RESULTS IN MEDIUM SENSOR NOISE (C6 CORRUPTED)

Class	Accuracy	Recall	Precision	F1 score	Uncertainty
C1	98.97%	100.00%	90.63%	95.09%	2.90%
C2	99.63%	100.00%	96.46%	98.20%	2.74%
C3	99.60%	99.33%	96.75%	98.03%	7.33%
C4	99.97%	100.00%	99.67%	99.83%	4.06%
C5	97.03%	100.00%	77.12%	87.08%	6.02%
C6	95.13%	51.33%	100.00%	67.84%	61.64%
C7	99.47%	100.00%	94.94%	97.40%	3.51%
C8	99.97%	100.00%	99.67%	99.83%	3.16%
C9	99.47%	95.33%	99.31%	97.28%	8.83%
C10	99.97%	100.00%	99.67%	99.83%	3.50%

To present the relationship between the classification probabilities and uncertainty estimations, ten samples are randomly selected from the 300 testing samples of the C6 under the medium noise disturbance, as shown in **Fig. 7**. In the plot, the different colored bars represent the classification probabilities of 10 classes for each sample. The red dashed

line is the uncertainty value. From the results, sample 1 and sample 5, which are incorrectly classified, have high uncertainty values of 78.19% and 70.95%, respectively. For the other 8 samples that are correctly classified, however, the mean value of uncertainty still reaches 31.12%. The main reason for this is that although samples 6 and samples 10 are properly classified, similar classification probabilities exist which indicates considerable uncertainty. For other correctly classified samples, sample 2 as an example, has a 0.9486 classification probability on C6 with a low uncertainty value of 5.52%. Observations suggest that when the diagnostic model has a hesitation in its output, as reflected in the case of similar classification probabilities, the results are low in evidence and high in uncertainty. The proposed method accurately distinguishes the unreliable classified samples from the uncertainty estimations when the trained model is implemented. It will give evidence supports to identifying misclassified samples and provide overconfidence alerts for correctly classified samples.

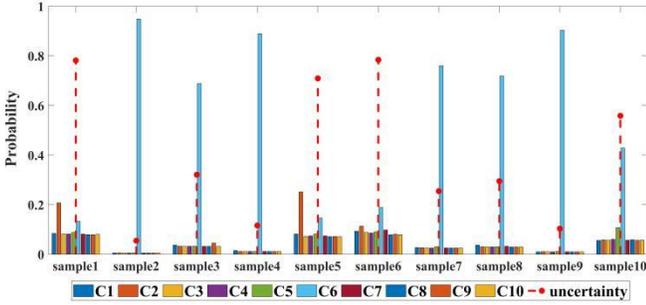


Fig. 7. Samples of diagnostic probability and uncertainty estimation of C6 under medium sensor noise

TABLE V

FAULT DIAGNOSIS PERFORMANCE OF 10 CLASSES UNDER SENSOR NOISE

Class	Accuracy			Uncertainty		
	Slight	Medium	Strong	Slight	Medium	Strong
C1	99.97%	99.77%	94.17%	12.33%	24.97%	67.46%
C2	99.77%	99.00%	94.37%	18.70%	45.79%	61.81%
C3	94.77%	89.97%	89.57%	58.84%	67.82%	70.89%
C4	95.23%	90.50%	90.00%	61.66%	63.30%	69.39%
C5	98.13%	97.23%	92.57%	36.02%	49.35%	63.65%
C6	99.40%	95.13%	91.53%	20.06%	61.64%	70.78%
C7	99.97%	99.80%	97.00%	12.96%	22.19%	56.16%
C8	97.37%	90.67%	89.93%	50.73%	64.37%	65.85%
C9	98.20%	94.23%	90.03%	38.79%	65.58%	67.62%
C10	98.10%	96.57%	93.60%	39.27%	59.90%	70.50%
Ave.	98.09%	95.29%	92.28%	34.94%	52.49%	66.41%

Testing results with three levels of Gaussian white noise are presented in **Table V**. The developed EVGG can achieve high diagnosis performance under noisy conditions. Under different noise levels, it obtains diagnosis accuracies of 98.09%, 95.29%, and 92.28%. And the corresponding mean uncertainty estimation values are 34.94%, 52.49%, and 66.41%, respectively. The prediction accuracy and uncertainty are consistent with the degree of disturbance. Overall, the proposed method achieves resistance to sensing noise interference with good performance and can provide predictive uncertainty estimation for trustworthy diagnosis.

C. Diagnosis Performance on Data with Sensor Failure

To verify the feasibility of the proposed method for sensor failure scenarios, testing samples corrupted with SAP noise are used. The experiment result of medium ball fault (C6)

under medium noise disturbance is shown in **TABLE VI**. It achieves satisfactory diagnosis performance for all the classes including C6 under medium noise conditions. But the uncertainty of the prediction for C6 is high which indicates the possible OOD of the observations. This demonstrates the effectiveness of the proposed EVGG in providing a valid estimation of the prediction uncertainty.

TABLE VI

FAULT DIAGNOSIS AND UNCERTAINTY ESTIMATION RESULTS IN MEDIUM SENSOR FAILURE (C6 CORRUPTED)

Class	Accuracy	Recall	Precision	F1score	Uncertainty
C1	95.63%	100.00%	69.61%	82.08%	3.26%
C2	100.00%	100.00%	100.00%	100.00%	3.04%
C3	99.67%	100.00%	96.77%	98.36%	4.86%
C4	99.97%	100.00%	99.67%	99.83%	3.36%
C5	98.97%	100.00%	90.63%	95.09%	5.32%
C6	94.50%	45.33%	99.27%	62.24%	55.82%
C7	100.00%	100.00%	100.00%	100.00%	4.57%
C8	100.00%	100.00%	100.00%	100.00%	3.74%
C9	99.33%	95.00%	98.28%	96.61%	8.61%
C10	100.00%	100.00%	100.00%	100.00%	5.44%

Here, 10 samples are randomly chosen from C6 to illustrate the difference between the correct and incorrect classifications, as shown in **Fig. 8**. The classification probability values for each class are represented by different colored bars and the corresponding uncertainty value for each sample is presented by the red dashed line. For the misclassified samples with similar output probabilities, including sample 4 and sample 5, the uncertainty values were 73.80% and 80.33%, respectively. Furthermore, the misclassifications of C6 are mostly concentrated in C1, the normal class, which will lead to severe system failure and downtime by ignoring the actual fault that has happened.

It can be seen that if the classification probabilities appear similar, the classifier has a greater likelihood of misclassifying the samples, thereby raising the level of uncertainty in the diagnosis. This situation can be also seen in samples 9 and 10, where the uncertainty is as high as 66.57% and 51.57% despite being correctly classified. For the other correct classified samples, sample 1 as the example, the probability of classifying C6 is 0.9416, and the sum of the probabilities for other categories is 0.0584, with an uncertainty of just 6.16%. The proposed method can distinguish the misclassified samples and alert the prediction with high uncertainty.

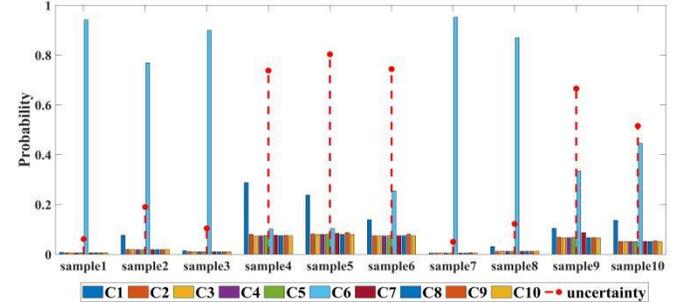


Fig.8. Samples of diagnostic probability and uncertainty estimation of C6 under medium sensor failure

Table VII presents the diagnosis accuracy and uncertainty estimation of 10 classes under variable SAP noise. The proposed EVGG obtains high diagnosis accuracy (mean values of 97.36%, 95.39%, and 93.29%) and provides a reliability indicator (mean values of 35.28%, 50.17%, and 63.82%), which are consistent with noise levels. The estimated

uncertainty values are high for misclassified samples, which can provide useful information in practice. The proposed method has been able to successfully warn of the possible untrustworthiness of the current prediction.

TABLE VII

FAULT DIAGNOSIS PERFORMANCE OF 10 CLASSES UNDER SAP NOISE

Class	Accuracy			Uncertainty		
	Slight	Medium	Strong	Slight	Medium	Strong
C1	100%	99.90%	97.83%	7.17%	12.69%	57.97%
C2	99.43%	98.60%	97.93%	39.93%	41.64%	51.35%
C3	89.87%	89.73%	89.70%	23.24%	53.46%	69.00%
C4	96.57%	90.97%	90.03%	57.42%	60.77%	62.87%
C5	97.60%	97.00%	92.17%	46.67%	58.21%	69.38%
C6	98.77%	94.50%	93.10%	23.15%	55.82%	69.32%
C7	99.67%	98.47%	97.20%	18.51%	38.36%	54.87%
C8	95.03%	91.67%	90.07%	63.55%	68.48%	70.50%
C9	99.53%	97.33%	92.00%	17.83%	45.18%	62.51%
C10	97.10%	95.70%	92.83%	56.35%	67.13%	70.47%
Ave.	97.36%	95.39%	93.29%	35.38%	50.17%	63.82%

D Comparison with The Different Noisy Conditions

a) Trustworthy diagnosis performance under sensor noise and sensor failure conditions.

Unlike other studies, the training data used in this paper are original data, while only data samples in the test set are corrupted by noise. This is a setting closer to practical application situations. The proposed method is shown to achieve high diagnosis performance together with valid uncertainty estimation values. The testing accuracy under sensor noise and sensor failure are satisfactory within 50 epochs. The average testing accuracy of the seven different noise disturbances is 95.94%, while for the original data it is over 99.90%. In the strong noise scenario, the sensor noise and sensor failure can maintain an accuracy of 92.28% and 93.29% respectively. **Fig. 9** shows diagnostic performance with different noise disturbances. It demonstrates the superior anti-noise performance of the proposed method.

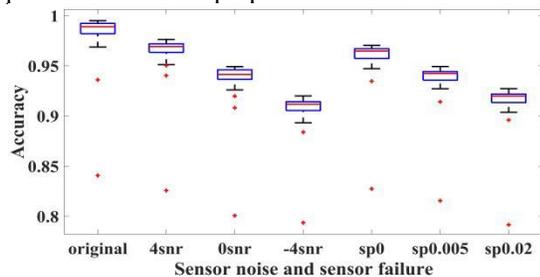


Fig. 9. Model diagnostic accuracy under different noise disturbances.

Uncertainty estimation offers trustworthy diagnosis results. **Fig. 10** presents the effectiveness of the correct classification and misclassification in the test process, where the red line indicates misclassification and the black line indicates correct classification. There are clear margins between the results of the two categories under all seven noise conditions. It effectively reduces the risk of overconfidence in misclassification and improves the accuracy of further decision-making in maintenance actions in practical application. Moreover, the subplots show significant differences between the distribution of the test data set (with added noise) and the training data set (raw data), identifying the former as belonging to OOD samples.

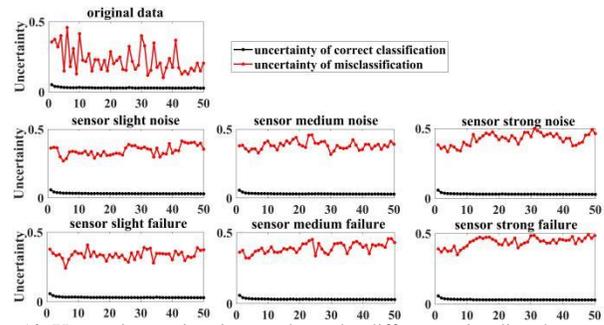


Fig. 10. Uncertainty estimation results under different noise disturbances.

b) Comparison with other methods for fault diagnosis performance under noise conditions

To verify the diagnosis performance of the proposed method, several state-of-the-art models are compared on the CWRU data set. The fault diagnosis accuracy and uncertainty estimation value under sensor noise are presented in **TABLE VIII**, where the uncertainty values are given in parentheses. Overall, the table indicates that the proposed method performs well comparably with other advanced approaches, with good noise immunity and robustness. With the original data, the proposed method achieves 99.90% accuracy, better than most advanced methods. Under strong noise conditions, while the proposed method is slightly less accurate than the methods of LDR-CNN [26] and AAnNet [27], it achieves a better performance than other compared methods. However, in this paper, variable noise interference is added only to a specific class of testing dataset, which is a more difficult task and closer to the real scenario.

Many state-of-the-art classification methods have been reported in rolling bearing fault diagnosis, however, limited work can be found on the uncertainty estimation under noise interference. The proposed method is more concerned with the reliability of the diagnostic results and provides uncertainty indicators to prevent overconfident predictions. In safety-critical applications, the proposed approach with more trustworthy prediction is advantageous.

TABLE VIII
COMPARISON RESULTS UNDER GAUSSIAN NOISE INTERFERENCE

Method	Original Data	Noise interference level		
		Slight	Medium	Strong
Proposed method	99.90% (4.30%)	98.09% (34.94%)	95.29% (52.49%)	92.28% (66.41%)
Ensemble TICNN[18]	99.50%	99.61%	98.22%	82.05%
LDR-CNN [26]	100%	99.82%	99.45%	98.88%
Deep Transfer Learning [28]	97.95%	96.72%	94.68%	90.32%
AAnNet [27]	99.48%	99.83%	99.83%	98.28%
MCNN-LSTM [29]	98.46%	88.19%	81.41%	77.27%

CASE 2 XJTU-SY Data

A. Diagnosis Performance on Original Data

Fig. 11 presents the effectiveness of reliable diagnostic performance in the training and testing process for the original data using the proposed method. The differences in evidence and uncertainty estimation between correct classification and misclassification are evident.

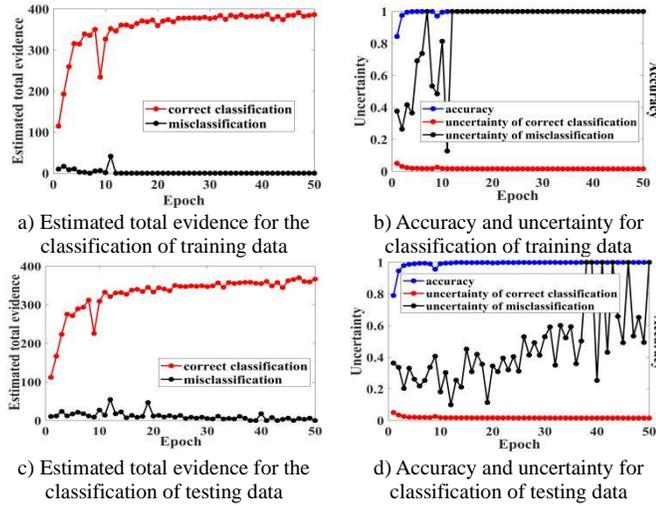


Fig. 11. Training and testing diagram of the original data from XJTU-SY.

B. Diagnosis Performance on Data with Sensing Noise

TABLE IX presents the diagnosis results of the inner race fault (C3) under the medium noise disturbance. The diagnosis accuracy of C3 reaches 96.67%, which is slightly lower than the other classes (without sensing noise). In addition, the uncertainty estimation value of C3 is 30.86%, significantly higher than other classes which demonstrates the effectiveness of the proposed method in estimating uncertainty.

TABLE IX

FAULT DIAGNOSIS AND UNCERTAINTY ESTIMATION RESULTS IN MEDIUM SENSOR NOISE (C3 CORRUPTED)

Class	Accuracy	Recall	Precision	F1score	Uncertainty
C1	97.33%	99.67%	86.42%	92.57%	3.02%
C2	100%	100%	100%	100%	2.15%
C3	96.67%	80.33%	99.59%	88.93%	30.86%
C4	100.00%	100%	100%	100%	2.05%
C5	99.61%	100%	97.72%	98.85%	1.61%
C6	99.72%	100%	98.36%	99.17%	2.14%

Ten samples are randomly selected from the 300 testing samples of the C3 under the medium noise disturbance, as shown in Fig. 12. The misclassified samples, including sample 3 and sample 8, have a high uncertainty value of 73.20% and 75.36%, respectively. For the other correctly classified samples, the average value of uncertainty is 34.95% due to the similar probability assigning among classes, such as sample 4 and sample 5. In the case of sample 2 and sample 9, the uncertainty estimation value is low since the samples have a clear probabilistic assigning for the correct class.

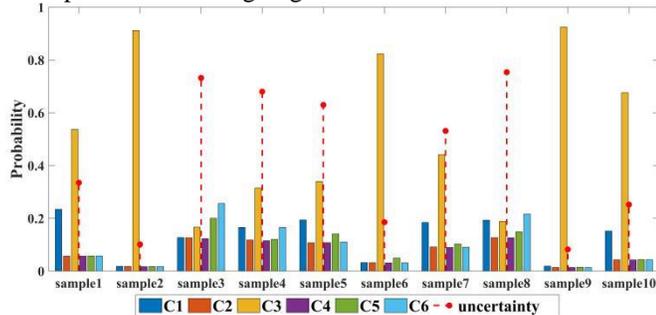


Fig. 12. Samples of diagnostic probability and uncertainty estimation of C3 under medium sensor noise.

The testing results under different sensing noise conditions are presented in Table X. With stronger noise interference, the average diagnostic accuracy reaches 99.06%, 96.17%, and

92.88%, respectively. Correspondingly, the average uncertain estimation values are 13.85%, 25.15%, and 36.83%, respectively. In summary, the proposed method achieves a good resistance to noise interference, even in a strong noise environment. In addition, the proposed method can provide credible diagnostic results with additional uncertainty estimation.

TABLE X

FAULT DIAGNOSIS PERFORMANCE OF 6 CLASSES UNDER SENSOR NOISE

Class	Accuracy			Uncertainty		
	Slight	Medium	Strong	Slight	Medium	Strong
C1	100.00%	99.83%	97.72%	4.11%	9.54%	22.13%
C2	97.67%	92.83%	86.67%	25.19%	47.06%	52.89%
C3	99.33%	96.67%	90.22%	14.31%	30.86%	46.49%
C4	97.67%	89.61%	86.56%	25.46%	36.23%	52.39%
C5	99.78%	99.56%	97.89%	6.27%	14.63%	26.09%
C6	99.89%	98.50%	98.22%	7.74%	12.57%	21.01%
Ave.	99.06%	96.17%	92.88%	13.85%	25.15%	36.83%

C. Diagnosis Performance on Data with Sensor Failure

The experiment results of inner race fault (C3) under medium sensor failure condition are shown in Table XI. The diagnosis accuracy of C3 is 99.22% and all classes have good diagnosis results. Furthermore, these classes differ significantly in uncertainty estimation values where the uncertainty value of C3 reaches 22.84% and others less than 2.5%.

TABLE XI

FAULT DIAGNOSIS AND UNCERTAINTY ESTIMATION RESULTS IN MEDIUM SENSOR FAILURE (C3 CORRUPTED)

Class	Accuracy	Recall	Precision	F1score	Uncertainty
C1	99.50%	100%	97.09%	98.52%	1.72%
C2	100%	100%	100%	100%	2.22%
C3	99.22%	95.33%	100%	97.61%	22.84%
C4	99.72%	100%	98.36%	99.17%	2.32%
C5	100%	100%	100%	100%	1.83%
C6	100%	100%	100%	100%	1.57%

The classification probabilities and uncertainty estimates of 10 randomly selected samples are shown in Fig. 13. For the correct classification samples (excluding sample 9 which is misclassified), the average uncertainty value is 31.03%. This phenomenon suggests a risk of overconfidence in diagnostic results despite accurate classification. For instance, sample 7 and sample 8 have a high uncertainty value of 61.45% and 57.31%, respectively, which is caused by similar class probabilities assignments. In addition, sample 6 has a specific probability in C3 with a low uncertainty value. Also, sample 9 belongs to the misclassification sample with similar class probabilities assigned and a high uncertainty value.

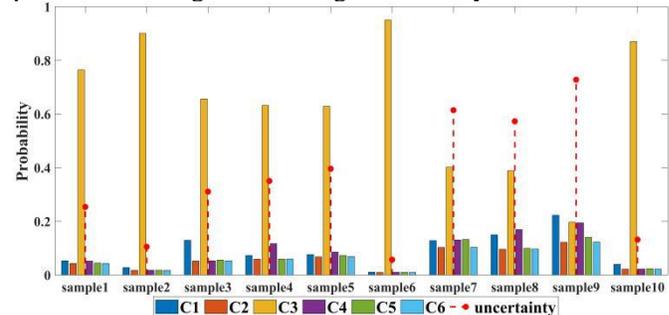


Fig. 13. Samples diagnostic probability and uncertainty estimation of C3 under medium sensor failure.

Under the different sensor failure scenarios, the diagnostic accuracy and uncertainty estimation are illustrated in Table

XII. The average diagnosis accuracy values under the different sensor failure levels are 99.33%, 98.60%, and 94.63%, correspondingly. The average uncertainty values are 13.57%, 21.46%, and 31.78%, respectively. The proposed method presents good resistance to noise interference and can identify OOD samples with high uncertainty estimation values. It realizes the reliable fault diagnosis in sensor noise and sensor failure conditions.

TABLE XII
FAULT DIAGNOSIS PERFORMANCE OF 6 CLASSES UNDER SPA NOISE

Class	Accuracy			Uncertainty		
	Slight	Medium	Strong	Slight	Medium	Strong
C1	99.33%	98.94%	98.89%	16.02%	20.11%	22.75%
C2	99.44%	98.83%	94.44%	14.89%	20.85%	41.00%
C3	99.44%	99.22%	88.94%	8.48%	22.84%	32.92%
C4	99.22%	96.72%	91.06%	16.80%	25.14%	42.04%
C5	99.72%	99.11%	97.89%	9.85%	17.61%	27.18%
C6	98.83%	98.78%	96.56%	15.40%	22.23%	24.79%
Ave.	99.33%	98.60%	94.63%	13.57%	21.46%	31.78%

D. Comparison with Other Methods

To demonstrate the superiority of the proposed diagnosis method, several state-of-the-art models are compared using the XJTU-SY data set. **TABLE XIII** presents the fault diagnosis accuracy and uncertainty values under different noise levels (uncertainty given in parentheses). With a low noise interference condition, some methods perform well. However, the performance of these methods under strong noise is significantly lower than the proposed method. Overall, the proposed EVGG model achieves excellent noise immunity, particularly when strong interference is present. Additionally, the proposed method provides quantitative uncertainty values that can be used to identify OOD samples in test data and ensure that predictions are reliable and trustworthy.

TABLE XIII
COMPARISON RESULTS UNDER GAUSSIAN NOISE INTERFERENCE

Method	Original Data	Noise interference level		
		Slight noise	Medium noise	Strong noise
Proposed method	100% (2.04%)	99.06% (13.85%)	96.17% (25.15%)	92.88% (36.83%)
SVM[30]	99.76%	95.29%	85.25%	70.85%
WT-CNNs[31]	96.92%	83.27%	80.61%	76.06%
MSCNN[32]	99.95%	98.88%	94.33%	85.81%
VMD-DCNNs[33]	99.99%	99.71%	98.02%	88.48%
NISTA-Net[34]	100%	94.38%	91.01%	83.71%

VI. CONCLUSIONS

The present paper proposed a trustworthy intelligent fault diagnosis approach with uncertainty estimation through a novel EVGG model. Because of the superiorities of the Dirichlet posterior model, which can directly learn its hyperparameters from the data, it can predict Dirichlet density as the *softmax* probability. The evidential loss function with KL divergence was applied in the EVGG model and achieved satisfactory uncertainty estimation without reducing the model diagnosis capacity. The effectiveness of the proposed method was verified on two public datasets, under sensing noise and sensor failure. It achieved excellent fault diagnosis accuracy in both cases with good noise immunity and robustness. In addition, the model produced a valid estimation of predictive uncertainty, which could detect the out-of-distribution samples

in the testing data, preventing overconfidence prediction. The proposed method was shown to largely enhance the trustworthiness of the prediction results, which is an important step forward in fault diagnosis for practical application.

The study has some limitations which could be addressed in future studies. Firstly, some hyperparameters in the EVGG model are determined manually. Bayesian optimization (BO) can be considered to provide an optimization strategy for hyperparameters. Secondly, the proposed approach can be further developed to detect OOD samples (new conditions or samples in the target domain in transfer learning [35]). Thirdly, varying operating conditions could be considered which is also common in practice.

REFERENCES

- [1] M. Xia, H. Shao, D. Williams, S. Lu, L. Shu, and C. W. de Silva, "Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning," *Reliability Engineering & System Safety*, vol. 215, p. 107938, 2021.
- [2] A. G. Roy, S. Conjeti, N. Navab, C. Wachinger, and A. s. D. N. Initiative, "Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control," *NeuroImage*, vol. 195, pp. 11-22, 2019.
- [3] Y. Ovadia *et al.*, "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift," *Advances in neural information processing systems*, vol. 32, 2019.
- [4] M. Abdar *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243-297, 2021.
- [5] R. L. Russell and C. Reale, "Multivariate uncertainty in deep learning," *IEEE Transactions on Neural Networks Learning Systems*, 2021.
- [6] A. Loquercio, M. Segu, and D. Scaramuzza, "A general framework for uncertainty estimation in deep learning," *IEEE Robotics Automation Letters*, vol. 5, no. 2, pp. 3153-3160, 2020.
- [7] A. Malinin and M. Gales, "Predictive uncertainty estimation via prior networks," *Advances in neural information processing systems*, vol. 31, 2018.
- [8] A. Amini, W. Schwarting, A. Soleimany, and D. Rus, "Deep evidential regression," *Advances in Neural Information Processing Systems*, vol. 33, pp. 14927-14937, 2020.
- [9] H. Shao *et al.*, "Fault diagnosis of a rotor-bearing system under variable rotating speeds using two-stage parameter transfer and infrared thermal images," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021.
- [10] M. Sensoy, L. Kaplan, F. Cerutti, and M. Saleki, "Uncertainty-aware deep classifiers using generative models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 04, pp. 5620-5627.
- [11] X. Pei, S. Dong, B. Tang, and X. Pan, "Bearing Running State Recognition Method Based on Feature-to-Noise Energy Ratio and Improved Deep Residual Shrinkage Network," *IEEE/ASME Transactions on Mechatronics*, 2021.
- [12] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 452-459, 2015.
- [13] M. Teye, H. Azizpour, and K. Smith, "Bayesian uncertainty estimation for batch normalized deep networks," in *International Conference on Machine Learning*, 2018: PMLR, pp. 4907-4916.
- [14] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," *Advances in neural information processing systems*, vol. 31, 2018.
- [15] J. Gast and S. Roth, "Lightweight probabilistic deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3369-3378.
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [17] Z.-G. Liu, G. Qiu, G. Mercier, and Q. Pan, "A transfer classification method for heterogeneous data based on evidence theory," *IEEE Transactions on Systems, Man, Cybernetics: Systems*, vol. 51, no. 8, pp. 5129-5141, 2019.

- [18] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mechanical Systems Signal Processing*, vol. 100, pp. 439-453, 2018.
- [19] H. Shang, C. Sun, J. Liu, X. Chen, and R. Yan, "Deep learning-based borescope image processing for aero-engine blade in-situ damage detection," *Aerospace Science and Technology*, vol. 123, p. 107473, 2022.
- [20] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7319-7327, 2020.
- [21] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Transactions on Reliability*, vol. 69, no. 1, pp. 401-412, 2018.
- [22] M. Xia, H. Shao, Z. Huang, Z. Zhao, F. Jiang, and Y. Hu, "Intelligent Process Monitoring of Laser-Induced Graphene Production with Deep Transfer Learning," *IEEE Transactions on Instrumentation and Measurement*, 2022.
- [23] M. Ourak, B. Tamadazte, O. Lehmann, and N. Andreff, "Direct visual servoing using wavelet coefficients," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 3, pp. 1129-1140, 2019.
- [24] M. Xia, T. Li, L. Xu, L. Liu, and C. W. De Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME transactions on mechatronics*, vol. 23, no. 1, pp. 101-110, 2017.
- [25] H. Han, H. Wang, Z. Liu, and J. Wang, "Intelligent vibration signal denoising method based on non-local fully convolutional neural network for rolling bearings," *ISA transactions*, vol. 122, pp. 13-23, 2022.
- [26] S. Ma, W. Liu, W. Cai, Z. Shang, and G. Liu, "Lightweight deep residual CNN for fault diagnosis of rotating machinery based on depthwise separable convolutions," *IEEE Access*, vol. 7, pp. 57023-57036, 2019.
- [27] G. Jin, T. Zhu, M. W. Akram, Y. Jin, and C. Zhu, "An adaptive anti-noise neural network for bearing fault diagnosis under noise and varying load conditions," *IEEE Access*, vol. 8, pp. 74793-74807, 2020.
- [28] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE transactions on industrial informatics*, vol. 16, no. 3, pp. 1688-1697, 2019.
- [29] X. Chen, B. Zhang, and D. Gao, "Bearing fault diagnosis base on multi-scale CNN and LSTM model," *Journal of Intelligent Manufacturing*, vol. 32, no. 4, pp. 971-987, 2021.
- [30] X. Chen, Y. Yang, Z. Cui, and J. Shen, "Vibration fault diagnosis of wind turbines based on variational mode decomposition and energy entropy," *Energy*, vol. 174, pp. 1100-1109, 2019.
- [31] P. Liang, C. Deng, J. Wu, Z. Yang, J. Zhu, and Z. Zhang, "Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform," *Computers in Industry*, vol. 113, p. 103132, 2019.
- [32] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3196-3207, 2018.
- [33] Z. Xu, C. Li, and Y. Yang, "Fault diagnosis of rolling bearing of wind turbines based on the variational mode decomposition and deep convolutional neural networks," *Applied Soft Computing*, vol. 95, p. 106515, 2020.
- [34] B. An, S. Wang, Z. Zhao, F. Qin, R. Yan, and X. Chen, "Interpretable Neural Network via Algorithm Unrolling for Mechanical Fault Diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-11, 2022.
- [35] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Few-shot transfer learning for intelligent fault diagnosis of machine," *Measurement*, vol. 166, p. 108202, 2020.



Hanting Zhou is currently a Ph.D. student at school of Economics and Management, Nanjing University of Science and Technology, China. She has been a visiting Ph.D. student at the School of Engineering at Lancaster University, UK. She received a B.S. degree (2016) and M.S. (2019) from China Jiliang University, China. Her research interests include smart manufacturing, signal processing, machine diagnosis and prognostics, and trustworthy AI.



Wenhe Chen is currently pursuing the Ph.D. degree at the Nanjing University of Science and Technology, China. He is a visiting Ph.D. student at the School of Engineering at Lancaster University, UK. He received the B.S. degree from Eastern Liaoning University, China (2016); the M.S. degree from Anhui University of Technology, China (2020). His current research interests include data mining, deep neural networks, renewable energy equipment, fault diagnosis, optimization algorithms.



Longsheng Cheng is currently a professor of management sciences and applied statistics at the School of Economics and Management, Nanjing University of Science and Technology, China. He received his bachelor's degree from Anhui Normal University, China (1985), M.S. degree from East China Institute of Technology, China (1989), and Ph.D. degree from Nanjing University of Science and Technology, China (1998). His current research interests include prognostics and health management, machine learning, quality engineering, and applied

statistics.



Darren Willimas is the Welding Systems Integration Team Leader at TWI, and the Professor and Director of the Joining 4.0 Innovation Centre at Lancaster University. He received Doctor of Engineering in System Engineering with Management and Accounting from University of Bath in 2013. He is a Chartered Engineer, and Fellow of the Institution of Engineering and Technology. His research interests include digital manufacturing, Industry 4.0, welding and joining

technologies.



Clarence W. de Silva (S'75-M'78-SM'85-F'98) received the first Ph.D. degree in mechanical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1978, the second Ph.D. degree in information engineering from the University of Cambridge, Cambridge, U.K., in 1998, the Honorary D.Eng. degree from the University of Waterloo, Waterloo, ON, Canada, in 2008, and the Higher Doctorate (Sc.D.) degree in mechatronics and automation from the University of Cambridge, in 2020.

He has been a Professor of Mechanical Engineering at the University of British Columbia, Vancouver, BC, Canada, since 1988. He is a fellow of: ASME, Canadian Academy of Engineering, and Royal Society of Canada. His appointments include the Tier 1 Canada Research Chair in Mechatronics and Industrial Automation, Professorial Fellow, Peter Wall Scholar, Mobil Endowed Chair Professor, and NSERCBC Packers Chair in Industrial Automation.



Min Xia (S'11-M'17-SM'22) is currently an Assistant Professor in the School of Engineering at Lancaster University, UK. He received B.S. degree in Industrial Engineering from Southeast University, China (2009); M.S. degree in Precision Machinery and Precision Instrumentation from the University of Science and Technology of China, China (2012); and Ph.D. degree in Mechanical Engineering from the University of British Columbia, Canada (2017). He has led 11 research projects in the UK, Canada, and Japan with total funding of £9 million. He has served various editorial roles including Associate Editor of IEEE Transactions on Instrumentation and Measurements. His research interests include smart manufacturing, machine diagnostics and prognostics, deep neural networks, process monitoring and optimization.