

Few-Shot Learning for Full Ceramic Bearing Fault Diagnosis with Acoustic Emission Signals

David He¹, Miao He², and Alessandro Taffari³

^{1,3}*The University of Illinois-Chicago, Chicago, IL, 60607, USA*
davidhe@uic.edu

²*Siemens Technology, Princeton, NJ, 08540, USA*
miao.he@siemens.com

ABSTRACT

Full ceramic bearings are critical components in many full ceramic and oil-free food processing and medical equipment. Developing effective full ceramic fault diagnostic methods is important. Supervised deep learning approaches have been considered promising for fault diagnosis in the era of big data where abundantly labelled datasets are available. However, in many industrial applications, datasets with fault labels are rare. This challenge has motivated the task for developing deep learning approaches for fault diagnosis with few training examples. To meet the challenge, one attractive direction is to use available pre-trained deep learning architectures to do fault diagnosis with only few examples. Specifically, this paper investigates the effectiveness of using pre-trained deep learning architectures successfully used in natural language processing to achieve few-shot learning for full ceramic bearing fault diagnosis using acoustic emission signals.

1. INTRODUCTION

The use of full ceramic bearings has grown significantly in recent years, with more and more industries adopting them for their superior performance and durability. Here are a few examples: (1) Semiconductor manufacturing equipment, where their non-conductive properties prevent electrical interference and ensure consistent, high-quality processing. (2) Chemical processing applications, where they can withstand high temperatures, corrosive chemicals, and other harsh operating conditions. (3) Ceramic bearings are widely used in high-speed rotating machinery due to their excellent mechanical properties. (4) Food and beverage processing equipment, where their non-toxic properties ensure product purity and prevent contamination. (5)

Medical equipment, such as surgical tools and imaging equipment, where their non-magnetic and non-conductive properties make them ideal for use in MRI machines and other sensitive medical applications. However, full ceramic bearings are vulnerable to faults such as cracks, spalls, and wear, which can cause catastrophic failures and downtime. Early fault diagnosis is crucial for preventing costly equipment failures and ensuring safe and reliable operation.

Acoustic emission (AE) signals generated by the bearing during operation have been proven to be effective in detecting bearing faults. However, traditional machine learning algorithms require a large amount of labeled data for training, which is time-consuming and expensive to collect. Few-shot learning, which aims to learn from a few labeled examples, has recently emerged as a promising solution to this problem.

In this paper, we propose a few-shot learning approach for full ceramic bearing fault diagnosis with AE signals. Specifically, we use pre-trained deep structures successfully used in natural language processing (NLP) like GPT2 to train a classifier with extracted AE features on a small number of labeled examples.

To the best of our knowledge, this is the first attempt to apply pre-trained NLP deep structures to few-shot learning for full ceramic bearing fault diagnosis with AE signals. We evaluate our approach on a full ceramic bearing seeded fault dataset collected from a bearing test rig in the laboratory.

The rest of the paper is organized as follows. Section II provides a brief overview of the related work. Section III describes the proposed approach in detail. Section IV presents the experimental setup and results. Finally, we conclude the paper in Section V.

2. RELATED WORK

2.1. Few-shot Learning for Fault Diagnosis

The goal of few-shot learning for fault diagnosis is to diagnose faults in a system or machine with high accuracy using only a small amount of labelled data. Typically, few-shot learning for fault diagnosis involves using transfer learning, domain adaptation, and meta-learning. Transfer learning can be used to transfer knowledge learned from one domain to another, while domain adaptation can be used to adapt a model trained on one domain to another. Meta-learning can be used to learn how to learn, enabling a model to quickly adapt to new domains with few samples.

One of the popular few-shot learning methods is K -way N -shot learning, which is used to classify new samples based on a small set of labeled samples (N) taken from K different categories.

Liu *et al.* (2021) conducted a survey on meta-learning for few-shot cross-domain fault diagnosis. Their paper provides an overview of various algorithms, including model-agnostic meta-learning (MAML), Reptile, and Prototypical Networks, which have been shown to be effective in adapting to new fault diagnosis tasks with limited labeled data. It suggests that meta-learning has great potential for few-shot cross-domain fault diagnosis. Feng *et al.* (2022) explored the use of meta-learning for fault diagnosis. The paper reviews various meta-learning algorithms and their applications in fault diagnosis across different domains, such as manufacturing, aerospace, and automotive industries. The authors also discussed the prospects of meta-learning for fault diagnosis and highlighted some of the challenges that need to be addressed to make this approach more practical and effective. For example, they pointed out that the lack of labeled data and the high cost of obtaining it are significant barriers to the adoption of meta-learning in industrial settings. Yan *et al.* (2021) proposed a few-shot learning framework for fault diagnosis in industrial machine. The proposed framework is based on a transformer architecture with attention mechanisms and uses contrastive learning to learn a feature representation that can discriminate between normal and faulty conditions. It is trained on a small labeled dataset and can quickly adapt to new machines with few labeled samples. Domain shift caused by changes in machine speed can be handled by the proposed framework. Hu *et al.* (2019) explored the use of external data and fine-tuning for improving the performance of few-shot learning pipelines. They found that fine-tuning the image encoder on the target task can improve performance of few-shot learning.

In our proposed approach, we use K -way N -shot learning with a pre-trained NLP deep structure to train a classifier with extracted AE features on a small number of labeled examples for ceramic bearing fault diagnosis. By leveraging the power of the pre-trained NLP deep structure and K -way

N -shot learning, we can effectively diagnose faults in ceramic bearings with minimal labeled data.

2.2. Pre-trained NLP Architectures

Recent advances in NLP have shown the effectiveness of pre-trained deep structures in various tasks such as text generation, translation, and sentiment analysis. The Generative Pre-trained Transformer 2 (GPT2) with a transformer structure is a state-of-the-art pre-trained language model that has achieved impressive results in many NLP benchmarks (Radford *et al.* 2018, Radford *et al.* 2019, Brown *et al.* 2020).

GPT2 uses a multi-layer transformer decoder to generate high-quality text. The model is trained on a large corpus of text data and learns to predict the next word in a sequence based on the context of the previous words (see GPT2 structure shown in Figure 1). This approach enables the model to capture the semantic and syntactic structures of natural language and generate coherent and diverse text.

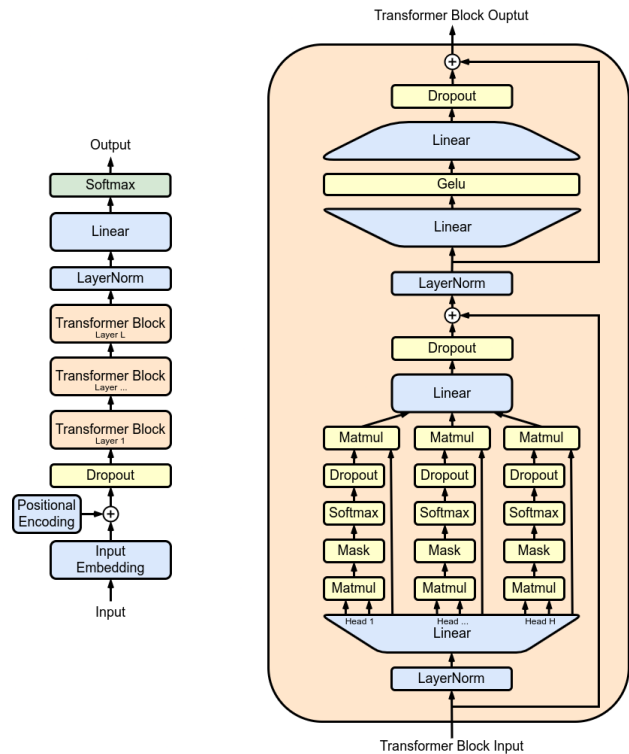


Figure 1. The structure of GPT2 (Wikipedia)

Inspired by the success of pre-trained deep structures in NLP, researchers have applied similar techniques to other domains such as computer vision and audio processing. Few-shot learning is one such area where pre-trained deep structures have shown promising results (Ren *et al.* 2020, Chen and Li 2020, Ko *et al.* 2020, Lee and Park 2021).

In our proposed approach, we leverage the power of GPT2 with extracted AE features to achieve few-shot learning for full ceramic bearing fault diagnosis. By using the pre-trained GPT2 deep structure, we can effectively transfer knowledge from a large corpus of unlabeled data to a small labeled dataset and improve the model's generalization ability.

There are a number of key factors attributed to the use of GPT2, and later models like GPT3 or ChatGPT in our method: (1) Model scale: GPT2 and its successors are built on an extremely large scale in terms of both model size (number of parameters) and amount of training data. This allows them to learn a broad representation of language. For instance, GPT-2 has 1.5 billion parameters, while GPT-3 has a staggering 175 billion. (2) Unsupervised learning: These models are trained using unsupervised learning on a large corpus of internet text. This means they are not specifically trained to answer questions or generate text in a specific style; rather, they learn to predict the next word in a sentence. It's through this task that they acquire an understanding of syntax, semantics, and some factual information. (3) Transfer learning: The GPT model can be considered an example of transfer learning. The model is initially trained on a broad task (predicting the next word in internet text), and this trained model is then fine-tuned on a

more specific task. This allows the model to apply the broad understanding of language it gained during pretraining to the specific task during fine-tuning. (4) Transformer architecture: The underlying architecture of GPT2 is a transformer model, which uses self-attention mechanisms to understand the context of words in a sentence. This allows it to generate more coherent and contextually appropriate text compared to older architectures like RNNs or LSTMs.

3. THE METHODOLOGY

The proposed few-shot learning for full ceramic bearing fault diagnosis framework is shown Figure 2. As shown in Figure 2, the collected AE signals are first processed using empirical mode decomposition (EMD) method into multiple intrinsic mode components (IMFs) which are used to generate the AE features (He *et al.* 2001). These AE features are then used to fine-tune the pre-trained GPT2 model as a classifier by K -way N -shot learning. Once the classifier is trained, it will be applied to incoming AE features with unknown faults to identify the bearing faults.

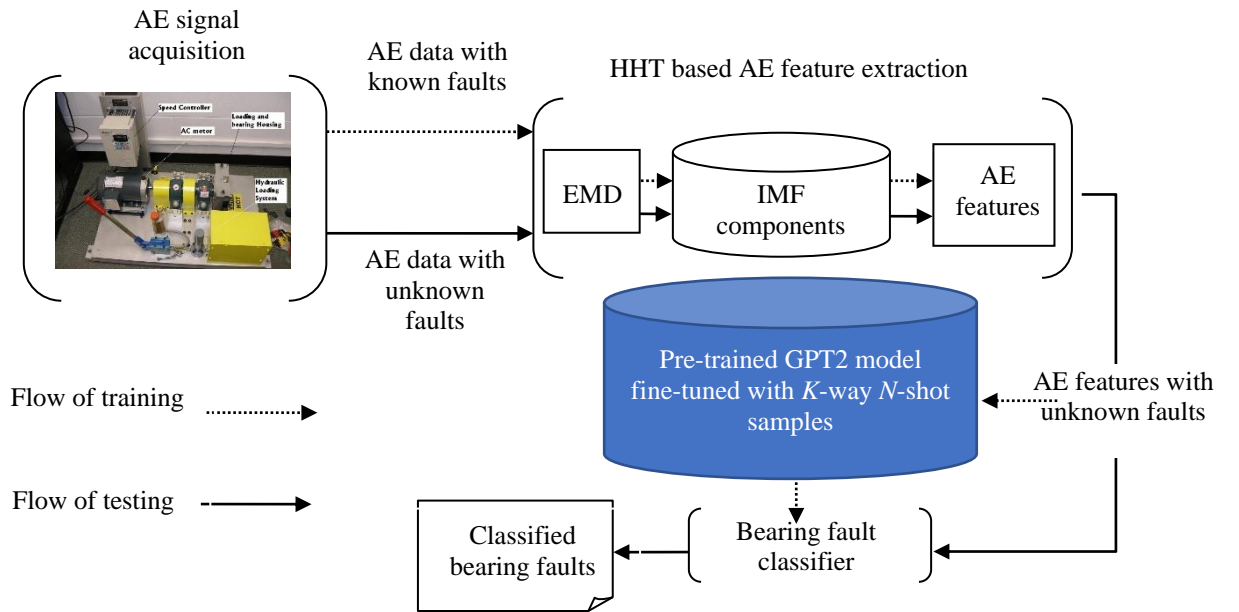


Figure 2. The framework of the proposed methodology

Figure 3 shows the details of the fine-tuning with K -way N -shot samples.

To achieve better transfer learning performance for fine-tuning the pre-trained deep structure, two effective

approaches have been suggested in the literature (Dhillon *et al.* 2020, Chen *et al.* 2020, Chen *et al.* 2019). One is to modify the loss function of the fine-tuning by adding a regularization. Another one is to modify the softmax activation function of the classify layer using cosine similarity.

Let

y_{ij} = the true label for class j of sample i .

p_{ij} = the predicted probability class j of sample i .

N = the total number of samples.

Then we can compute the mean cross-entropy loss over all samples (N) as:

$$L_{mean} = -(1/N) * \sum \sum y_{ij} * \log(p_{ij}) \quad (1)$$

Since sample size in few-shot learning is small, the fine-tuning of the pre-trained model could lead to overfitting problems. To prevent overfitting during fine-tuning, it is suggested that an entropy regularization should be added to the cross-entropy function. Let \mathbf{p} be the probability distribution as the output of the softmax function in the classification layer. Then the entropy of \mathbf{p} can be computed as:

$$e(\mathbf{p}) = \sum \sum p_{ij} * \log(p_{ij}) \quad (2)$$

The entropy regularization is defined as the average of $e(\mathbf{p})$ as: $\frac{\sum e(\mathbf{p})}{N}$. Therefore, modified loss function with the entropy regularization can be computed as:

$$L'_{mean} = -(1/N) * \sum \sum y_{ij} * \log(p_{ij}) - (1/N) * \frac{\sum \sum p_{ij} * \log(p_{ij})}{\sum \sum p_{ij}} \quad (3)$$

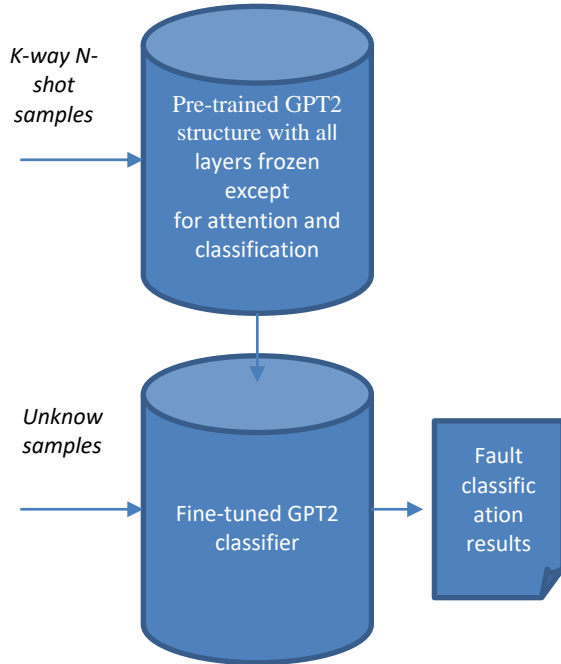


Figure 3. The fine-tuning of GPT2 model

To modify the softmax activation function, define:

\mathbf{x} = the test samples

\mathbf{W} = weight vector from the classification layer

\mathbf{b} = the bias

The softmax activation function can be modified with the cosine similarity as:

$$\mathbf{p} = \text{Softmax}\left(\frac{\mathbf{w}^T \mathbf{x}}{\|\mathbf{w}\|_2 \|\mathbf{x}\|_2} + \mathbf{b}\right) \quad (4)$$

4. EXPERIMENT SETUP AND RESULTS

4.1. The Dataset

To evaluate the performance of the proposed methodology for full ceramic bearing fault diagnosis, AE signal dataset collected during bearing seeded fault tests performed on a bearing test rig in the laboratory are used. Figure 4 shows the bearing test rig and the AE sensors on the bearing housing.

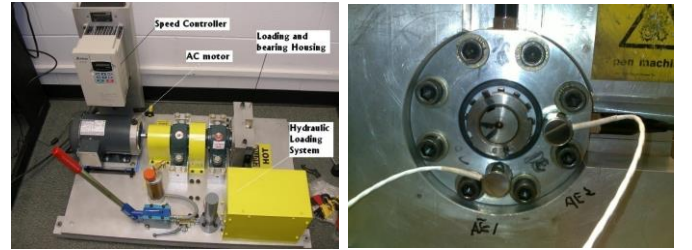


Figure 4. Bearing test rig (left) and AE sensors (right)

Two wide band (WD) type AE sensors and a 2-channel data acquisition card with 18-bit resolution and a maximum sampling rate up to 40 MHz were used to collect the AE burst data. The AE sensors were attached to the bearing housing by instant glues.

During the test, bearings with following seeded faults were run on the test rig to collect the AE signals: inner race fault, outer race fault, ball fault, and cage fault (see Figure 5). The speed of the motor shaft was controlled at 10Hz (600 rpm) and the AE signals were collected at a sampling rate of 5 MHz.



Figure 5. Bearing seeded faults

For each collected AE signal, the EMD method was applied to decompose it into a number of IMF components. An example of the bearing inner race fault AE signal and the first 4 IMF components of the signal are provided in Figure 6.

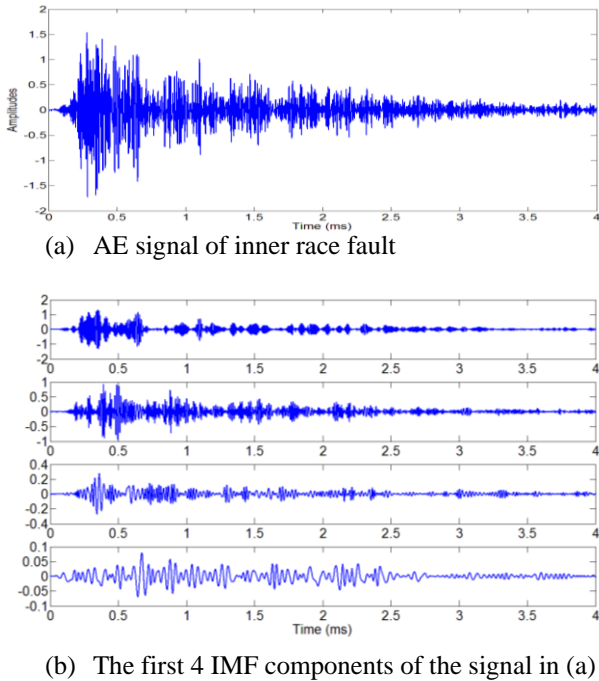


Figure 6. An example of the bearing inner race fault AE signal and its IMF components

To extract the AE features, the three IMF components were summed and then the following three values were extracted from the summed IMF components: *rms*, *peak value*, and

kurtosis. From these values, 7 AE features were formed as shown in Table 1.

Table 1. The extracted AE features

| AE feature | Explanation |
|-----------------------|--|
| rms-average | Average of the root mean squared values |
| rms-stdev | Standard deviation of the root mean squared values |
| rms-average+rms-stdev | Sum of rms-average and rms-stdev |
| kurtosis-average | Average of the kurtosis values |
| kurtosis-stdev | Standard deviation of the kurtosis values |
| peak-average | Average of the peak values |
| peak-stdev | Standard deviation of the peak values |

For each type of bearing faults, a total of 40 data points were generated. Therefore, a total of 160 data points were available for the data analysis.

4.2. The Analysis Results

To fine-tune the pre-trained GPT2 model, K -way N -shot samples were randomly generated without replacement from the dataset of 160 samples. Then remaining samples were split with an 80-20 ratio. 20% of the remaining samples were used as the validation set for the K -way N -shot learning.

Since GPT2 is developed for text learning only. In order to use the pre-trained GPT2 for fault classification, the AE feature data were first converted into text using text normalization. Text normalization is the process of converting numbers, symbols, and other non-textual data into their corresponding textual representations. For example, for a numerical value like "98.6", it can be converted to "ninety-eight point six" using a text normalization technique.

For the fine-tuning of the GPT2, the following parameters were set up as shown in Table 2.

Table 2. Parameter settings of the fine-tuning

| Name | Value |
|--------------------------------------|---|
| Fully connected classification layer | Number of neurons = 768 |
| Optimizer | AdamW: beta1=0.9, beta2=0.999, epsilon = 1e-8 |
| Learning rate | 3e-5 |

Since there were 4 types of faults in the dataset, $K = 4$. 5 different 4-Way N -shot learnings for $N = 1, \dots, 5$ were performed. The results are present in Tables 3-6.

Table 3. 4-Way 1-Shot learning results

| Old loss +softmax | | New loss+softmax | |
|-------------------|--------------------|------------------|--------------------|
| Val accuracy (%) | Train accuracy (%) | Val accuracy (%) | Train accuracy (%) |
| 23.13 | 25.00 | 53.34 | 57.5 |

Table 4. 4-Way 2-Shot learning results

| Old loss+softmax | | New loss+softmax | |
|------------------|--------------------|------------------|--------------------|
| Val accuracy (%) | Train accuracy (%) | Val accuracy (%) | Train accuracy (%) |
| 28.39 | 35.00 | 62.10 | 67.5 |

Table 5. 4-Way 3-Shot learning results

| Old loss+softmax | | New loss+softmax | |
|------------------|--------------------|------------------|--------------------|
| Val accuracy (%) | Train accuracy (%) | Val accuracy (%) | Train accuracy (%) |
| 33.33 | 45.00 | 74.33 | 75.17 |

Table 6. 4-Way 4-Shot learning results

| Old loss+softmax | | New loss+softmax | |
|------------------|--------------------|------------------|--------------------|
| Val accuracy (%) | Train accuracy (%) | Val accuracy (%) | Train accuracy (%) |
| 36.59 | 53.00 | 83.13 | 83.50 |

Table 5. 4-Way 5-Shot learning results

| Old loss+softmax | | New loss+softmax | |
|------------------|--------------------|------------------|--------------------|
| Val accuracy (%) | Train accuracy (%) | Val accuracy (%) | Train accuracy (%) |
| 42.29 | 61.89 | 92.34 | 92.71 |

From the results presented in Tables 3-6, we can see that modifying the loss function with entropy regularization and softmax function with cosine similarity improves the few-shot learning for full ceramic bearing fault diagnosis significantly. As we increase the number of shots, the diagnostic accuracy increases.

5. CONCLUSION

In this paper, a few-shot learning approach for full ceramic bearing fault diagnosis with AE signals was presented. Specifically, we used the pre-trained deep NLP structure GPT2 to train a fault classifier with extracted AE features on a small number of labeled examples. The fine-tuning of the GPT2 model was involved with a modified loss function and a modified softmax activation function. 4-way N -shot learning was performed on a set of full ceramic bearing seeded fault data collected from a bearing test rig in the laboratory. The evaluation results have shown that the presented method was able to diagnose the bearing faults with satisfied accuracies.

REFERENCES

- He, D., Li, R., Zade, M., & Zhu, J. (2011). A data mining based full ceramic bearing fault diagnostic system using AE sensors. *IEEE Transactions on Neural Networks*, 22(12), 2022-2031. <https://doi.org/10.1109/TNN.2011.2172803>.
- Chen, Y., Wang, Y., Liu, Y., Xu, Z., & Darrell, T. (2020). A new Meta-Baseline for few-shot learning. In

- Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11934-11943).
<https://doi.org/10.1109/CVPR42600.2020.01193>
- Chen, T., Liu, S., Kira, Z., Wang, Y., & Huang, J. (2019). A closer look at few-shot classification. In Proceedings of the International Conference on Learning Representations (ICLR). Retrieved from <https://openreview.net/forum?id=HkxLXnAcFQ>.
- Feng, Y., Chen, J., Xie, J., Zhang, T., Lv, H., & Pan, T. (2022). Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects. *Knowledge-Based Systems*, 237, 107165. <https://doi.org/10.1016/j.knosys.2021.107165>.
- Hu, S. X., Li, D., Stühmer, J., Kim, M., & Hospedales, T. M. (2019). Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 11803-11812). doi: 10.1109/CVPR.2019.01206
- Ren, M., Zhai, D., & Yang, B. (2020). Few-shot learning with global class representations. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, 9195-9204.
- Chen, Q., & Li, Z. (2020). Few-shot learning with auxiliary classifier generative adversarial networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, 10660-10669.
- Ko, T., Krishnamurthy, J., Aluru, S., & Chang, S. F. (2020). Learning representations for few-shot learning with deep convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 31(7), 2342-2355.
- Lee, J. Y., & Park, J. (2021). Few-shot learning with meta-transformer. *IEEE Transactions on Multimedia*, 23, 2573-2583.
- Liu, X., Zhang, Y., Liu, Y., Li, H., & Li, Z. (2021). Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects. *Neurocomputing*, vol. 449, pp. 93-103. <https://doi.org/10.1016/j.neucom.2021.05.086>.
- Yan, Y., Liu, Y., Liu, X., Chen, Y., Peng, Y., & Li, X. (2021). Few-shot learning under domain shift: Attentional contrastive calibrated transformer of time series for fault diagnosis under sharp speed variation. *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2834-2844.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. URL https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf.
- G. S. Dhillon, P. Chaudhari, A. Ravichandran, and S. Soatto, "A Baseline for Few-Shot Image Classification," in Proceedings of the International Conference on Learning Representations (ICLR), 2020.



David He received a Ph.D. in Industrial Engineering from The University of Iowa. Dr. He is a Professor and Director of the Intelligent Systems Modeling & Development Laboratory in the Department of Mechanical and Industrial Engineering at The University of Illinois-Chicago.

Dr. He is also a Fellow of the Prognostics and Health Management (PHM) Society. Dr. He's research areas include: PHM, Industrial AI, smart manufacturing systems modeling and analysis, quality and reliability engineering.



Miao He received a Ph.D. in Industrial Engineering and Operations Research from University of Illinois at Chicago, a B.E. degree in Vehicle Engineering from University of Science and Technology Beijing, Beijing, China and a M.S. degree in Mechanical Engineering from Purdue University, Hammond, IN. She is

currently a senior research scientist at Siemens Corporation. Her research interests include digital signal processing, machinery health monitoring and fault diagnosis, prognostics, artificial neural network applications, and edge computing on factory automation.