



ARTICLE

## Full Ceramic Bearing Fault Diagnosis with Few-Shot Learning Using GPT-2

David He<sup>1,\*</sup>, Miao He<sup>2</sup> and Jay Yoon<sup>3</sup>

<sup>1</sup>Department of Mechanical and Industrial Engineering, College of Engineering, University of Illinois Chicago, Chicago, IL 60607, USA

<sup>2</sup>Siemens Corporation, Princeton, NJ 08540, USA

<sup>3</sup>NOV, Inc., Houston, TX 77042, USA

\*Corresponding Author: David He. Email: davidhe@uic.edu

Received: 31 January 2025; Accepted: 22 April 2025; Published: 30 May 2025

**ABSTRACT:** Full ceramic bearings are mission-critical components in oil-free environments, such as food processing, semiconductor manufacturing, and medical applications. Developing effective fault diagnosis methods for these bearings is essential to ensuring operational reliability and preventing costly failures. Traditional supervised deep learning approaches have demonstrated promise in fault detection, but their dependence on large labeled datasets poses significant challenges in industrial settings where fault-labeled data is scarce. This paper introduces a few-shot learning approach for full ceramic bearing fault diagnosis by leveraging the pre-trained GPT-2 model. Large language models (LLMs) like GPT-2, pre-trained on diverse textual data, exhibit remarkable transfer learning and few-shot learning capabilities, making them ideal for applications with limited labeled data. In this study, acoustic emission (AE) signals from bearings were processed using empirical mode decomposition (EMD), and the extracted AE features were converted into structured text for fine-tuning GPT-2 as a fault classifier. To enhance its performance, we incorporated a modified loss function and softmax activation with cosine similarity, ensuring better generalization in fault identification. Experimental evaluations on a laboratory-collected full ceramic bearing dataset demonstrated that the proposed approach achieved high diagnostic accuracy with as few as five labeled samples, outperforming conventional methods such as *k*-nearest neighbor (KNN), large memory storage and retrieval (LAMSTAR) neural network, deep neural network (DNN), recurrent neural network (RNN), long short-term memory (LSTM) network, and model-agnostic meta-learning (MAML). The results highlight LLMs' potential to revolutionize fault diagnosis, enabling faster deployment, reduced reliance on extensive labeled datasets, and improved adaptability in industrial monitoring systems.

**KEYWORDS:** LLMs; GPT-2; few-shot learning; fault diagnosis; full ceramic bearing; acoustic emission

### 1 Introduction

Full ceramic bearings have grown significantly in recent years, with an increasing number of industries adopting them for their superior performance and durability. These bearings are crucial in several applications: (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) high-speed rotating machinery, benefiting from their excellent mechanical properties; (4) food and beverage processing equipment, where their non-toxic properties ensure product purity and prevent contamination; and (5) medical equipment, such as surgical tools and imaging systems, where their non-magnetic and



non-conductive properties make them ideal for use in magnetic resonance imaging (MRI) machines and other sensitive medical devices. Despite these advantages, full ceramic bearings are vulnerable to faults such as cracks, spalls, and wear, which can cause catastrophic failures and operational downtime. Early fault diagnosis is essential to prevent costly equipment failures and ensure safe and reliable operation.

Acoustic emission (AE) signals generated by the bearings during operation have proven effective for detecting faults. However, traditional machine learning algorithms often require extensive labeled datasets for training, which are time-consuming and expensive to collect. Few-shot learning has recently emerged as a promising approach to overcome this limitation by enabling models to generalize effectively from only a few labeled examples.

Recent advancements in artificial intelligence, particularly the development of large language models (LLMs), offer an exciting new avenue for few-shot learning. LLMs, such as GPT-2 and GPT-3, pre-trained on vast and diverse datasets, demonstrate exceptional transfer learning capabilities. Their ability to adapt to various tasks with minimal additional training makes them ideal for applications where labeled data is scarce. Leveraging LLMs for fault diagnosis has the potential to revolutionize the field by reducing reliance on extensive labeled datasets, accelerating deployment, and enabling more adaptive, cost-effective, and accurate monitoring solutions.

In this paper, we propose a novel few-shot learning approach for full ceramic bearing fault diagnosis using AE signals. Specifically, we harness pre-trained LLM structures, such as GPT-2, to train a classifier with extracted AE features using only a small number of labeled examples. The ability of LLMs to capture complex patterns and relationships in diverse data domains positions them as a transformative tool for this application. We evaluate our approach on a full ceramic bearing seeded fault dataset collected from a laboratory bearing test rig. The results demonstrate the potential of LLMs to address critical challenges in fault diagnosis, paving the way for more efficient and reliable maintenance strategies in industrial applications.

## 2 Related Work

### 2.1 Full Ceramic Bearing Fault Diagnosis

Even though full ceramic bearings play an important role in many industrial settings, unlike their metal counterparts, little research on full ceramic bearing fault diagnosis using vibration and acoustic emission signals has been reported in the literature. Most of the research focuses on conventional machine learning techniques such as  $k$ -nearest neighbor (KNN) [1,2], neural networks [2], stacked autoencoders [3], and large memory storage and retrieval (LAMSTAR) neural network [4]. He et al. [5] first proposed using few-shot learning for full ceramic bearing fault diagnosis. However, an established link between  $K$ -way  $N$ -shot learning and LLMs for full ceramic bearing fault diagnosis is missing in [5]. A comprehensive comparison of the LLMs integrated with  $K$ -way  $N$ -shot learning with other methods is not provided in [5].

### 2.2 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 labeled data. One of the popular few-shot learning methods is  $K$ -way  $N$ -shot learning. It refers to a scenario where a model must learn to recognize  $K$  different classes (fault types) using only  $N$  examples per class. This approach is particularly useful when labeled data is scarce, such as in bearing fault diagnosis.  $K$ -way  $N$ -shot learning for fault diagnosis typically involves transfer learning [6–9], meta-learning [10–14], Siamese networks [15,16], contrastive learning [17,18], task-agnostic meta-learning [19], and joint class representation learning [20]. Although powerful, each of these  $K$ -way  $N$ -shot learning techniques has its limitations. The limitations of transfer learning are: (1) Domain

shift sensitivity: models fine-tuned on one domain may perform poorly when applied to different but related tasks, e.g., fine-tuning on jet engine diagnostics but struggling with car engine diagnostics. (2) Fine-tuning still requires a moderate amount of labeled data, which is challenging in few-shot learning. (3) Catastrophic forgetting: when fine-tuning with a small dataset, the model may forget prior knowledge from its pre-training phase. Meta-learning has the following limitations: (1) Computationally expensive: training meta-learning models requires multiple inner-loop optimizations, making it resource-intensive. (2) Task-specificity: the meta-trained model performs well on tasks it has seen but struggles with highly unrelated new tasks. (3) Slow adaptation: meta-learning requires gradient updates for adaptation, making it slower than in-context learning. The limitations of Siamese networks include: (1) Limited to pairwise comparisons: Siamese networks excel at comparing two examples but struggle with multi-class classification. (2) Feature extraction dependency: the quality of embeddings determines performance, making it less flexible for diverse classification tasks. (3) No self-learning ability: Siamese networks require labeled positive/negative pairs for effective training. The following are the limitations of contrastive learning: (1) It requires massive negative samples for effective contrastive learning. (2) Embedding quality dependence: if the initial feature representations are poor, contrastive learning won't help much. (3) Lack of task awareness: it works well for similarity-based tasks but struggles with structured reasoning tasks. Task-agnostic meta-learning can suffer from the following limitations: (1) Weak generalization to unseen domains: since tasks are unknown during training, the model may fail to generalize effectively. (2) High variability in task distributions: without predefined task structures, learning optimal initialization is challenging for task-agnostic meta-learning. (3) It works best with a structured task space, but real-world tasks are often messy and unstructured.

### **2.3 Large Language Models for Fault Diagnosis**

Recent advancements in artificial intelligence have highlighted the transformative potential of pre-trained LLMs across various domains. LLMs, such as the Generative Pre-trained Transformer 2 (GPT-2), GPT-3, and ChatGPT, have set new benchmarks in tasks ranging from text generation and summarization to question answering and contextual understanding [21–23].

LLMs are built on the transformer architecture, which employs self-attention mechanisms to capture intricate patterns and dependencies within data. The effectiveness of LLMs can be attributed to several key factors: (1) Scale of models and data: LLMs are built on a massive scale in terms of both model size (number of parameters) and the amount of training data. For instance, GPT-2 has 1.5 billion parameters, while its successor, GPT-3, incorporates 175 billion parameters. This large scale enables these models to learn broad and nuanced representations of language and other structured data, making them highly versatile. (2) Unsupervised pretraining: LLMs are trained using unsupervised learning on extensive datasets that encompass diverse topics and styles. This training paradigm involves predicting the next token in a sequence, enabling the models to acquire a deep understanding of syntax, semantics, and general world knowledge without requiring labeled datasets. (3) Transfer learning: LLMs exemplify transfer learning by applying knowledge gained from pretraining on large, diverse datasets to downstream tasks with minimal additional data. Fine-tuning these models on domain-specific tasks allows them to adapt their broad understanding to specialized applications, even with limited labeled examples. (4) Transformer architecture: The transformer architecture, which underpins LLMs, leverages self-attention mechanisms to understand relationships and context within data. Compared to older machine learning architectures like recurrent neural networks (RNNs) or long short-term memory (LSTM) networks, transformers are more effective at capturing long-range dependencies, leading to coherent and contextually appropriate outputs.

Inspired by the success of LLMs in various fields, researchers have begun exploring their application to domains beyond natural language processing (NLP), such as bearing fault diagnosis. Two recent papers

have reported using LLMs for fault diagnosis [24,25]. Tao et al. [24] presented an LLM-based framework for bearing fault diagnosis with vibration signal processing. In their approach, time-domain features such as mean, standard deviation, skewness) and frequency-domain features such as power spectrum and cepstrum are first extracted from vibration signals. These vibration signal features are then converted into structured textual data to fine-tune the LLMs. Their results have shown that the trained LLM cross-dataset models improved accuracy by about 10%, providing the adaptability of LLMs to input patterns. Qaid et al. [25] proposed a fault diagnosis method based on a multimodal large language model (MM-LLM) to improve the diagnosis of complex equipment faults. The proposed fault diagnosis large language model (FD-LLM) aims to enhance fault classification accuracy by integrating domain-specific knowledge, modal alignment training, and fuzzy semantic embedding into a fine-tuned LLM. In their approach, they aligned time-series sensor data with text descriptions using an encoder-based approach and used contrastive learning to map time-series data into a feature space that aligns with natural language descriptions. They assigned probability weights to different fault categories instead of hard classification using fuzzy semantic embedding (FSE) to address pattern aliasing. This would allow faults with overlapping features to be diagnosed more accurately. They also introduced learnable prompt embeddings to integrate fault diagnosis knowledge into LLM reasoning. This would enhance contextual fault understanding while reducing fine-tuning costs. Their method was compared against convolutional neural network (CNN), ResNet18, Transformer, SimCLR, Attention-LSTM, and Attention-gated recurrent unit (GRU). FD-LLM outperformed all of them in terms of accuracy, precision, and generalization capability.

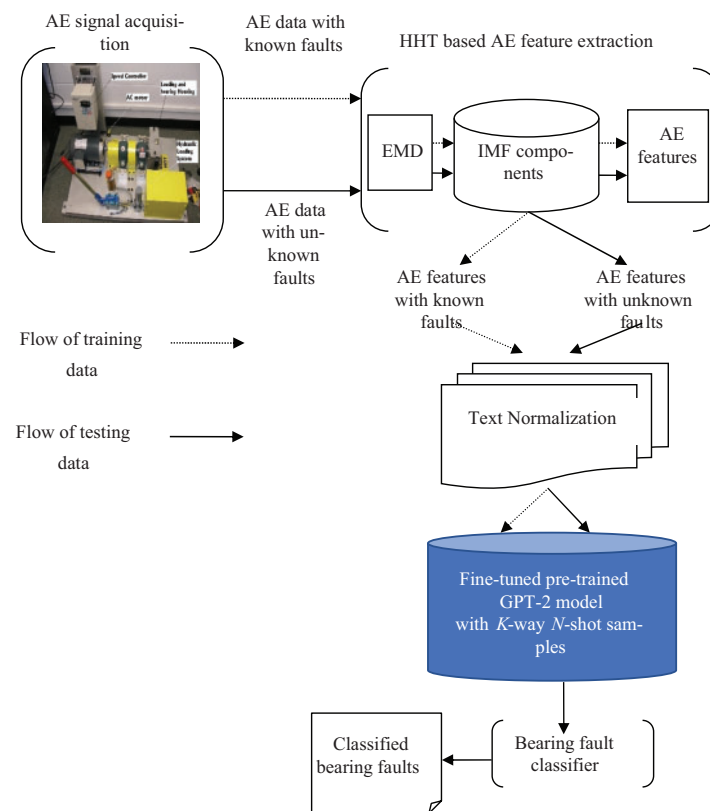
Few-shot learning is one area where LLMs have demonstrated significant potential, as their ability to generalize from limited examples aligns well with the challenges posed by data-scarce environments. Integrating LLMs with  $K$ -way  $N$ -shot learning techniques has emerged as a promising solution to overcome the limitations of the current  $K$ -way  $N$ -shot learning techniques and enhance the accuracy and adaptability of fault diagnosis systems. When integrating  $K$ -way  $N$ -shot learning techniques with LLMs, the ease of integration depends on how well each method aligns with LLMs' strengths in processing text-based tasks, generalizing from limited data, and leveraging pre-trained knowledge. Among the  $K$ -way  $N$ -shot learning techniques, transfer learning stands out as a good candidate for the job. LLMs are fundamentally built on transfer learning. Models like GPT-2, Bidirectional Encoder Representations from Transformers (BERT), and T5 are pre-trained on large corpora and fine-tuned for specific tasks. One can fine-tune an LLM on a domain-specific dataset, e.g., few-shot bearing data, while leveraging its pre-trained knowledge. Although meta-learning can be considered a possible candidate for LLMs integration, its high computational effort involving multiple inner-loop optimizations in model training makes it less attractive than transfer learning. Other  $K$ -way  $N$ -shot learning techniques, such as Siamese networks, contrastive learning, and task-agnostic meta-learning, cannot be considered good candidates for LLM integration. Siamese networks rely on pairwise comparisons of embeddings rather than direct token-level processing. While LLMs can generate embeddings (e.g., using BERT sentence embeddings), standard LLM architectures are not optimized for similarity learning at scale. LLMs can leverage contrastive learning (e.g., OpenAI's Contrastive Language-Image Pre-Training (CLIP)), but this requires additional embedding-based training. Therefore, adaptation is needed to integrate contrastive learning with LLMs. Task-agnostic meta-learning assumes task distributions are unknown during training, making it less compatible with current LLM fine-tuning methods. LLMs would need hierarchical task embeddings to generalize across unseen classification tasks.

In summary, the review of related work has shown that the research on using LLMs for fault diagnosis reported in the literature discussed little, utilizing the power of pre-trained LLMs to achieve  $K$ -way  $N$ -shot learning for bearing fault diagnosis. This paper aims to bridge the gap by leveraging the power of pre-trained LLMs such as GPT-2 to achieve  $K$ -way  $N$ -shot learning for full ceramic bearing fault diagnosis. By

extracting AE features and utilizing the pre-trained GPT-2 model, we effectively transfer knowledge from a broad pretraining phase to a highly specialized fault diagnosis task. This approach reduces reliance on large, labeled datasets, enhances model generalization, and demonstrates the potential of LLMs to address critical challenges in industrial applications

### 3 The Methodology

Our proposed approach uses  $K$ -way  $N$ -shot learning with a pre-trained LLM to train a classifier with extracted AE features from a few labeled examples for ceramic bearing fault diagnosis. By leveraging the power of the pre-trained LLM deep structure and  $K$ -way  $N$ -shot learning, we can effectively diagnose faults in ceramic bearings with minimal labeled data. The framework of the proposed methodology is shown in Fig. 1. As shown in Fig. 1, the collected AE signals are first processed using the empirical mode decomposition (EMD) method into multiple intrinsic mode components (IMFs), which are used to generate the AE features. Since GPT-2 was developed for text learning only, the AE features are converted into text to fine-tune the pre-trained GPT-2 model as a classifier using  $K$ -way  $N$ -shot learning. Text normalization is the process of converting numbers, symbols, and other non-textual data into their corresponding textual representations. For example, a numerical value like “98.6” can be converted to “ninety-eight point six” using a text normalization technique. Once the classifier is trained, it will be applied to incoming AE features with unknown faults to identify the bearing faults. The pre-trained GPT-2 model will be fine-tuned using  $K$ -way  $N$ -shot learning. In few-shot learning,  $K$ -way  $N$ -shot learning is a typical experimental setup used to evaluate a model’s ability to generalize from very few examples. This notation refers to  $K$  number of classes and  $N$  number of examples per class used during the learning process.



**Figure 1:** The framework of the proposed methodology

The structure of GPT-2 used in the proposed methodology is shown in Fig. 2. GPT-2 uses a multi-layer transformer decoder to model the context of sequences and predict the next token in a given input. This approach enables GPT-2 to learn language’s semantic and syntactic structures while generalizing across diverse tasks. Fig. 3 shows the fine-tuning of the pre-trained GPT-2 model.

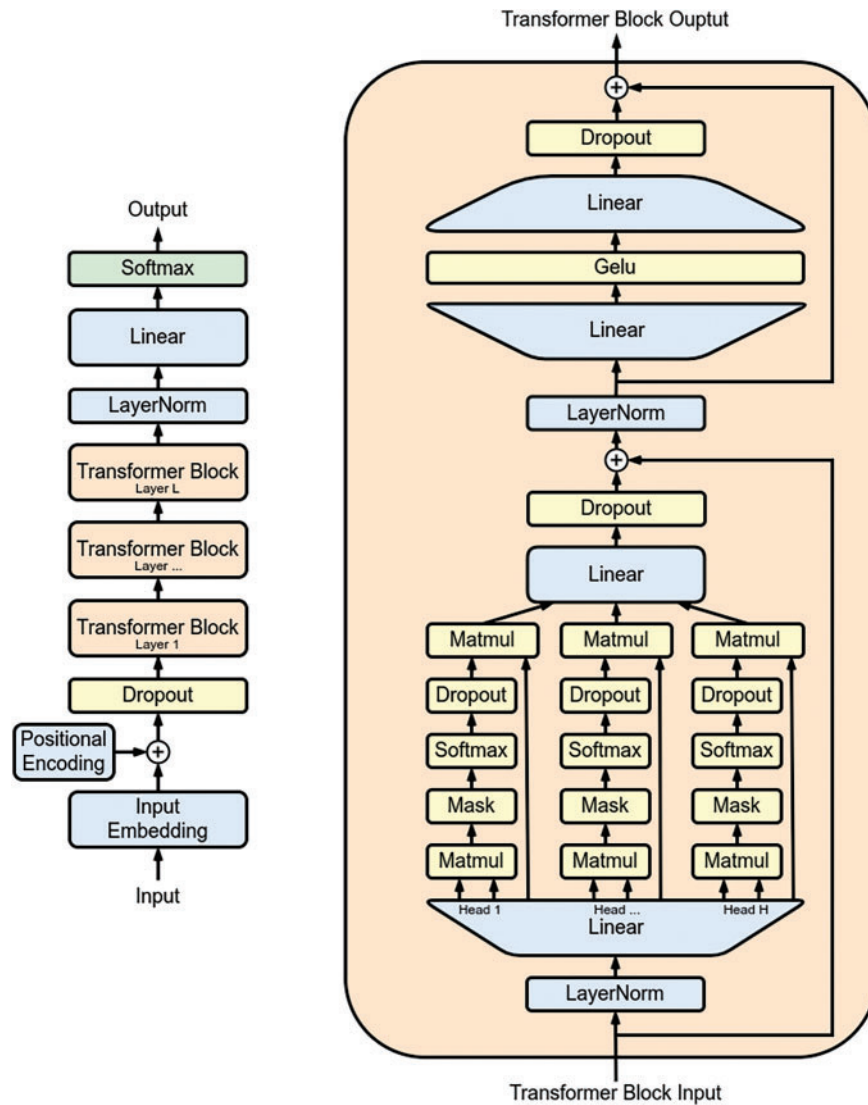
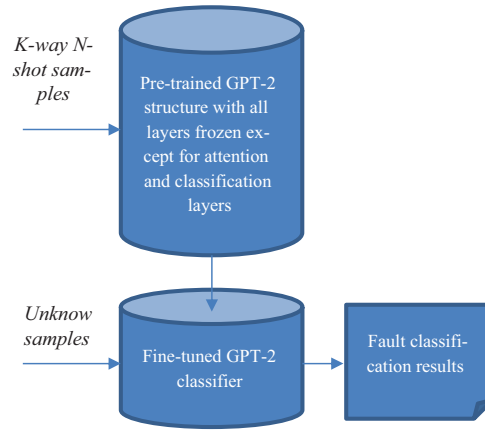


Figure 2: The structure of GPT-2



**Figure 3:** The fine-tuning of GPT-2 model

Two effective approaches have been suggested in the literature to achieve better transfer learning performance for fine-tuning the pre-trained deep structure [26–28]. 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 classification 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:

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

Since the sample size in few-shot learning is small, fine-tuning 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:

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

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

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

To modify the softmax activation function, define:

$\mathbf{x}$  = the test samples

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

$\mathbf{b}$  = the bias

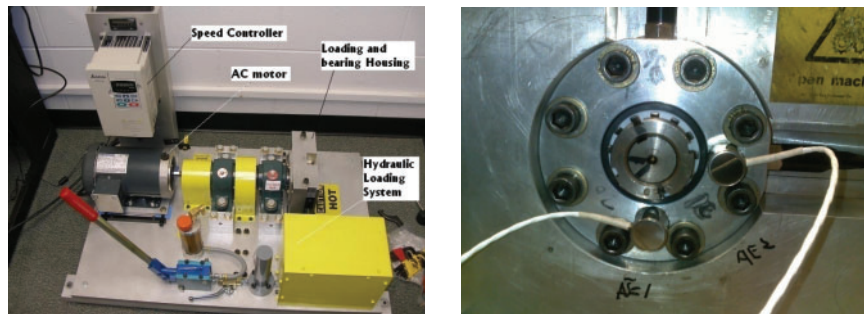
The softmax activation function can be modified with the cosine similarity as:

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

## 4 Experiment Setup and Results

### 4.1 The Dataset

To evaluate the performance of the proposed methodology for full ceramic bearing fault diagnosis, the AE signal dataset collected during seeded fault tests performed on a bearing test rig in the laboratory is used. Fig. 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 wideband (WD) type AE sensors and a 2-channel data acquisition card with 18-bit resolution and a maximum sampling rate of up to 40 MHz were used to collect the AE burst data. The AE sensors were attached to the bearing housing with instant glue.

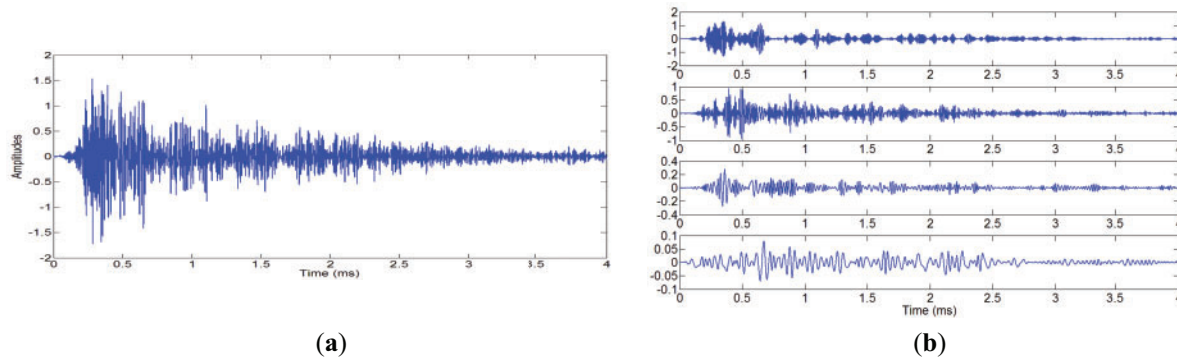
During the test, bearings with the following seeded faults were run on the test rig to collect the AE signals: inner race fault, outer race fault, ball fault, and cage fault (see Fig. 5). The speed of the motor shaft was controlled at 10 Hz (600 rpm), and the AE signals were collected at a sampling rate of 5 MHz. A more detailed description of the equipment parameters used in the experiment can be found in [1,2,4].



**Figure 5:** Bearing seeded faults



The EMD method decomposed each collected AE signal into several IMF components. An example of the bearing inner race fault AE signal and the first 4 IMF components of the signal are provided in Fig. 6.



**Figure 6:** An example of the bearing inner race fault AE signal and its IMF components. (a) AE signal of inner race fault; (b) The first 4 IMF components of the signal in (a)

To extract the AE features, the three IMF components were summed and then the following values were extracted from the summed IMF components: *rms*, *peak value*, and *kurtosis*. 7 AE features were formed from these values, 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 bearing fault and a normal bearing, a total of 40 data points were generated, with each data point represented as a feature described in Table 1. Therefore, 200 data points were available for the data analysis.

#### 4.2 The Analysis Results

To fine-tune the pre-trained GPT-2 model, *K*-way *N*-shot samples were randomly generated without replacement from 80% of the dataset, while the remaining 20% was used for validation.

The following parameters were set up for the fine-tuning of GPT-2, 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 = $1 \times 10^{-8}$
Learning rate	$3 \times 10^{-5}$

In this paper, since we have five bearing types: normal, ball fault, cage fault, outrace fault, and inner race fault, we set  $K = 5$ . In few-shot learning, the goal is to use a minimal number of samples to achieve the desired diagnosis accuracy. In our experimental analysis, we started with  $N = 1$ . Then, we increased  $N$  until a satisfactory accuracy was achieved. At this point, the  $N$  value represents the minimal number of samples to achieve satisfactory diagnosis accuracy. The results are presented in Table 3. Note that the values in Table 3 represent the average of 10 computational tests with the margin of error for its 90% confidence interval.

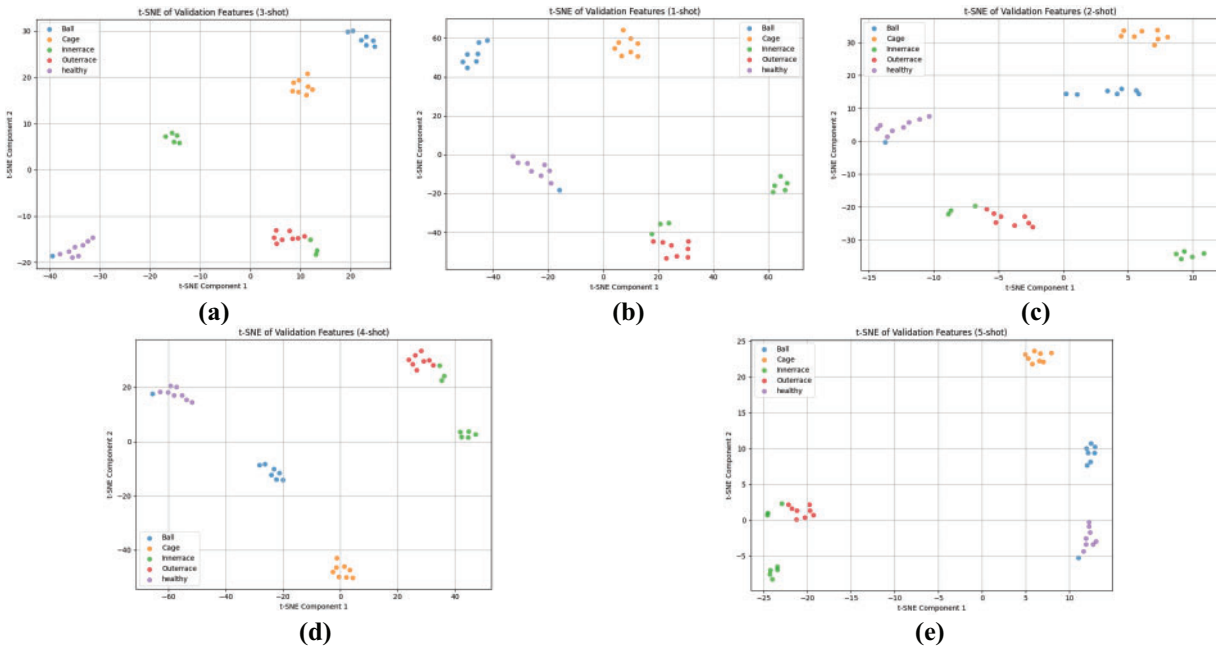
**Table 3:** Results of 5-way  $N$ -shot learning with different  $N$  values

$N$	Old loss + Softmax		New loss + Softmax	
	Validation accuracy (%)	Training accuracy (%)	Validation accuracy (%)	Training accuracy (%)
1	23.13 ± <b>0.30</b>	25.00 ± <b>0.28</b>	53.34 ± <b>0.31</b>	57.5 ± <b>0.27</b>
2	28.39 ± <b>0.28</b>	35.00 ± <b>0.26</b>	62.10 ± <b>0.29</b>	67.5 ± <b>0.27</b>
3	33.33 ± <b>0.26</b>	45.00 ± <b>0.25</b>	74.33 ± <b>0.27</b>	75.17 ± <b>0.26</b>
4	39.26 ± <b>0.22</b>	51.00 ± <b>0.21</b>	81.54 ± <b>0.21</b>	82.32 ± <b>0.20</b>
5	42.29 ± <b>0.22</b>	61.89 ± <b>0.21</b>	92.34 ± <b>0.23</b>	92.71 ± <b>0.22</b>

The results in Table 3 show that modifying the loss function with entropy regularization and the softmax function with cosine similarity significantly improves the few-shot learning for full ceramic bearing fault diagnosis. As we increase the number of shots, the diagnostic accuracy increases. Using up to as few as 5 samples, the LLM-based fault diagnostic model could generate as good performance as other machine learning methods, with 160 samples reported in [2]. Fig. 7 shows the improvement in the features obtained by the fine-tuned GPT-2 model using the  $t$ -SNE plots as  $N$  increases from 1 to 5. As shown in Fig. 7, as  $N$  increases from 1 to 5, the 5 bearing classes formed by the features extracted from the inputs into the classification layer become increasingly clear.

The performance of the presented approach was compared with a few competitive machine-learning algorithms. These algorithms were KNN, LAMSTAR neural network, deep neural network (DNN), RNN, and LSTM network. The main reason for choosing these machine learning models is that they have been reported to give good classification performance [2,4,29–31]. For a fair comparison, 5-way 5-shot samples were used to train the compared algorithms since the presented approach gave the best performance using the 5-way 5-shot samples. To set up the computational experiment for comparison, models were constructed based on the information provided in the literature: KNN in [2], LAMSTAR in [4], DNN in [29], RNN in [30], and LSTM in [31]. In addition, 5-way 5-shot meta-learning implemented with model-agnostic meta-learning (MAML) [13,14] and MAML + GPT-2 were compared with the presented method. The same parameters for fine-tuning GPT-2 in the proposed method shown in Table 2, were used to fine-tune the MAML +

GPT-2 model. A moderate number of tasks was set as 5 for both MAML and MAML + GPT-2. Note that the values in Table 4 represent the average of 10 computational tests with the margin of error for its 90% confidence interval.



**Figure 7:** The *t*-SNE plots on the inputs to the classification layer for different *N* shot values. (a) 3-shot; (b) 1-shot; (c) 2-shot; (d) 4-shot; (e) 5-shot

**Table 4:** Diagnostic performance comparison results with 5-way 5-shot learning

Models	Diagnostic accuracy (%)
Presented method	92.34 ± 0.23
MAML + GPT-2	92.13 ± 0.21
MAML [14]	90.12 ± 0.22
LAMSTAR neural network [4]	90.26 ± 0.24
KNN [2]	90.12 ± 0.30
DNN [29]	84.24 ± 0.26
RNN [30]	85.31 ± 0.25
LSTM [31]	86.72 ± 0.25

As shown in Table 4, compared with conventional machine learning methods such as KNN, DNN, RNN, and LSTM, the proposed method outperforms these methods for 5-way 5-shot learning. Also, the proposed method outperforms MAML, a major meta-learning player for *K*-way *N*-shot learning. When compared with MAML + GPT-2, the proposed method gives a comparable performance. This shows the power of integrating few-shot learning techniques with LLMs. However, as discussed in Section 2, integrating meta-learning techniques such as MAML with LLMs demands higher computational effort for multiple inner-loop optimizations in model training, which makes it less attractive than the proposed method if the computational resource is a concern.

## 5 Discussion

It is interesting to note that even though, given the AE features extracted, the proposed method improves the fault diagnosis accuracy by integrating the LLMs, there is still some room for accuracy improvement. One direction is to improve the feature extraction process by adopting a better approach. Chauhan et al. [32] introduced the adaptive feature mode decomposition (FMD) technique to process vibration signals for bearing fault diagnosis. Their results have shown that the FMD technique generates better feature extraction results than ensemble EMD (EEMD), which is an improved version of EMD. The FMD could be used to extract more effective AE features for the proposed method in this paper.

Some limitations of the proposed method are worth discussing. To leverage the power of the pre-trained GPT-2 model for full ceramic bearing fault diagnosis, numerical AE features must be converted into text using text normalization. However, this process can introduce several types of information loss. These include: (1) Loss of structure and readability: for example,  $x = 3(2 + 4) \rightarrow$  “ $x$  equals three times the sum of two and four”. (2) Loss of computational efficiency: text is harder to process computationally. Operations like sorting, filtering, and arithmetic become impractical. (3) Increased data redundancy and storage overhead: words take more space than numerical values. For example, 1,000,000 (requires 4 bytes in storage)  $\rightarrow$  “one million” (requires more than 4 bytes in storage). A tradeoff exists between gaining the pre-trained LLMs’ power and losing information. This tradeoff between harnessing pre-trained LLMs and minimizing information loss depends on the specific context and application. Our study demonstrates that the advantages of the pre-trained GPT-2 model outweigh these limitations, effectively compensating for potential data loss.

Integrating few-shot techniques with LLMs like GPT-2 offers significant advantages in bearing fault diagnosis tasks. However, it’s essential to understand how it compares in terms of computational costs to traditional methods such as KNN, DNN, RNN, LSTM, and MAML. It can lead to higher training costs due to the complexity and size of the LLMs. Once trained, LLMs can perform inference efficiently. However, the large number of parameters in LLMs can still pose challenges in resource-constrained environments.

## 6 Conclusion

This paper introduced a novel few-shot learning approach for full ceramic bearing fault diagnosis using acoustic emission (AE) signals and large language models (LLMs). Specifically, we leveraged the pre-trained GPT-2 model, integrating  $K$ -way  $N$ -shot learning, to train a classifier capable of diagnosing bearing faults with minimal labeled data. To further enhance the model’s effectiveness, we employed modified loss functions and softmax activation with cosine similarity, improving the generalization of GPT-2 in this domain.

The experimental evaluation on a laboratory-collected full ceramic bearing dataset demonstrated that the proposed method could achieve high diagnostic accuracy with as few as five labeled samples. Compared to conventional machine learning approaches such as KNN, LAMSTAR, DNN, CNN, LSTM, and meta-learning approach such as MAML, our approach not only outperformed these methods in terms of accuracy but also reduced the dependency on extensive labeled datasets. Furthermore, the results showed that LLMs, when integrated with few-shot learning techniques, can effectively capture complex fault patterns, providing a cost-effective and adaptive solution for industrial applications.

While the proposed approach significantly enhances fault diagnosis performance, some challenges remain, including potential information loss during text normalization and computational overhead in fine-tuning large-scale LLMs. Future work will explore alternative feature extraction techniques, such as adaptive feature mode decomposition (FMD), to improve AE signal representations, as well as investigate

more efficient ways of integrating LLMs with structured numerical data. Additionally, optimizing the computational efficiency of fine-tuning LLMs for industrial fault diagnosis will be an important area of future research.

Overall, this study underscores the transformative potential of LLMs in predictive maintenance and fault diagnosis, demonstrating their ability to learn from limited data, adapt to diverse fault scenarios, and reduce reliance on extensive supervised learning datasets. By bridging the gap between LLMs and  $K$ -way  $N$ -shot learning, this work lays the foundation for next-generation AI-driven diagnostic systems that can operate with higher adaptability, efficiency, and accuracy in real-world industrial applications.

**Acknowledgement:** Not applicable.

**Funding Statement:** The authors received no specific funding for this study.

**Author Contributions:** The authors confirm contribution to the paper as follows: Conceptualization, David He, Miao He and Jay Yoon; methodology, David He; validation, David He; formal analysis, David He; investigation, David He, Miao He and Jay Yoon; resources, Miao He and Yay Yoon; data curation, David He; writing—original draft preparation, David He; writing—review and editing, Miao He and Jay Yoon; visualization, David He; supervision, David He; project administration, Miao He and Yay Yoon; funding acquisition, Miao He and Yay Yoon. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest to report regarding the present study.

## References

1. Li R, He D, Zhu J. Investigation on full ceramic bearing fault diagnostics using vibration and AE sensors. In: 2012 IEEE Conference on Prognostics and Health Management; 2012 Jun 18–21; Denver, CO, USA. doi:10.1109/ICPHM.2012.6299517.
2. He D, Li R, Zhu J, Zade M. Data mining based full ceramic bearing fault diagnostic system using AE sensors. IEEE Trans Neural Netw. 2011;22(12):2022–31. doi:10.1109/TNN.2011.2169087.
3. Saucedo-Dorantes JJ, Arellano-Espitia F, Delgado-Prieto M, Osornio-Rios RA. Diagnosis methodology based on deep feature learning for fault identification in metallic, hybrid and ceramic bearings. Sensors. 2021;21(17):5832. doi:10.3390/s21175832.
4. Yoon JM, He D, Qiu B. Full ceramic bearing fault diagnosis using LAMSTAR neural network. In: 2013 IEEE Conference on Prognostics and Health Management (PHM); 2013 Jun 24–27; Gaithersburg, MD, USA. doi:10.1109/ICPHM.2013.6621427.
5. He D, He M, Taffari A. Few-shot learning for full ceramic bearing fault diagnosis with acoustic emission signals. PHMAP23 Conf. 2023;4(1):1–7. doi:10.36001/phmap.2023.v4i1.3752.
6. Xie Z, Zhan H, Wang Y, Zhan C, Wang Z, Jia N. Meta-learning-based fault diagnosis method for rolling bearings under cross-working conditions. Meas Sci Technol. 2025;36(1):16218. doi:10.1088/1361-6501/ad916a.
7. Wang X, Jiang B, Xiao L, Ma L. Enhanced meta-transfer learning for few-shot fault diagnosis of bearings with variable conditions. In: 2023 6th International Symposium on Autonomous Systems (ISAS); 2023 Jun 23–25; Nanjing, China. doi:10.1109/ISAS59543.2023.10164544.
8. Liu S, Chen J, He S, Shi Z, Zhou Z. Few-shot learning under domain shift: attentional contrastive calibrated transformer of time series for fault diagnosis under sharp speed variation. Mech Syst Signal Process. 2023;189(10):110071. doi:10.1016/j.ymsp.2022.110071.

9. Hu SX, Li D, Stühmer J, Kim M, Hospedales TM. Pushing the limits of simple pipelines for few-shot learning: external data and fine-tuning make a difference. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. p. 9058–67. doi:10.1109/CVPR52688.2022.00886.
10. Li X, Su H, Xiang L, Yao Q, Hu A. Transformer-based meta learning method for bearing fault identification under multiple small sample conditions. *Mech Syst Signal Process*. 2024;208(12):110967. doi:10.1016/j.ymsp.2023.110967.
11. Zeng L, Jian J, Chang X, Wang S. A meta-learning method for few-shot bearing fault diagnosis under variable working conditions. *Meas Sci Technol*. 2024;35(5):056205. doi:10.1088/1361-6501/ad28e7.
12. Chen J, Hu W, Cao D, Zhang Z, Chen Z, Blaabjerg F. A meta-learning method for electric machine bearing fault diagnosis under varying working conditions with limited data. *IEEE Trans Indus-Trial Inform*. 2023;19(3):2552–64. doi:10.1109/TII.2022.3165027.
13. Feng Y, Chen J, Xie J, Zhang T, Lv H, Pan T. Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: algorithms, applications, and prospects. *Knowl Based Syst*. 2022;235(8):107646. doi:10.1016/j.knosys.2021.107646.
14. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning; 2017 Aug 6–11; Sydney, Australia: PMLR. p. 1126–35. doi:10.48550/arXiv.1703.03400.
15. Zheng X, Feng Z, Lei Z, Chen L. Few-shot learning fault diagnosis of rolling bearings based on Siamese network. *Meas Sci Technol*. 2024;35(9):095018. doi:10.1088/1361-6501/ad57d9.
16. Wu G, Zhao L, Wu H, Zhang N, Zhang X. Few-shot bearing fault diagnosis based on ResAttn-Siamese. In: 2024 8th CAA International Conference on Vehicular Control and Intelligence (CVCI); 2024 Oct 25–27; Chongqing, China. doi:10.1109/CVCI63518.2024.10830042.
17. An Z, Wang H, Yan Y, Jia S, Wang L, Yang R. Contrast learning with hard example mining for few-shot fault diagnosis of rolling bearings. *Meas Sci Technol*. 2024;35(10):106121. doi:10.1088/1361-6501/ad5fac.
18. Qiu C, Tang T, Yang T, Chen M. Learning to generalize with latent embedding optimization for few- and zero-shot cross domain fault diagnosis. *Expert Syst Appl*. 2024;254(11):124280. doi:10.1016/j.eswa.2024.124280.
19. Yang X, Zhang L, Wang J. Task-agnostic generalized meta-learning based on MAML for Few-shot bearing fault diagnosis. In: Lu H, Ouyang W, Huang H, Lu J, Liu R, Dong J, et al., editors. *Image and graphics*. Berlin/Heidelberg, Germany: Springer; 2023. p. 118–29. doi:10.1007/978-3-031-46305-1\_10.
20. Li A, Luo T, Xiang T, Huang W, Wang L. Few-shot learning with global class representations. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 9714–23. doi:10.1109/iccv.2019.00981.
21. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. San Francisco, CA, USA: OpenAI; 2018. [cited 2025 Mar 12]. Available from: [https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf).
22. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. *OpenAI Blog*. 2019;1(8):9.
23. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst*. 2020;3:1877–901. doi:10.5555/3495724.3495883.
24. Tao L, Liu H, Ning G, Cao W, Huang B, Lu CY. LLM-based framework for bearing fault diagnosis. *Mech Syst Signal Process*. 2025;224(3):112127. doi:10.1016/j.ymsp.2024.112127.
25. Qaid HAAM, Zhang B, Li D, Ng SK, Li W. FD-LLM: large language model for fault diagnosis of machines. *arXiv: 2412.01218*. 2024.
26. Dhillon GS, Chaudhari P, Ravichandran A, Soatto S. A baseline for few-shot image classification. *arXiv: 1909.02729*. 2019.
27. Chen T, Liu S, Kira Z, Wang Y, Huang J. A closer look at few-shot classification. In: International Conference on Learning Representations (ICLR); 2019 May 6–9; New Orleans, LA, USA.
28. Chen Y, Liu Z, Xu H, Darrell T, Wang X. Meta-baseline: exploring simple meta-learning for few-shot learning. *arXiv:2003.04390*. 2021.

29. Wen X, Xu Z. Wind turbine fault diagnosis based on ReliefF-PCA and DNN. *Expert Syst Appl.* 2021;178(11):115016. doi:10.1016/j.eswa.2021.115016.
30. An Z, Li S, Wang J, Jiang X. A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network. *ISA Trans.* 2020;100:155–70. doi:10.1016/j.isatra.2019.11.010.
31. Zou P, Hou B, Lei J, Zhang Z. Bearing fault diagnosis method based on EEMD and LSTM. *Int J Comput Commun Control.* 2020;15(1):1–14. doi:10.15837/ijccc.2020.1.3780.
32. Chauhan S, Vashishtha G, Kumar R, Zimroz R, Gupta MK, Kundu P. An adaptive feature mode decomposition based on a novel health indicator for bearing fault diagnosis. *Measurement.* 2024;226:114191. doi:10.1016/j.measurement.2024.114191.