Intelligent Fault Diagnosis of Rolling Bearing based on Incremental Learning

Zhiyu Liu
SJTU Paris Elite Institute of Technology
Shanghai Jiao Tong University
Shanghai, China
lizhyuxi@outlook.com

Zhiyi Zhang
SJTU Paris Elite Institute of Technology
Shanghai Jiao Tong University
Shanghai, China
robin1412@sjtu.edu.cn

Mohamed Sallak
Laboratoire Heudiasyc UMR CNRS 7253
Université de Technologie de Compiègne
Compiègne, France
mohamed.sallak@hds.utc.fr

Siqi Qiu*
SJTU Paris Elite Institute of Technology
Shanghai Jiao Tong University
Shanghai, China
siqiqiu@sjtu.edu.cn
**Corresponding author*

***Abstract*—Fault diagnosis in rolling bearings presents considerable challenges due to the dynamic and complex nature of machine operations. In real-world scenarios, systems continually accumulate new data, often including previously unknown fault types. Traditional diagnostic methods frequently struggle to adapt to these evolving fault patterns, making it difficult to maintain accuracy and reliability in fault detection. To address these challenges, this paper proposes the VMD-Enhanced Gradient Episodic Memory Model (VEGEM). Firstly, Variational Mode Decomposition (VMD) is utilized for data preprocessing, which effectively decomposes signals into key frequency components, isolating essential features and reducing noise. Secondly, the Wide Deep Convolutional Neural Network (WDCNN) is employed as the baseline framework for fault diagnosis, leveraging its deep learning capabilities to handle vibration signals with high precision. Thirdly, the incremental learning framework is enhanced using Gradient Episodic Memory (GEM), a constrained replay technique that manages gradient updates to prevent catastrophic forgetting. This method ensures new learning does not disrupt previously acquired knowledge. The proposed VEGEM model was rigorously tested on the CWRU dataset, demonstrating great robustness and adaptability. It effectively incorporated new fault categories and maintained an accuracy rate above 90% across five incremental learning phases. These results confirm that the VEGEM model can effectively resolve the issues of catastrophic forgetting in fault diagnosis, providing a reliable, efficient, and adaptable solution for the incremental learning requirements of modern industrial environments.**

Keywords—Rolling Bearings, Intelligent Fault Diagnosis, Continuous Wavelet Transform, Mode Decomposition, Incremental Learning

I. Introduction

Rolling bearings play a pivotal role in the seamless operation of industrial machinery, underpinning the success of modern industrial automation. As essential components, the reliability of rolling bearings directly influences the efficiency and safety of entire production processes. However, the inherent complexities of mechanical operations and external environmental factors contribute to frequent bearing failures, presenting substantial challenges for maintenance and operational continuity.

Traditional fault diagnosis methods, such as vibration analysis and acoustic diagnosis, rely heavily on manual interpretation of data and often fail to adapt to the dynamic nature of industrial environments. These methods struggle with the rapid evolution of machinery conditions and the emergence of new types of faults. This research aims to address these challenges by developing an incremental learning model that can continuously adapt to new fault patterns and evolve with the machinery it monitors, thereby enhancing diagnostic accuracy and system reliability.

Incremental learning is a pivotal approach in modern machine learning, uniquely designed to update diagnostic models incrementally as new data becomes available. This method allows for fine-tuning models with new information without the necessity to retrain them from scratch. A fundamental advantage of incremental learning is its capability to mitigate catastrophic forgetting—a prevalent challenge in traditional learning models where integrating new knowledge often results in the loss of previously acquired information. By preserving existing knowledge while integrating new data, incremental learning significantly enhances the adaptability of systems to evolving fault conditions.

Incremental learning encompasses various scenarios [1] such as Class Incremental Learning, where the model's output space expands over time to include new categories without forgetting old ones; Task Incremental Learning, focusing on the model's ability to learn different sets of categories across successive tasks without sharing information between them; and Domain Incremental Learning, which deals with changes in the input data distribution while the output categories remain constant. These scenarios are crucial in defining the adaptability of incremental learning models to new information while preserving existing knowledge, whether adapting to new classes, distinguishing between tasks, or responding to changing environments without relearning known categories.

Methods of incremental learning are broadly classified into three types: regularization methods, which mitigate forgetting by adding constraints to the loss function to preserve old knowledge [2]; rehearsal methods, which involve retraining the model on a mix of old and new data [3]; and parameter isolation methods, where some of the model parameters are frozen during the training process [4]. These methods enable the model to assimilate new information while retaining previously acquired knowledge, effectively emulating continual human learning.

National Natural Science Foundation of China (Grant No. 52175476 and 51805326), Designing and Verifying Key Technologies of Marine Engines High Reliability in Major Special Project Ship Research Program of MIIT.

This paper proposes VMD-Enhanced Gradient Episodic Memory Model (VEGEM), which focuses on Class Incremental Learning and adopts the Gradient Episodic Memory [5] (GEM), a constrained replay approach that strategically manages gradient updates to prevent catastrophic forgetting. This technique involves restricting gradient adjustments during training to ensure new learning does not disrupt previously acquired knowledge. Incorporating Variational Mode Decomposition [6] (VMD) significantly enhances the model's performance by effectively decomposing signals into key frequency components. This preprocessing step isolates essential features and reduces noise, enabling the GEM model to focus on relevant data. Consequently, VMD helps mitigate catastrophic forgetting, ensuring stable and distinctive feature extraction. The synergy between VMD and GEM boosts the robustness and accuracy of the fault diagnosis system across incremental learning phases.

II. METHODOLOGY

A. Data Collection and Preprocessing

To ensure the reliability and robustness of the fault diagnosis model, comprehensive data collection and meticulous preprocessing are crucial. The data used in this study were primarily derived from bearing fault experiments conducted using a motor test bench setup. This setup is equipped with various sensors that capture vibration signals under different operational conditions to simulate real-world scenarios.

In the approach to data preprocessing, besides VMD, two more methods have been tried to enhance the model's input data quality: Continuous Wavelet Transform [7] (CWT) and Empirical Mode Decomposition [8] (EMD). The comparative results of these preprocessing techniques are presented in the third section.

VMD decomposes a signal into a predefined number of band-limited Intrinsic Mode Functions (IMFs). VMD enforces a bandwidth constraint on the mode functions to enhance the separation of modal components, as in (1), where u_k is the mode, $\hat{u}_k(\omega)$ is the corresponding frequency domain representation, ω_k is the central frequency, and τ is the delay of the data. It is particularly useful for isolating and identifying overlapping frequencies that may indicate faults.

$$min_{\{u_k\},\{\omega_k\}}\left\{\sum_{k=1}^{K}\int\left(\delta(t)-\frac{\partial}{\partial t}\left[\left(\frac{\hat{u}_k(\omega)}{e^{-jw\tau}}\right)\right]\right)^2 d\omega\right\} \quad (1)$$

These preprocessing techniques can extract meaningful features from raw vibration data. The features extracted are then utilized to train the incremental learning model, ensuring that it can effectively learn and adapt to new fault conditions without being misled by noise or irrelevant variations in the data.

B. Model Development

To tackle the challenges of intelligent fault diagnosis in rolling bearings, a robust incremental learning model was developed. This model leverages a combination of deep learning techniques and incremental learning strategies to ensure high adaptability and accuracy over time.

As shown in Fig. 1, the backbone of the model is a Wide Deep Convolutional Neural Network [9] (WDCNN), meticulously crafted to handle vibration signals from rolling bearings with high precision. The architecture comprises seven weight layers, including six convolutional layers and one fully connected layer, arranged to progressively refine feature maps and enhance fault detection capabilities. The initial convolutional layer utilizes large 64x64 kernels to capture a broad spectrum of features from the input data, while subsequent layers employ 3x3 kernels to extract finer details. Each convolutional layer is followed by a 2x2 max pooling layer that reduces data dimensionality and highlights essential features for accurate fault classification. The activation functions across all hidden layers are Rectified Linear Units (ReLU), which introduce necessary non-linearity efficiently, aiding the network in learning complex fault patterns without compromising training speed. Following the convolutional and pooling layers, the network transitions into a fully connected layer that culminates in a softmax output layer, categorizing input signals into various fault types based on the features learned. The model employs cross-entropy as its loss function, as shown in (2), where $\hat{y}_n$ is the softmax output of the model, and y_n is the one-hot true label vector of the input data.. Cross-Entropy loss can effectively quantify the discrepancy between predicted probabilities and the actual classifications, thereby optimizing for minimal classification errors in fault diagnosis.

$$\mathcal{L}_{CE}(\mathbf{y}_n,\hat{\mathbf{y}}_n) = -\mathbf{y}_n \cdot \log\hat{\mathbf{y}}_n \quad (2)$$

The incremental learning component of the model leverages the GEM algorithm. As shown in Fig. 2, it utilizes a replay approach where the model is periodically trained on a subset of old training data alongside new data. This process ensures that the model retains familiarity with historical fault patterns while continuously adapting to new ones. The key aspect of GEM is its ability to maintain performance on previously learned tasks by managing how new knowledge is integrated. This is achieved through a unique gradient projection method, which is central to the GEM approach.

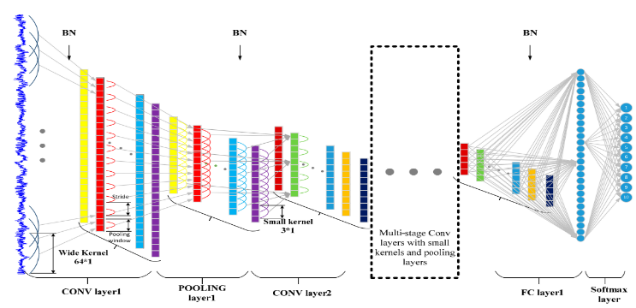

Fig. 1. WDCNN architecture [9]

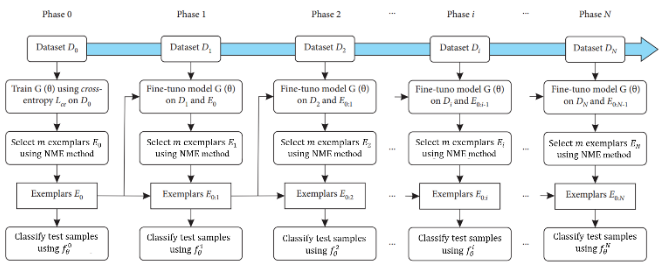

Fig. 2. Replay model training pipeline

Initially, WDCNN model $G(\theta)$ is trained on dataset D_0 using the cross-entropy loss L_{CE}. Subsequently, using the Nearest-Mean-of-Exemplars (NME), an exemplar set E_0 is constructed with the aid of the model′s feature extractor φ. For each incremental phase i, the CNN′s output layer randomly initializes weights for each new class. The model is then fine-tuned on D_i and $E_{0:i-1}$, where $D_i \cup E_{0:i-1}$ is inherently unbalanced because $E_{0:i-1}$ contains only a small sample of previous classes. All parameters θ within the network are updated in each incremental process. At the end of each incremental phase i, the fine-tuned model predictor f_θ^i is used to predict the test set samples.

The Nearest-Mean algorithm [3] constructs the exemplar set using the feature extractor φ, where the mean is the average of the feature vectors, and the feature vectors are obtained from the feature extraction layer of the model (φ, excluding the final decoding layer). Whenever the model encounters a new class, its exemplar set is adjusted. In this scenario, all classes are treated equally; thus, when T classes have been observed thus far, and M is the total number of samples that can be stored, the model uses $m = M/T$ samples for each class (rounded). This ensures that the memory budget for M samples is always fully utilized but never exceeded.

Algorithm 1 NME CONSTRUCT EXEMPLAR SET

input image set $X = \{x_1, \dots, x_n\}$ of class y
input m target number of exemplars
require current feature function $\varphi: \mathcal{X} \to \mathbb{R}^d$

$$\mu \leftarrow \frac{1}{n}\sum_{x\in X} \varphi(x) \quad \text{//current class mean}$$

for $k = 1, \dots, m$ **do**

$$p_k \leftarrow \underset{x\in X}{\operatorname{argmin}} \left\| \mu - \tfrac{1}{k}[\varphi(x) + \textstyle\sum_{j=1}^{k-1} \varphi(p_j)] \right\|$$

end for
$P \leftarrow (p_1, \dots, p_m)$
output exemplar set P

Algorithm 2 REDUCE EXEMPLAR SET

input m // target number of exemplars
input $P = (p_1, \dots, p_{|P|})$ // current exemplar set
$P \leftarrow (p_1, \dots, p_m)$ // i.e. keep only first m
output exemplar set P

The construction and management of the exemplar sets are governed by two algorithms: one selects exemplars for new classes, and the other reduces the size of the exemplar sets for previous classes. Algorithm 1 outlines the steps for exemplar selection. Exemplars $p_1, \dots, p_m$ are chosen and stored until the target number m is satisfied. At each iteration, a sample from the current training set is added to the exemplar set such that the average feature vector of all exemplars closely matches the average feature vector of all training data. Thus, the exemplar set essentially acts as a priority list where the order of elements signifies their importance. The details of updating the exemplar set are further elaborated in Algorithm 2, where to reduce the number of exemplars for a class from any m' to m, exemplars $p_{m+1}, \dots, p_{m'}$ are discarded, retaining only $p_1, \dots, p_m$.

In the GEM algorithm, gradients computed from the new data are projected onto the gradients from the old data. This projection ensures that the updates made to the model's weights do not increase the loss on previous tasks, thereby mitigating any potential decline in performance on old data. Specifically, when the model is trained on new tasks, the gradient of the loss function with respect to the old tasks is calculated. If this gradient indicates that performance on the old tasks would worsen, the gradient for the new tasks is adjusted so that it lies within the space that does not harm the old knowledge. This approach not only preserves previously acquired information but also allows the model to continue learning new patterns effectively.

III. EXPERIMENTS

A. Experiment Setup

Experiments were conducted in an environment equipped with an NVIDIA GeForce RTX1660Ti GPU, operating on Windows 10, using Python 3.9 as the programming language, primarily relying on the PyTorch 11.1 library. The model's hyperparameters were set as follows: batch_size = 32, learning_rate = 0.001, epochs = 100, increment_class = 2.

TABLE I shows ten types of data information extracted from the Case Western Reserve University dataset [10], which serve as inputs for the model.

TABLE I. CWRU TEN-CLASS DATA INFORMATION

Class Label	Motor Load	Fault Diameter	Fault Location
C1	0	Health	/
C2	0	0.007"	Inner Race
C3	0	0.007"	Ball
C4	0	0.007"	Outer Race (@6:00)
C5	0	0.014"	Inner Race
C6	0	0.014"	Ball
C7	0	0.014"	Outer Race (@6:00)
C8	0	0.021"	Inner Race
C9	0	0.021"	Ball
C10	0	0.021"	Outer Race (@6:00)

B. *Application and Result Analysis*

All experimental results in this chapter are based on data preprocessed with Variational Mode Decomposition (VMD), as after testing five rounds of training with different preprocessing methods, the GEM model performed best with VMD. TABLE II shows the final model accuracy achieved with different data preprocessing methods. The parameters used for the Continuous Wavelet Transform (CWT) testing were as follows: sampling period of 1/12000, scale length of 128, and the wavelet base function was the complex Morlet wavelet. The number of intrinsic modes for EMD was set to 7, and for VMD, it was set to 4.

VMD led to the most prominent effects in each round of incremental training with 100 epochs, maintaining final accuracies above 90%; EMD was next, with each round of incremental training with 100 epochs resulting in a final accuracy of 89%, while CWT was the lowest, with each round of incremental training with 30 epochs, ending with a final accuracy of 81%. The experimental results show that the VMD preprocessing method can effectively extract features beneficial for model learning. VMD optimizes signal decomposition by minimizing the bandwidth of each mode, possibly capturing key time-frequency characteristics in the signal more effectively, thereby improving the model's adaptability and accuracy to new data. Moreover, features processed by VMD may be more effective in mitigating the forgetting problem in incremental learning, as they provide more stable and distinctive feature representations.

The model's performance during VEGEM model's training, including average accuracy and Cross-Entropy Loss (CE Loss) is presented in Fig. 3. Throughout the model updating process, each incremental training step was set to 100 epochs. The training process demonstrated that the original model trained on the Phase 0 dataset could quickly converge to peak accuracy. In the subsequent four incremental training steps, cross-entropy loss gradually decreased as the epochs progressed, but the model consistently maintained an accuracy above 90%, eventually stabilizing at around 91%.

TABLE II. ACCURACY ACHIEVED BY CWT, EMD, VMD

Preprocessing Method	Proposed(VEGEM)
CWT	81.2%
EMD	87.8%
VMD	90.6%

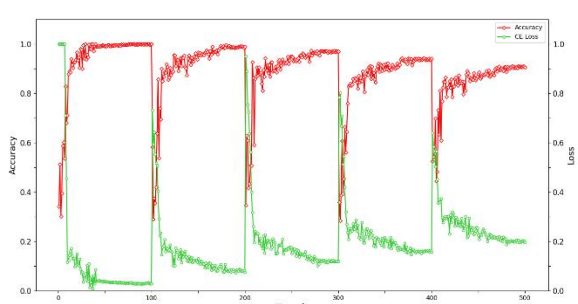

Fig. 3. VEGEM model training process

To analyze the performance of VEGEM in the fault diagnosis of rolling bearings, comparative experiments were conducted, and the method was compared with three types of non-incremental learning and two different variants. TABLE III displays the accuracy changes of various models over 5 phases, with 4 phases of incremental training (P 1 - P 4).

R1, R2 and R3 are related non-incremental learning methods. R1 involves training the model with all known fault data alongside new fault data. This training method is used when fault diagnosis tasks increase without the use of incremental learning techniques, and results from this method are often considered the upper bound for incremental learning models. R2 and R3 employ transfer learning strategies for model training. R2 involves fine-tuning the classifier after freezing the feature extraction layers post training phase 0, whereas R3 fine-tunes the entire network, both experiencing catastrophic forgetting. TABLE III shows that the accuracy of R1 remains consistently at 100%, demonstrating the effectiveness of preprocessing and the WDCNN baseline model in extracting features from bearing fault signals. The accuracies for R2 and R3 drop significantly after each incremental round, with R2 decreasing more slowly because all classifiers in R2 share the same feature extractor with fixed parameters, allowing stable gradient descent across new and old tasks. In contrast, R3, based on new task data, fine-tunes all parameters in the network, resulting in predictions for old tasks being nearly random.

Variation1 and Variation2 are two ablation experiments concerning regularization and restricted replay to validate the advantages of the VEGEM model. Variation1 does not store old task data during training and relies solely on Knowledge Distillation Loss [2] (KD Loss), for incremental training. Variation2 combines KD Loss with Replay for incremental training. TABLE III shows that Variation1 performs the worst with a final accuracy of only 70.2%, due to its reliance solely on KD Loss. Variation2, on the other hand, builds on Variation1 by constructing a typical dataset based on the Nearest-Mean-of-Exemplars algorithm, thus performing old data replay and further reducing catastrophic forgetting. Therefore, Variation2 is more effective than Variation1.

TABLE III. ACCURACY ACROSS 5 PHASES FOR ABLATION STUDY MODELS

Method	P 0	P 1	P 2	P 3	P 4
R1	100%	100%	100%	100%	100%
R2	100%	74.2%	59.6%	48.8%	42.1%
R3	100%	62.5%	44.4%	34.4%	28.8%
Variation1 (CE+KD)	100%	94.1%	87.7%	79.0%	70.2%
Variation2 (CE+KD+Replay)	100%	96.6%	92.7%	88.2%	83.7%
Proposed (VEGEM)	100%	**98.9%**	**97.1%**	**93.9%**	**90.6%**

TABLE IV shows the relationship between final accuracy and memory size in the experiments for VEGEM and Variation2. It is apparent that the final accuracy for VEGEM is an increasing function of memory size, thus not requiring meticulous adjustment of this hyperparameter. VEGEM consistently outperforms Variation2 across multiple memory size ranges

In terms of final accuracy, the R1 method proves that the WDCNN baseline model can effectively handle ten-class bearing fault diagnosis tasks. However, its flaw lies in the need to continuously store all data and retrain the model every round, which is extremely time-consuming. TABLE V shows the total time consumed by each method. For R2 and R3, which fine-tune within the transfer learning context, fewer epochs are actually needed. However, to compare training durations, the time for all methods is noted when the epoch count is set to 100. Variation2 and VEGEM, both replay models, set their memory size to 200. The results indicate that the R1 method is the most time-consuming, requiring 11071 seconds, due to its time consumption being quadratic in relation to the total data volume. Although R1's model performance is stable, its inefficiency may pose a serious obstacle in rapidly changing real industrial environments. R2 and R3, as fine-tuning methods within transfer learning, generally have an advantage in time efficiency. Since R2 freezes the feature extraction layer parameters, its gradient computation is much faster than other methods, requiring only 720 seconds. Variation2, due to the need to compute gradients for additional regularization terms in the loss function and store old data for training, is second only to R1 in full training duration, requiring 7730 seconds. The VEGEM method, by making maximum use of old data to constrain updates, optimizes the training process and takes even less time than the two Variation models, only 4173 seconds.

C. Comparison

TABLE VI showcases a comparison between the VEGEM model and three other selected incremental models. The first model is the Dynamic Weight Allocation [11] (DWA) model. The second model, FT + NCC + H [12], represents a multi-class incremental learning framework that combines Fine-Tuning (FT), Nearest Centroid Classifier (NCC), and a herding method for exemplar selection. The third model is a Lifelong Learning-based Diagnosis Method [13] (LLDM), which integrates a Dual-branch Aggregate Residual Network (DARN) with retained exemplars to overcome catastrophic forgetting and address the dilemma between stability and plasticity.

TABLE IV. ACCURACY WITH DIFFERENT MEMORY SIZES

Memory Size	20	50	100	200	400
Variation2 (CE+KD+Replay)	71.8%	76.3%	80.2%	83.7%	87.2%
Proposed (VEGEM)	74.2%	80.9%	86.3%	**90.6%**	**92.3%**

TABLE V. TRAINING TIME FOR DIFFERENT MODELS

Method	Time(second)
R1	11071
R2	720
R3	2762
Variation1(CE+KD)	5217
Variation2(CE+KD+Replay)	7730
Proposed(VEGEM)	4173

TABLE VI. ACCURACY FOR DIFFERENT MODELS

Method	Accuracy	Phase Numbers
DWA [11]	99.2%	2
	97.7%	3
	95.8%	4
FT+NCC+H [12]	94.8%	4
LLDM [13]	96.5%	2
	96.3%	3
	92.9%	4
Proposed(VEGEM)	**98.9%**	2
	97.1%	3
	93.9%	4
	90.6%	5

As shown in TABLE VI, with the increase in incremental stages, the accuracy of all models declines. However, the Proposed (VEGEM) model still maintains an accuracy of over 90% after five incremental stages. Besides, the VEGEM model consistently outperforms the LLDM model across multiple phases, highlighting its superior learning and retention capabilities. The results for the three phases show that the VEGEM model is within less than 1% difference from the DWA model. Among all models, the DWA model performs best after four phases, thanks to its ability to dynamically allocate weights [11].

IV. CONCLUSION

This research has demonstrated a comprehensive approach to fault diagnosis in rolling bearings, integrating advanced data preprocessing, robust model architecture, and sophisticated incremental learning techniques. Initially, the effectiveness of various data preprocessing methods was assessed. A rigorous comparison of CWT, EMD, and VMD revealed that VMD significantly optimizes the model's ability to process and learn from complex signal data, thereby improving diagnostic accuracy substantially.

Following the preprocessing stage, WDCNN was employed as the baseline model for fault diagnosis. Leveraging its deep learning capabilities, the WDCNN model was meticulously crafted to handle the intricacies of vibration signals from rolling bearings with high precision, establishing a solid foundation for subsequent incremental learning.

The incremental learning framework was then enhanced using the GEM technique, a constrained replay approach that effectively manages gradient updates to prevent catastrophic forgetting. This method ensures that new learning does not disrupt previously acquired knowledge. When applied to the CWRU dataset, VEGEM model demonstrated exceptional adaptability and robustness, effectively incorporating new fault categories without losing accuracy on previously learned categories. Impressively, even after five phases of incremental learning, the VEGEM model maintained an accuracy rate above 90%.

The superior training efficiency of the VEGEM model was also highlighted. It not only outperformed traditional models in terms of diagnostic accuracy but also proved to be more time-efficient during training. By leveraging stored data and employing a method of constraining gradient updates through projection, the VEGEM model efficiently minimized the need for extensive derivations typically required for loss function modifications. This streamlined approach to handling gradients significantly reduces computational demands, making it highly advantageous for industrial applications where time and resource conservation are crucial.

In conclusion, the integration of VMD preprocessing techniques with the robust WDCNN architecture and the innovative GEM incremental learning method in the VEGEM model offers a comprehensive solution to the challenges of intelligent fault diagnosis in rolling bearings. Future research could further this work by exploring hybrid models that combine VEGEM with other advanced machine learning strategies to tackle increasingly complex diagnostic scenarios, paving the way for more adaptive, efficient, and reliable fault diagnosis systems in industrial settings.

REFERENCES

[1] De Lange M, Aljundi R, Masana M, et al. A continual learning survey: Defying forgetting in classification tasks[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(7): 3366-3385.

[2] Li Z, Hoiem D. Learning without forgetting[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(12): 2935-2947.

[3] Rebuffi S A, Kolesnikov A, Sperl G, et al. icarl: Incremental classifier and representation learning[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017: 2001-2010.

[4] Fernando C, Banarse D, Blundell C, et al. Pathnet: Evolution channels gradient descent in super neural networks[J]. arXiv preprint arXiv:1701.08734, 2017.

[5] Lopez-Paz D, Ranzato M A. Gradient episodic memory for continual learning[J]. Advances in neural information processing systems, 2017, 30.

[6] Dragomiretskiy K, Zosso D. Variational mode decomposition[J]. IEEE transactions on signal processing, 2013, 62(3): 531-544.

[7] Morlet J, Arens G, Fourgeau E, et al. Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media[J]. Geophysics, 1982, 47(2): 203-221.

[8] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, 1998, 454(1971): 903-995.

[9] Zhang W, Peng G, Li C, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals[J]. Sensors, 2017, 17(2): 425.

[10] Smith W A, Randall R B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study[J]. Mechanical systems and signal processing, 2015, 64: 100-131.

[11] Hu K, He Q. A New Incremental Training Framework Based on Dynamic Weight Allocation for Intelligent Fault Diagnosis[C]//2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD). IEEE, 2022: 1-6.

[12] Jiménez-Guarneros M, Grande-Barreto J, Rangel-Magdaleno J J. Multiclass incremental learning for fault diagnosis in induction motors using fine-tuning with a memory of exemplars and nearest centroid classifier[J]. Shock and Vibration, 2021, 2021: 1-12.

[13] Chen B, Ding C, Jiang X, et al. Lifelong Learning for Bearing Fault Diagnosis with Incremental Fault Types[C]//2022 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, 2022: 114-12