IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 December 2023, accepted 10 December 2023, date of publication 14 December 2023,
date of current version 22 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343157

RESEARCH ARTICLE

Bidirectional Current WP and CBAR Neural Network Model-Based Bearing Fault Diagnosis

JIANGUANG ZHU AND TANG LIU, (Student Member, IEEE)
School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China
Corresponding author: Jianguang Zhu (zhujg@sut.edu.cn)

ABSTRACT In the era of artificial intelligence, the development of an efficient bearing, fault diagnosis method is of vital importance to ensure smooth production operations and avoid major economic losses. To this end, this paper proposes a bearing fault diagnosis method based on biphasic currents. The method first performs wavelet denoising on the biphasic current signal, then extracts its features by simple vector representation and algebraic operations, and finally, combines the CBAR model of Convolutional Block Attention Module (CBAM) and Residual Network (ResNet) for bearing fault diagnosis. The experimental results show that the highest accuracy rate reaches 100% in both single-point fault and single-point mixed with multiple faults conditions on the open source current bearing fault diagnosis dataset, respectively. Compared with other methods, the method proposed in this paper has the advantage of simple data processing, concise model structure, and high-fault diagnosis accuracy, which provides an effective way for dual-phase current-based bearing fault diagnosis. It is worth emphasizing that based on wavelet denoising, this paper uses the simplest vector representation and algebraic operations to preprocess the signal (WP), making the method more efficient and easy to implement. (Some experiment-related code is posted on the Code Open Source Repository website. https://github.com/LTbig/LT_Bearing_Fault)

INDEX TERMS Fault diagnosis, WP post-processing, biphasic current, CBAR model.

I. INTRODUCTION

The operating condition of the motor plays an essential role in the operation of the entire rail transportation system in the field of rail transportation, and the bearing is an important component indispensable to the operation of various types of equipment, and its health status directly affects the performance and service life of the motor. Failures are unavoidable, however, since bearings are subjected to high-load operation for a prolonged amount of time. Failure to repair deficiencies in a timely way can result in financial losses as well as significant safety issues. As a result, bearing diagnosis is quite crucial.

Li et al. used continuous wavelet transform to convert vibration signals into 2D time-frequency images. They then applied a convolutional neural network to extract features and used maximum mean difference (MMD) to align the global distribution of features for fault diagnosis [1]. Kafeel et al. obtained a three-dimensional vibration signal dataset in a significant induction motor representing both healthy and faulty states, and a hybrid combination of temporal and spectral features using a support vector machine with a Gaussian kernel for classification yielded superior diagnostic results [2]. Altaf et al. developed the time, frequency, and spectral power domain feature vectors of vibration signals and transmitted these feature vectors to the K-nearest neighbor, support vector machine, and kernel linear discriminant analysis for identifying and classifying bearing faults [3]. Zuo et al. introduced a fault diagnosis method for rolling bearings using multi-scale weighted visibility graph (MSWVG) and multi-channel graph convolution network (MCGCN). The method converts vibration signals into multiple weighted graphs and uses MCGCN to extract local node feature and global topology information [4]. These technologies,

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

however, can only identify problems that have already occurred. Additionally, the vibration signal-based method cannot be used in machine systems when additional sensors are too expensive, inaccessible, or both. Using current signals for defect diagnosis, on the other hand, offers various advantages; it allows for online detection and is quite affordable. As a result, there have been studies that have used current signals to diagnose faults. Current signals offer several advantages over vibration signals in the context of equipment monitoring and fault diagnosis:

- **Robustness:** Current signals are generally less susceptible to environmental noise and disturbances compared to vibration signals. For example, in a noisy industrial environment, vibration signals may be affected by background noise, while current signals may remain more stable and reliable.
- **Real-time monitoring:** Current sensors typically offer faster response times compared to vibration sensors, enabling more real-time monitoring of equipment performance. This non-intrusive approach to monitoring is especially beneficial for promptly detecting sudden changes or abnormalities in equipment operation.
- **Fault diagnosis:** Current signals can provide valuable insights into the internal operating status and fault characteristics of the equipment. For instance, abnormal current patterns may indicate issues such as bearing wear, electrical faults, or mechanical problems, allowing for more accurate and timely fault diagnosis.
- **Ease of data acquisition:** In some cases, acquiring current signals may be more convenient and cost-effective than obtaining vibration signals. Current signals can be directly accessed through the equipment's electrical interface, eliminating the need for additional sensor equipment and simplifying the data acquisition process.

Overall, the robustness, real-time monitoring capabilities, diagnostic potential, and ease of acquisition make current signals a valuable and reliable source of information for equipment condition monitoring and fault diagnosis. Thus, there have been studies that have employed current signals for fault diagnosis. An et al. proposed an unsupervised contrast domain adaptive network (UCDAN) for bearing fault diagnosis under variable operating conditions [5]. Tang et al. propose a discriminative fault diagnosis method that integrates Robust Principal Component Analysis (RPCA) and multiple kernels into deep neural networks. The method outperforms traditional machine learning and classical deep learning approaches when tested on laboratory signals and structural failure data [6]. Zhukovskiy et al. presented induction motor bearing defect detection based on stator current singular value decomposition [7].

Chen et al. proposed a new deep learning based classification of bearing fault diagnosis methods from the perspective of target domain data attributes divided by labels, machines and faults [8]. Combining the comprehensive attention mechanism and the characteristics of domain adaptive neural networks, Cai et al. proposed a multi-fault detection method for series battery packs based on category-reinforced domain adaptive networks, which can diagnose multiple types of faults [9]. Sun et al. developed a method for diagnosing mechanical faults in induction motors using stator current signals. Their approach enhances the signal-to-noise ratio through intelligent noise cancellation and reconstruction modeling, leading to high diagnostic accuracy, even with sample variations [10]. Yin et al. introduced a novel method, Fault Feature Proxy Transfer (FFPT), to transfer fault features between different operating condition domains using raw current signals [11]. Zhang et al. proposed a new nonconvex penalty called generalized logarithmic (G-log) penalty, which enhances sparsity and reduces noise interference [12]. Mohammed et al. developed a semi-supervised domain adaptation method for cross-machine troubleshooting, combining model optimization and Generative Adversarial Networks (GANs) to connect the source and target domains [13]. Chen et al. developed a phase space reconstruction algorithm for fault diagnosis in current signals, selecting reconstruction parameters using the first-order forward difference method and mutual information method. The method constructs a fault feature set using overall and local inflection point features [14]. Wu et al. proposed a real-time open-circuit fault diagnosis method for T-type rectifiers based on median current analysis. The method uses three-phase currents to obtain the median current for diagnosis [15]. In addition to that, there is some remarkable work in the field. In 2023, Qian et al. proposed a novel Relational Transfer (RT) diagnostic framework aimed at indirectly measuring and reducing distributional differences between source and unseen target domains [16]. The authors utilized the widely used Maximum Mean Difference (MMD) metric, which is based on the mean statistic, as well as a novel metric called Maximum Mean Square Deviation (MMSD), to comprehensively capture regenerated kernel Hilbert space mean and variance information of the data samples in the regenerated kernel Hilbert space in order to enhance domain confounding [17]. Additionally, Qian et al. introduced the Improved Joint Distribution Adaptation (IJDA) mechanism to enhance distributional alignment, as well as a new I-Softmax loss that optimizes feature learning and facilitates the learning of more separable features. They combined the IJDA mechanism and the I-Softmax loss to construct a Deep Discriminative Transfer Learning Network (DDTLN) for fault diagnosis [18].

Bearing fault detection methods have primarily focused on vibration signals for fault identification, with limited exploration of current-based methods, particularly for biphasic currents, which are less common in the literature. Current signal-based methods are increasingly prevalent but often rely heavily on specialized knowledge and are perceived as challenging to implement. Simplicity in method design is crucial for practical engineering applications. To address this gap, this study proposes a novel approach for diagnosing bearing

defects. The approach incorporates wavelet denoising [19] and biphasic current preprocessing, along with the CBAR neural network model adapted from the CBAM-ResNet model [20]. The method aims to enhance the signal quality by applying wavelet denoising to the current signal. It further utilizes biphasic current vector representation and algebraic computation as data preprocessing (WP post-processing) before extracting features. This preprocessing step aims to optimize the robustness and accuracy of feature extraction in fault diagnosis. Finally, the CBAR model is employed to analyze and categorize the obtained data, enabling efficient identification of bearing faults.

Another common challenge is ensuring communication security and addressing noise interference issues in data. Kou et al. proposed an offshore wind power encryption algorithm based on a two-dimensional lag complex logic mapping (2D-LCLM) and Zhouyi Bagua [21]. Additionally, Al-Hazaimeh et al. proposed a new image encryption algorithm that utilizes the Chen system and Bogdanov graph anti-synchronization [22]. Fortunately, however, current signal characterization is a non-intrusive bearing fault diagnosis method, using the motor control system has sensors through the drive process to obtain the signal containing bearing fault information, not only can avoid the installation of additional acceleration or acoustic emission sensors, etc., and the sensor by external factors and the installation location of the influence of the smaller, to a large extent, to reduce the investment and maintenance workload. In addition, this type of method can be well combined with embedded systems, which is conducive to engineering implementation.

In summary, this study presents an improved approach for bearing defect diagnosis by combining wavelet denoising, biphasic current preprocessing, and the CBAR neural network model. The improved CBAR model achieves high accuracy with an uncomplicated model architecture compared to other complex models. In conclusion excellent diagnostic results were achieved on the bi-directional current dataset through uncomplicated modeling and data preprocessing (WP post-processing).

II. FAULT DIAGNOSIS METHOD

A. WAVELET DENOISING

The wavelet transform [23] and thresholding [24] are the foundations of wavelet denoising. The wavelet transform decomposes the signal into numerous wavelet coefficients [25] of varying scales and frequencies and depicts it as a linear combination of a collection of wavelet basis functions (Daubechies, Db) [26]. The wavelet transform can convey information about a signal in both the frequency and temporal domains [27] allowing for a better comprehension of the signal's characteristics. Thresholding is an essential step in wavelet denoising. It is based on the statistical features of wavelet coefficients, with bigger coefficients retained and lower coefficients set to 0. This processing method effectively removes noise while preserving local features and detailed signal information.

$$y(t) = \sum_{j=1}^{J} \sum_{k}^{2^j-1} d_{j,k}\psi_{j,k}(t) + \sum_{k=0}^{N-1} \omega_k \psi_{J,k}(t) \tag{1}$$

The wavelet denoising mathematical theory is particularly expressed in equation, where $y(t)$ is the original signal, $d_{j,k}$ is the wavelet coefficient with a scale j and frequency k, ω_k is the high-frequency noise coefficient, and $\psi_{J,k}(t)$ is the wavelet basis function.

The denoised completed signal is acquired by inverse transformation after a new wavelet coefficient $d'_{j,k}$ is obtained by thresholding the wavelet coefficients, as described by equation (2).

$$y'(t) = \sum_{j=1}^{J} \sum_{k}^{2^j-1} d'_{j,k}\psi_{j,k}(t) + \sum_{k=0}^{N-1} \omega_k \psi_{J,k}(t) \tag{2}$$

The threshold processing formula is given by equation (3). where λ_j is the threshold value, and either a fixed or adaptive threshold can be chosen. As a result, wavelet denoising is an excellent signal processing method for removing noise while retaining signal detail information.

$$d'_{j,k} = \begin{cases} d_{j,k}, & |d_{j,k}| > \lambda_j \\ 0, & |d_{j,k}| \le \lambda_j \end{cases} \tag{3}$$

B. TWO-PHASE CURRENT

Biphasic current is a type of alternating current that consists of two sinusoidal currents with a phase difference of 90 degrees, and Figure 1 depicts a simplified diagram of a two-phase alternator. A complex number can be used to represent a biphasic current, with the real part representing one sinusoidal current and the imaginary part representing the other.

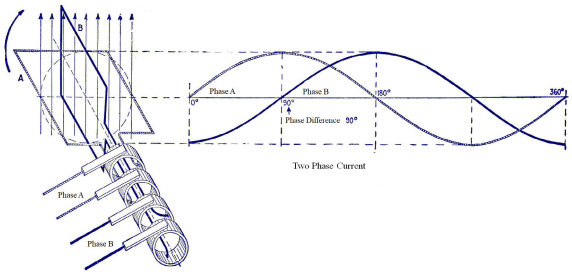

FIGURE 1. A simplified diagram of a two-phase alternator.

For example, a biphasic current can be written as $I1 + jI2$, where $I1$ and $I2$ are the amplitudes of the two sinusoidal currents, respectively, and j is an imaginary unit. Biphasic currents are often calculated in the complex plane [28], where this complex number can be composed as a vector whose mode length indicates the amplitude of the current and the phase angle represents the phase of the current.

If there is a normal biphasic current with magnitude $I1$ is u(A) and phase angle 0 degrees and magnitude $I2$ is v(A)

and phase angle 90 degrees, the biphasic current I may be expressed as equation (4).

$$I = u + jv \tag{4}$$

It has a mode length of $\sqrt{u^2 + v^2}$ and a phase angle of (arctan v/u) degrees. The origin serves as the beginning point for this vector, and the position of the endpoint in the complex plane reflects the amplitude and phase of the biphasic current.

First of all, it is known that when the bearing fails, this vector will change correspondingly, and then it is feasible to identify whether the bearing has failed based on this change. Further, a defect in a two-phase circuit generally results in a phase change or amplitude change of the current, which may be stated as follows:

- **1. Phase difference:** The phase difference between two currents in a two-phase circuit is normally 90 degrees. If a fault occurs in the circuit, the phase of one of the currents may change, causing the phase difference to change, and the magnitude and direction of the phase shift may be described in terms of the phase angle.
- **2. Change in amplitude:** A defect can also produce a change in the amplitude of the current.

A combination of phase and amplitude variations is required to determine the location and kind of fault in a biphasic circuit. However, in data-oriented analysis, the problem may be simplified by assuming that both phase and amplitude fluctuations are based on the two-phase current data $I1$ and $I2$. So, disregarding previous inherent a priori knowledge and physical terminology, we can simply express the amplitude change by adding $I1$ and $I2$, subtracting $I1$ and $I2$, and multiplying $I1$ and $I2$, whereas the phase change is expressed by the phase angle of $I1$ and $I2$. If these values vary in comparison to the usual condition, the bearing failure can be identified.

Why not utilize one of the two-phase currents for problem diagnosis? This is undoubtedly conceivable, and there are academics working on it. However, much as humans can, in certain situations, employ only a single eye to monitor a target and presume that this eye combines several complicated statistical discriminative qualities in order to properly assess the object, However, even the most powerful monocular is sometimes not as easy to sweep as the simplest binocular. Thus, bi-phase currents may be better suited for fault diagnostics than single-phase currents. Furthermore, because there may be significant differences in the signal characteristics of phase currents in the case of a bearing fault, these differences may not be fully reflected when using one-phase current alone; however, a calculation using two-phase currents can more clearly reflect these differences and thus more accurately determine whether a bearing is faulty.

In summary, as illustrated in Figure 2, the vector representation and algebraic operation of the biphasic current can reflect the bearing's fault situation and so aid in fault diagnosis.

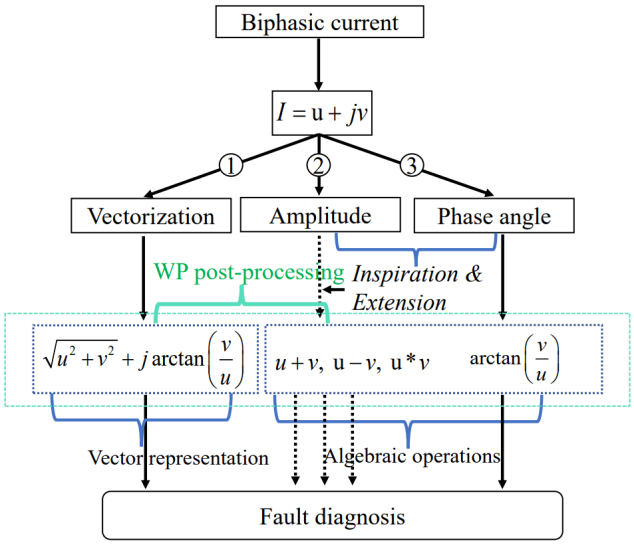

FIGURE 2. Vector representation and algebraic operations (WP post-processing).

C. CBAR NETWORK MODEL

CBAR consists of 3 main components. including traditional Convolutional Neural Networks (CNNs) [29], Convolutional Block Attention Module (CBAM) [30], [31], and (Deep Residual Learning for Image Recognition, ResNet) [32]. Among them, CNNs are used to extract image features, CBAMs are used to enhance the expressiveness of features and distinguish different features, while ResNet is used to reduce the gradient disappearance problem during model training.

Convolutional neural networks include three primary structures: convolution, pooling, and fully connected layers [33].

CBAM is a computer vision attention method that increases model performance and stability by adaptively altering the relevance of each channel in the feature map. Figure 3 depicts the operation of CBAM, which is separated into two steps: channel attention [34] and spatial attention [35]. The attention mechanism [36] is a key strategy for increasing the performance and resilience of convolutional neural networks.

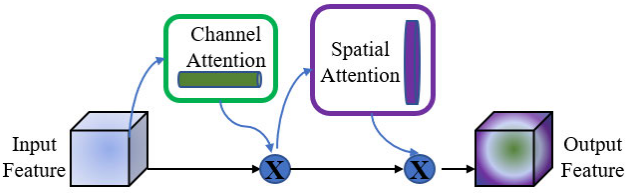

FIGURE 3. Theory of CBAM.

The channel attention module is primarily used to learn the weights of several channels in order to better capture the correlation between them. Each convolutional layer of a convolutional neural network generates a feature map, which typically has many channels, each corresponding to a

distinct feature. The channel attention module can improve network performance by learning the weights of each channel in order to boost valuable characteristics and lessen the effect of worthless features.

The channel attention mechanism is often implemented in CBAM by global pooling of the feature graph [37]. Each channel in the feature graph is compressed into a vector that indicates its relevance via global pooling. To generate the weight coefficients for each channel, these vectors are fed into a sigmoid [38] function through a fully linked layer. To achieve channel attention, these weight factors are applied to each channel in the feature map.

The spatial attention module is primarily used to learn weights between various places in order to better capture spatial connections. Each convolutional layer in a convolutional neural network generates a feature map, which typically has numerous spatial locations, each corresponding to a distinct pixel. The spatial attention module can improve network performance by learning the weights of each location in order to boost relevant pixels and diminish the effect of unnecessary pixels.

The particular implementation of its spatial attention module will change depending on the job. In this study, we simply utilized a convolutional layer to extract each position in the feature map and then executed Softmax [39] operations on each position to obtain the weight coefficients of the corresponding locations. This method of obtaining weight coefficients can minimize calculation time and the number of parameters.

ResNet, as seen in Figure 4, is a deep neural network architecture whose main purpose is to overcome the gradient disappearance and gradient explosion [40] problems in deep neural networks.

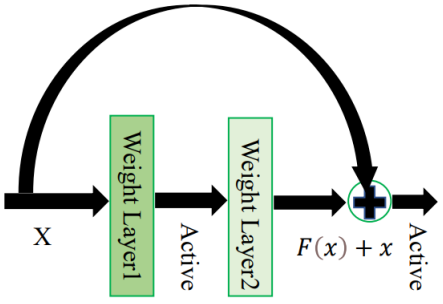

FIGURE 4. ResNet.

D. BEARING FAULT DIAGNOSIS METHOD

The bearing defect detection technique is made up of various components, including wavelet denoising, biphasic current, the CBAR model, an optimization function, and a loss function, among others. It may also be separated into two parts: data pre-processing and network modeling in general.

The data pre-processing section contains two parts, as illustrated in Figure 5. The first step is to use wavelets to denoise the biphasic current signal. To generate the wavelet coefficients (WC) array, the input raw signal is decomposed into four layers based on Db8. The high-frequency components of the wavelet coefficients are then adjusted to zero to reduce noise and high-frequency interference before reconstructing the WC.

After wavelet denoising, the current signal is preprocessed by biphasic current algebraic operations or vector (WP post-processing) in the second stage to acquire more accurate bearing fault features. These data are standardized, as shown in equation (5), to make the model more stable during training.

$$y = \frac{2\,(x - x_{\min})}{x_{\max} - x_{\min}} - 1 \tag{5}$$

We choose to normalize the data in this work to a value between plus and minus 1, where is the value after the current algebraic operations, $x_{\min}$ is the smallest value, $x_{\max}$ is the highest value, and y is the normalized value.

The CBAR model is employed in the network model component, as illustrated in Figure 6, to feature extract and categorize the pre-processed biphasic current data in order to perform automated detection of bearing problems. The model is made up of two components: a feature extractor and a classifier. The preprocessed one-dimensional vector is fed into a two-dimensional convolutional neural network by the feature extractor, and the $1*2560$-dimensional vector is up-dimensioned [41] to produce a $40*64$-dimensional vector. The feature extractor also performs two independent feature extraction processes.

A convolutional layer, a batch layer, a parametric rectified linear unit (PReLu) [42] activation function, a CBAM attention mechanism, a maximum pooling layer, and residual blocks are used in the autonomous feature extraction process. The retrieved features are then submitted to the classifier (completely connected layer and Softmax classifier), and the defect diagnostic category is returned.

The batch layer is utilized for feature normalization [43] on top of the convolutional layer for feature extraction to boost network stability.

As illustrated in equation (6), the PReLU(parametric rectified linear unit) function is a modified linear unit function. When $x < 0$, the PReLU function multiplies by a tiny positive integer to keep the gradient of the negative portion from disappearing and the linearity of the positive part maintained, and when $x \geq 0$, the PReLU function is the same as the (Rectified Linear Unit, ReLU) [44] function. When compared to the commonly used ReLU function, the PReLu activation function not only boosts nonlinearity but also allows the network to retain more neurons, allowing it to solve complicated problems more effectively.

$$f\,(x) = \begin{cases} ax, & if\ x < 0 \\ x, & otherwise \end{cases} \tag{6}$$

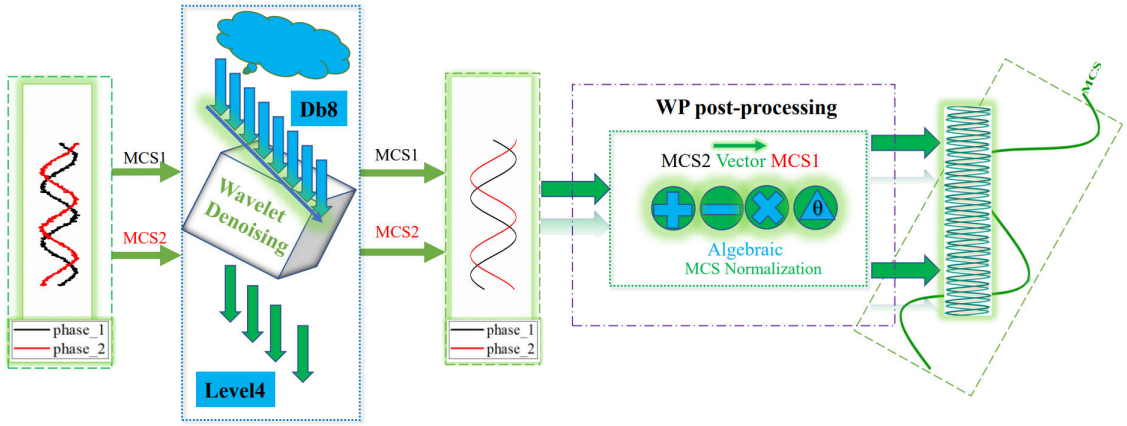

FIGURE 5. Signal preprocessing.

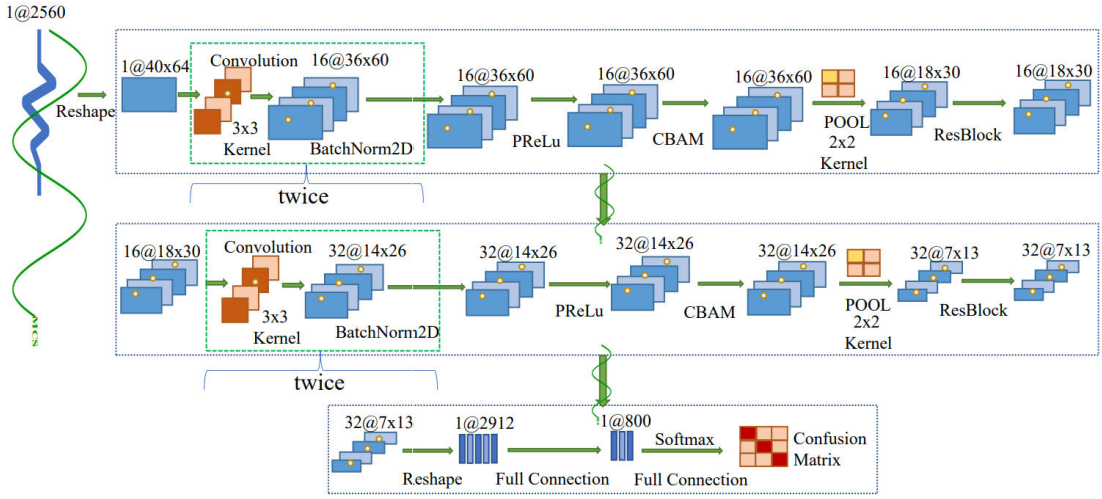

FIGURE 6. CBAR network model.

The model was trained using optimization algorithms Stochastic Gradient Descent (SGD) [45] and Cross Entropy Loss Function (CLF) [46].

III. EXPERIMENT

A. DATA SET PROCESSING

The experimental data utilized in this research from the rolling bearing condition monitoring bench at the University of Paderborn(UPB) [47], Germany.

Figure 7 shows the experimental setup, which contains the test motor, measurement shaft, bearing module, flywheel, and load motor. The bearings were all type 6203 and the current signal was captured at 64 kHz for 4 seconds per sample. The experimental damage types were divided into single damage, repetitive damage, and multiple damages, where multiple damages includes repetitive harm. Fatigue, wear, corrosion, galvanic corrosion, plastic deformation, fracture, and cracking were the six primary damage mechanisms identified. In the UPB data literature [47], the precise UPB

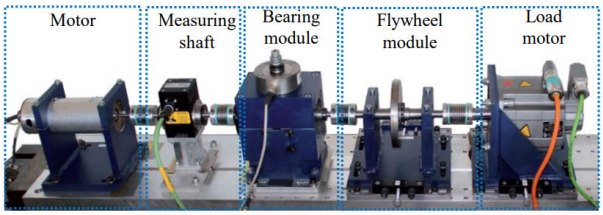

FIGURE 7. Experimental platform.

data-related experimental setup and operating techniques are given in depth.

For method testing, this work cross-selects an artificial single-point fault in the UPB current signal and the multi-damage fault bearing derived after an accelerated lifetime test to enrich the experimental condition. K001 denotes the normal state; KI01 (artificial single damage [47]) denotes inner ring failure; KA01 (artificial single damage) denotes outer ring failure; KB24 (multiple damage for accelerated life test) and KB27 (multiple damage

TABLE 1. Manual single point of failure and real multiple failure information.

Bearing code	Fault Location	Damage type and loss level
KA01	OR	EDM 1
KI01	IR	EDM 1
KB27	OR+IR	Plastic deform:Indentations 1
KB24	IR+(OR)	Fatigue:pitting 3

for accelerated life test) denote combined inner and outer ring failure. The inner ring of KB24 was the most damaged, whereas KB27's inner and outer rings were about equally damaged. Table 1 shows the specifics.

OR, IR, OR+(IR), IR+OR, and NORMAL denote the outer ring, inner ring, outer ring+(inner ring), inner ring+outer ring, and normal, respectively. The diagnostic categories of the faults are shown in Table 2. Table 3 shows how the experiments were divided into four groups based on the testing conditions' speed, torque, and radial load.

TABLE 2. Fault Category.

Bearing code	Fault Location	Category
KA01	OR	0
KI01	IR	1
KB27	OR+IR	2
KB24	IR+(OR)	3
K001	NORMAL	4

TABLE 3. Experimental conditions.

Setting	Rotational speed/rpm	Torque/Nm	Radial force/N
N15-M07-F10	1500	0.7	1000
N15-M01-F10	1500	**0.1**	1000
N15-M07-F04	1500	0.7	**400**
N09-M07-F10	**900**	0.7	1000

2560 data points are selected as samples for each set of tests, and the raw current signal data are sliced and diced to obtain 10014 cases of N09-M07-F10 data, 10028 cases of N15-M01-F10 data, 10015 cases of N15-M07-F04 data, and 10012 cases of N15-M07-F10 data. As shown in Figure 8, the data were assigned at random under these distinct working

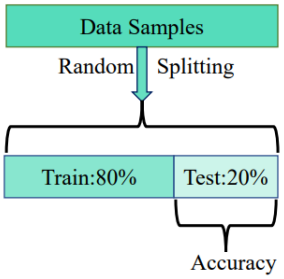

FIGURE 8. Random splitting.

circumstances, with 80% of the data serving as the training set and 20% serving as the test set.

Figure 9 depicts the waveforms of the biphasic current signal for each fault type under N15-M07-F10 operating conditions with one unit sample for algebraic multiplication and normalized to between plus and minus 1.

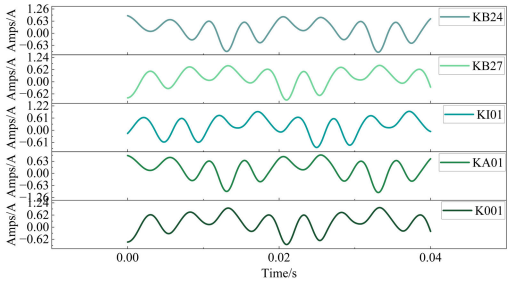

FIGURE 9. Algebraic multiplication operation waveform.

Figure 9 shows that the waveforms of the current signals following the biphasic current algebraic operations have more visible changes between fault and normal situations under this operating environment. As a result, the biphasic current algebraic procedures are highly beneficial for extracting the characteristics of the current signal, which can subsequently be utilized to better train the model in the experiment.

B. EXPERIMENTAL PLATFORMS

The experiment is based on windows 10 system for training the model and validating the method. The processor is 12th Gen Intel(R) Core(TM) i7-12700H 2.30 GHz. Then the GPUs are RTX-3060 6GB and RTX-3090 24GB.

C. EXPERIMENTAL RESULTS

For each operational state, experiment with one algebraic operation and one vector representation. To confirm the trustworthiness of the experimental results, 20 sets of experiments were undertaken, with 10 trials completed for each set, for a total of 200 experiments, each containing 200 iteration steps. The test set was subjected to an accuracy examination. The results of all tests are presented in Figure 10's radar map,

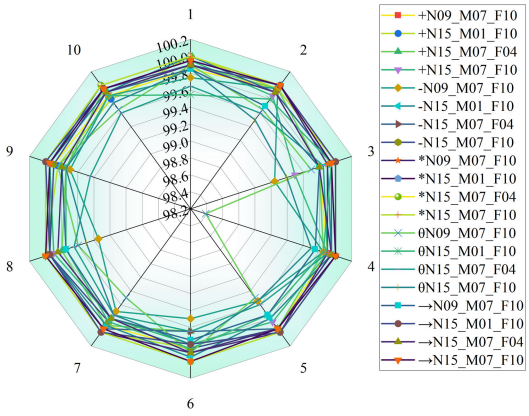

FIGURE 10. Radar chart of experimental results.

TABLE 4. Experimental results of algebraic addition (WP-A).

Setting	AvAcc/%	MaAcc/%	SD
N09-M07-F10	99.88	99.95	0.04
N15-M01-F10	**99.94**	**100**	0.07
N15-M07-F04	99.84	99.95	0.09
N15-M07-F10	99.83	99.95	0.12

TABLE 5. Experimental results of algebraic subtraction (WP-S).

Setting	AvAcc/%	MaAcc/%	SD
N09-M07-F10	99.62	99.95	0.21
N15-M01-F10	99.93	**100**	0.05
N15-M07-F04	99.92	**100**	0.09
N15-M07-F10	**99.94**	**100**	0.05

TABLE 6. Experimental results of algebraic multiplication (WP-M).

Setting	AvAcc/%	MaAcc/%	SD
N09-M07-F10	99.96	**99.98**	0.02
N15-M01-F10	99.98	**100**	0.02
N15-M07-F04	99.97	**100**	0.03
N15-M07-F10	99.98	**100**	0.02

TABLE 7. Experimental results of algebraic phase angle (WP-P).

Setting	AvAcc/%	MaAcc/%	SD
N09-M07-F10	99.55	99.90	0.41
N15-M01-F10	**99.79**	**100**	0.10
N15-M07-F04	99.68	99.90	0.12
N15-M07-F10	99.65	99.85	0.15

TABLE 8. Experimental results of vector representation (WP-V).

Setting	AvAcc/%	MaAcc/%	SD
N09-M07-F10	99.78	99.90	0.06
N15-M01-F10	99.96	**100**	0.06
N15-M07-F04	99.87	99.95	0.06
N15-M07-F10	**99.97**	**100**	0.02

and it can be observed macroscopically that the results of each experiment are almost over 99.0%, and most of them are close to 100% accuracy, with the greatest accuracy rate approaching 100%.

Tables 4-8 contain the complete experimental statistics. The statistical results of each group of experiments show that all four algebraic operations and vector representations (WP post-processing) are effective in diagnosing bearing faults, with the algebraic multiplication (WP-M) method achieving the highest accuracy of 100% under all four working conditions.

To see the model's convergence during training and the diagnosis findings for each sort of defect more clearly. Figures 11, 12, and 13 illustrate the accuracy variation curve, loss curve, and confusion matrix at the greatest accuracy for the algebraic multiplication operation(WP-M).

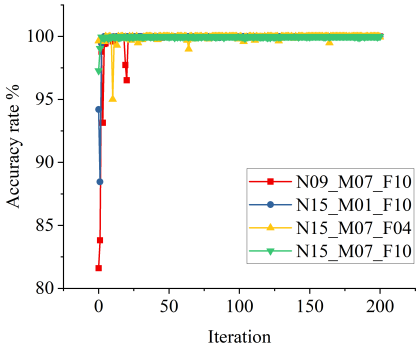

FIGURE 11. Algebraic multiplication operation accuracy curve.

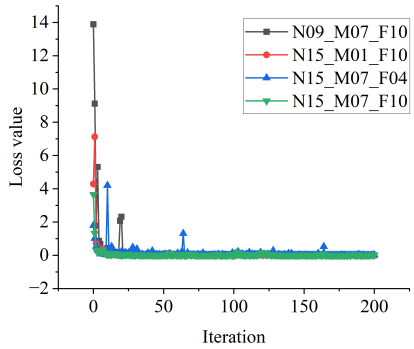

FIGURE 12. Algebraic multiplication operation loss curve.

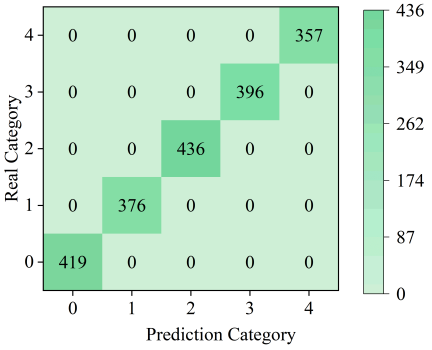

FIGURE 13. Confusion matrix for algebraic multiplication operations under N09-M07-F10.

D. EXPERIMENTAL WORKING CONDITION DISCUSSION

It may be required to examine the N09-M07-F10 operating state independently while doing the analysis. When the experimental data is obtained in this circumstance, the speed

reduces dramatically, and the recorded bearing fault information is conveyed to the current signal more weakly.

In the N09-M07-F10 operating condition, the feature distinction between different fault signals is not so obvious compared with other operating conditions. Simply calculate the average correlation coefficient between each fault, and the average correlation coefficient is calculated as shown in equation (7).

$$\overline{\mathrm{Cor}} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\rho_{ij}\right| \tag{7}$$

In (7), n specifies the size of the correlation coefficient matrix, ρ_{ij} denotes the i-th row and j-th column entries, and $\overline{\mathrm{Cor}}$ is the average correlation coefficient. The greater the average correlation coefficient, the more similar the series are.

N09-M07-F10, N15-M01-F10, N15-M07-F04, and N15-M07-F10 had correlation values of 0.74, 0.63, 0.68, and 0.71, respectively. The strongest correlation coefficient is N09-M07-F10, which helps explain why the link between the characteristics of this condition is so strong and difficult to separate. Figure 14 depicts the correlation coefficient matrix of N09-M07-F10. Of course, the above description is based only on the CBAR model for training; if other models are employed, the results may change.

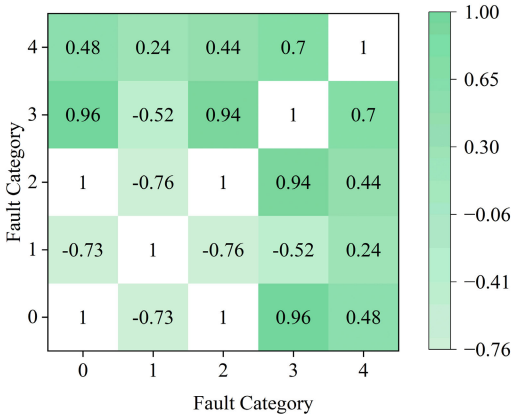

FIGURE 14. N09-M07-F10 correlation coefficient matrix.

And in order to consider most working situations, the number of data points for each data sample was set to 2560 in the experiment, whereas the number of data points gathered for one rotation of the bearing under N09-M07-F10 working conditions was roughly 4266 [48]. Despite the drawbacks faced, this disadvantage may be significantly addressed by the algebraic operations or vector representation (WP post-processing) of the two-phase current, which strongly link the two-phase currents to each other, and the experimental results obtained by preprocessing the current data are still strong.

E. HYBRID FEATURE EXTRACTION

Although single-feature extraction is a straightforward and efficient approach for data preprocessing, it may have limitations in capturing the full complexity of the underlying data. To overcome this, feature fusion techniques have gained attention in recent years [49]. Feature fusion involves combining information from multiple feature sources to enhance model performance and resilience, while mitigating the risks of overfitting and underfitting [50].

In this study, we conducted experiments to explore the potential of feature fusion by utilizing four algebraic operational features (WP-A, WP-S, WP-M, WP-P) proposed in WP, along with vector representation (WP-V). The feature fusion is illustrated in Fig. 16, and our objective is to improve the robustness of the classification model by leveraging the complementary information from each feature source. The experimental results are promising, as shown in Figure 15 demonstrating the effectiveness of feature fusion in improving the accuracy of the classification model. Classification model accuracy. In the four operational scenarios was 99.60%. This significant improvement is attributed to the ability of feature fusion to capture different aspects of the data and provide a more comprehensive characterization and provide a more comprehensive representation.

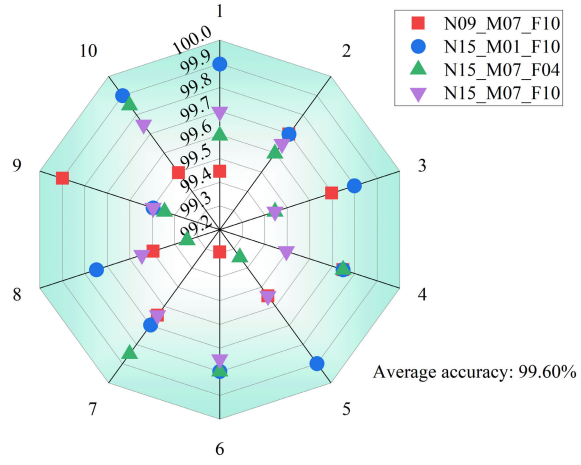

FIGURE 15. Feature fusion accuracy distribution.

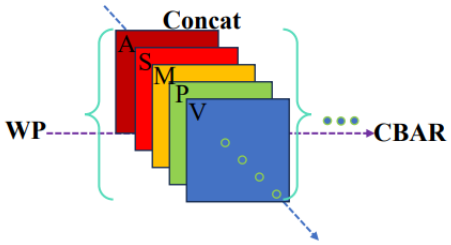

FIGURE 16. Schematic of feature fusion.

Furthermore, the use of hybrid feature extraction approaches, which combine both single-feature extraction and feature fusion techniques, offers a more robust and scalable solution. By incorporating the strengths of both approaches, hybrid feature extraction can further enhance the performance of the classification model and ensure its adaptability to different datasets and operating conditions.

Overall, the findings of this study highlight the importance of feature fusion as a powerful technique for data pretreatment. It not only maximizes the utilization of information from multiple feature sources but also improves the overall performance and resilience of the classification model. This research contributes to the advancement of feature engineering methods and provides valuable insights for future studies in the field of machine learning and data analysis.

IV. EXPERIMENTAL COMPARISON

It is important to note that the focus of this paper is on the diagnosis of two-phase current bearing faults. Therefore, other fault diagnostic signals, such as vibration signal comparisons, have been widely recognized, vibration signals will be covered in this paper, but are not the focus of the comparison in this paper. As mentioned in the introduction, vibration signals have some limitations,so the comparison experiments in this paper primarily focus on bidirectional current and some classic basic network models. Additionally, the effectiveness of wavelet denoising techniques will be evaluated.

A. PHASE CURRENT COMPARISON EXPERIMENT

N09-M07-F10, which is somewhat challenging in this paper, was selected as the working condition for the experiments. In the earlier tests of two-phase current algebraic operations and vector representation (WP post-processing), various damage faults KB24 and KB27 in the dataset were picked in order to enrich the diversity of faults. while in the comparative studies, these many damage flaws were omitted and the data size was extended in order to make the experiments more intuitive, intelligible and convincing. The artificial faults in the UPB dataset were selected as experimental data and were simply classified as outer ring faults (OR), inner ring faults (IR), and normal (NORMAL).

TABLE 9. Test bearings with artificial damage.

Bearing code	Fault Location	Damage type loss level	Category
KA01	OR	EDM 1	0
KA03	OR	Electric engraver 2	0
KA05	OR	Electric engraver 1	0
KA06	OR	Electric engraver 2	0
KA07	OR	Drilling 1	0
KA08	OR	Drilling 2	0
KA09	OR	Drilling 2	0
KI01	IR	EDM 1	1
KI03	IR	Electric engraver 1	1
KI05	IR	Electric engraver 1	1
KI07	IR	Electric engraver 2	1
KI08	IR	Electric engraver 2	1
K001	NORMAL	-	2
K002	NORMAL	-	2
K003	NORMAL	-	2
K004	NORMAL	-	2
K005	NORMAL	-	2
K006	NORMAL	-	2

Table 9 shows the bearing defect information and fault categories, and the three sets of tests were run 10 times each with 200 training cycles. Meanwhile, the algebraic multiplication operation(WP-M) that produced the greatest experimental results in this research was chosen as data preprocessing.

TABLE 10. Phase current comparison experimental results.

Setting	AvAcc/%	MaAcc/%	SD
Algebraic operation (**WP**)	**99.61**	**99.73**	0.08
No algebraic operation (**NO-WP**)	96.39	96.74	0.30
Single-phase current	72.06	72.94	0.72

Table 10 shows the detailed experimental statistics, Fig. 17 and Fig. 18 show the comparison curves of the change of accuracy and the change of loss value, respectively, and then Fig. 19 shows the distribution of the experimental accuracy. From these plots, it can be seen that the two-phase current after algebraic multiplication operation is obviously better than the single-phase current, and the accuracy of the two-phase current after preprocessing by algebraic multiplication operation is improved by about 3 percentage points.

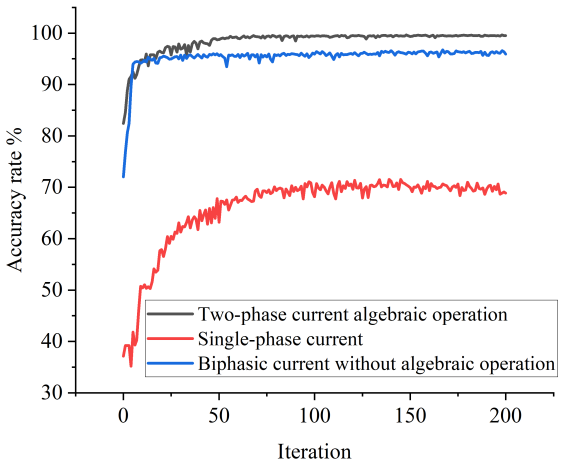

FIGURE 17. Phase current accuracy change comparison curve.

B. NEURAL NETWORK MODEL COMPARISON EXPERIMENT

Model comparison experiments are separated into two experiments. The models evaluated in the first comparison experiment are listed as follows:

(1) Shallow neural network with one hidden layer using original signals as input (BPNN)

(2) Deep belief network with 4 hidden layers using original signals as input (DBN)

(3) Deep belief network with more hidden layers using WP-M signals as input (1D-CNNs) [51]

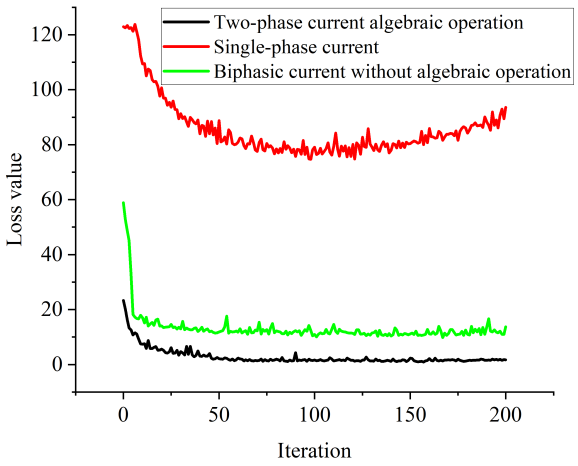

FIGURE 18. Phase current loss value change comparison curve.

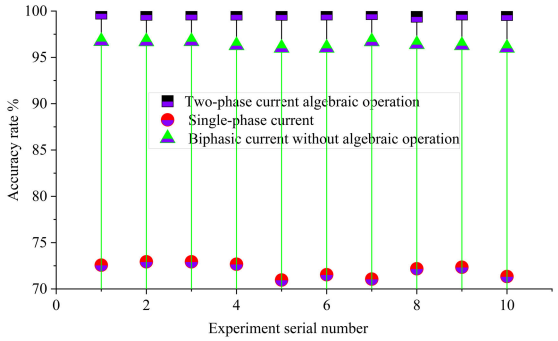

FIGURE 19. Phase current accuracy distribution.

(4) Signals are converted to time-frequency distribution (TFD), deep Convolutional Neural Network (DCNN) and fusion of vibration and current signals(Multi-signal) [52]

(5) CBAR+LSTM [53]

(6) Vision Transformer [54] (VIT)

(7) Swin Transformer [55] (ST)

The experimental settings of the first comparison experiment are identical to those of the phase current comparison, and the experimental findings are reported in Table 11. And the confusion matrix of the CBAR model is shown in Figure 20.

According to the experimental data, the CBAR model is more accurate than the CBAR+LSTM, CNNs, and VIT models. Based on the experimental data, the CBAR model demonstrates higher accuracy compared to the BPNN, DBN, CBAR+LSTM, CNNs, TFD + DCNN (multi-signal), and VIT models. Although the CBAR+LSTM model shows similar accuracy to the CBAR model, the inclusion of LSTM does not significantly improve accuracy and demands additional training resources. The VIT model also exhibits limited accuracy and requires extensive training resources. Standard BPNN, DBN, and 1D-CNNs models offer faster processing but lower accuracy. The TFD + DCNN method appears to closely match the accuracy of CBAR, but it is relatively more complex.

TABLE 11. Comparison of experimental results of network models under N09-M07-F10 operating conditions.

Model	AvAcc/%	MaAcc/%	ET/s	SD
BPNN	82.25	85.22	3.15	0.18
DBN	90.35	91.55	5.21	0.36
1D-CNNs	93.09	94.13	5.42	0.15
CBAR	**99.61**	**99.73**	6.27	0.08
CBAR+LSTM	99.29	99.52	9.30	0.17
TFD + DCNN(multi-signal)	99.55	99.69	6.85	0.13
VIT(encoder)	97.20	97.36	13.55	0.52
$ST_{224*224}$(encoder)	**99.93**	**100**	18.07	0.12

NOTE
ET:Time taken for one epoch of model training.
SD:Standard deviation of accuracy.

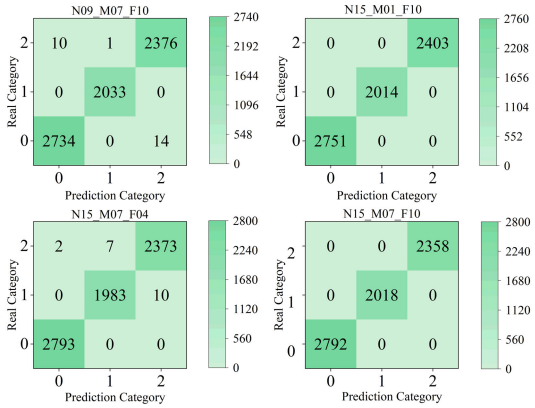

FIGURE 20. Confusion matrix for the CBAR model.

TABLE 12. CBAR and ST models compared experimental results.

Model	Setting	AvAcc/%	MaAcc/%	SD
CBAR	N09-M07-F10	99.61	99.73	0.08
	N15-M01-F10	100	100	0
	N15-M07-F04	99.82	99.90	0.06
	N15-M07-F10	100	100	0
$ST_{224*224}$	N09-M07-F10	99.93	100	0.12
	N15-M01-F10	100	100	0
	N15-M07-F04	94.40	97.55	1.67
	N15-M07-F10	99.81	100	0.26
$ST_{160*160}$	N09-M07-F10	99.95	100	0.07
	N15-M01-F10	100	100	0
	N15-M07-F04	95.88	97.36	1.39
	N15-M07-F10	100	100	0
ST_{96*96}	N09-M07-F10	99.61	99.70	0.14
	N15-M01-F10	100	100	0
	N15-M07-F04	98.20	98.70	0.37
	N15-M07-F10	100	100	0

It is worth mentioning that the accuracy of the ST model under N09-M07-F10 operating circumstances is higher than that of the CBAR model, up to 100%. For this reason, the second model comparison experiment between CBAR and ST models under four operating circumstances is done in this research, and the results are provided in Table 12.

The experimental findings reveal that the accuracy of the ST model is drastically lowered under N15-M07-F04 operating circumstances. In addition, the ST model requires more costly training compared to the CBAR model. Although the ST model works well in many projects, especially in large-scale data tasks, the CBAR model-based bearing fault detection approach is more stable and cost-effective in the current fault diagnosis task.

C. COMPARATIVE EXPERIMENTS ON DENOISING TECHNIQUES

This paper employs wavelet denoising as the signal denoising technique. To validate the effectiveness of this method, a comparison experimental group and a processing experimental group are established: the comparison group comprises the original signal without wavelet denoising(WD), while the processing group includes the signal processed with wavelet denoising.

The results presented in Table 13 demonstrate an enhancement in diagnostic outcomes following the application of wavelet denoising. Although the improvement is modest, it does not have a negative impact. This provides further evidence that bidirectional current preprocessing (WP-M) remains effective even in the presence of noise. Additionally, it reaffirms the robustness of fault diagnosis based on current signals compared to methods based on other signals.

TABLE 13. Comparative experimental results of wavelet denoising.

Setting	AvAcc/%	MaAcc/%	SD
Wavelet denoising **(WD)**	**99.61**	**99.73**	0.08
No wavelet denoising **(NO-WD)**	99.25	99.45	0.15

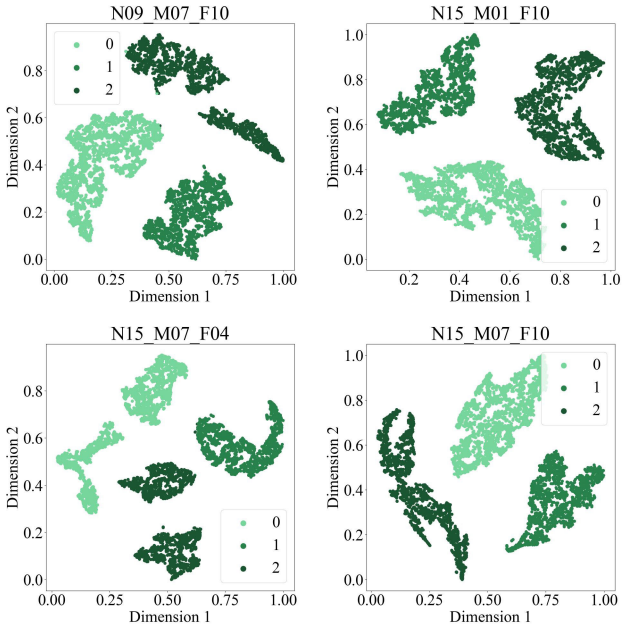

FIGURE 21. Two-dimensional visualization of fault characteristics.

D. FEATURE VISUALIZATION

T-SNE [56], [57] (t-distributed stochastic neighbor embedding) is a dimensionality reduction technique that is commonly used for visualizing high-dimensional data. In the context of the CBAR model, T-SNE is applied to the features extracted from the last hidden layer of the model after normalization. Figure 21 displays the results of the T-SNE visualization, showcasing the patterns and relationships present in the defect data. This visualization enables a quick and intuitive understanding of the data, making it easier to identify clusters or similarities among the data points. By reducing the dimensionality of the data while preserving its structure, T-SNE allows for a more comprehensive analysis and interpretation of the CBAR model's outputs.

V. CONCLUSION

This work proposes a biphasic current-based bearing defect diagnostic technique. The approach begins by pre-processing the current signal with features using wavelet denoising, biphasic current algebraic operation, or vector representation(WP). Then, a CBAR neural network model is built to intelligently recognize the characteristics and, ultimately, diagnose bearing defects.

The limitations of bidirectional current detection for bearing fault diagnosis include:

- *Dependency on bearing's working state:* Detection accuracy may be affected if the bearing is not in operation or under a light load.
- *Environmental interference:* Surrounding factors such as electromagnetic interference and temperature variations can impact the accuracy of the detection results.
- *Inability to pinpoint fault location:* The method can indicate the presence of a fault but cannot precisely locate its specific position, necessitating complementary diagnostic techniques.
- *Inability to differentiate fault types:* The detection method can identify the presence of faults within the bearing but cannot distinguish between different types of faults, such as wear, cracks, or loosening.

The successful experience gained from the fault diagnosis approach using two-phase current signals can provide valuable insights and ideas for fault detection in three-phase AC [58], [59], and multi-phase AC systems. Specifically, the approach of linking the collected synchronous signals together instead of analyzing individual signal aspects in isolation offers important insights and directions for future research. In addition, there is a vast scope for future research in the areas of feature extraction methods, transferability, model migration learning and unsupervised deep migration learning. These areas have the potential to drive further progress and innovation in the field of bearing defect diagnosis.

REFERENCES

[1] Y. Li, H. Wan, and L. Jiang, "Alignment subdomain-based deep convolutional transfer learning for machinery fault diagnosis under different working conditions," *Meas. Sci. Technol.*, vol. 33, no. 5, May 2022, Art. no. 055006. [Online]. Available: https://api.semanticscholar.org/CorpusID:245067727

[2] A. Kafeel, S. Aziz, M. Awais, M. A. Khan, K. Afaq, S. A. Idris, H. Alshazly, and S. M. Mostafa, "An expert system for rotating machine fault detection using vibration signal analysis," *Sensors*, vol. 21, no. 22, p. 7587, Nov. 2021.

[3] M. Altaf, T. Akram, M. A. Khan, M. Iqbal, M. M. I. Ch, and C.-H. Hsu, "A new statistical features based approach for bearing fault diagnosis using vibration signals," *Sensors*, vol. 22, no. 5, p. 2012, Mar. 2022.

[4] D. Zuo, T. Tang, and M. Chen, "Rolling bearing fault diagnosis based on multi-scale weighted visibility graph and multi-channel graph convolution network," *Meas. Sci. Technol.*, vol. 34, no. 11, Nov. 2023, Art. no. 115019. [Online]. Available: https://api.semanticscholar.org/CorpusID:259957396

[5] Y. An, K. Zhang, Y. Chai, Q. Liu, and X. Huang, "Bearing fault diagnosis under variable working conditions base on contrastive domain adaptation method," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[6] H. Tang, Z. Liao, P. Chen, D. Zuo, and S. Yi, "A robust deep learning network for low-speed machinery fault diagnosis based on multikernel and RPCA," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 3, pp. 1522–1532, Jun. 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:236684258

[7] Y. Zhukovskiy, A. Buldysko, and I. Revin, "Induction motor bearing fault diagnosis based on singular value decomposition of the stator current," *Energies*, vol. 16, no. 8, p. 3303, Apr. 2023.

[8] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang, "Deep transfer learning for bearing fault diagnosis: A systematic review since 2016," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–21, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:256982412

[9] L. Cai, H. Wang, Z. Dong, Z. He, M. Gao, and Y. Song, "A multi-fault diagnostic method based on category-reinforced domain adaptation network for series-connected battery packs," *J. Energy Storage*, vol. 60, Apr. 2023, Art. no. 106690. [Online]. Available: https://api.semanticscholar.org/CorpusID:256157223

[10] M. Sun, H. Wang, P. Liu, Z. Long, J. Yang, and S. Huang, "A novel data-driven mechanical fault diagnosis method for induction motors using stator current signals," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 1, pp. 347–358, Mar. 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:247874793

[11] K. Yin, C. Chen, B. Luo, and J. Deng, "A bearing fault feature cross-domain transfer method based on motor current signals," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:264113132

[12] Z. Zhang, W. Huang, Y. Liao, Z. Song, J. Shi, X. Jiang, C. Shen, and Z. Zhu, "Bearing fault diagnosis via generalized logarithm sparse regularization," *Mech. Syst. Signal Process.*, vol. 167, Mar. 2022, Art. no. 108576. [Online]. Available: https://api.semanticscholar.org/CorpusID:244093208

[13] M. Alabsi, L. Pearlstein, and M. Franco-Garcia, "Cross domain fault diagnosis based on generative adversarial networks," *J. Vib. Control*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:260643176

[14] Y. Chen, Q. Li, Y. Zou, G. Long, N. Yan, and R. Fan, "Circuit breaker fault diagnosis method based on coil current time sequence phase trajectory characteristics," *Processes*, vol. 11, no. 4, p. 1241, Apr. 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:258223066

[15] Z. Wu, J. Zhao, H. Luo, and Y. Liu, "Real-time open-circuit fault diagnosis method for T-type rectifiers based on median current analysis," *IEEE Trans. Power Electron.*, vol. 38, pp. 8956–8965, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:258016407

[16] Q. Qian, J. Zhou, and Y. Qin, "Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions," *IEEE Trans. Ind. Informat.*, vol. 19, pp. 9898–9908, 2023, doi: https://api.semanticscholar.org/CorpusID:256184870.

[17] Q. Qian, Y. Wang, T. Zhang, and Y. Qin, "Maximum mean square discrepancy: A new discrepancy representation metric for mechanical fault transfer diagnosis," *Knowl.-Based Syst.*, vol. 276, Sep. 2023, Art. no. 110748. [Online]. Available: https://api.semanticscholar.org/CorpusID:259566511

[18] Q. Qian, Y. Qin, J. Luo, Y. Wang, and F. Wu, "Deep discriminative transfer learning network for cross-machine fault diagnosis," *Mech. Syst. Signal Process.*, vol. 186, Mar. 2023, Art. no. 109884. [Online]. Available: https://api.semanticscholar.org/CorpusID:253139925

[19] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[20] R. U. Khan, H. Khattak, W. S. Wong, H. AlSalman, M. A. A. Mosleh, and S. M. M. Rahman, "Intelligent Malaysian sign language translation system using convolutional-based attention module with residual network," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–12, Dec. 2021.

[21] L. Kou, J. Wu, F. Zhang, P. Ji, W. Ke, J. Wan, H. Liu, Y. Li, and Q. Yuan, "Image encryption for offshore wind power based on 2D-LCLM and Zhou Yi Eight Trigrams," *Int. J. Bio-Inspired Comput.*, vol. 22, no. 1, pp. 53–64, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259165359

[22] N. Tahat, M. F. Al Jamal, M. J. Bawaneh, A. K. Alomari, and O. M. Al Hazaimeh, "Image encryption using anti-synchronisation and Bogdanov transformation map," *Int. J. Comput. Sci. Math.*, vol. 15, no. 1, p. 43, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:248091791

[23] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.

[24] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[25] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4290–4300, May 2018.

[26] R. Sabir, D. Rosato, S. Hartmann, and C. Guehmann, "LSTM based bearing fault diagnosis of electrical machines using motor current signal," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 613–618.

[27] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains," *ISRN Neurosci.*, vol. 2014, pp. 1–7, Feb. 2014.

[28] M. R. Alam, F. Bai, R. Yan, and T. K. Saha, "Classification and visualization of power quality disturbance-events using space vector ellipse in complex plane," *IEEE Trans. Power Del.*, vol. 36, no. 3, pp. 1380–1389, Jun. 2021.

[29] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976.

[30] S. Woo, J. Park, J.-Y. Lee, and I.-S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[31] S.-H. Wang, S. L. Fernandes, Z. Zhu, and Y.-D. Zhang, "AVNC: Attention-based VGG-style network for COVID-19 diagnosis by CBAM," *IEEE Sensors J.*, vol. 22, no. 18, pp. 17431–17438, Sep. 2022.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[33] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[34] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. R. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 294–310.

[35] X. Chu, Z. Tian, Y. Wang, B. Zhang, H. Ren, X. Wei, H. Xia, and C. Shen, "Twins: Revisiting the design of spatial attention in vision transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 1–12.

[36] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[37] S. Zhang, S. Zhang, C. Zhang, X. Wang, and Y. Shi, "Cucumber leaf disease identification with global pooling dilated convolutional neural network," *Comput. Electron. Agricult.*, vol. 162, pp. 422–430, Jul. 2019.

[38] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Netw.*, vol. 107, pp. 3–11, Nov. 2018.

[39] H. Sun, J. Ren, H. Zhao, P. Yuen, and J. Tschannerl, "Novel Gumbel–Softmax trick enabled concrete autoencoder with entropy constraints for unsupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:236705541

[40] Y. Guo, R. Liu, A. Arab, Q. Zhou, B. Guo, Y. Ren, W. Chen, C. Ran, and P. Chen, "Dynamic behavior and adiabatic shearing formation of the commercially pure titanium with explosion-induced gradient microstructure," *Mater. Sci. Eng., A*, vol. 833, Jan. 2022, Art. no. 142340.
[41] M. Morimoto, K. Fukami, K. Zhang, A. G. Nair, and K. Fukagata, "Convolutional neural networks for fluid flow analysis: Toward effective metamodeling and low dimensionalization," *Theor. Comput. Fluid Dyn.*, vol. 35, no. 5, pp. 633–658, Oct. 2021.
[42] X. Wang, X. Liu, J. Wang, X. Xiong, S. Bi, and Z. Deng, "Improved variational mode decomposition and one-dimensional CNN network with parametric rectified linear unit (PReLU) approach for rolling bearing fault diagnosis," *Appl. Sci.*, vol. 12, no. 18, p. 9324, Sep. 2022.
[43] J. Wang, S. Li, Z. An, X. Jiang, W. Qian, and S. Ji, "Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines," *Neurocomputing*, vol. 329, pp. 53–65, Feb. 2019.
[44] G. Niu, X. Wang, M. Golda, S. Mastro, and B. Zhang, "An optimized adaptive PReLU-DBN for rolling element bearing fault diagnosis," *Neurocomputing*, vol. 445, pp. 26–34, Jul. 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233647542
[45] F. T. Admojo and Y. I. Sulistya, "Analisis performa algoritma stochastic gradient descent (SGD) dalam mengklasifikasi tahu berformalin," *Indonesian J. Data Sci.*, vol. 3, no. 1, pp. 1–8, Mar. 2022.
[46] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.
[47] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," in *Proc. PHM Soc. Eur. Conf.*, vol. 3, no. 1, 2016, pp. 1–17.
[48] A. S. Barcelos and A. J. M. Cardoso, "Current-based bearing fault diagnosis using deep learning algorithms," *Energies*, vol. 14, no. 9, p. 2509, Apr. 2021.
[49] D. T. Hoang and H. J. Kang, "A motor current signal-based bearing fault diagnosis using deep learning and information fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3325–3333, Jun. 2020.
[50] T. Koren, R. Livni, Y. Mansour, and U. Sherman. (2022). *Benign Underfitting of SGD in Stochastic Convex Optimization*. [Online]. Available: https://api.semanticscholar.org/CorpusID:247159078
[51] W.-C. Lin and Y.-R. Yeh, "Efficient malware classification by binary sequences with one-dimensional convolutional neural networks," *Mathematics*, vol. 10, no. 4, p. 608, Feb. 2022.
[52] S. Shao, R. Yan, Y. Lu, P. Wang, and R. X. Gao, "DCNN-based multi-signal induction motor fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2658–2669, Jun. 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:199089543
[53] H. Zheng, F. Lin, X. Feng, and Y. Chen, "A hybrid deep learning model with attention-based Conv-LSTM networks for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 6910–6920, Nov. 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:226474922
[54] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 548–558. [Online]. Available: https://api.semanticscholar.org/CorpusID:232035922
[55] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.
[56] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[57] V. Spiwok and P. Kříž, "Time-lagged t-distributed stochastic neighbor embedding (t-SNE) of molecular simulation trajectories," *Frontiers Mol. Biosci.*, vol. 7, p. 132, Jun. 2020.
[58] Y. Li, X. Song, X. Zhou, Z. Huang, and S. Zheng, "A sensorless commutation error correction method for high-speed BLDC motors based on phase current integration," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 328–338, Jan. 2020.
[59] W. Wang, H. Yan, Y. Xu, J. Zou, X. Zhang, W. Zhao, G. Buticchi, and C. Gerada, "New three-phase current reconstruction for PMSM drive with hybrid space vector pulsewidth modulation technique," *IEEE Trans. Power Electron.*, vol. 36, no. 1, pp. 662–673, Jan. 2021.

JIANGUANG ZHU was born in 1970. He received the Ph.D. degree. He is currently a Full Professor. His research interests include intelligent power supply and fault diagnosis.

TANG LIU (Student Member, IEEE) was born in 1995. He received the M.S. degree. His research interests include signal processing and fault diagnosis.