Research Article

Bearing Fault Diagnosis Method Based on Multidomain Heterogeneous Information Entropy Fusion and Model Self-Optimisation

Renwang Song,^{1,2} Xiaolu Bai ^{1,2} Rui Zhang ^{1,2} You Jia,^{1,3} Lihu Pan,² and Zengshou Dong ^{1,2}

¹Shanxi Province Engineering Research PHM, Taiyuan 030024, Shanxi, China

²College of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China

³Department of Engineering Mechanics, Taiyuan University of Science and Technology, Taiyuan 030024, China

Correspondence should be addressed to Rui Zhang; zhangrui@tyust.edu.cn and Zengshou Dong; dongzs@tyust.edu.cn

Received 28 September 2021; Revised 1 December 2021; Accepted 12 January 2022; Published 5 February 2022

Academic Editor: Nibaldo Rodríguez

Copyright © 2022 Renwang Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Incomplete diagnostic information, inadequate multisource sensor information, weak diagnosis models, and subjective experience result in difficulty in predicting rotating machinery faults. To overcome these limitations, we proposed a multiple domain and heterogeneous information entropy fusion model based on an optimisation of bearing fault diagnosis. The spatiotemporal approach uses a multiscale domain fusion strategy based on heterogeneous sensors (HSMSF) to extract feature fusion strategies and analyses the characteristics of the bearing fault features by multichannel processes with convolutional neural networks to vibration signals. After the mapping of multiple quality characteristics, the high-quality features are combined with each other, and the adaptive entropy weighted fusion method is used to analyse and make decisions on sensor information from different detection points. Nineteen key model parameters that were required for HSMSF construction were selected by adaptive optimisation using the chaos elitist modified sparrow search algorithm (CEI-SSA), and a self-learning diagnostic model that is suitable for multiple detection points was constructed. The validity and feasibility of the proposed fault diagnosis method were verified experimentally on two common reference-bearing datasets, CWRU and IMS, and compared with other fault diagnosis methods.

1. Introduction

The development of intelligent manufacturing technology makes rotating machinery systems increasingly complicated. As a core rotary component of rotating machinery, rolling bearings often experience failure because of alternating impact load, mechanical wear, and thermal fatigue [1]. Improper handling of these faults affects the normal operation of rotating machinery, which results in serious cost losses. Therefore, it is important to carry out effective, timely, accurate, and early diagnosis of the status of rolling bearings.

Vibration signal analysis is a commonly used method in mechanical fault diagnosis to extract valuable features that provide internal machine information from collected chaotic

signals [2]. For example, Ahmed et al. [3] used discrete wavelet and wavelet packet transforms to filter the acquired signals, processed wavelet coefficients in the time and frequency domains, extracted various features, and classified the surface roughness of mechanical faults in cutting tests. Cheng et al. [4] proposed a fully set empirical mode decomposition method of adaptive noise complementarity, which improved decomposition performance by reducing the reconstruction error and the influence of mode aliasing and enabled bearing fault detection. However, during rolling bearing operation, fault characteristic information is affected by the coupling vibration signals of multiple components and is submerged in noise or other signal components. The above research methods are aimed at analysing the signal of a single measuring point

sensor, which may lose part of the fault information and affect the extraction and expression of bearing fault features.

In recent years, scholars have introduced the idea of information fusion into the field of rotating machinery fault diagnosis to make use of the fault information of different sensitivities at multiple measuring points and to improve diagnostic accuracy. Yuan et al. [5] used the multivariate empirical mode decomposition method to extract dynamic characteristics from multivariate information that was collected by multiple sensors, analysed the multiscale sample entropy, and combined this information with a back-propagation neural network to classify faults in rolling bearings. Gao et al. [6] proposed a rolling bearing fault diagnosis method based on a complementary ensemble empirical mode decomposition entropy fusion feature. Kernel principal component analysis was used for feature fusion and dimension reduction, and fault classification was carried out by an optimised least-squares support vector machine. The information fusion method improves the reliability and accuracy of fault diagnosis to a certain extent, but it has the following limitations in feature extraction: (1) during the processing of multisource information, the above algorithms rely largely on feature qualities extracted by hand, and they have poor nonlinear fitting ability, which makes it difficult to deal with current complex, massive, heterogeneous, and real-time bearing fault information. (2) In the fusion of information collected by multiple sensors, the above algorithm only splices the data features of the measured points directly. Because of differences in the original data, mechanical data superposition leads to a heavier algorithm load and a lower model diagnostic performance.

With the continuous expansion of the concept of "intelligent manufacturing," the convolutional neural network (CNN) has found a wide range of applications to solve the above problems. Chen et al. [7] proposed a fault diagnosis method for rotating machinery based on parameter transmission, using a designed wide-core one-dimensional (1D) CNN to automatically learn the transferable features of the fault signals and to eliminate the influence of manual features. Li et al. [8] proposed a fault diagnosis method based on multiscale permutation entropy and a multichannel fusion convolutional neural network. Multiscale permutation entropy was used to analyse vibration signals of rotating machinery at different scales, and a multichannel fusion convolutional neural network was used to fuse multichannel features and identify faults. Intelligent fault diagnosis methods based on CNN retain rich and effective fault feature information to a large extent and improve diagnostic accuracy, but some challenges remain: (1) Most of the existing studies are based on information from analysis in the time, frequency, and time-frequency domains to determine the fault type. Few studies exist on other mapping methods of time sequence signals and the complementary and simultaneous correlation of spatial components such as hue, lightness, and purity in the postmapping spatial domain. (2) The establishment of a CNN model requires designers to have rich professional and domain knowledge, to weigh model size and computational efficiency, and to conduct repeated experiments

during model construction to obtain the best model parameters. With new practical problems, it is often necessary to reconstruct the model. These problems may be difficult to interpret and generalise.

In summary, research was carried out on a bearing fault diagnosis method based on multidomain heterogeneous information entropy fusion and model self-optimisation. The spatiotemporal multiscale domain fusion strategy based on heterogeneous sensors (HSMSF) was used to analyse and assess the characteristics of bearing fault signals, and the optimal feature extractor that was obtained by the chaos elitist modified sparrow search algorithm (CEI-SSA) adaptive optimisation was used to reconstruct the multidomain fusion space and decompose features in the space-time dimension. The multisensor predictive value fusion strategy was combined with information entropy theory to identify and diagnose bearing faults.

The main contributions in this paper are summarised as follows:

- (1) We propose a spatiotemporal multiscale domain feature fusion strategy based on heterogeneous sensors. From the perspective of sequence periodicity, correlation, and spatiotemporal transformation, a variety of spatial domain mapping methods were explored. Based on the method of channel cascade and the application of long short-term memory (LSTM) networks, characteristic information of multidomain samples in time and space dimensions was enriched, and decision reasoning regarding sensor information from multiple measurement points was carried out by adaptive entropy weight. This approach can improve the accuracy and numerical stability of bearing faults diagnosis and provide ideas for information fusion in heterogeneous multiscale domains.
- (2) In this work, we propose a feature extraction model and adaptive optimisation method based on CEI-SSA. To improve the global optimisation efficiency and pioneering ability of the original SSA, a chaotic initialisation and reverse elite strategy were introduced to improve the original SSA, and the improved CEI-SSA was applied to the HSMSF model to achieve adaptive optimisation reconstruction of 19 structural parameters. The interpretability and generalisation of the model design were enhanced, and the time and labour costs of the model designs were reduced, while construction of a model by subjective experience was avoided.

The remainder of this article is arranged as follows. Theoretical background is described in Section 2. The method proposed in this paper is described in detail in Section 3. In Section 4, a benchmark function was used to test the advancement of an improved intelligent optimisation algorithm, and the validity of the proposed method was verified compared with other diagnostic methods on two experimental datasets of rolling bearings. Section 5 provides the conclusions and future research directions of this study.

2. Theoretical Background

2.1. Long Short-Term Memory Network. LSTM, as an improved recurrent neural network (RNN), adds three gating units to the original RNN structure to control information transmission, namely, input, forgetting, and output gates, which can capture time-related information from input data and avoid gradient explosion defects in RNNs in long sequences [9]. The functions of the input and output gates are to read data and transfer the processed data to the next point in time. The main task of the forgetting gate is to save the useful information in the output unit of the last module and to avoid the backward propagation of useless information at the last moment. The internal updates of the three are expressed as in the six following equations:

$$f_t = \sigma(W_f x_t + V_f h_{t-1} + b_f), \quad (1)$$

$$i_t = \sigma(W_i x_t + V_i h_{t-1} + b_i), \quad (2)$$

$$g_t = \tanh(W_g x_t + V_g h_{t-1} + b_g), \quad (3)$$

$$o_t = \sigma(W_o x_t + V_o h_{t-1} + b_o), \quad (4)$$

$$c_t = g_t \odot i_t + f_t \odot c_{t-1}, \quad (5)$$

$$h_t = o_t \odot \tanh(c_t), \quad (6)$$

where W and V represent the input and hidden status weights, respectively; b represents the deviation, and σ represents the sigmoid function. The input variable at each moment contains the input x_t at the current moment, the unit state c_{t-1} at the previous moment, and the intermediate state h_{t-1} . Intermediate variables include the outputs i_t and o_t of the input and output gates, the output f_t of the forgetting gate, and the output g_t of the input node. The output variables include the current moment unit state c_t and the intermediate state h_t .

2.2. Sparrow Search Algorithm. The SSA algorithm is a new swarm intelligence optimisation algorithm that was proposed by Xue and Shen [10] in 2020. The algorithm has the characteristics of few parameters that require adjustment and good stability and has been applied to battery stack parameter optimisation [11], CT image location detection [12], and other practical problems. The specific algorithm model includes the producer, scrounger, and guard. The producer looks for food and provides foraging directions for the population, the scrounger uses the producer to obtain food, and the guard provides antipredation behaviour when the sparrow population becomes aware of danger. The iterative equations of the three are as follows:

$$X_{i,d}^{t+1} = \begin{cases} X_{i,d}^t \cdot \exp\left[\frac{-i}{\alpha \cdot T}\right], & \text{if } R_2 < ST \\ X_{i,d}^t + \Gamma \cdot L, & \text{if } R_2 > ST \end{cases}, \quad (7)$$

$$X_{i,d}^{t+1} = \begin{cases} \Gamma \cdot \exp\left[\frac{Xw_d^t - X_{i,d}^t}{i^2}\right], & \text{if } i > \frac{n}{2} \\ Xp_d^{t+1} + |X_{i,d}^t - Xp_d^{t+1}|A^+ \cdot L, & \text{otherwise} \end{cases}, \quad (8)$$

$$X_{i,d}^{t+1} = \begin{cases} Xb_d^t + \beta(X_{i,d}^t - Xb_d^t), & \text{if } f_i \neq f_g \\ X_{i,d}^t + K \left[\frac{X_{i,d}^t - Xw_d^t}{|f_i - f_w| + \delta} \right], & \text{if } f_i = f_g \end{cases}, \quad (9)$$

where $X_{i,d}$ represents the position of the i -th sparrow in the d -dimension, t represents the current iteration number, T is the maximum number of iterations, Xw_d^t represents the worst position of the sparrow in dimension d at iteration t , Xp_d^{t+1} represents the optimal position of sparrows in the d -dimension at the $t+1$ iteration of the population, α is a random number between (0, 1), Γ is a random number that follows a standard normal distribution, $R_2 \in [0, 1]$ and $ST \in [0.5, 1]$ indicate the warning value and safety value, respectively, and f_g and f_w are the optimal and worst fitness values of the current sparrow population, respectively.

3. Model Building

The overall framework of this paper is shown in Figure 1, which can be divided into two parts. The HSMSF was constructed, sensor information of multiple measurement points was mapped to multiple high-dimensional domains, and several domains with a strong characterisation ability for information were selected from the basic CNN model as input data. By using the multichannel characteristics of CNN combined with the complexity of each domain, the corresponding domain information was sent to the corresponding CNN channel in chronological order to fuse into a feature map with time characteristics, and the LSTM layer disassembled its space-time characteristics and made fault predictions based on information entropy theory. The CEI-SSA algorithm in the second part served as the feature extraction component of the HSMSF and used its population iterative characteristics to conduct adaptive optimisation of model components and establish an optimal feature model for fault diagnosis.

3.1. Bearing Fault Signal Classifier Based on HSMSF. Bearing fault diagnosis uses only vibration signals that are collected by a single sensor for analysis, and the collected data samples are 1D time-domain sequences, which do not contain obvious fault characteristic information and

FIGURE 1: Flow chart of bearing fault diagnosis based on the proposed method.

insufficient correlation expression between features. This paper focuses on a study of the HSMSF. From multiple perspectives of bearing components, heterogeneous sensors were used to collect sequence signals in multiple time scales, and the collected data were mapped into high-dimensional domains. The multidomain fusion space of the initial signal features was reconstructed in a multichannel cascade mode, and LSTM was introduced into the fusion space for feature disassembly. The complementarity of the features in the spatial and temporal dimensions of the data samples was explored. The probability of the sample labels was solved according to effective features that were obtained by disassembly, and the adaptive entropy weighted fusion of the predicted values of different source sensors was carried out combined with the information entropy theory to achieve efficient diagnosis of the bearing faults. The overall HSMSF framework is shown in Figure 2.

The 1D signal sample was mapped into a high-dimensional space and was presented as a two-dimensional (2D) image by a special coding method. The purpose was to optimise the problem of feature singularity in the original 1D domain by describing details such as the point, line, image colour, or cross boundary. In this HSMSF study, the high-dimensional mapping of data samples has five conversion modes: time domain (TD), continuous wavelet transform (CWT), time-frequency domain [13], Gramian angular field [14], Markov transition field (MTF) [15], and recurrence plot field (RPF) [16].

The CWT is used to transform the signal to the time and frequency domain, which does not require setting the width of the time-frequency window in advance and has good local analysis ability for time and frequency. If the signal $x(t) \in L^2(R)$ is set, the wavelet basis transformation function is shown in the following equation:

FIGURE 2: Architecture of HSMF.

$$\varphi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \varphi\left(\frac{t-\tau}{s}\right). \quad (10)$$

The Gramian angular field encodes the time series in the polar coordinate system to show different information granularity of the samples. Each element in the Gramian matrix is the trigonometric function value of angles. By

considering the difference of the sum or difference operation of trigonometric functions, two different implementations are derived, namely, the Gramian angular summation field (GASF) and the Gramian angular difference field (GADF). The transformation expression function is shown as follows:

$$G_{\text{gasf}} = \begin{bmatrix} \cos(\theta_1 + \theta_2) & \cdots & \cos(\theta_1 + \theta_n) \\ \vdots & \ddots & \vdots \\ \cos(\theta_n + \theta_1) & \cdots & \cos(\theta_n + \theta_n) \end{bmatrix} = \tilde{X}^T \tilde{X} - \sqrt{I - \tilde{X}^2}^T \sqrt{I - \tilde{X}^2}, \quad (11)$$

$$G_{\text{gadf}} = \begin{bmatrix} \sin(\theta_1 - \theta_2) & \cdots & \sin(\theta_1 - \theta_n) \\ \vdots & \ddots & \vdots \\ \sin(\theta_n - \theta_1) & \cdots & \sin(\theta_n - \theta_n) \end{bmatrix} = \sqrt{I - \tilde{X}^2}^T \tilde{X} - \tilde{X}^T \sqrt{I - \tilde{X}^2}, \quad (12)$$

where $\tilde{X}$ is the transformed polar coordinate sequence and I is the unit row vector $[1, 1, \dots, 1]$.

The Markov transition field is a variant based on a first-order Markov chain that is used to solve the problem that the Markov transfer matrix is insensitive to the time characteristics of the sequential samples and that the time and position relationship of the sequential samples is given

priority during transformation. The signal sequence $x(t)$ with length n is mapped to each corresponding range q_j , and the corresponding transition matrix is constructed by combining q_j with the frequency w_{ij} immediately adjacent to each sample. Then the Markov transition field image can be obtained by transforming the following equation:

$$M = \begin{bmatrix} w_{ij}|x_1(t) \in q_i, x_1(t) \in q_j & \cdots & w_{ij}|x_1(t) \in q_i, x_n(t) \in q_j \\ \vdots & \ddots & \vdots \\ w_{ij}|x_n(t) \in q_i, x_1(t) \in q_j & \cdots & w_{ij}|x_n(t) \in q_i, x_n(t) \in q_j \end{bmatrix}. \quad (13)$$

The recurrence plot field is an effective method to analyse the periodicity, chaos, and nonstationary nature of signal sequences and can separate the internal structure, similarity, and predictability of the data samples. The expression of the transformation function is shown as follows:

$$R = \Theta\left(\varepsilon - \|\vec{x}_i - \vec{x}_j\|\right), \quad (14)$$

where Θ is the Heaviside function and ε is the recursive threshold.

The HSMSF considers the correlation and complementarity of different domains in the feature space. Two domains were selected with a strong ability to represent the fault information of the corresponding dataset based on the diagnostic performance of the basic CNN in a single domain. According to the complexity of the different mapping domains, two types of domain data exist with different resolution sizes at different times and CNN site node inputs. In the process of convolution with multicore CNN characteristics, the use of a multichannel cascade mode for multidomain fault characteristic information provides significant overlays, and space-time characteristics are included in the multidomain fault feature fusion. We assume that signal i maps to data vectors $s^k(i)$ and $c^k(i)$ of the above two optimal domains. Equation (15) can be obtained for M preset data channels in the basic CNN structure:

$$\begin{cases} s^k(i) = [s_1^k(i), s_2^k(i), \dots, s_m^k(i)]^T, \\ z^k(i) = [z_{m+1}^k(i), z_{m+2}^k(i), \dots, z_M^k(i)]^T, \end{cases} \quad (15)$$

where m is an integer in the interval $(1, M)$; that is, the channel proportion of different mapping domains will be adjusted dynamically according to the spatial complexity of its mapping domain. To make full use of the information represented by a domain with more complex features, it is necessary to input the information at the front end of the network with higher resolution. Because only two domains with better performance were selected in this study for comparison of the two types of data, when they were input into the network, the domain with a higher degree of spatial complexity was input one convolution layer ahead of the other, and the resolution was set to twice that of the other. Based on the above settings, the channel proportion of the domain with higher spatial complexity was two-thirds (if the channel value is not an integer, it should be rounded down). Convolution and pooling operations were carried out, and a full connection layer was added after the last convolution block for data dimension reduction to eliminate redundant information between different domains. The stack form was used in the splicing of spatial features of different mapping domains. Unlike the traditional *concat* and *add* methods, stack can retain the time step information of the feature tensor. The data after dimension reduction of the two types of domains were spliced into a 2D matrix form, which was synchronised with the number of samples into a three-dimensional (3D) block and then transmitted to the LSTM layer as an input.

The LSTM layer was equipped with a special gating mechanism, which prevents the content of feature units within multiple time steps from being tampered with, enables learning of long-term dependent features between multiple domains, and enables processing of global features that have not been explored in the CNN layer. Therefore, the LSTM has a superior solution space capability for the above splicing data blocks. The solution process is shown in equations (1)–(6). During this process, the cross-entropy loss algorithm was used to reverse the propagation of training errors and gradually update model parameters layer by layer. When the model parameters were updated to an optimum, the fault features with a strong spatiotemporal characterisation ability could be disintegrated and input into the SoftMax function to solve for the probability value. A corresponding output probability exists for the u group heterogeneous sensor. To ensure the stability and classification accuracy of the diagnostic model, decision-level fusion was carried out by the adaptive entropy weighted fusion method for group u values, as follows.

The probability output matrix of the heterogeneous sensor group can be constructed from the output probability value of group u :

$$P(x) = \begin{bmatrix} p_{11}(x) & p_{12}(x) & \cdots & p_{1n}(x) \\ \vdots & \vdots & \vdots & \vdots \\ p_{u1}(x) & p_{u2}(x) & \cdots & p_{un}(x) \end{bmatrix}, \quad (16)$$

where each line represents the final diagnostic prediction probability of a group of sensor data x and n represents the number of fault categories. The analysis shows that if the difference in probabilities from each group of sensors was small, the classification uncertainty was large. If the difference between the maximum probability and other probabilities was large, the classification result was more reliable. Therefore, the information entropy $H_i(x)$, as shown in (17), was used to represent the uncertainty of the classification:

$$H_i(x) = - \sum_{j=1}^n p_{ij}(x) \log_2 p_{ij}(x), \quad i = 1, 2, \dots, u, \quad (17)$$

where $p_{ij}(x)$ represents the probability that the classification predicted value of the i -th sensor group is identified as belonging to j . According to the information entropy value of each group of sensors, the adaptive fusion weight is given by (18). Combined with the probability output matrix $P(x)$ of (16), the final probability output matrix $P'(x)$ is obtained as shown in (19):

$$\omega_i = \frac{\exp(-H_i(x))}{\sum_{p=1}^n \exp(-H_p(x))}, \quad i = 1, 2, \dots, u, \quad (18)$$

$$P'(x) = \begin{bmatrix} \omega_1 p_{11}(x) & \omega_1 p_{12}(x) & \cdots & \omega_1 p_{1n}(x) \\ \vdots & \vdots & \vdots & \vdots \\ \omega_u p_{u1}(x) & \omega_u p_{u2}(x) & \cdots & \omega_u p_{un}(x) \end{bmatrix}. \quad (19)$$

The maximum label is obtained by summing the matrices by column weight to provide the result according to

$$\text{label}(x) = \arg \max \left[\sum_{j=1,2,\dots,n} \omega_j p_{ij}(x) \right]. \quad (20)$$

3.2. Adaptive Optimisation of Feature Extractor Model Based on CEI-SSA. In the HSMSF feature extractor, the basic CNN structure was inspired by application of the CNN network to diagnose bearing faults in the literature [17]. In addition, $32 \ 3 \times 3$ convolution blocks and $32 \ 2 \times 2$ max-pooling blocks were added to the front end as a preconvolution layer in the complex feature domain, and the LSTM that was used for the spatiotemporal feature decomposition adopted a single-layer network structure. However, such feature extractors are mostly models selected based on human experience, and the determination of internal parameters depends on extensive expert experience for debugging. Because of differences in the form of bearing data sampling, it is likely that the model will need to be reconstructed for different diagnostic problems. Therefore, this work proposes an adaptive optimisation method for feature extractor models based on CEI-SSA; the algorithm's flow chart is shown in Figure 3.

Initial sparrow individuals in the original SSA were generated randomly within the search space. It was difficult to obtain good population diversity, which led to poor convergence and local algorithm development problems. The reverse learning strategies by chaotic sequence initialisation and the elite strategy provided two approaches to improving the quality of the initial SSA solution and enhanced the global and local search capabilities of the algorithm.

The chaotic sequence has the advantages of a uniform ergodic space and fast convergence. It can improve the global search ability of sparrows in the SSA and avoid a decrease in diversity of parameters to be optimised by the population in later iterations. To consider the randomness and regularity of a chaotic operator, a chaotic sequence is generated by Tent mapping to initialise the population. The Tent mapping expression is as follows:

$$z_d^{(i+1)} = \begin{cases} 2z_d^{(i)}, & 0 \leq z \leq \frac{1}{2} \\ 2(1 - z_d^{(i)}), & \frac{1}{2} < z \leq 1 \end{cases} \quad d = 1, 2, \dots, D. \quad (21)$$

After the population as initialised by chaos is generated, chaotic individuals need to be transformed into the corresponding search space. The transformation formula is as follows:

$$X_{i,d} = X_{lb,d} + (X_{ub,d} - X_{lb,d})z_d^{(i)}, \quad d = 1, 2, \dots, D, \quad (22)$$

where $X_{ub,d}$ and $X_{lb,d}$ are the upper and lower boundaries of the individuals.

The main idea of a reverse learning strategy is to find a feasible solution to a problem, find its reverse solution,

FIGURE 3: Flow chart of CEI-SSA algorithm.

evaluate the original solution, reverse the solution, and then select the optimal solution as the next generation of individuals. This strategy can improve the individual quality of sparrows in the SSA and improve the local search performance and exploration ability of the algorithm. Let the feasible solution of individual No. 1 in the above D-dimensional space be $X_{1,d} = (X_{11}, X_{12}, \dots, X_{1D})$ ($X_{1,d} \in [X_{lb,d}, X_{ub,d}]$); then its reverse solution is $\bar{X}_{1,d} = (\bar{X}_{11}, \bar{X}_{12}, \dots, \bar{X}_{1D})$, where $\bar{X}_{1,d} = r(X_{lb,d} + X_{ub,d}) - X_{1,d}$ and r are random numbers on the interval $[0, 1]$.

The reverse solution generated by the reverse learning strategy does not necessarily make it easier to find the global optimal solution compared with the original solution in the current search space. To solve this problem, an elitist reverse solution was used to improve the algorithm. It has been proven that elite reverse solution individuals carry more effective information than ordinary individuals [18]. The excellent individuals that were selected from the current population and the reverse population that was formed by the elite individuals are of great search value and can avoid algorithm prematurity and search out the required structural parameters more effectively.

The elite individual is the extreme point that corresponds to the general individual; namely, $X_{i,d}^E = (X_{i1}^E, X_{i2}^E, \dots, X_{iD}^E)$ ($i = 1, 2, \dots, N; d = 1, 2, \dots, D$), and its reverse solution $\bar{X}_{i,d}^E = (\bar{X}_{i1}^E, \bar{X}_{i2}^E, \dots, \bar{X}_{iD}^E)$ is defined by

$$\overline{X}_{id}^E = \delta \times (lb_d + ub_d) - X_{id}^E, \quad (23)$$

where δ is the random value of interval $[0, 1]$, $X_{id}^E \in [lb_d, ub_d]$, $lb_d = \min(X_{id}^E)$, $ub_d = \max(X_{id}^E)$, and lb_d and ub_d are the lower and upper bounds of the dynamic boundary, respectively. The dynamic boundary overcomes the problem that the fixed boundary makes it difficult to preserve the search experience, so that the elite reverse solution is in the space of the fine range, which is conducive to algorithm convergence. The reset equation (equation (26)) is used to reset when $\overline{X}_{id}^E$ goes out of bounds.

$$\overline{X}_{id}^E = \text{rand}(lb_d + ub_d) \quad (24)$$

Constructing a feature extractor model usually requires human professional experience and repeated experiments to determine the optimal parameter structure, resulting in insufficient model generalisation and the need to redesign the structure in the face of new diagnostic problems. Moreover, some parameters of the feature extraction model are crucial to its performance. For example, if the batch size is too small, the randomness of the model will become greater, making it difficult to achieve convergence. If the batch size is too large, the training solution will fall into local optima. When the learning rate is too slow, the model is prone to overapproximation and slow convergence, and when the learning rate is too fast, the training process will have large oscillations. For different training tasks, the selection of different kinds of optimisers also affects diagnostic model performance. For example, although the well-known Adam algorithm does not need artificial intervention to adjust the learning rate, the learning rate will rise suddenly in a certain parameter space, resulting in nonconvergence. The size of the convolution kernel, the number of channels, the number of hidden layer units, and the pooling mode are also crucial to the model. Improper collocation and combination will lead to misidentification or loss of the interlayer receptive field, leading to degradation of model diagnostic performance.

In summary, this paper uses CEI-SSA modified by the above mechanism to conduct structural optimisation operations on the feature extractor model, with the aim of reducing the difficulty of model design and improving the adaptive recombination and adaptation ability of the model. The 19 dimensions of each sparrow corresponded to 18 undetermined parameters of the CNN layer and one undetermined parameter of the LSTM layer in the feature extractor. The convolution kernel had four convolution layers (Ck_s_1 , Ck_s_2 , Ck_s_3 , and Ck_s_4), four convolution core layers ($channel_1$, $channel_2$, $channel_3$, and $channel_4$), four activation functions in the convolution layer (Act_1 , Act_2 , Act_3 , and Act_4), a pooling mode of the other three layers except for the third layer ($pool_1$, $pool_2$, and $pool_4$), an LSTM layer hidden unit number (Hid_s), a learning rate (L_rate), an optimiser (Op), and a batch size ($Batch$). Except for the activation function, pooling mode, and learning rate, 11 undetermined parameters needed rounding of their first decimal because of

integer programming of information. The feature extractor model was established according to the position information of each sparrow after chaotic initialisation. The initial feature extractor model was trained, and its prediction error was calculated on the test set using equation (27):

$$f = \frac{1}{\text{batch}} \sum_{i=1}^{\text{batch}} \frac{|t(i) - y(i)|}{t(i)}, \quad (25)$$

where $y(i)$ is the prediction label, $t(i)$ is the real label, and batch is the batch size. f was used as a fitness function in CEI-SSA for population updating, and the optimal individual information was used to build the final feature extractor model. Specific optimisation steps are shown in Algorithm 1.

4. Experimental Validation

In this section, we evaluate the optimisation performance of CEI-SSA on 12 test functions and test it on two fiducial bearing datasets, Case Western Reserve University (CWRU) and Intelligent Maintenance Systems (IMS), to verify the effectiveness of the proposed method. The 12 test functions are Sphere (F1), Schwefel N2.22 (F2), Schwefel N1.2 (F3), Schwefel N2.21 (F4), Rosenbrock (F5), Step (F6), and Quartic (F7), which are seven high-dimensional single-peak test functions, and Schwefel (F8), Rastrigin (F9), Ackley (F10), Griewank (F11), and Penalized (F12) [19], which are five high-dimensional multipeak test functions.

4.1. Performance Verification of CEI-SSA Algorithm. To verify the optimisation performance of the CEI-SSA algorithm, CEI-SSA was compared with the SSA [10, 20], Dragonfly Algorithm (DA) [21], Grey Wolf Optimiser (GWO), Artificial Bee Colony (ABC), and Particle Swarm Optimisation (PSO) algorithms on the 12 test functions. The population size of each algorithm was set to 100, the maximum number of iterations was set to 1000, and the numbers of producers and guards were set to 20% of the population size. To avoid the contingency of optimisation results and prove CEI-SSA stability, each benchmark function was run independently 30 times, the experimental results were selected as experimental data, and the mean (Mean) and standard deviation (Std) of each algorithm were taken as the final evaluation indices. The comparative results are shown in Tables 1 and 2.

According to the analysis in Table 1, for the same test constraints, the statistical results of CEI-SSA on five groups of high-dimensional single-peak test functions were significantly better than those of the other five comparison algorithms. For the test functions F1, F2, and F3, CEI-SSA yielded its theoretical optimal solution stably. When solving F3, F5, F6, and F7, CEI-SSA failed to find the theoretically optimal solution, but it was several orders of magnitude better than the other algorithms on the two evaluation indices of mean and standard deviation. Compared with the other algorithms, the improved CEI-SSA had a stronger algorithm development ability and stability on single-peak functions.

Input: the number of sparrows N
Output: optimal fitness value F_g ; Optimal value of feature extractor structure y_{best}

- (1) **initialization:** initialize sparrow population via (21) and (22);
- (2) **repeat**
- (3) **while** $k \in$ Population size **do**
- (4) Calculate fitness values of the each individual via (25);
- (5) Sort each individual according to its fitness value;
- (6) Random R_2 value in the interval $[0, 1]$;
- (7) **for** each producer $i = 1$ to $(N * PD)$ **do**
- (8) Update the location of the producer via (7).
- (9) **end for**
- (10) **for** each producer $j = (N * PD) + 1$ to N **do**
- (11) Update the location of the scroungers via (8).
- (12) **end for**
- (13) Use the elite reverse strategy to reverse solution and update outstanding individual via (23) and (24);
- (14) Update F_g and y_{best} ;
- (15) $k = k + 1$;
- (16) **end while**
- (17) **until** a fixed number of iterations
- (18) **return** F_g, y_{best}

ALGORITHM 1: The framework of optimising feature extractor.

TABLE 1: Results achieved by the algorithms on unimodal functions (dimension = 30).

	Mean SSA	Std.	Mean DA	Std.	Mean GWO	Std.
F1	3.472E-35	1.654E-34	1.200E-02	3.200E-02	2.061E-34	3.192E-33
F2	1.153E-20	4.559E-20	1.818E+00	1.497E+00	1.155E-20	3.784E-19
F3	1.648E-26	8.040E-26	2.137E+02	4.291E+02	1.569E-08	1.515E-08
F4	2.652E-14	1.221E-13	1.895E+00	1.012E+00	2.024E-08	1.173E-08
F5	8.360E-02	1.988E-06	2.197E+00	9.501E-01	2.645E+01	1.376E-04
F6	1.580E-03	1.541E-10	1.167E+00	1.408E+00	2.223E-01	1.848E-01
F7	5.110E-04	1.660E-04	1.790E-03	1.051E-01	8.752E-04	4.663E-04
	Mean ABC	Std.	Mean PSO	Std.	Mean CEI-SSA	Std.
F1	2.689E+04	2.408E+04	2.240E-24	5.938E-22	0.0	0.0
F2	1.732E+01	7.150E+00	1.491E-12	4.727E-13	0.0	0.0
F3	1.558E+01	7.743E+00	0.075E+00	0.064E+00	6.276E-93	1.492E-93
F4	1.321E+05	9.865E+04	1.262E+00	1.085E+00	0.0	0.0
F5	3.532E+01	1.465E+01	4.564E+01	3.543E+01	5.014E-08	1.443E-07
F6	1.736E+01	1.736E+01	7.229E-04	6.384E-04	1.350E-12	1.777E-12
F7	1.019E-04	5.536E-05	3.024E-03	2.641E-03	5.835E-05	4.013E-05

Table 2 shows the high-dimensional multipeak test functions F9 and F11. CEI-SSA and SSA converged stably to a global optimum. In the F10 test function, CEI-SSA and SSA reached a value closest to the optimal fitness and had strong stability compared with other algorithms. When solving the F8 function, there was no obvious difference between the results of several algorithms; CEI-SSA performed slightly better than the other algorithms in optimisation and was slightly less stable than SSA. When solving F12, CEI-SSA was slightly worse than SSA on the two evaluation indices but was greatly superior to the other comparison algorithms. Therefore, CEI-SSA showed good performance in multiple optimisation processes and on single-peak or multipeak test functions. Therefore, CEI-SSA has an efficient global optimisation local exploration ability and can fully and efficiently explore the search space with strong stability and robustness;

hence, it can be applied to the adaptive optimisation problem of the HSMSF model in this paper.

4.2. Case 1: CWRU-Bearing Dataset. CWRU datasets were provided by the Case Western Reserve University's Loaded Data Center [22] and have been used extensively in rolling bearing fault diagnosis. The experimental device is shown in Figure 4. It is mainly composed of a 2-HP motor, a bearing accelerometer, a torque sensor, and a power tester. The driving end and fan end test bearings were 6205-2RS JEM SKF and 6203, respectively. By EDM, single-point faults were arranged on the rolling body, inner raceway, and outer raceway of the bearings with damage levels of 7, 14, and 21 mm. An acceleration sensor was placed on the driving end of the motor and the bearing seat of the fan end of the motor,

FIGURE 4: Experimental platform for CWRU dataset. (a) Signal acquisition device. (b) Schematic of the test rig with details.

respectively, to collect vibration acceleration signals at motor speeds from 1730 RPM to 1797 RPM and motor loads from 0 HP to 3 HP. Vibration signals were collected by a 16-channel data recorder with a sampling frequency of 12 kHz.

The main purpose of this example was to verify the accuracy of the HSMSF model based on CEI-SSA adaptive optimisation in the proposed bearing fault diagnosis method for diagnosis of bearing defect location and damage degree. Therefore, in the process of type label assignment, the same fault types at different motor speeds and loads under the background of signal data were given the same type label. According to the standard, ten faults with different severities in the rolling element, inner raceway, and outer raceway of the driving-end rolling bearing were studied for sensors with different sources on the driving-end and fan-end bearing seats. Figure 5 shows the time-domain waveform of a 10 s vibration signal that was collected from sensors at the driving end and the fan end when the internal raceway fault damage was 14 mm.

The vibration signals of each fault state monitored by the sensor were divided into samples of equal length. The sample set of 10 fault samples was generated according to a window size of 1024 time points and a window moving step of 1000 time points. For the same fault type, 300 samples were collected under four operating conditions (3 HP/1730 RPM, 2 HP/1750 RPM, 1 HP/1772 RPM, and 0 HP/1797 RPM). Twelve hundred samples in each state were obtained, for a total of 12,000 samples. Two types of sensors at the driver end and fan end were used to construct datasets DE and FE, and the ratio of training set to test set in each dataset was 8 : 2. A detailed description of the experimental data is given in Table 3. To adapt to the data input format of the feature extractor and carry out high-quality domain information fusion, multispace domain transformation was carried out for various fault samples in the DE and FE datasets. Figure 6 lists the 2D images of five corresponding domains after transformation of some samples.

When the validity of the fault sample spatial domain transformation had been verified, the basic CNN structure mentioned in Section 3.2 was used to compare the diagnostic accuracy of the DE dataset. The detailed configuration structure is shown in Table 4. To explore the influence of the resolution of the input layer and the number of network

layers on the ability of each spatial domain to characterise fault features, the input layer of each domain was set to resolutions of 32×32 , 112×112 , and 224×224 , and eight 3×3 convolution blocks and eight 2×2 max-pooling blocks were added to the front end of the basic CNN structure for testing through five experiments. The same parameter configuration was used for all experiments, with a learning rate of 0.001, a batch size of 32, and 50 iterations. Specific comparison results are shown in Figures 7(a)–7(d).

Compared with the time domain method, the average accuracy from all domain mapping methods was improved, which proves the advancement of the proposed domain transformation method. By comparing the results in Figures 7(a) and 7(b), when the CNN structure was a three-layer convolution, the average diagnostic accuracy of samples in each domain with an input resolution of 112×112 was improved by $\sim 2\%$ – 9% compared with the original 32×32 input size. This means that resolution loss cannot be ignored in diagnosis. However, higher resolution does not imply a more accurate diagnosis. Figures 7(b) and 7(c) show that an increase in the resolution of the input layer decreases the average accuracy of diagnosis when using CWT time-frequency domain samples. This result may originate from differences in domain mapping methods. The pixel intensity of the noise in such image samples becomes more obvious with an increase in resolution, which affects diagnostic accuracy. Shallow network layers may not be able to fully characterise the features of domain samples with complex structures, and therefore the CNN structure was changed to four layers based on the constant resolution of the input layer, 224×224 . Figures 7(c) and 7(d) show that the diagnostic accuracy of each domain was improved to some extent after the network structure of one layer was deepened. Except for the CWT time-frequency domain, the results obtained in the other domains after combination in this way were the best among the four combinations of verification methods. These results indicate that the network model performs optimally only when it matches the corresponding data. Too deep or too shallow a layer leads to a decrease in diagnostic performance.

To combine the advantages between domains, and for better application to the research method in this work, the CWT (3-layer structure, 112×112 resolution) with the

FIGURE 5: 10 s vibration signal time-domain waveform at the driving end and fan end.

TABLE 2: Results achieved by the algorithms on multimodal functions (dimension = 30).

	Mean SSA	Std.	Mean DA	Std.	Mean GWO	Std.
F8	$-7.953E+03$	$7.204E+02$	$-1.99E+03$	$1.701E+03$	$-1.702E+03$	$1.473E+02$
F9	0.0	0.0	$1.218E+01$	$5.908E+00$	$3.047E-00$	$3.716E+00$
F10	$8.882E-16$	0.0	$2.204E+00$	$2.167E+00$	$7.021E-00$	$2.751E-14$
F11	0.0	0.0	$3.381E-01$	$1.243E-02$	$1.325E+01$	$4.741E-03$
F12	$3.352E-12$	$2.853E-11$	$5.014E-01$	$3.313E-01$	$1.826E-01$	$1.422E-04$
	Mean ABC	Std.	Mean PSO	Std.	Mean CEI-SSA	Std.
F8	$2.365E+00$	$1.677E+00$	$3.299E+00$	$5.017E-01$	$-1.250E+04$	$6.492E+02$
F9	$5.342E-05$	$1.216E-04$	$1.291E+00$	$1.182E-01$	0.0	0.0
F10	$3.299E+0$	$5.017E-01$	$0.175E+00$	$0.267E+00$	$8.882E-16$	0.0
F11	$1.291E+00$	$1.182E-01$	$8.594E-01$	$1.121E+00$	0.0	0.0
F12	$4.752E-06$	$4.443E-06$	$7.153E-02$	$4.502E-01$	$1.124E-11$	$3.204E-11$

FIGURE 6: Images of partial samples transformed by five domains.

TABLE 3: Experimental data in case 1.

Dataset	DE/FE									
Number of samples	1200/ 1200	1200/ 1200	1200/ 1200	1200/ 1200	1200/ 1200	1200/ 1200	1200/ 1200	1200/ 1200	1200/ 1200	1200/ 1200
Fault location	Rolling element			Inner raceway			Outer raceway			Normal
Fault size (mils)	7	14	21	7	14	21	7	14	21	0
Class label	1	2	3	4	5	6	7	8	9	10

highest average accuracy in the above verification experiment was selected as the back-end input of the improved structure described in Section 3.1, and the second GADF (4-layer structure, 224×224 resolution) was selected as the front-end input. A multichannel cascade method combined with an LSTM temporal and spatial feature decomposition was used in the diagnostic experiment verification. The diagnostic accuracy curve of the test data is shown in Figure 8.

In Figure 8, CWT-CNN refers to the original basic network input by CWT, GADF-CNN refers to the original basic network input by GADF, and P-Net refers to the

network structure as improved by the multichannel cascade mode and by LSTM. To analyse the improvement in feature representation ability of the improved structure, a t-distributed stochastic neighbour embedding (T-SNE) [23] was used to visualise the feature distribution of the three contrast modes, as shown in Figures 9(a)–9(c). Each point in the diagram represents a sample, and different types of faults are represented in different colours. The results show that, compared with the previous two methods, P-Net has better feature separation ability and classification performance, and the extracted features can be clustered. The verification accuracy was 96.2%, which was $\sim 2.7\%$ higher than that of the

FIGURE 7: Diagnostic accuracy of experiments conducted by four combinations in case 1. Network structure and Input resolution size: (a) 3 layers and 32×32 , (b) 3 layers and 112×112 , (c) 3 layers and 224×224 , and (d) 4 layers and 224×224 .

previous optimal CWT-CNN. The multidomain feature fusion method developed in this research makes good use of the advantages of each domain and helps the model to separate different fault types to a greater extent after fusion of the multidomain feature information.

Figure 9(c) shows that although P-Net has a certain performance improvement in diagnostic accuracy, overlap remains among categories 1, 3, and 8. This occurred largely because an insufficient number of basic components in the model were set for the artificial experience, and the original model parameter settings affected the quality of the feature extraction. The preset convolution channel in the multi-channel model is likely to appear during the cascade channel redundancy phenomenon, which increases the required

calculations and reduces fault classification accuracy. Therefore, it is necessary to study adaptive optimisation of the structure of diagnostic models.

CEI-SSA, which had been verified experimentally, was used to conduct model adaptive optimisation for P-Net. To reduce the time and cost required, when the loss function of the test set did not decrease for 10 consecutive cycles, the program stopped and saved the model that corresponded to the minimum loss function of the test set. The initial parameter settings of CEI-SSA are shown in Table 5, and the 19 parameters were attached when individuals obtained the global optimal solution, as shown in Table 6. The 3D diagram of the iterative process of each sparrow in the population is shown in Figure 10, with horizontal and vertical coordinates that

FIGURE 8: Diagnostic accuracy of single-domain method and multidomain fusion model in case 1.

correspond to channel_1 and channel_2 parameters. All individuals converged to the optimal fitness of the population according to the optimisation mechanism, which indicates that the CEI-SSA used in this work can perform good model adaptive optimisation and deal adaptively with the corresponding fault diagnosis problems to reduce the cost of manpower and design time. Figure 11 shows the CEI-SSA adaptive optimisation of the model curve after AP-Net classification accuracy. The diagnostic accuracy increased to 97.7%. Furthermore, because the original partially redundant structure was abandoned during model adaptive optimisation, the accuracy curve convergence was faster and more stable, which verified the effectiveness of the optimisation method of the feature extractor model component in this paper.

The signals detected by sensors at different measuring points were complementary, and therefore AP-Net was used to extract fault features from the FE dataset collected at the fan end, and the category probability prediction was obtained from the two types of datasets. The adaptive entropy weighted fusion method described in Section 3.1 was used for verification and analysis. Figure 12 shows that, through the confusion matrix, the model identified the results of each health state in the test set data after fusing information from different sensors. The average diagnostic accuracy increased to 99.8%, the error sample size decreased to 6, and the accuracy of each type of sample exceeded 99%. Therefore, the HSMSF adaptive optimisation based on CEI-SSA in this work can apply different sources of information to improve diagnostic accuracy. Figure 13 shows a visual analysis of the feature dimension reduction of the final model. Compared with Figure 9(c), the model showed good same-type cohesion and a strong ability to separate different types of states.

4.3. Case 2: IMS-Bearing Dataset. It is difficult to verify the effectiveness and practical application of the proposed

method to CWRU dataset with artificially damaged parts. Therefore, the method described in this paper was also tested on the IMS dataset, which was collected from the natural evolution of bearing faults and provided by the Intelligent Maintenance Systems Center of the University of Cincinnati [24]. The bearing fault experimental platform is shown in Figure 14 and consists of the main body part, the transmission part, the loading system, the lubrication system, and the control circuit. The speed of the AC motor is 2000 RPM. The test shaft consisted of four Rexnord ZA-2115 double-row bearings with pitch diameter of 2.815 in., roller diameter of 0.331 in., and pressure angle of 15.17°. Thermocouple sensors placed on the bearing together with the lubrication system monitored and regulated the bearing temperature and lubrication. The shaft and bearing carried a radial load of 26.67 kN applied by the spring mechanism. Each bearing was connected to two high-precision accelerometers (PCB 253B33 High-Sensitivity Quartz ICP®), both measured in the X- and Y-directions. The sampling frequency was 20 KHz, and 20480 data points were sampled each time.

The IMS dataset records three groups of data according to different situations. The first group of experimental data was used to construct the dataset. A total of 2156 experimental data points from four sets of bearings were recorded. When the test bench stopped, bearings 1 and 2 were normal, bearing 3 experienced inner ring failure, and bearing 4 experienced rolling body failure. Three data categories therefore existed. For each type of fault sample, the sample set was constructed by moving the window by steps of 600 points with a window length of 1024 time points. Finally, 2,155 samples were generated for each state; a total of 6,465 samples were generated. The bearings in the X- and Y-directions were divided into two datasets, *IMS_X* and *IMS_Y*. Each dataset was divided into a training set and a test set in the ratio of 8:2. A detailed description of the experimental data is provided in Table 7. For the *IMS_X* dataset, the

FIGURE 9: t-SNE visualisation of features extracted from three comparison models in case 1. (a) GADF-CNN model. (b) CWT-CNN model. (c) P-Net model.

TABLE 4: Basic CNN structure parameters.

Layer	Number/size/activate	Number of nodes
Convolutional layer 1	16/3/-	
Max-pooling layer 1	16/2/Relu	
Convolutional layer 2	32/3/Relu	
Convolutional layer 3	48/3/-	
Max-pooling layer 3	48/2/Relu	
Fully connected layer		512
Output layer		10

TABLE 5: Initial setting of CEI-SSA parameters.

Parameter	Scope
Population	20
Dimension	19
Iterations	6
Optimiser	“Adam”, “SGD”, “Adamx”, “Adadelata”, “Adagrad”, “AdamW”, “ASGD”, “RMSprop”
Learning rate	$[1e-5, 1e-1]$
Batch size	$[16, 64]$
Activation function	“Relu”, “Sigmoid”, “Relu6”, “Tanh”, “LRelu”, “Softsign”
Channel number	$[3, 15]$, $[16, 30]$, $[31, 45]$, $[46, 60]$
Kernel size	$[2, 3, 3, 5, 6, 6, 7, 11]$
Pooling form	“Maxpool”, “Avgpool”
Hidden unit in LSTM	$[20, 80]$

FIGURE 10: CEI-SSA adaptive optimisation process for individual iteration 3D graph. (a) The first iteration. (b) The sixth iteration.

FIGURE 11: Diagnostic accuracy curve of AP-Net model in case 1.

Predicted label	1	300 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	2	2 0.1%	298 9.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.3% 0.7%
	3	0 0.0%	1 0.0%	298 9.9%	0 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	99.3% 0.7%
	4	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	299 10.0%	0 0.0%	0 0.0%	1 0.0%	0 0.0%	99.7% 0.3%
	6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	0 0.0%	100% 0.0%
	8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%	0 0.0%	100% 0.0%
	9	0 0.0%	1 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	299 10.0%	99.7% 0.3%
	10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	300 10.0%
		99.3% 0.7%	99.3% 0.7%	100% 0.0%	100% 0.0%	99.7% 0.3%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	99.8% 0.2%
	1	2	3	4	5	6	7	8	9	10	
	Ture label										

FIGURE 12: Confusion matrix of the proposed method in case 1.

FIGURE 13: Visualisation of features extracted of the proposed method in case 1.

FIGURE 14: Experimental platform for IMS dataset. (a) A bearing with the connected accelerometers. (b) Schematic of the test rig with details.

TABLE 6: The optimal model components after CEI-SSA optimisation in case 1.

Variable	Value
<i>Ck_s_1</i>	9
<i>Ck_s_2</i>	3
<i>Ck_s_3</i>	3
<i>Ck_s_4</i>	2
<i>channel_1</i>	12
<i>channel_2</i>	24
<i>channel_3</i>	43
<i>channel_4</i>	52
<i>Act_1</i>	"Relu"
<i>Act_2</i>	"LRelu"
<i>Act_3</i>	"LRelu"
<i>Act_4</i>	"Relu"
<i>pool_1</i>	"Maxpool"
<i>pool_2</i>	"Maxpool"
<i>pool_4</i>	"Avgpool"
<i>Hid_s</i>	50
<i>L_rate</i>	0.0006
<i>Op</i>	"Adamx"
<i>Batch</i>	44

diagnostic accuracy comparison results of each domain are shown in Figures 15(a)–15(d) according to the parameter setting of the domain transformation mode in Experiment 1.

According to the comparison results, compared with the time-domain transformation, the other five domain transformation methods in the *IMS_X* dataset improved the diagnostic accuracy by 1 to 30 points, and the optimal average accuracy was 93.5%. The conclusions from analysing the influence of resolution and network layer on the diagnostic performance are consistent with those from Experiment 1. The average accuracies of the CWT (3-layer structure, 112×112 resolution) and GADF (4-layer structure, 224×224 resolution) models were the top two. When input into the above P-Net model, the diagnostic accuracy curve in Figure 16 resulted. The results verify the P-Net processing method compared with that for a single domain. The fault features of the samples were separated, their

inherent information was explored, and the fault identification accuracy improved to 95.6%.

CEI-SSA was used to conduct model adaptive optimisation for P-Net on the IMS dataset with the same parameter settings as in Experiment 1. Table 8 shows the 19 optimal component parameters that were obtained after individual adaptive optimisation of the algorithm on the IMS dataset. The classification accuracy curve of the optimised AP-Net model is shown in Figure 17. The diagnostic accuracy was improved to 96.3%, and the model convergence was rapid and stable, with a reduction in manpower and time cost consumed in the design. The above verifies the good generalisation ability of CEI-SSA in this work and the effectiveness of the model self-optimisation method.

The *IMS_Y* datasets of sensors in different directions were combined to verify the validity of the method. AP-Net was used to extract the features of the two datasets, and the diagnostic performance was verified by using an adaptive entropy-weighted fusion method. The final confusion matrix of the three health states of the bearings on the IMS dataset is shown in Figure 18. The average diagnostic accuracy improved to 99.8%, and the performance improved significantly. In all test samples, only three samples in the inner ring fault class were classified incorrectly belonging to the normal class. A visualisation of the feature dimension reduction of the model is shown in Figure 19, which proves that the model can delete redundant information and has reliable feature separation ability during feature learning. The CEI-SSA optimisation with the HSMSF method improved the accuracy of fault diagnosis and had a strong generalisation ability.

To explore the influence of the selection and number of training samples on algorithm's performance, experiments were carried out on the CWRU and IMS datasets according to five segmentation ratios of training set and test set (5:5, 6:4, 7:3, 8:2, and 9:1). The CWT-CNN, GADF-CNN, and AP-Net algorithms were selected for comparison and verification. Both CWT-CNN and GADF-CNN were set according to the same fixed structural parameters, whereas the parameter setting in AP-Net used an adaptive selection

FIGURE 15: Diagnostic accuracy of experiments conducted by four combinations in case 2. Network structure and input resolution size: (a) 3 layers and $32 * 32$, (b) 3 layers and $112 * 112$, (c) 3 layers and $224 * 224$, and (d) 4 layers and $224 * 224$.

value. The diagnostic accuracy of the two datasets is shown in Figure 20. It is apparent that when the segmentation ratio was 5 : 5, all algorithms performed poorly. The main reason for this was that the training and test sets could not provide reasonable learnable parameters for the algorithm model under the segmentation ratio of 5 : 5, resulting in poor algorithm fitting. In contrast, the other four segmentation methods could basically meet the training and diagnosis requirements of the algorithm. Among them, the diagnostic accuracy of the AP-Net model was always better than those of the other two methods and showed high robustness.

To illustrate the superiority of the proposed method, the performances of other deep and shallow learning

algorithms were compared with that of the proposed method according to the same data partitioning method. The deep learning algorithms included ST-CNN [17], the Deep Belief Network, and CNNEPDNN [25]. ST-CNN is composed of three convolution layers with filter sizes of 3×3 , the Deep Belief Network structure is 1024-512-256-128/10, and CNNEPDNN is composed of two convolution layers with filter sizes of 5×1 and four hidden-layer DNN structures with 20, 40, 80, and 160 nodes. Shallow learning methods included the support vector machine (SVM), the PSO least-squares support vector machine (PSO-LSSVM) [6], and the back-propagation (BP) neural network. During the comparison of the algorithms' performances, the

FIGURE 16: Diagnostic accuracy curve of P-Net model in case 2.

TABLE 7: Experimental data in case 2.

Dataset	Number of samples	Fault location	Class label
<i>IMS_X/IMS_Y</i>	2155/2155	Rolling element	1
	2155/2155	Inner raceway	2
	2155/2155	Normal	3

FIGURE 17: Diagnostic accuracy curve of AP-Net model in case 2.

influence of feature set selection on diagnostic accuracy was considered. In the deep learning algorithm, three feature sets, CWT, GADF, and FD, were selected for experiments. FD was the frequency-domain image converted from timing signals at constant speed, and data samples are shown in Figure 21. Because the shallow learning method cannot extract the sequence features automatically, the mean, variance, root mean square, peak factor, kurtosis coefficient, and waveform factor of the signal were extracted manually in the experiment as six time-domain

features, and the centre frequency, root-mean-square frequency, frequency standard deviation, correlation factor, spectral moment, and total power spectrum were extracted manually in the experiment as the six characteristics of the frequency domain. All models were verified five times, and the initial learning rate and number of iterations were set to 0.001 and 50, respectively. A diagnostic performance comparison of the various models is shown in Table 9. For the two large CWRU and IMS datasets, the comparative convergence curves of the deep learning

FIGURE 18: Confusion matrix of the proposed method in case 2.

FIGURE 19: Visualisation of features extracted of the proposed method in case 2.

FIGURE 20: Comparison of the influence of selection and number of training samples on algorithm’s performance (based on CWRU and IMS datasets).

FIGURE 21: Continued.

FIGURE 21: Frequency-domain image data sample (based on CWRU dataset). (a) Fault status of rolling body. (b) Inner ring fault status. (c) Outer ring fault status. (d) Normal state.

TABLE 8: The optimal model components after CEI-SSA optimisation in case 2.

Variable	Value
<i>Ck_s_1</i>	11
<i>Ck_s_2</i>	5
<i>Ck_s_3</i>	3
<i>Ck_s_4</i>	3
<i>channel_1</i>	15
<i>channel_2</i>	22
<i>channel_3</i>	38
<i>channel_4</i>	55
<i>Act_1</i>	"Relu"
<i>Act_2</i>	"Relu"
<i>Act_3</i>	"Relu6"
<i>Act_4</i>	"Relu"
<i>pool_1</i>	"Maxpool"
<i>pool_2</i>	"Avgpool"
<i>pool_4</i>	"Avgpool"
<i>Hid_s</i>	50
<i>L_rate</i>	0.001
<i>Op</i>	RMSprop
<i>Batch</i>	48

algorithms are shown in Figures 22(a) and 22(b) (with the CWT feature set as an example), and a visualisation of the SVM classification results of the shallow learning algorithm is shown in Figures 22(c) and 22(d). This proves that the unique HSMSF design and the CEI-SSA adaptive model

optimisation method proposed in this paper can use multisensor information for fault analysis, remove redundant structures, and improve diagnostic efficiency and stability for practical application with reduced labour, time, and other costs.

FIGURE 22: Comparison of diagnostic performances of some models. (a) Diagnostic convergence curves of four deep learning methods (based on CWRU dataset). (b) Diagnostic convergence curves of four deep learning methods (based on IMS dataset). (c) Classification results of SVM in shallow learning algorithm (based on CWRU dataset). (d) Classification results of SVM in shallow learning algorithm (based on IMS dataset).

TABLE 9: Comparison of diagnostic performances of different models.

Methods	Feature sets	CWRU test accuracy (%)	CWRU test time (s)	IMS test accuracy (%)	IMS test time (s)
ST-CNN	CWT	97.42 ± 0.22	0.27	96.75 ± 0.15	0.24
	GADF	94.64 ± 1.33	0.42	95.41 ± 0.72	0.33
	FD	94.79 ± 0.81	0.31	93.05 ± 0.28	0.27
DBN	CWT	93.78 ± 1.54	0.68	92.66 ± 2.41	0.53
	GADF	92.43 ± 1.80	0.91	92.11 ± 2.79	0.94
	FD	93.33 ± 2.04	0.85	92.47 ± 0.68	0.77
	CWT	94.14 ± 1.47	0.33	90.42 ± 1.87	0.29

TABLE 9: Continued.

Methods	Feature sets	CWRU test accuracy (%)	CWRU test time (s)	IMS test accuracy (%)	IMS test time (s)
CNNEPDNN	GADF	92.70 ± 1.66	0.71	92.02 ± 0.28	0.64
	FD	92.16 ± 0.95	0.44	92.69 ± 0.77	0.45
SVM	Time	89.47 ± 1.17	0.16	90.22 ± 0.46	0.07
	Frequency	90.35 ± 0.75	0.14	90.64 ± 0.70	0.16
PSO-LSSVM	Time	94.50 ± 0.13	0.57	94.32 ± 0.58	0.44
	Frequency	94.80 ± 0.34	0.70	94.45 ± 0.14	0.26
BP	Time	89.66 ± 1.21	0.21	88.27 ± 2.63	0.10
	Frequency	91.34 ± 0.25	0.24	91.85 ± 1.20	0.25
Our method	CWT + GADF	99.88 ± 0.10	0.30	99.60 ± 0.29	0.21

5. Conclusions

We proposed a bearing fault diagnosis method based on multidomain heterogeneous information entropy fusion and model self-optimisation. A method of multiscale domain fusion strategy for spatiotemporal features based on heterogeneous sensors was designed. After mapping the information collected by multiple sensors into multiple high-dimensional domains, the multidomain information was found to be complementary and simultaneous by channel cascade and spatiotemporal feature decomposition, and the information collected by multiple sensors was determined by an adaptive entropy-weighted fusion method. The CEI-SSA improved by chaotic initialisation and elite reverse solution was used to carry out adaptive optimisation of the model components of the feature extractor model, and end-to-end bearing fault diagnosis was achieved without manual design of the model. The effectiveness and superiority of the proposed method were verified experimentally and by comparative analysis. Compared with five other optimisation algorithms, CEI-SSA had better search performance on the 12 benchmark test functions, and the feature extraction model after CEI-SSA adaptive optimisation improved the diagnostic accuracy. Compared with other deep and shallow learning algorithms on the CWRU and IMS bearing datasets, the identification accuracy of the proposed method for bearing fault categories exceeded 99%, and the influence of artificial parameter design on diagnostic performance was reduced. Advantages of the approach included learning ability, reduced cost burden, and generalisation performance.

The method in this work forms part of supervised learning, and the evaluation data were visible-type samples consistent with the training set; therefore, the model was trained with labelled data. In the next step, unsupervised data processing will be considered in fault diagnosis research in a zero-small sample environment to diagnose and recognise complex samples that were not present in the training set.

Data Availability

The datasets supporting the conclusions of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank International Science Editing (<http://www.internationalscienceediting.com>) for editing this manuscript. This work was supported by the Foundation of Shanxi Province Engineering Research Center for Equipment Digitization and PHM (no. ZBPHM20201104), in part by the Science and Technology Innovation Project of Higher Education in Shanxi Province (no. 2019L0653), in part by the Basic Research Project of Shanxi Province (nos. 20210302123216 and 201901D111259), in part by the Shanxi Key Laboratory of Advanced Control and Equipment Intelligence (no. ACEI202002), in part by the Excellent Innovation Project for Graduate students in Shanxi Province (no. 2021Y699), and in part by the Project Commissioned by Shanxi Mechanical Product Quality Supervision and Inspection Station Co., Ltd. (no. 2021063).

References

- [1] M. Xia, T. Li, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE*, vol. 23, no. 1, pp. 101–110, 2018.
- [2] Z. C. Qiao, Y. Q. Liu, and Y. Y. Liao, "An improved method of EWT and its application in rolling bearings fault diagnosis," *Shock and Vibration*, vol. 2020, Article ID 4973941, 13 pages, 2020.
- [3] Y. S. Ahmed, A. F. M. Arif, and S. C. Veldhuis, "Application of the wavelet transform to acoustic emission signals or built-up edge monitoring in stainless steel machining," *Measurement*, vol. 154, Article ID 107478, 2020.
- [4] Y. Cheng, Z. Wang, B. Chen, W. Zhang, and G. Huang, "An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis," *ISA Transactions*, vol. 91, pp. 218–234, 2019.
- [5] R. Yuan, Y. Lv, H. Li, and G. Song, "Robust fault diagnosis of rolling bearings using multivariate intrinsic multiscale entropy analysis and neural network under varying operating conditions," *IEEE Access*, vol. 7, Article ID 130804, 2019.
- [6] S. Gao, T. Li, and Y. Zhang, "Rolling bearing fault diagnosis of PSO-LSSVM based on CEEMD entropy fusion," *Transactions*

- of the Canadian Society for Mechanical Engineering, vol. 44, no. 3, pp. 405–418, 2020.
- [7] Z. Chen, K. Gryllias, and W. Li, “Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 339–349, 2020.
 - [8] H. Li, J. Huang, X. Yang, J. Luo, L. Zhang, and Y. Pang, “Fault diagnosis for rotating machinery using multiscale permutation entropy and convolutional neural networks,” *Entropy*, vol. 22, no. 8, Article ID 851, 2020.
 - [9] L. Yu, J. L. Qu, F. Gao, and Y. Tian, “A novel hierarchical algorithm for bearing fault diagnosis based on stacked LSTM,” *Shock and Vibration*, vol. 2019, Article ID 2756284, 10 pages, 2019.
 - [10] J. Xue and B. Shen, “A novel swarm intelligence optimization approach: sparrow search algorithm,” *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020.
 - [11] Y. Zhu and N. Yousefi, “Optimal parameter identification of PEMFC stacks using adaptive sparrow search algorithm,” *International Journal of Hydrogen Energy*, vol. 46, no. 14, pp. 9541–9552, 2021.
 - [12] J. Zhang, K. Xia, Z. He, Z. Yin, and S. Wang, “Semi-Supervised ensemble classifier with improved Sparrow Search Algorithm and its application in pulmonary nodule detection,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 6622935, 18 pages, 2021.
 - [13] G. Q. Hou and C. M. Lee, “Estimation of the defect width on the outer race of a rolling element bearing under Time-Varying speed conditions,” *Shock and Vibration*, vol. 2019, Article ID 8479395, 11 pages, 2019.
 - [14] S. Barra, S. M. Carta, A. Corrigan, A. S. Podda, and D. R. Recupero, “Deep learning and time series-to-image encoding for financial forecasting,” *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 3, pp. 683–692, 2020.
 - [15] C. L. Yang, Z. X. Chen, and C. Y. Yang, “Sensor classification using convolutional neural network by encoding multivariate time series as Two-Dimensional colored images,” *Sensors*, vol. 20, no. 1, 2020.
 - [16] X. Meng, S. Qiu, S. Wan, K. Cheng, and L. Cui, “A motor imagery EEG signal classification algorithm based on recurrence plot convolution neural network,” *Pattern Recognition Letters*, vol. 146, pp. 134–141, 2021.
 - [17] T. Han, C. Liu, L. Wu, S. Sarkar, and D. Jiang, “An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems,” *Mechanical Systems and Signal Processing*, vol. 117, pp. 170–187, 2019.
 - [18] Y. Zhang, R. Liu, A. A. Heidari et al., “Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis,” *Neurocomputing*, vol. 430, pp. 185–212, 2021.
 - [19] H. R. Boveiri and M. Elhoseny, “A-COA: an adaptive cuckoo optimization algorithm for continuous and combinatorial optimization,” *Neural Computing & Applications*, vol. 32, no. 3, pp. 681–705, 2020.
 - [20] R. Zhang, N. Zhao, L. H. Fu, X. Bai, and J. Cai, “Recognizing defects in stainless steel welds based on multi-domain feature expression and self-optimization,” *Journal of Intelligent Manufacturing*, 2021.
 - [21] L. Wang, R. Shi, and J. Dong, “A hybridization of dragonfly algorithm optimization and angle modulation mechanism for 0-1 knapsack problems,” *Entropy*, vol. 23, no. 5, 2021.
 - [22] Y. Li, X. Wang, S. Si, and S. Huang, “Entropy based fault classification using the case western reserve university data: a benchmark study,” *IEEE Transactions on Reliability*, vol. 69, no. 2, pp. 754–767, 2020.
 - [23] Y. Han, S. Liu, D. Cong et al., “Resource optimization model using novel extreme learning machine with t-distributed stochastic neighbor embedding: application to complex industrial processes,” *Energy*, vol. 225, Article ID 120255, 2021.
 - [24] S. Mushtaq, M. M. M. Islam, and M. Sohaib, “Deep learning aided Data-Driven fault diagnosis of rotatory machine: a comprehensive review,” *Energies*, vol. 14, no. 16, Article ID 5150, 2021.
 - [25] H. Li, J. Huang, and S. Ji, “Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network,” *Sensors*, vol. 19, no. 9, Article ID 2034, 2019.