

A Novel OLTC Fault Diagnosis Method Based on Optimized Long Short-Term Memory Parameters

Yan YAN¹, Hongzhong MA

College of Energy and Electrical Engineering, Hohai University, China

Abstract. Recently, long short-term memory (LSTM) networks have been widely adopted to help with fault diagnosis for power systems. However, the parameters of LSTM networks are determined by prior knowledge and experience and thereby not capable of dealing with unexpected faults in volatile environments. In this paper, we propose and apply an improved grey wolf optimization (IGWO) algorithm to optimize the parameters of LSTM networks, aiming to circumvent the drawback of empirical LSTM parameters and enhance the fault diagnosis accuracy for on-load tap changers (OLTCs). The composite multiscale weighted permutation entropy and energy entropy yielded by the grasshopper optimization algorithm and variational mode decomposition (GOA-VMD) method are used as the inputs of LSTM networks. The IGWO algorithm is applied in an iterative manner to optimize the relevant super arithmetic of the LSTM. In this way, an IGWO-LSTM combination model is constructed to classify different faults diagnosed in OLTCs. Experimental results verify the diagnosis performance superiority of the proposed method over several widely used comparison benchmarks

Keywords. Long short-term memory (LSTM) network, fault diagnosis, on-load tap changer (OLTC), parameter optimization, improved grey wolf optimization (IGWO)

1. Introduction

On-load tap changers (OLTCs) are important components in transformers but are relatively fault-prone compared to other components. In recent years, it has been reported that the faults in OLTCs have accounted for more than 20% of transformer faults, and this percentage keeps increasing [1]. OLTCs are mechanical components in transformers, and therefore when switching OLTCs, complex friction and collision among components generate random and nonlinear vibration signals [2]. As a result, one can process and study these vibration signals to extract features, summarize patterns, and finally diagnose operational statuses [3]. Time-frequency methods are suitable for the analysis and processing of non-stationary signals and can characterize the relationship between signal frequency series and time [4]. Among them, the wavelet analysis is widely used for OLTC fault diagnosis by virtue of its multi-resolution, high accuracy, and small reconstruction error [5]. However, due to the arbitrary selection of wavelet bases, the analytical results by time-frequency analysis can hardly characterize mechanical statuses in an accurate manner. As an alternative, mode decomposition methods, e.g., empirical

¹ Corresponding author; Email: yanyan503986706@163.com

mode decomposition [6] and variational mode decomposition [2], can extract vibration characteristic information of OLTCs quickly and efficiently. However, it is still challenging to cope with randomness and real-time diagnosis requirements for OLTCs by these mode decomposition methods. Recently, with the advances in artificial intelligence (AI) technologies, researchers also resort to hidden Markov model (HMM) [7], support vector machine (SVM) [8], convolutional neural network (CNN) [9] to help with OLTC fault diagnosis. Among these technologies, HMM cannot yield synchronized data, and thereby features are easily submerged in noise. SVM can be greatly affected by internal parameters and relies on an extra optimization process with higher computational complexity. The main issues with CNN are the low training efficiency and manually adjusted parameters. Different from the aforementioned AI technologies, the long short-term memory (LSTM) is a special architecture of recurrent neural network (RNN) and has been proven effective for fault diagnosis of power systems [10]. On the other hand, the LSTM parameters are determined empirically based on previous knowledge and experience and thereby not capable of dealing with unexpected faults in volatile environments. In this regard, we propose and apply an improved grey wolf optimization (IGWO) algorithm to optimize the parameters of LSTM networks, aiming to circumvent the drawback of empirical LSTM parameters and enhance the fault diagnosis accuracy for OLTCs. In particular, the proposed IGWO algorithm inputs composite multiscale weighted permutation entropy and energy entropy yielded by the grasshopper optimization algorithm and variational mode decomposition (GOA-VMD) method to LSTM networks and optimizes relevant super arithmetic of the LSTM on an iterative basis. In this way, we can construct an IGWO-LSTM model to efficiently classify different faults diagnosed in OLTCs. The rest of the paper is organized as follows. We first propose the novel OLTC fault diagnosis method and expatiate on the optimization of LSTM parameters in Section II. Experimental setups are presented in Section III, analysis of experimental results in Section III. Finally, the paper is concluded in Section V.

2. IGWO-LSTM fault diagnosis based on multi-feature fusion

2.1. Feature fusion based on CMWPE

The GOA-VMD method lies on the core of the proposed the OLTC fault diagnosis method. Specifically, the GOA [11] is used to optimize the number of intrinsic mode function (IMF) components and a penalty factor yielded by VMD, which are denoted as K and γ . It is stipulated by that the minimum envelop entropy of a vibration signal detected from the OLTC denoted as $\min E_p$ is adopted as the fitness function the fault diagnosis process

In order to further verify the effectiveness of the proposed method, CMWPE is introduced into the multi-scale analysis. CMWPE [12] is also a measure of system complexity. The specific calculation steps are as follows:

① input vibration signal $x(i)(i = 1, 2, 3, \dots, N)$, the number of embedded bits is τ , and the delay time is m ;

② initialization $CMWPE = \emptyset, s = 1$;

③ for $x(i)(i = 1, 2, 3, \dots, N)$ coarse-grained time series $y^{s,d} = \{y^{s,d}(y(j))\}_{j=1}^{\lfloor (N+1)/s \rfloor - 1}$ can be expressed as:

$$y^{s,d}(j) = \frac{1}{S} \sum_{i=(j-1)S+d}^{jS+d-1} x(i) \tag{1}$$

Where, s is scale factors, $[N + 1]/s$ represents the a ceiling function, $j = 1, 2, \dots, [(N + 1) / s] - 1$, and $d < s \leq S_{max}$, $d = 1, 2, \dots, s, i = 1, 2, \dots, jS + d - 1$

④CMWPE can be expressed as:

$$CWMPE(Y, m, \tau, s) = \frac{1}{S} \sum_{d=1}^s WPE(y^{s,d}, m, \tau) \tag{2}$$

Where, CMWPE is the average WPE value

⑤ CMWPE = CMWPE \cup CMWPE(Y, m, τ, s), $s=s+1$, repeat steps 2-4 ,until $s > S_{max}$.

The vibration signal of OLTC is decomposed by GOAVMD to obtain l IMF. The CMWPE value and energy value are calculated respectively, and then two of them are fused to form the feature vector. The matrix of CMWPE is as follow:

$$H = \begin{bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,S} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ H_{L,1} & H_{L,2} & \cdots & H_{L,S} \end{bmatrix} \tag{3}$$

Where, $H_{\eta,\eta}$ represents the η value of the IMF element with scale η . $\forall \eta \in \{1, 2, \dots, l\}$ and $\forall \eta \in \{1, 2, \dots, s\}$, then add an average of 1 to l elements in each H column vector,

which is expressed as follows: $\bar{H} = [\bar{H}_1, \bar{H}_2, \dots, \bar{H}_s]$, $\bar{H}_\eta = \frac{1}{l} \sum_{\eta=1}^l \bar{H}_\eta$. The energy eigenvector is expressed as $B = [E_{1,1}, E_{1,2}, \dots, E_{l,j}]$. Where $E_{i,j}$ represents the i -th energy

eigenvalue corresponding to the j -th IMF component, and $E_i = \int_{-\infty}^{+\infty} |c_j|^2 dt$. where $c_i(t)$ represents the magnitude of each scale coefficient or natural mode function at time t . The energy characteristic B is standardized, and is expressed as follows:

$$B' = [E_{1,1} / E_\Sigma, E_{1,1} / E_\Sigma, \dots, E_{l,j} / E_\Sigma] \tag{4}$$

Where, $E_\Sigma = \sqrt{\sum_{j=1}^l |E_j|^2}$

All the features are fused to form the feature matrix $D = [\bar{H}, B']$. In this paper, the sample signal with length N is divided into z segments. Each segment has 1024 samples so that the characteristic samples can be expressed as

$$D = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_z \end{bmatrix} = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,J} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ d_{z,1} & d_{z,2} & \cdots & d_{z,J} \end{bmatrix} \tag{5}$$

Where, $J = s + l$ represents the dimension of solving the problem, and Z represents the number of the Grey Wolf population, which is applied in the next scene.

2.2. The improved Grey Wolf Optimization

The Grey Wolf Optimization(GWO) [13] algorithm simulates the social system and hunting behavior of a wolf pack, and the optimal solution is set as α wolf; the second

optimal solution is β wolf, the third optimal solution is σ wolf, and the remaining solution is ω wolf. The hunting process of wolves can be divided into three stages: the stage of tracking and approaching the prey; surrounding and harassing; and the stage of attacking the prey, and the mathematical model of surround behavior is:

$$Q(t+1)=Q_p - A|C \cdot Q_p(t) - Q(t)| \tag{6}$$

Where t is the current iteration number, C and a are coefficients, a is the distance control parameter, $C = 2r_1$, $a = 2ar_1 - a$, Q_p is the position vector of the prey, Q is the position of the grey wolves, and r_1 and r_2 are random variables uniformly distributed in $[0,1]$. The mathematical model of the tracking target is as follows:

$$\begin{cases} Q_1(t+1)=Q_\alpha(t) - A_1 |C_1 \cdot Q_\alpha(t) - A_1 \cdot Q(t)| \\ Q_2(t+1)=Q_\beta(t) - A_2 |C_2 \cdot Q_\beta(t) - A_2 \cdot Q(t)| \\ Q_3(t+1)=Q_\delta(t) - A_3 |C_3 \cdot Q_\delta(t) - A_3 \cdot Q(t)| \end{cases} \tag{7}$$

$$Q(t+1)=\frac{Q_1(t)+Q_2(t)+Q_3(t)}{3} \tag{8}$$

In the hunting process, α , β and σ wolves locate the position of the prey and form an enclosure. Under the guidance of α , β and σ wolves, ω wolf wolves surround the prey. According to the literature, the GWO algorithm can easily fall into the local optimum [14]. This paper improves the GWO algorithm. Inspired by particle swarm optimization (PSO), a strategy of adjusting parameters by random distribution is proposed according to the literature [15]: Assuming that the distance parameter a obeys a random number of a certain distribution, the value of a control parameter is adjusted by using the characteristics of random variables to make the algorithm jump out of the local optimum; if the optimal individual is near, the random distribution control parameter produces a relatively small value, which is conducive to accelerating the convergence speed of the algorithm. The specific formula is as follows:

$$a(t) = a_{\text{initial}} - (a_{\text{initial}} - a_{\text{final}}) \times \text{rand}() + \lambda \times \text{randn}() \tag{9}$$

Where, a_{initial} and a_{final} represent the initial value and the end value of the distance control parameter, respectively. $\text{rand}()$ is $[0, 1]$, $\text{randn}()$ is a random number that obeys the normal distribution, λ (variance) is used to measure the deviation degree between the control parameter and the random variable and the mathematical expectation (mean value). It is to control the error of parameter a in the value selection and make the control parameter evolve to the expected control parameter a .

2.3. Parameter optimization of optimized LSTM based on IGWO

Long short-term memory (LSTM) neural network is a variant of recurrent neural network (RNN). It can learn the long-term dependence of input data and alleviate the problems of gradient disappearance and gradient explosion in model training. It has obvious advantages in processing data with nonlinear time series.

As multiple parameters need to be adjusted in the process of using LSTM, the parameter values have a significant impact on the model prediction and classification effect, and the manual search of parameters is time-consuming and laborious, so the goal programming algorithm is needed to assist the parameter optimization. Therefore, the IGWO algorithm in this paper is used to optimize the LSTM. In the neural network, the number of hidden layer neurons, and the learning rate are optimized as follows:

①The number and range of optimization parameters are determined. Set the values of the parameters to be optimized in LSTM, and set the upper and lower limit arrays ub and lb .

The parameters are expressed as follows:

$$Lb=[lb_1,lb_2,lb_3,\dots,lb_{v-1},lb_v] \tag{10}$$

$$Ub=[ub_1,ub_2,ub_3,\dots,ub_{v-1},ub_v] \tag{11}$$

Where v is the number of parameters to be optimized, lb_v is the lower limit of the v th parameter to be optimized, and ub_v is the lower limit of the v th parameter to be optimized.

② The model index accuracy of LSTM is used as the fitness function output, and the fitness function is as follows:

$$Fitness = \frac{\varepsilon}{J} \times 100\% \tag{12}$$

Where ε represents the number of correct classifications.

③ Set algorithm parameters: population size m , maximum iteration parameters t_{max} , $a_{initial}$, a_{final} ;

④ Initialize the grey wolf population $\{d_i, i=1,2,\dots,Z\}$, and calculate the fitness value of each grey wolf individual using equation (12);

⑤The fitness value is sorted in non-decreasing order, and the first, second and third individuals of the wolf population are found, which are α , β and δ . In the iterative process, the position of each individual is updated using (7) and (8) in the wolf population;

⑥ Use formula (9) to generate random parameters a , A and C ;

⑦ Judge whether the algorithm reaches the termination condition ($t \leq t_{max}$). If the conditions are not met, repeat steps ④~⑧. If the termination condition is reached, the iteration is stopped, and the optimal solution is output. The test flow chart is shown in as Fig. 1.

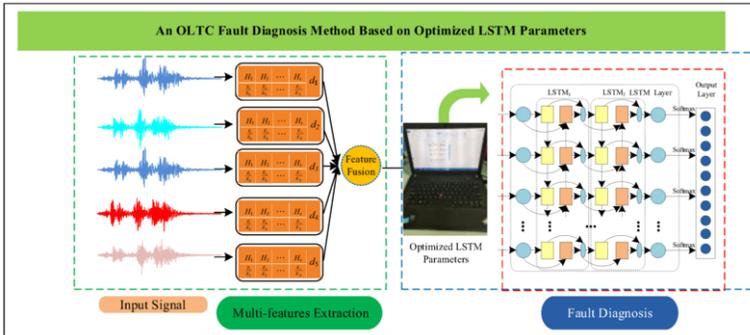


Fig. 1: Test flow chart



Fig. 2: Contact fault setting



Fig. 3: Acceleration sensor location

In this paper, the model CM111-50-63B-10193W Huaming on-load tap changer was adopted, and lc0151 was used in the experiment. The advantage of this collector is that the traditional piezoelectric sensor and charge amplifier are integrated, which simplifies the test system and improves the test accuracy and reliability. A typical mechanical fault of the OLTC switch was simulated. Specifically, it was divided into to normal operation marked as 1, contact burning marked as 2, contact loosening marked as 3, and contact shedding marked as 4. for comparison. A total of 60 groups were collected in each state, and the sampling frequency was 50kHz. The contact fault setting is shown in Fig. 2. The acceleration sensor is shown in Fig. 3.

3. Analysis of Experimental Results

To verify the superiority of this method, the optimal value (the maximum accuracy) and the standard deviation are used as the evaluation indexes. Comparing the support vector machine(SVM)assigned marked as 1. kernel extreme learning machine(KELM) marked as 2,Convolutional Neural Networks(CNN) marked as 3, LSTM marked as 4.GWO-LSTM marked as 5 compared with the proposed method in this paper which marked as 6.The proposed method network simulation is composed of the input layer, one hidden layer, and the output layer, and the Adaam algorithm is used to train the internal parameters of LSTM. The activation function in the hidden layer adopts the tanh function, and the abandonment rate of the network node is 0.2. In order to prevent excessive fitting, the number of iterations is selected as 100. Fig. 4 shows the parameter iteration diagram of IGWO-LSTM. As can be seen from Fig. 4a, the convergence of IGWO is earlier than that of GWO, indicating that the convergence performance of IGWO is stronger than that of GWO, and the convergence value is finally stabilized at 0.0038. Fig. 4b shows the change of hidden neurons with the number of iterations and finally stabilized at 128. LSTM network, SVM network, and KLEM network are constructed by training set samples. CNN network, combined with literature, the learning rate of the LSTM model is 0.01, and the number of hidden neurons is 152. Fig. 4c shows that the learning rate is stable at 0.0012 with the change of iteration times. The kernel parameter g of SVM is 1, wand the penalty factor c is 2. The number of network layers of CNN is seven layers (3 convolution layers, three pooling layers, and one full connection layer), the kernel parameter c of KELM is 0.001.

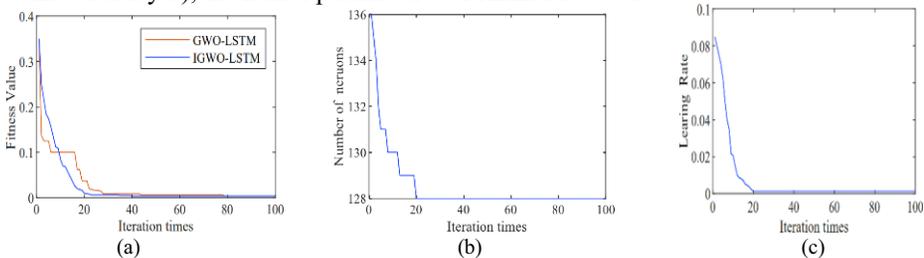


Fig. 4: Changes of c , RLR and T_{max} in IGWO-LSTM optimal grey wolf individual

It is compared with the other five methods in the case of CMWPE only. It can be seen from Fig. 5 that the average value of IGWO-LSTM is the highest when $\tau = 3; 4; 5; 6$. It can be seen from Fig. 7 that the STD value of the proposed is the smallest of the six methods, which indicates that the accuracy of OLTC fault diagnosis based on LSTM can be improved by optimizing LSTM parameters by IGWO.

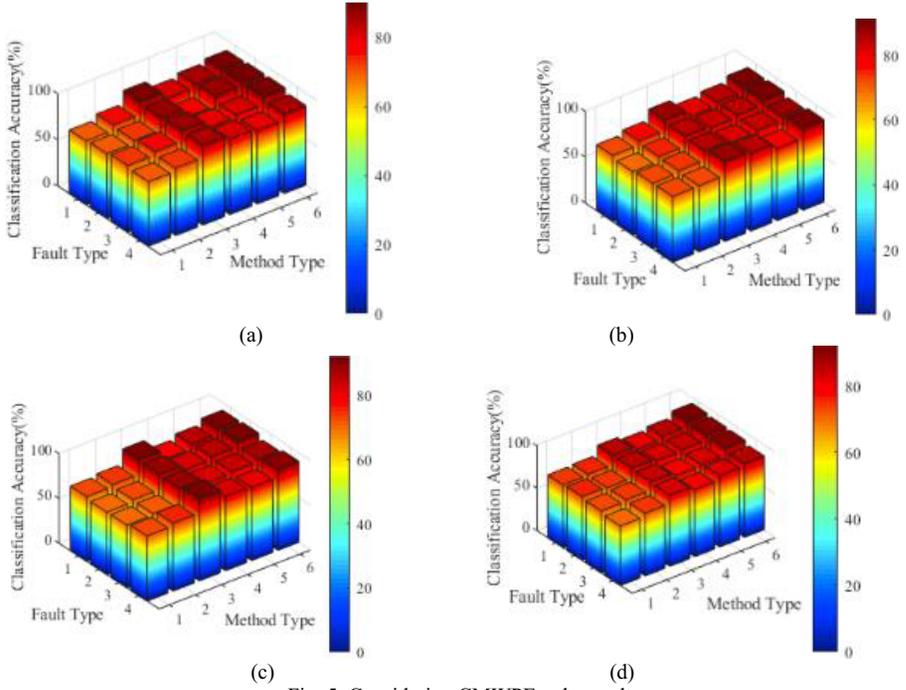


Fig. 5: Considering CMWPE only results

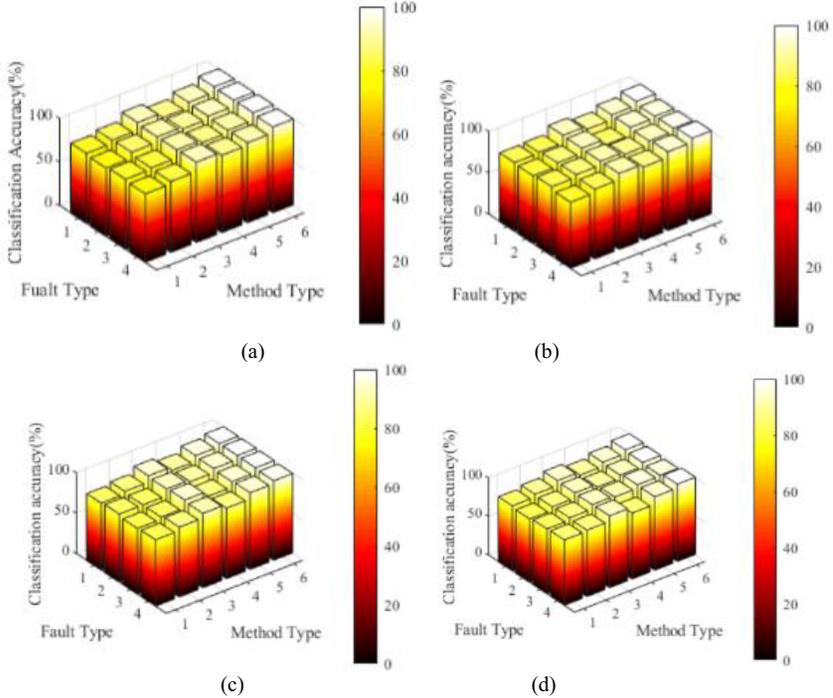


Fig.6 Feature fusion results

But because based on single scale proposed method there is still a false negative, there can still be a misjudgment in OLTC fault diagnosis. To further improve the fault diagnosis accuracy of OLTC, CMWPE and energy entropy of fusion, the proposed

method in the same place with six kinds of method is compared. As can be seen from Fig. 6, compared with proposed method based on a single feature, when $\tau = 3; 4; 5; 6$, the average value of the proposed method is 8%-10% higher than that of GWO-LSTM and about 20%-30% higher than that of the other 4 methods. As can be seen from Fig. 8, the STD value of proposed method is the smallest, so it can be concluded that the method proposed in this paper has the highest fault diagnosis accuracy and the best classification effect among the six methods. This shows that the fault diagnosis accuracy of OLTC based on IGWO-LSTM can be greatly improved by feature fusion.

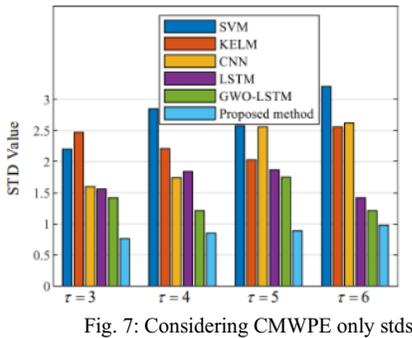


Fig. 7: Considering CMWPE only stds

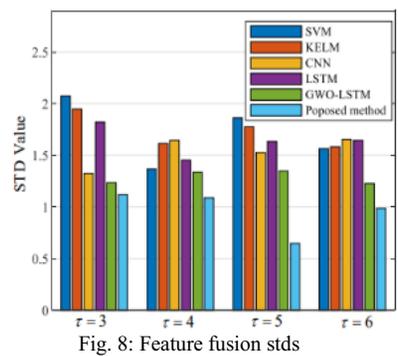


Fig. 8: Feature fusion stds

4. Conclusion

In this paper, an improved grey wolf optimization algorithm is proposed to optimize the relevant parameters of LSTM, and it is applied to OLTC fault diagnosis. The test is carried out by single feature and multi-feature fusion. The specific results are as follows: under the condition of CMWPE only, the parameters of LSTM can be optimized by IGWO, which can improve the fault diagnosis accuracy of OLTC based on LSTM. Under the condition of CMWPE and energy entropy fusion, the effect of the method proposed in this paper is obvious, compared with the accuracy of OLTC fault diagnosis based on single feature IGWO-LSTM. The stability of the method is better, which shows that the method proposed in this paper is relatively good, and further provides a theoretical basis for online OLTC fault monitoring.

References

- [1] P. Kang and D. Birtwhistle, "Condition monitoring of power transformer on-load tap-changers. part 1: Automatic condition diagnostics," *Generation, Transmission and Distribution, IEE Proceedings-*, vol. 148, no. 4, pp. 301–306, 2001.
- [2] G. Qian and S. Wang, "Fault diagnosis of on-load tap-changer based on the parameter-adaptive vmd and sa-elm," in *2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, 2020, pp. 1–4.
- [3] Y. Yan, H. Ma, M. Wen, S. Dang, and H. Xu, "Multi-feature fusion based mechanical fault diagnosis for on-load tap changers in smart grid with electric vehicles," *IEEE Sensors Journal*, 2020.
- [4] R. Duan and F. Wang, "Fault diagnosis of on-load tap-changer in converter transformer based on time-frequency vibration analysis," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 6, pp. 3815–3823, 2016.
- [5] P. Kang and D. Birtwhistle, "Characterisation of vibration signals using continuous wavelet transform for condition assessment of on-load tap changers," *Mechanical Systems and Signal Processing*, vol. 17, no. 3, pp. 561–577, 2003.

- [6] Y. Xu, C. Zhou, J. Geng, S. Gao, and P. Wang, "A method for diagnosing mechanical faults of on-load tap changer based on ensemble empirical mode decomposition, volterra model and decision acyclic graph support vector machine," *IEEE Access*, vol. 7, pp. 84 803–84 816, 2019.
- [7] Q. Li, T. Zhao, L. Zhang, and J. Lou, "Mechanical fault diagnostics of on-load tap changer within power transformers based on hidden Markov model," *IEEE Transactions on Power Delivery*, vol. 27, no. 2, pp. 596–601, 2012.
- [8] H. Gu, Y. Wang, B. Li, Y. Chen, and Z. Li, "Mechanical fault diagnosis of transformer on-load tap-changer based on improved variational mode decomposition and support vector machine," pp. 1–4, 2020.
- [9] X. Liang, Y. Wang, and H. Gu, "A mechanical fault diagnosis model of on-load tap changer based on same-source heterogeneous data fusion," *IEEE Transactions on Instrumentation and Measurement*, pp. 1–1, 2021.
- [10] G. S. Lee, S. S. Bang, H. A. Mantooth, and Y. J. Shin, "Condition monitoring of 154 kv hts cable systems via temporal sliding LSTM networks," *IEEE Access*, vol. 8, pp. 144 352–144 361, 2020.
- [11] D. Wang, H. Chen, T. Li, J. Wan, and Y. Huang, "A novel quantum grasshopper optimization algorithm for feature selection," *International Journal of Approximate Reasoning*, vol. 127, pp. 33–53, 2020.
- [12] H. A. Cheng, W. B. Tao, B. Cl, and C. C. Tong, "A novel method of composite multiscale weighted permutation entropy and machine learning for fault complex system fault diagnosis - sciencedirect," *Measurement*, vol. 158.
- [13] Y. Fu, H. Xiao, L. H. Lee, and M. Huang, "Stochastic optimization using grey wolf optimization with optimal computing budget allocation," *Applied Soft Computing*, vol. 103, p. 107154, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568494621000776>
- [14] M. H. Nadimi-Shahraki, S. Taghian, and S. Mirjalili, "An improved grey wolf optimizer for solving engineering problems," *EXPERT SYSTEMS WITH APPLICATIONS*, vol. 166, MAR 15 2021.
- [15] W. Long, S. Cai, J. Jiao, and et al, "An improved grey wolf optimization algorithm," *ACTA ELECTRONICA SINICA*, vol. 47, no. 1, pp. 169–175, 2019. [Online]. Available: <http://www.cnki.com.cn/Article/CJFDTotol-DZXU201901022.htm>