

Article

AVOA-LightGBM Power Fiber Optic Cable Event Pattern Recognition Method Based on Wavelet Packet Decomposition

Xiaojuan Chen ^{1,*} , Wenbo Cui ¹ and Tiantong Zhang ²

¹ College of Electrical and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China; aixigua1232021@126.com

² Computer Network Project, Dalian Maritime University, Dalian 116026, China

* Correspondence: cxj001@cust.edu.cn

Abstract: The type of power fiber optic cable fault event obtained by analyzing the optical time domain reflectometer (OTDR) detection curve is an important basis for ensuring the operation quality of communication lines. To address the issue of low accuracy in recognizing fault event patterns, this research proposes the AVOA-LightGBM method for optical cable fault event pattern recognition based on wavelet packet decomposition. Initially, a three-layer wavelet packet decomposition is performed on different fault events, resulting in eight characteristic signals. These signals are then normalized and used as input for each recognition model. The Light Gradient Boosting Machine (LightGBM) is optimized using the African vulture optimization algorithm (AVOA) for pattern recognition. The experimental results demonstrate that this method achieves a recognition accuracy of 98.24%. It outperforms LightGBM, support vector machine (SVM), and extreme learning machine (ELM) by 3.7%, 19.15%, and 5.67%, respectively, in terms of accuracy. Moreover, it shows a 1.8% improvement compared with the combined model PSO-LightGBM.

Keywords: power fiber optic cable; pattern recognition; OTDR; African vulture optimization algorithm; LightGBM



Citation: Chen, X.; Cui, W.; Zhang, T. AVOA-LightGBM Power Fiber Optic Cable Event Pattern Recognition Method Based on Wavelet Packet Decomposition. *Electronics* **2023**, *12*, 3743. <https://doi.org/10.3390/electronics12183743>

Academic Editor: George A. Papakostas

Received: 8 August 2023

Revised: 24 August 2023

Accepted: 28 August 2023

Published: 5 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to their advantages in terms of good anti-interference performance, long distance transmission, large range coverage, and intrinsic safety [1,2], power optical cables have become an essential means of information transmission in power systems [3]. The stability of their transmission plays a crucial role in the overall safety of the power system [4,5]. However, due to their unique installation method, power optical cables are exposed to the wild for extended periods and are vulnerable to failures caused by severe weather conditions such as strong winds and icing. These failures directly jeopardize the safety and stability of the power system. Therefore, identifying the common fault event types of power optical cables is crucial for the timely resolution of fault issues and ensuring the reliability of the power communication system [6,7].

The OTDR is an instrument that utilizes Rayleigh scattering and Fresnel reflection phenomena of light as it propagates through optical fibers. It extracts valuable signals from optical attenuation information and detects and analyzes the condition of optical fibers [8,9]. OTDR offers several advantages including easy operation, convenience, low energy consumption, and compatibility with portable power sources. It only requires access to one end of the optical fiber link. In power optical cables, OTDR plays a crucial role [10–12]. However, the current conventional operation and maintenance management methods heavily rely on the manual inspection of physical damage to the cables, which contradicts the expectations of the smart grid. The maintenance of optical cables is challenging and inefficient due to the influence of materials and natural factors [13]. This paper proposes a method to obtain an optical power attenuation curve containing the status information of

the optical cable by connecting and testing the OTDR with the optical cable under test. By analyzing this curve, it is possible to identify the fault event type of the optical cable under test [14].

Numerous studies have been conducted on OTDR event recognition methods. H. Wu et al. employed a four-layer wavelet decomposition to extract event features and combined it with a BP neural network for recognition, achieving a recognition accuracy rate of 91.1% [15]. Similarly, C. Xu et al. utilized SVM to identify time-domain features of OTDR events. Their results demonstrated that SVM can effectively identify types of optical cable faults with an accuracy of 93.8% [16]. However, this method overlooks the frequency-domain information present in OTDR signals. B. M. Tabi Fouda et al. proposed a two-level target recognition method. In this approach, the signal's short-term energy and short-term cross-threshold rate are compared with a dynamic threshold in the first level, followed by the estimation of the power spectrum of the initially screened samples. Features are then extracted and input to SVM for secondary recognition [17]. This two-level recognition method significantly improves the accuracy of recognition. J. Wu et al. proposed a multi-scale one-dimensional convolutional neural network for identifying event types [18]. This method simplifies the algorithm complexity and maintains a comparative advantage by directly inputting preprocessed data into the network, resulting in good recognition performance. Y. Wang et al. conducted wavelet energy spectrum analysis on the original vibration signal to extract feature vectors. They determined that the optimal number of decomposition layers for wavelet energy spectrum analysis was six. The extracted features were then classified using a correlation vector machine (RVM) to achieve the recognition of OTDR events [19]. However, the research results indicate an average recognition accuracy of only 88.60%. X. Wang et al. employed the random forest classifier to identify data after time-domain feature extraction, achieving a recognition accuracy of 96.58% [20]. However, the random forest algorithm is complex and computationally expensive. M. H. Li et al. first preprocessed the signal using an average filter and then extracted time- and frequency-domain signals for classification using a support vector machine [21]. However, this method is unable to simultaneously identify multiple events.

This paper focuses on addressing the issue of low recognition accuracy of optical cable fault event types. The research is conducted in three main areas: data feature extraction, recognition algorithm, and optimization algorithm. Initially, the original signal undergoes wavelet packet decomposition and energy spectrum construction processing. The resulting signal is then used as a feature vector for training and testing in the LightGBM algorithm. Additionally, four parameters are optimized to enhance the accuracy and performance of the LightGBM algorithm. Contributions of this paper can be summarized as follows:

- (1) In order to address the issue of low recognition accuracy of power optical cable fault event types, we propose the AVOA-LightGBM optical cable fault event pattern recognition method based on wavelet packet decomposition. We simulated various types of optical cable faults in the laboratory to validate the feasibility of this identification method. Furthermore, field data demonstrate the high superiority and accuracy of our proposed method.
- (2) Traditional feature extraction methods may encounter problems such as information loss, inaccurate feature extraction, and loss of time series information when dealing with non-stationary data sets in the time domain and frequency domain. To address these issues, we use the 'rbio3.1' wavelet basis function for three-layer wavelet packet decomposition on the data set. This decomposition allows us to obtain eight frequency subband signals from the time-domain signal, and the energy value of each subband signal serves as the fault characteristic value of the signal. To improve the performance and generalization ability of the recognition model, we introduce the AVOA to determine the four parameters of the LightGBM algorithm used for classifying fault events. Experimental results show that the average accuracy rate of the model reached 98.24%, demonstrating the success of our approach.

The rest of this paper is composed as follows: Section 2 introduces the algorithm principle, including LightGBM, AVOA, and wavelet packet decomposition; Section 3 is the data processing, conducting various analyses on the experimental results, and comparing them with other models; Section 4 is the conclusion of this paper.

2. Related Theory

2.1. LightGBM

Gradient boosting decision tree (GBDT) is a well-established machine-learning algorithm [22] that has demonstrated exceptional performance and effectiveness in numerous practical applications. However, GBDT requires building a new tree for each iteration, which makes the training process slower and consumes more memory due to the large number of trees. Additionally, if the depth of the tree is large and the parameter settings are improper, it is more likely to encounter overfitting issues. Moreover, GBDT needs to deal with missing values when faced with them in the data. The LightGBM model is an efficient implementation of GBDT where a set of weak learners are combined using a boosting strategy to progressively enhance the accuracy of the overall learner. In contrast to GBDT, LightGBM incorporates a range of optimization algorithms, enabling significant reductions in memory usage and training time without compromising model accuracy; when there are missing values in the data, it can also automatically handle these without additional preprocessing steps.

The histogram algorithm discretizes continuous eigenvalues into k integers and constructs a histogram with a width of k . The algorithm uses the discretized values as indexes to accumulate statistics in the histogram during traversal. It then finds the optimal split point based on the index. This approach only requires saving the discrete value instead of the entire training set, resulting in significantly reduced memory consumption. Figure 1 illustrates the histogram algorithm.

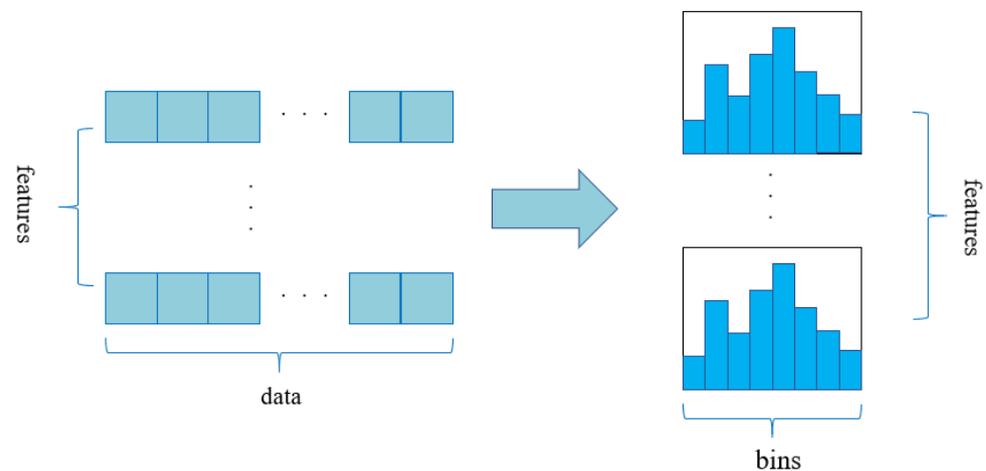


Figure 1. Histogram algorithm.

LightGBM utilizes the leaf-wise decision tree growth strategy [23]. This strategy involves finding the leaf with the highest splitting gain among all the current leaves for splitting, repeating the process, and continuously iterating. This approach offers higher precision compared with the level-wise decision used in traditional GBDT. The leaf-wise decision tree growth strategy is illustrated in Figure 2.

In the traditional gradient sampling method, samples with small gradients are typically selected, which can impact the training speed and accuracy of the model. To address this issue, LightGBM employs a technique called gradient-based one-side sampling (GOSS). This method prioritizes the retention of instances with larger gradients and randomly samples instances with smaller gradients. By doing so, accurate information gain estimates can be obtained using a smaller amount of data. This approach effectively reduces computational requirements while maintaining the predictive performance of the model.

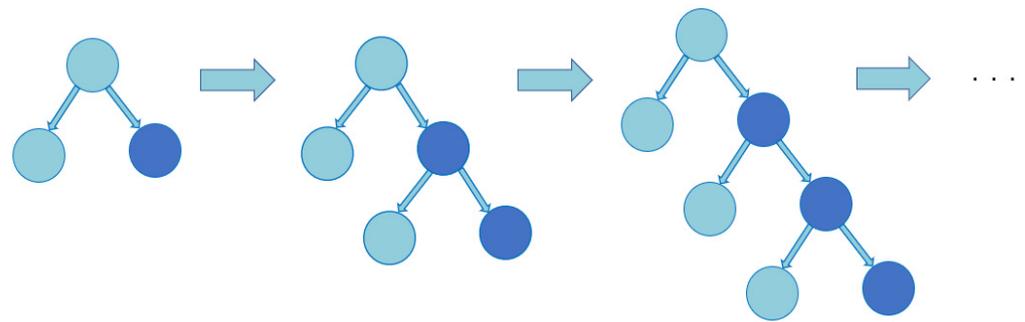


Figure 2. The leaf-wise decision tree growth strategy.

2.2. The African Vulture Optimization Algorithm

The AVOA is a novel intelligent optimization algorithm that draws inspiration from the foraging behavior and living habits of African vultures. The AVOA exhibits strong optimization ability and fast convergence speed [24]. The process is illustrated in Figure 3. The position of an individual African vulture within the population can be represented by a mathematical vector. The vulture population, which consists of n vultures, is represented by Formula (1):

$$X = [X_1, X_2, \dots, X_i, \dots, X_n]^T \tag{1}$$

where n represents the number of vultures in the population, and X_i represents the position of the i -th vulture in the population. The AVOA is an iterative process that can be expressed in the following four steps:

Phase 1: Population grouping and identification of optimal vultures. After the initial population is formed, the fitness of all vultures is calculated. The vulture position that corresponds to the optimal fitness value is considered the optimal vulture position, while the vulture position that corresponds to the suboptimal fitness value is considered the suboptimal vulture position [25]. The remaining vultures move toward the optimal and suboptimal positions using Formula (2). Formula (3) is the calculation method for the probability of selecting the best vulture. This process is repeated in each fitness iteration, recalculating the entire population.

$$R(t) = \begin{cases} BestVulture_1 & p_i = L_1 \\ BestVulture_2 & p_i = L_2 \end{cases} \tag{2}$$

$$p_i = \frac{F_i}{\sum_{i=1}^n F_i} \tag{3}$$

where $BestVulture_1$ represents the optimal vulture position, while $BestVulture_2$ represents the suboptimal vulture position. The probability of selecting the best vulture is denoted by p_i . The fitness value of other vultures is represented by F_i . Lastly, t denotes the current iteration number. L_1 and L_2 are algorithm setting parameters that range from 0 to 1. The sum of L_1 and L_2 is always equal to 1. When L_1 approaches 1 and L_2 approaches 0, the algorithm’s search ability is enhanced, making it more effective in finding local solutions. On the other hand, when L_1 approaches 0 and L_2 approaches 1, the algorithm’s search diversity increases, making it more effective in finding the global optimal solution.

Phase 2: Determine the hunger rate of vultures. Vultures that are full have more energy, enabling them to fly longer distances in search of food. On the other hand, vultures in a hungry state conserve energy by staying close to stronger vultures and can only search for food nearby. Additionally, vultures in a hungry state exhibit more aggressive behavior [26], which can be represented by Formula (4):

$$F = (2 \times rand_1 + 1) \times z \times \left(1 - \frac{t}{T}\right) + h \left[\sin^\omega \left(\frac{\pi}{2} \times \frac{t}{T}\right) + \cos \left(\frac{\pi}{2} \times \frac{t}{T}\right) - 1 \right] \tag{4}$$

where F represents the hunger rate of the vulture; T is the maximum number of iterations; $rand_1 \in (0, 1)$; $z \in (-1, 1)$ will change with each iteration. When the z value drops below 0, it indicates that the vulture is in a hungry state. When it increases to 0, it indicates that the vulture is full. Meanwhile, ω is a parameter set before the optimization operation. Increasing the value of ω increases the probability of entering the exploration stage in the final optimization stage; otherwise, the probability decreases.

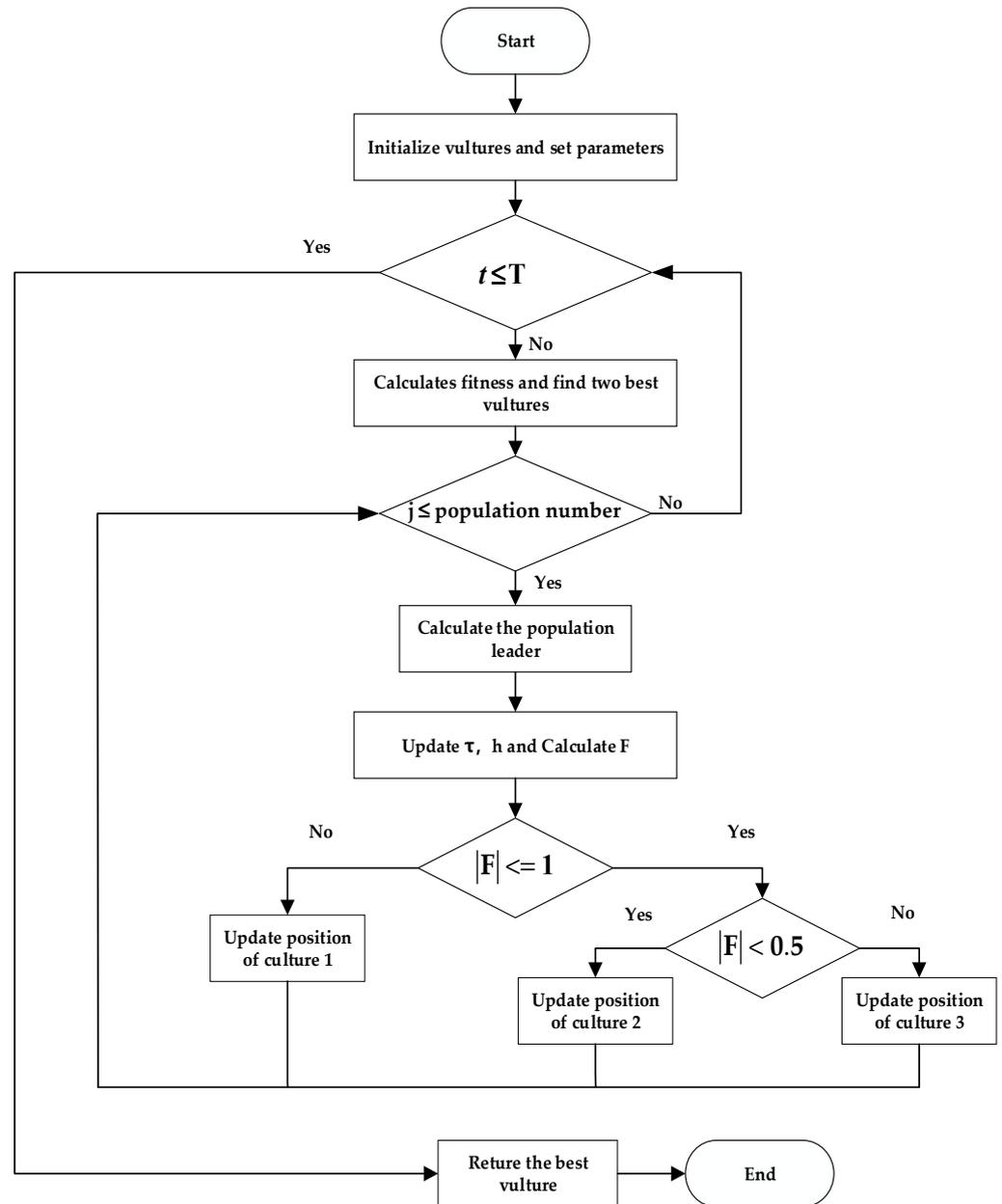


Figure 3. AVOA flow chart.

Phase 3: Exploration stage. Vultures can forage using two different strategies and select the strategy mode based on the user-defined parameter $p_1 (p_1 \in (0, 1))$ [27]. This stage can be represented by Formulas (5) and (6).

$$P(i + 1) = \begin{cases} R(i) - D(i)F & P_1 \geq rand_{p_1} \\ R(i) - F + rand_2 & \\ [(ub - lb)rand_3 + lb] & P_1 \leq rand_{p_1} \end{cases} \quad (5)$$

$$D(i) = |XR(i) - P(i)| \tag{6}$$

where $P(i + 1)$ and $P(i)$ represent the position of the vulture in the t -th and $t + 1$ iterations, respectively. X is the coefficient vector, which aims to increase the random motion and changes in each iteration. X is calculated as 2 multiplied by a random number between 0 and 1. P_1 is the selection parameter of the exploration stage. $rand_2, rand_3, rand_{p_1}$ are random numbers between 0 and 1. u_b represents the upper limit of the search space, while l_b represents the lower limit of the search space.

Phase 4: Exploitation stage. When the value of $|F|$ is less than 1, the AVOA enters the exploitation stage. In this stage, there are two development strategies. The critical value for choosing which strategy is $|F| = 0.5$. When $|F| \geq 0.5$, the mathematical model imitates the characteristics of the vulture’s spiral flight and appears as Formulas (7)–(10):

$$P(i + 1) = \begin{cases} D(i)(F + rand_4) & P_2 \geq rand_{p_2} \\ -d(t) & P_2 \leq rand_{p_2} \\ R(i) - (S_1 + S_2) & \end{cases} \tag{7}$$

$$d(t) = R(i) - P(i) \tag{8}$$

$$S_1 = R(i) \frac{rand_5 P(i)}{2\pi} \cos P(i) \tag{9}$$

$$S_1 = R(i) \frac{rand_6 P(i)}{2\pi} \sin P(i) \tag{10}$$

where $rand_4, rand_5, rand_6, rand_{p_2}$ are random numbers between 0 and 1, and the rest of the parameters have the same meaning as above.

When the value of $|F|$ is less than 0.5, the vulture experiences fatigue and hunger, which simulates its gathering behavior and triggers an offensive behavior. This behavior can be represented by Formulas (11)–(13):

$$P(i + 1) = \begin{cases} (A_1 + A_2)/2 & P_3 \geq rand_{p_3} \\ R(i) - |d(t)| & \\ FLevy(d) & P_3 \leq rand_{p_3} \end{cases} \tag{11}$$

$$A_1 = BestVulture_1(i) - \frac{BestVulture_1(i) \times P(i)}{BestVulture_1(i) \times P(i)^2} \times F \tag{12}$$

$$A_2 = BestVulture_2(i) - \frac{BestVulture_2(i) \times P(i)}{BestVulture_2(i) \times P(i)^2} \times F \tag{13}$$

where $rand_{p_3}$ is a random number between 0 and 1. $BestVulture_1(i)$ and $BestVulture_2(i)$ denote the optimal and suboptimal positions of vultures in the current iteration. $Levy(d)$ is a stochastic process that simulates a random walk, influencing the step size of each individual in the population across dimensions. A larger value of $Levy(d)$ corresponds to a larger step size, increasing the algorithm’s exploratory nature but making it harder to converge to the global optimal solution. Conversely, a smaller $Levy(d)$ implies a smaller step size, making the algorithm more inclined toward local search but potentially overlooking the global optimal solution within the search space.

2.3. AVOA-LightGBM Power Fiber Optic Cable Event Pattern Recognition Method Based on Wavelet Packet Decomposition

The LightGBM algorithm offers flexibility to cater to various requirements such as model complexity control, computing efficiency, memory control, and different fields [28]. To optimize the LightGBM algorithm and enhance recognition accuracy, this paper utilizes

the AVOA's strong global search ability to optimize four parameters. The optimized parameter settings are presented in Table 1.

Table 1. Optimized parameter settings.

Parameters	Meaning of Parameters	Value Ranges
learning_rate	Controlling the update degree of the tree model during each iteration is crucial. Using a small learning rate can enhance the stability of the model, but it may also necessitate more iterations to attain optimal performance.	[0.01, 0.3]
feature_fraction	Control the proportion of features used by each tree during training.	[0.2, 0.95]
num_leaves	Determines the complexity of the model; taking a small value can reduce the risk of overfitting, but it may also lead to underfitting of the model.	[20, 3000]
max_depth	Limit the depth of decision tree, and a reasonable value can reduce the risk of overfitting [29].	[3, 12]

The fitness function of the AVOA is constructed as shown in Formula (14):

$$Fitness^t = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y(i) - f(x_i))^2} \quad (14)$$

where n is the number of test datasets, $Y(i)$ is the classification value, and $f(x_i)$ is the real value.

The AVOA-LightGBM power fiber optical cable fault event identification model based on wavelet packet decomposition, is illustrated in Figure 4. It comprises three main components: Data Preprocessing, AVOA-LightGBM model optimization, and Model classification. The specific steps are as follows:

1. Data preprocessing: The OTDR curve data are subjected to wavelet packet decomposition to extract the feature vector. This vector is then normalized and used as the input for the LightGBM recognition model. Furthermore, the data are divided into a training set and a test set for model evaluation.
2. AVOA-LightGBM model optimization: The AVOA-LightGBM optical cable event recognition model is constructed by setting and initializing the initial value of LightGBM and the hyperparameter optimization range. The algorithm continuously calculates the fitness value of the AVOA and updates the individual and population parameters until the termination condition is reached. Finally, the optimal parameters are out-putted.
3. Model classification: Input the data set into the model for identification and output the classification result.

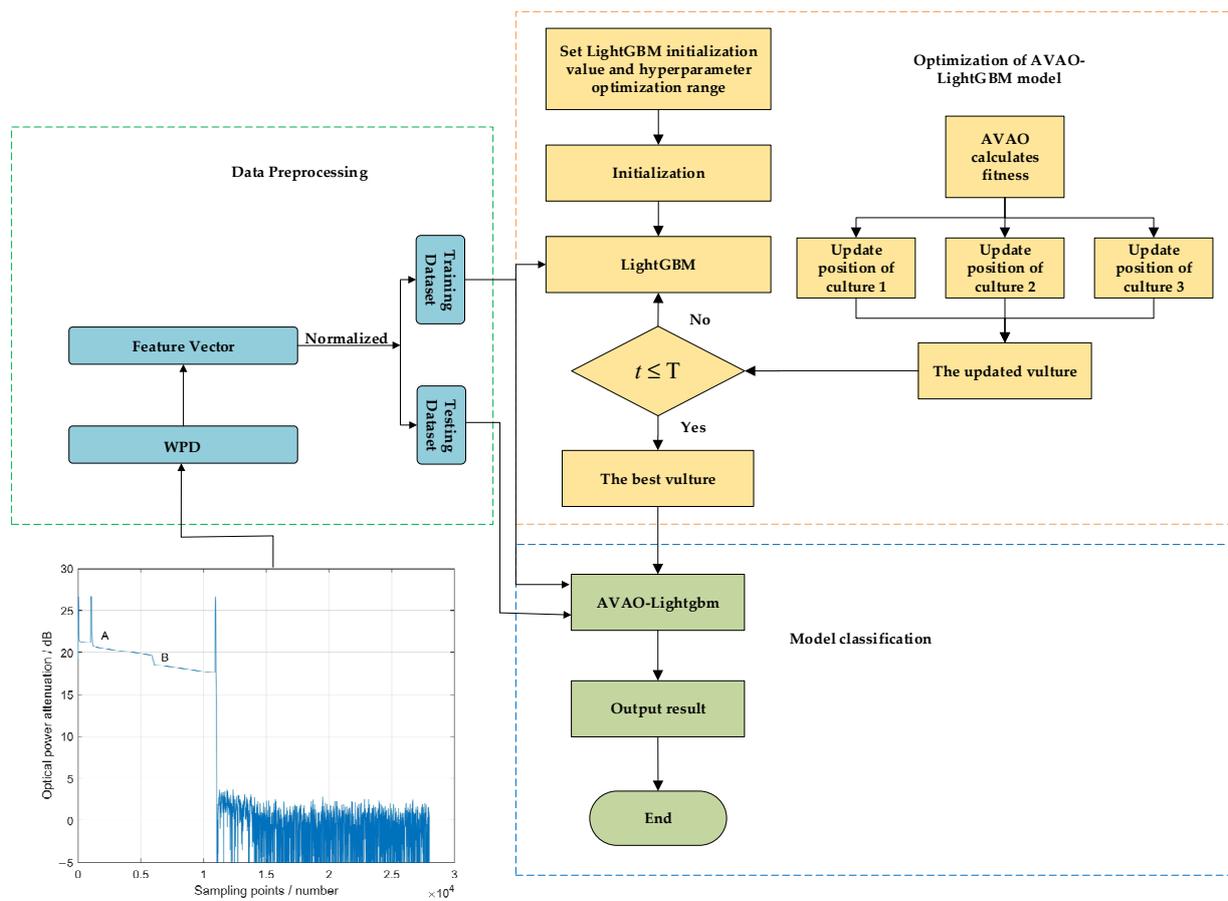


Figure 4. A Framework for pattern recognition of AVOA-LightGBM power fiber optic cable fault events based on wavelet packet decomposition.

3. Experimental Results and Data Analysis

3.1. Experimental Data Acquisition

In this paper, an OTDR is utilized to collect the signal of the faulty power fiber optical cable. The acquisition device is depicted in Figure 5. The pulse generator is employed to control the laser diode (LD), injecting an optical pulse of specific width and period into the optical fiber via the circulator. Subsequently, after the optical signal is injected into the fiber, certain signals return due to Rayleigh backscattering and Fresnel reflection. These returning signals then pass through the coupler once again. The coupler separates the returned optical signal from the emitted one, directing it to the photodiode. The optical signal is then converted into an electrical signal, amplified by the amplifier, sampled by the DAQ sampler, and output and displayed as an optical power attenuation curve on the PC. Additionally, this curve is stored for subsequent analysis.

3.2. Experimental Data Processing

In this paper, the above-mentioned OTDR device is utilized to capture two common types of optical cable faults, namely reflection events and non-reflection events. The reflection event is depicted as point A on the left side of Figure 5, appearing as a bulge on the curve where the relative optical power experiences a sudden increase followed by attenuation. This paper collects reflection events by conducting destructive experiments such as man-made fiber breaks and poor connections at splice points. The non-reflective events are represented as B in the graph, and the curve demonstrates a stepwise decrease in relative optical power. Additionally, the paper collects non-reflective events through destructive experiments involving improper fiber fusion and excessive bending. As each

event occurs at a different location, this paper segments the OTDR fault events by allocating 1000 sampling points to each event.

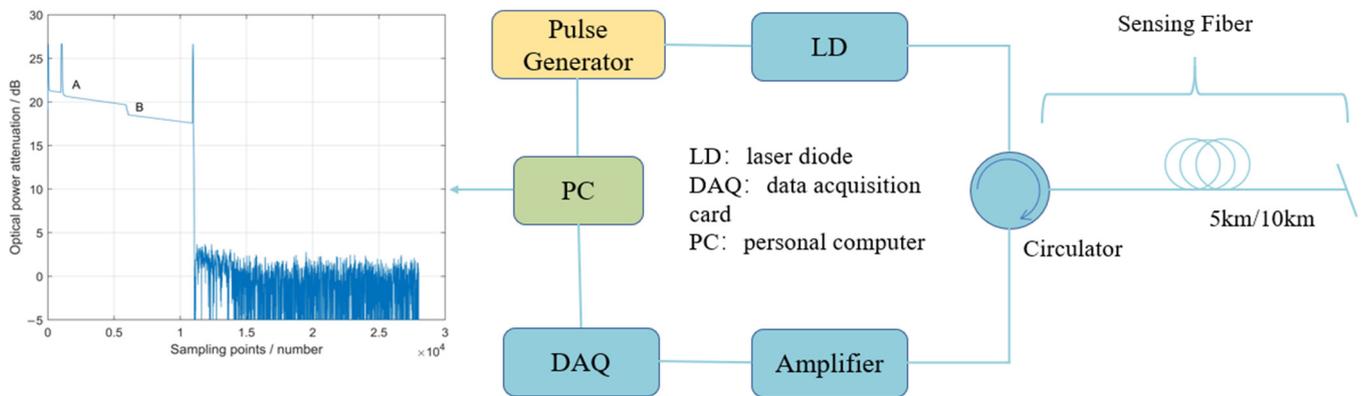


Figure 5. Data acquisition schematic diagram.

The OTDR signal curve undergoes reflection and attenuation under different conditions of the optical cable line. It is non-stationary in both the time domain and the frequency domain. The frequency components of the reflection event are typically concentrated in the high-frequency band, while the non-reflection event signal usually has a lower-frequency element. Wavelet packet decomposition is a time–frequency analysis method that decomposes the original signal into a series of frequency bands. This decomposition allows for a better reflection of the signal resolution in each frequency band [30]. The steps for feature extraction using wavelet packet are as follows:

1. Select “rbio3.1” as the wavelet base to decompose the OTDR signal curve in layers, and frequency bands can be obtained.
2. To obtain $D_j^p(k)$, reconstruct the wavelet packet coefficient d_j^p at node (j, p) . Here, j represents the number of decomposition layers, and $p = 0, 1, \dots, 2^j - 1, k = 1, 2, \dots, m$. $D_j^p(k)$ denotes the amplitude of the k -th sampling point corresponding to the p -th sub-band of the j -th layer after decomposition.
3. Calculate the energy value of each frequency band of the j -th layer according to Formula (15):

$$E_j^p = \sum_{k=1}^N |D_j^p(k)|^2 \tag{15}$$

where N is the sum of the coefficients of each frequency band.

4. To reconstruct the eigenvector, the total energy of each node is used as an element. The energy ratio of each node is then calculated. The X eigenvector can be expressed in Formula (16):

$$X = [P_0, P_1, P_2, \dots, P_{2^j-1}] \tag{16}$$

$$P_i = \frac{E_j^i}{\sum_{i=0}^{2^j-1} E_j^i} \tag{17}$$

In Equation (17), P_i represents the total energy of the signal, while E_j^i represents the total energy of each node.

5. The wavelet packet energy feature extraction is separately conducted on the reflection event and the non-reflection event. This process yields two sets of feature vectors, each containing 8 features. These vectors are then normalized. The feature energy and the enlarged energy of the two types of event points are depicted in Figure 6.

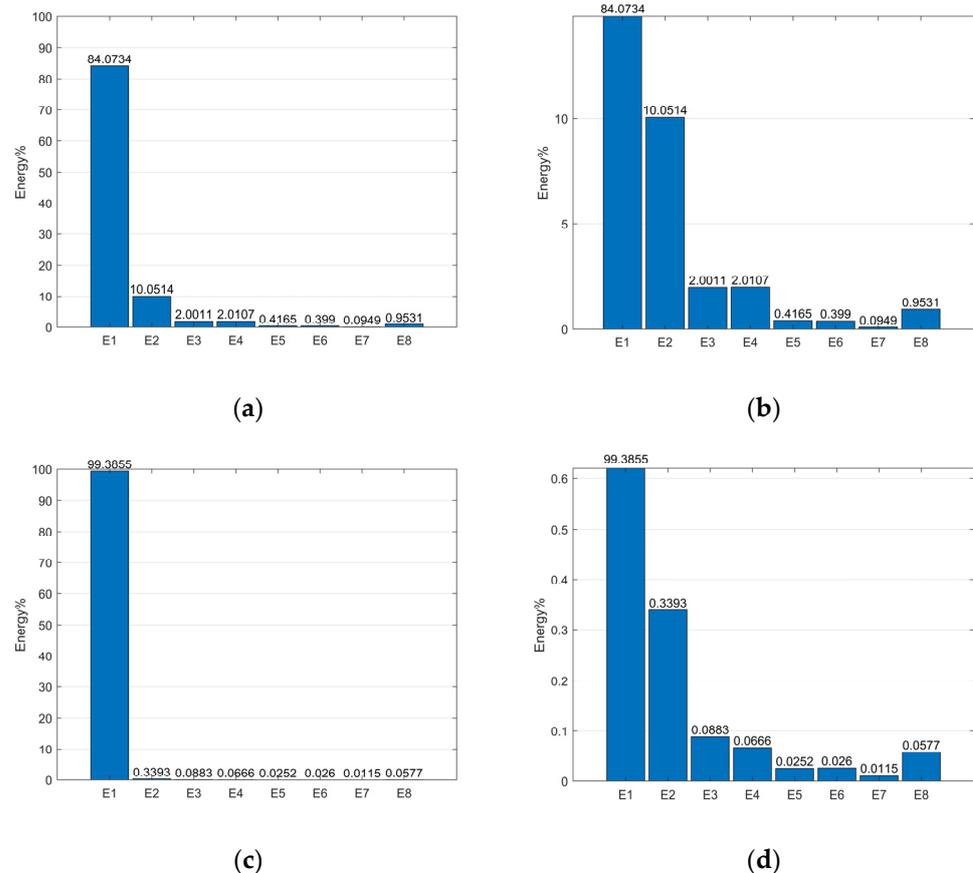


Figure 6. The feature energy of two fault events. (a) The original energy map of reflection event. (b) Reflection event amplification energy map. (c) The original energy map of non-reflection event. (d) Non-reflection event amplification energy map.

3.3. Result Analysis

This paper utilizes 500 groups of fault events collected by OTDR as the training set. It also incorporates a total of 57 actual optical cable lines from the new building of China Grid Jilin Electric Power Co., Ltd. to the Sanjiazi 220 kV substation of State Grid Jilin Electric Power Co., Ltd. as the test set for optical cable fault events. In this test set, reflection events are marked as 0, while non-reflection events are marked as 1, with the order of events being randomly disturbed. To compare and evaluate the performance of the LightGBM algorithm, this paper constructs two additional models using classic machine-learning algorithms: SVM and ELM. These models are trained and tested using the same training and test sets. The experimental results, as depicted in Figure 7, show that the LightGBM algorithm achieves the highest recognition accuracy at 94.74%. The ELM algorithm achieves a recognition accuracy of 92.98%, surpassing the SVM algorithm. The SVM algorithm exhibits the lowest recognition accuracy at 82.46%. Therefore, it can be concluded that utilizing the LightGBM algorithm to develop the recognition model is reasonable and can yield the expected recognition effect.

Based on the previous analysis, it can be observed that LightGBM offers several parameters that can be optimized. By optimizing some of these parameters, it is possible to improve the recognition results. In this study, the values of p_1 , p_2 , and p_3 in the AVOA were set to 0.6, 0.4, and 0.6, respectively. The number of populations was set to 50, and the maximum number of iterations was set to 50. After optimization, the learning_rate value was determined to be 0.3, the feature_fraction value was set to 0.6, the num_leaves value was set to 25, and the max_depth value was set to 3.

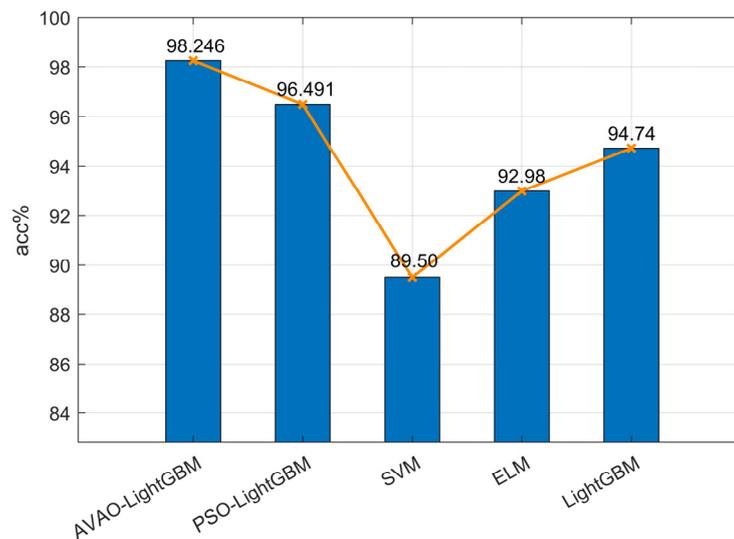


Figure 7. Comparison of recognition results of five models.

In order to compare and study the optimization effect of the AVOA, this paper also uses the classic particle swarm optimization algorithm (PSO) to construct the PSO-LightGBM recognition model. Comparative experiments were conducted to evaluate the recognition accuracy of the two combined models: AVOA-LightGBM and PSO-LightGBM. The results in Figure 7 show that the recognition accuracy of AVOA-LightGBM is 98.25%, while the recognition accuracy of PSO-LightGBM is 96.49%. Both models have higher recognition accuracy compared with the unoptimized recognition model. The fitness curve of the two combined models is depicted in Figure 8. The AVOA exhibits a strong global search ability, allowing for extensive exploration of the search space. This reduces the risk of getting stuck in a local optimal solution. However, the convergence speed of the algorithm may vary depending on the complexity of the problem and parameter settings. On the other hand, the PSO algorithm also possesses some global search ability, but it is less sensitive to parameters and more prone to falling into local optimal solutions when dealing with problems.

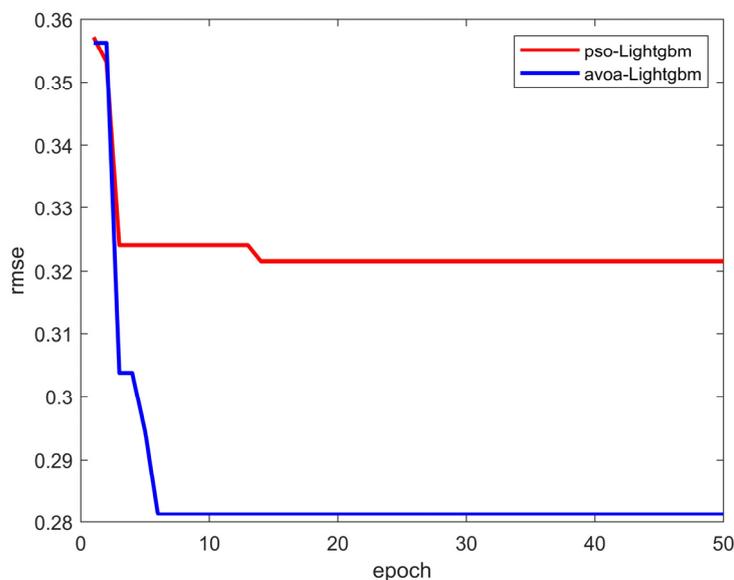


Figure 8. Fitness curve.

3.4. Model Evaluation

In order to further validate the effectiveness of the model proposed in this paper, four indicators—accuracy, precision, recall, and F1 value—are used to evaluate the performance of the proposed model and compare it with the baseline model. The accuracy rate represents the proportion of correctly predicted samples out of the total number of samples, providing an overall measure of recognition accuracy. The precision rate measures the proportion of samples predicted as positive examples that are actually positive, while the recall rate measures the proportion of true positive examples that are correctly predicted as positive. The F1 value, which is the harmonic mean of precision and recall, provides a comprehensive assessment of the model's performance. The experimental results comparing the recognition models are presented in Table 2.

Table 2. Identification model comparison experimental results.

Model	Accuracy	Precision	Recall	F1 Value
LightGBM	94.74	90	94.74	92.31
ELM	92.98	90	90	90
SVM	82.46	95	67.86	79.17
AVOA-LightGBM	98.25	95.45	100	97.67
PSO-LightGBM	96.49	90	100	94.74

Table 2 demonstrates that directly using the basic LightGBM as a model for optical cable event recognition yields favorable results when processing the same data set. Thanks to the AVOA's exceptional optimization ability, the AVOA-LightGBM model outperforms others in terms of accuracy, precision, recall, and F1 value. Its accuracy rate reaches 98.25%, which is 3.7% higher than that of the single LightGBM model and 19.15% higher than that of the SVM recognition algorithm. This further validates the superiority of the algorithm employed in this paper for optical cable event identification.

4. Conclusions

In order to enhance the accuracy of power optical cable fault event recognition, this paper presents a novel method called AVOA-LightGBM for pattern recognition of optical cable fault events based on wavelet packet decomposition. The feature vector used in the identification model is obtained by performing wavelet packet decomposition on fault events and extracting the energy vector. This enables accurate identification of reflection events and non-reflection events in the OTDR curve. Experimental results demonstrate that LightGBM achieves the highest recognition accuracy among the four models tested (LightGBM, SVM, ELM, and PSO-LightGBM). Additionally, the AVOA demonstrates the strongest global search capability. The proposed AVOA-LightGBM model achieves a recognition accuracy of 98.24%, which is a 3.7% improvement over the single LightGBM model and 19.15% more accurate than the SVM recognition algorithm. In practical applications of power fiber optic cables, utilizing the AVOA-LightGBM model proposed in this study for identifying the fault event of power fiber optic cable event points holds significant importance for fault diagnosis and maintenance of the fiber optic cable.

This study only identified two types of fault events: reflective events and non-reflective events. However, the OTDR detection curve also includes start events and end events. Therefore, future studies should focus on increasing the detection of these two types of events. Additionally, the AVOA, a newly proposed optimization algorithm in recent years, has several limitations such as parameter sensitivity, limited application range, and effectiveness on various problems. In our future research, we will aim to improve the algorithm itself to address or minimize these limitations.

Author Contributions: Conceptualization, X.C. and W.C.; methodology, X.C. and W.C.; software, W.C. and T.Z.; validation, W.C. and T.Z.; formal analysis, W.C.; investigation, X.C.; resources, X.C.; data curation, W.C. and T.Z.; writing—original draft preparation, W.C.; writing—review and editing, X.C.; visualization, W.C.; supervision, X.C.; project administration, X.C.; funding acquisition, X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Jilin Province Science and Technology Development Plan Project, grant number 20210203044SF, and the Jilin Province Budget Capital Construction Funds (Innovation Ability Construction) Plan Project, grant number 2022C045-8.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bao, X.; Chen, L. Recent progress in distributed fiber optic sensors. *Sensors* **2012**, *12*, 8601–8639. [[PubMed](#)]
2. Masoudi, A.; Newson, T.P. Contributed Review: Distributed optical fibre dynamic strain sensing. *Rev. Sci. Instrum.* **2016**, *87*, 011501. [[PubMed](#)]
3. Zhou, H.; Xu, H.; Duan, J.-A. Review of the technology of a single mode fiber coupling to a laser diode. *Opt. Fiber Technol.* **2020**, *55*, 102097.
4. Chou, M.-H.; Su, C.-L.; Lee, Y.-C.; Chin, H.-M.; Parise, G.; Chavdarian, P. Voltage-drop calculations and power cable designs for harbor electrical distribution systems with high voltage shore connection. *IEEE Trans. Ind. Appl.* **2016**, *53*, 1807–1814. [[CrossRef](#)]
5. Xie, K.; Zhang, H.; Zhao, Y.; Tian, Y.; Yuan, L. Structural health monitoring of power transmission system based on optical fiber sensor under transmission line galloping. *Laser Optoelectron. Prog.* **2018**, *55*, 070606.
6. Jiang, X.; Ma, J.; Wang, S.; Sun, C.; Shu, L. Transmission lines' ice accidents and analysis of the formative factors. *Electr. Power* **2005**, *38*, 27–30.
7. Lu, L.; Sun, X.; Bu, X.; Li, B. Study on passive, wide area and multi-state parameter monitoring and diagnosis for power transmission lines. In Proceedings of the 2018 International Conference on Power System Technology (POWERCON), Guangzhou, China, 6–8 November 2018; pp. 3903–3907.
8. Yan, Y.; Kui, Z.; Wei, L. A new method for OTDR signal denoising. *Sci. Technol. Eng.* **2012**, *20*, 3486–3489.
9. Peng, F.; Duan, N.; Rao, Y.-J.; Li, J. Real-time position and speed monitoring of trains using phase-sensitive OTDR. *IEEE Photonics Technol. Lett.* **2014**, *26*, 2055–2057. [[CrossRef](#)]
10. Stopiński, S.; Anders, K.; Szostak, S.; Piramidowicz, R. Optical time domain reflectometer based on application specific photonic integrated circuit. In Proceedings of the European Conference on Lasers and Electro-Optics, Munich, Germany, 23–27 June 2019; p. ch_p_38.
11. Tian, K.; Farrell, G.; Wang, X.; Yang, W.; Xin, Y.; Liang, H.; Lewis, E.; Wang, P. Strain sensor based on gourd-shaped single-mode-multimode-single-mode hybrid optical fibre structure. *Opt. Express* **2017**, *25*, 18885–18896. [[CrossRef](#)]
12. Qiang, W.; Xiao, R.; Shui, Y.; Jie, Y.; Pu, Z.; Wei, L. Long-distance and High-precision LFM Optical Time Domain Reflectometer. *Study Opt. Commun.* **2020**, *46*, 7.
13. Xin, P.; Zou, J.; Zhang, S.; Xiao, Q.; Lu, J.; Wei, Y. OPGW Cable Risk Identification Method Based on K-means Clustering. In Proceedings of the 2022 IEEE 22nd International Conference on Communication Technology (ICCT), Nanjing, China, 11–14 November 2022; pp. 1452–1457.
14. Lu, L.; Liang, Y.; Li, B.; Guo, J. Maintenance of the OPGW using a distributed optical fiber sensor. In Proceedings of the 2014 International Conference on Power System Technology, Chengdu, China, 20–22 October 2014; pp. 1251–1256.
15. Wu, H.; Qian, Y.; Zhang, W.; Tang, C. Feature extraction and identification in distributed optical-fiber vibration sensing system for oil pipeline safety monitoring. *Photonic Sens.* **2017**, *7*, 305–310.
16. Xu, C.; Guan, J.; Bao, M.; Lu, J.; Ye, W. Pattern recognition based on enhanced multifeature parameters for vibration events in ϕ -OTDR distributed optical fiber sensing system. *Microw. Opt. Technol. Lett.* **2017**, *59*, 3134–3141.
17. Tabi Fouda, B.M.; Han, D.; An, B.; Lu, X.; Tian, Q. Events detection and recognition by the fiber vibration system based on power spectrum estimation. *Adv. Mech. Eng.* **2018**, *10*. [[CrossRef](#)]
18. Wu, J.; Guan, L.; Bao, M.; Xu, Y.; Ye, W. Vibration events recognition of optical fiber based on multi-scale 1-D CNN. *Opto-Electron. Eng.* **2019**, *46*, 180493.
19. Wang, Y.; Wang, P.; Ding, K.; Li, H.; Zhang, J.; Liu, X.; Bai, Q.; Wang, D.; Jin, B. Pattern recognition using relevant vector machine in optical fiber vibration sensing system. *IEEE Access* **2019**, *7*, 5886–5895. [[CrossRef](#)]
20. Wang, X.; Liu, Y.; Liang, S.; Zhang, W.; Lou, S. Event identification based on random forest classifier for Φ -OTDR fiber-optic distributed disturbance sensor. *Infrared Phys. Technol.* **2019**, *97*, 319–325.
21. Li, M.; Gu, J.; Luo, X.; Xiong, B.; Li, J.; Wang, F.; Dou, R. Pattern recognition of fiber disturbance based on support vector machine in polarization optical time domain reflectometry. In Proceedings of the 2019 International Conference on Optical Instruments and Technology: Optical Sensors and Applications, Beijing, China, 26–28 October 2019; pp. 241–249.

22. Wang, J.; Wang, Z.; Li, J.; Peng, Y. An Interpretable Depression Prediction Model for the Elderly Based on ISSA Optimized LightGBM. *J. Beijing Inst. Technol.* **2023**, *32*, 168–180.
23. Yan, P.; Chen, F.; Zhao, T.; Zhang, H.; Kan, X.; Liu, Y. Transformer fault diagnosis research based on LIF technology and IAO optimization of LightGBM. *Anal. Methods* **2023**, *15*, 261–274.
24. Abdollahzadeh, B.; Gharehchopogh, F.S.; Mirjalili, S. African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems. *Comput. Ind. Eng.* **2021**, *158*, 107408.
25. Salah, B.; Hasanien, H.M.; Ghali, F.M.; Alsayed, Y.M.; Abdel Aleem, S.H.; El-Shahat, A. African Vulture Optimization-Based Optimal Control Strategy for Voltage Control of Islanded DC Microgrids. *Sustainability* **2022**, *14*, 11800. [[CrossRef](#)]
26. He, Z.; Tang, B.; Luan, F. An improved African vulture optimization algorithm for dual-resource constrained multi-objective flexible job shop scheduling problems. *Sensors* **2022**, *23*, 90. [[PubMed](#)]
27. Nayak, S.R.; Khadanga, R.K.; Panda, S.; Sahu, P.R.; Padhy, S.; Ustun, T.S. Participation of Renewable Energy Sources in the Frequency Regulation Issues of a Five-Area Hybrid Power System Utilizing a Sine Cosine-Adopted African Vulture Optimization Algorithm. *Energies* **2023**, *16*, 926.
28. Liu, X.; Liu, Y.; Kong, X.; Ma, L.; Besheer, A.H.; Lee, K.Y. Deep neural network for forecasting of photovoltaic power based on wavelet packet decomposition with similar day analysis. *Energy* **2023**, *271*, 126963.
29. Dhar, J. An adaptive intelligent diagnostic system to predict early stage of parkinson's disease using two-stage dimension reduction with genetically optimized lightgbm algorithm. *Neural Comput. Appl.* **2022**, *34*, 4567–4593.
30. Li, Z.; Chen, Z. Short-term load forecasting based on CEEMDAN-FE-ISSA-LightGBM model. *Front. Energy Res.* **2023**, *11*, 1111786.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.