

Article

Transmission Line Fault-Cause Identification Based on Hierarchical Multiview Feature Selection

Shengchao Jian ¹, Xiangang Peng ^{1,*}, Haoliang Yuan ^{1,*}, Chun Sing Lai ^{1,2,*}  and Loi Lei Lai ^{1,*} 

¹ Department of Electrical Engineering, School of Automation, Guangdong University of Technology, Guangzhou 510006, China; 2111904005@mail2.gdut.edu.cn

² Brunel Interdisciplinary Power Systems Research Centre, Department of Electronic and Electrical Engineering, Brunel University London, London UB8 3PH, UK

* Correspondence: epvg@gdut.edu.cn (X.P.); haoliangyuan@gdut.edu.cn (H.Y.); chunsing.lai@brunel.ac.uk (C.S.L.); l.l.lai@gdut.edu.cn (L.L.L.)

Abstract: Fault-cause identification plays a significant role in transmission line maintenance and fault disposal. With the increasing types of monitoring data, i.e., micrometeorology and geographic information, multiview learning can be used to realize the information fusion for better fault-cause identification. To reduce the redundant information of different types of monitoring data, in this paper, a hierarchical multiview feature selection (HMVFS) method is proposed to address the challenge of combining waveform and contextual fault features. To enhance the discriminant ability of the model, an ϵ -dragging technique is introduced to enlarge the boundary between different classes. To effectively select the useful feature subset, two regularization terms, namely $l_{2,1}$ -norm and Frobenius norm penalty, are adopted to conduct the hierarchical feature selection for multiview data. Subsequently, an iterative optimization algorithm is developed to solve our proposed method, and its convergence is theoretically proven. Waveform and contextual features are extracted from yield data and used to evaluate the proposed HMVFS. The experimental results demonstrate the effectiveness of the combined used of fault features and reveal the superior performance and application potential of HMVFS.

Keywords: fault-cause identification; transmission line; sparse learning; multiview learning; feature selection

check for
updates

Citation: Jian, S.; Peng, X.; Yuan, H.; Lai, C.S.; Lai, L.L. Transmission Line Fault-Cause Identification Based on Hierarchical Multiview Feature Selection. *Appl. Sci.* **2021**, *11*, 7804. <https://doi.org/10.3390/app11177804>

Academic Editor: Jordi Cusido

Received: 1 August 2021

Accepted: 21 August 2021

Published: 25 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Transmission lines cover a wide area and work in diverse outdoor environments to achieve long-distance, high-capacity power transmission. In order to maintain stable power supply, high-speed fault diagnosis is indispensable for line maintenance and fault disposal.

Traditional fault diagnosis technologies concerning fault detecting, fault locating, and phase selection are well developed [1,2], while diagnosis on external causes is still underdeveloped. Operation crews attach great importance to fault location for line patrol and manual inspection. However, on-site inspection is labor-intensive and depends on subjective judgment. Moreover, cause identification after inspection is too late for dispatchers to give better instructions according to the external cause, such as forced energization. Fault-cause identification is expected to help dispatch and maintenance personnel make a proper and speedy fault response.

Transmission line faults are more often triggered by external factors due to environmental change or surrounding activities. Though the cause categories are slightly different between regions or institutions, the common causes can be listed as lighting, tree, animal contact, fire, icing, pollution and external damage [3]. Considering complexity and variability of open-air work, it is hard to model fault scenarios for diverse root causes [4,5]. Thus, these existing studies on line fault-cause identification have been developed based on data-driven methods rather than physical modeling.

The early identification methods were rule-based, such as statistical analysis, CN2 rule induction [6] and fuzzy inference system (FIS) [7–9]. Their identification frameworks are finally presented in the form of logic flow, demanding a great degree of robustness and generality for their rules or thresholds. In recent years, various machine learning (ML) techniques that attach great importance to hand-crafted features have been applied to diagnose external causes [10–14], such as logistic regression (LR), artificial neural network (ANN), k-nearest neighbor (KNN) and support vector machine (SVM). Deep learning (DL) provides a more efficient way in the field of fault identification. In [15], deep belief network (DBN) is used as the classification algorithm after extracting time–frequency characteristics from traveling wave data. Even when using DL methods, feature engineering is still an inevitable part to achieve high accuracy.

Feature signature study provides knowledge about fault information and plays a critical role in fault-cause identification. On the one hand, when fault events happen, power quality monitors (PQMs) enable us to have easy access to electrical signals and time stamps [16]. Time-domain features extracted from fault waveform and time stamp were used to construct logic flow to classify lightning-, animal- and tree-induced faults [6]. To exploit transient characteristics in the frequency domain, signal processing techniques such as wavelet transform (WT) and empirical mode decomposition (EMD) are used for further waveform characteristic analysis [17–20]. In [21], a fault waveform was characterized based on the time and frequency domain to develop an identification logic. However, a fault waveform is easily affected by the system operation state, and there is no direct connection between these characteristics and external causes. On the other hand, weather condition is directly relevant to many fault-cause categories such as lightning, icing and wind. With the development of monitoring equipment and communication technology, dispatchers now can make judgments with more and more outdoor information [22]. These nonwaveform characteristics such as time stamps, environment attributes and other textual data are called contextual characteristics in this paper. Table 1 lists and compares the characterization and classification methods in existing works.

Table 1. A summarized list of characterization and classification methods used for fault-cause identification.

Article	Waveform Characteristics				Time Characteristics	External Characteristics	Classification Methods
	Signal Amplitude	Sequence Component	Spectrum Analysis	Phase or Phase Angle			
* Núñez, Meléndez [6]	✓	✓		✓	✓		CN2
Liang, Li [7]	✓	✓					FIS
* Xu, Chow [8–10]				✓	✓	✓	FIS/LR/ANN
* Cai, Chow [11]				✓	✓	✓	LR
Chang, Hong [12]	✓		✓				SVM
* Jiang, Liu [14]	✓		✓	✓	✓	✓	KNN
Liang, Liu [15]	✓		✓		✓		DBN
Asman, Aziz [20]	✓		✓				decision tree
* Qin, Wang [21]	✓	✓	✓	✓	✓		logic flow
* Dehbozorgi, Rastegar [22]					✓	✓	decision tree
Minnaar, Nicolls [23]	✓	✓		✓	✓	✓	KNN

Articles with * concern faults on distribution network but their work is still inspiring for transmission network.

Studies have shown that waveform and contextual features can achieve high accuracy without each other, but there are high data requirements. For economic and operational reasons, data condition will not change significantly in the short term. It is necessary to study performance improvement for fault-cause identification based on current data conditions. One of the challenges is determining how to combine waveform features and multisource contextual features. This is an information fusion problem, and the simplest approach is feature concatenation. The authors of [23] tried to combine contextual features and waveform features as a mixed vector, but concatenated features reduce performance. Moreover, in contrast to focusing on either side, a few studies use both waveform and contextual characteristics for higher classification performance.

To tackle the fusion challenge, multiview learning (MVL) is introduced in this paper because waveform and contextual features describe the same fault event in different views. MVL aims to integrate multiple-view data properly and overcome biases between multiple views to obtain satisfactory performance. One of typical MVL methods is canonical correlation analysis (CCA), which maps multiview features into a common feature space [24]. Instead of mapping features, multiview feature selection that selects features from each view is preferred in fault-cause identification. Unlike traditional feature selection, multiview feature selection treats multiview data as inherently related and ensures that complementary information from other views is exploited [25,26]. In [27], a review on real-time power system data analytics with wavelet transform is given. The use of discrete wavelet transform was used to identify the high impedance fault and heavy load conditions [28]. The authors of [29] propose a fault diagnosis approach for the main drive chain in a wind turbine based on data fusion. To deal with the kind of multivariable fault diagnosis problem for which input variables need to be adjusted for different typical faults, the deep autoencoder model is adopted for the fault diagnosis model training for different typical fault types.

In this paper, we propose a hierarchical multiview feature selection (HMFVS) method for transmission line fault-cause identification. Two view datasets are composed of the waveform features and the contextual features. Our proposed HMFVS is applied to conduct the feature selection for the optimal feature combination. In our model, to enhance the discriminant ability of regression, an ε -dragging technology is used to enlarge the margin between classes. Next, two regularization terms, namely $l_{2,1}$ -norm and Frobenius norm (F -norm) penalty, are adopted to perform the hierarchical feature selection. Here, the $l_{2,1}$ -norm realizes the row sparsity to reduce the unimportant features of each view and the F -norm realizes the view-level sparsity to reduce the diversity between these two-view data. Hence, these two penalties can be viewed as low-level and high-level feature selection, respectively. At last, the fault-cause identification is carried out using ML classifiers and integrated features. The contributions of this paper are highlighted as follows:

- To the best of our knowledge, this is the first time that multiview learning is introduced for transmission line fault-cause identification in view of the nature of multiview fault data.
- We propose a novel approach, HMFVS, based on the ε -dragging and two regularization terms to select the discriminative features across views. We also develop an iterative algorithm to solve the optimization problem and prove its convergence theoretically.
- The performance of HMFVS is evaluated on field data and compared with classical feature selection methods. Experimental results prove the effectiveness of combining waveform and contextual features and demonstrate the feasibility and superiority of HMFVS.

The rest of this paper is organized as follows: Section 2 presents the proposed HMFVS algorithm and its convergence analysis. Section 3 outlines the real-life line fault dataset and extracts features in terms of waveform and nonwaveform. The empirical study is provided and discussed in Section 4. Section 5 presents concluding remarks.

2. Hierarchical Multiview Feature Selection (HMFVS)

2.1. Notation

Sparsity-based multiview feature selection can be formulated as an optimization problem and denoted by loss functions and regularization items. Before introducing our formulation, the notation is stated.

Matrices are denoted by boldface uppercase letters, and vectors are denoted by boldface lowercase letters. Given original feature matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T \in \mathbb{R}^{n \times d}$, each row of which corresponds to a fault instance, n is the total number and d denotes the size of features. $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$ and $\mathbf{x}_i^{(v)} \in \mathbb{R}^{d^{(v)}}$ denote a feature matrix and a vector in the

v th view. There are two views in this paper; thus, $\mathbf{X} = [\mathbf{X}^{(1)}, \mathbf{X}^{(2)}]$. Suppose there are c categories, the label matrix will be represented as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^T \in \{0, 1\}^{n \times c}$. Weight matrix \mathbf{W} can be derived as $\mathbf{W} = [\mathbf{W}^{(1)}, \mathbf{W}^{(2)}]^T = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d]^T \in \mathbb{R}^{d \times c}$.

2.2. The Objective Function

Given the notation defined and a fault dataset (\mathbf{X}, \mathbf{Y}) , the problem of HMFVS is transformed into determining weight matrix \mathbf{W} and then ranking features for selection. We formulate the optimization problem as

$$\min_{\mathbf{W}, \mathbf{M}} \Psi(\mathbf{W}, \mathbf{M}) + \alpha \Phi(\mathbf{W}) + \beta \Omega(\mathbf{W}) = \min_{\mathbf{W}, \mathbf{M}} \|\mathbf{X}\mathbf{W} - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}\|_F^2 + \alpha \|\mathbf{W}\|_{2,1} + \beta \sum_{v=1}^m \|\mathbf{w}_v\|_F, \quad (1)$$

where m is the view number; $m = 2$ in this paper.

In this formulation, $\Psi(\mathbf{W}, \mathbf{M})$ is the loss function that measures the calculation distance to achieve minimum regression error, which is derived from the least square loss function. Furthermore, the ε -dragging is introduced to drag binary outputs in \mathbf{Y} away along two opposite directions. The outputs for positive digits will become $1 + \varepsilon_i$ and the outputs for negative digits will be $-\varepsilon_i$, in which all of the ε s are nonnegative. The treatment that enlarges the distance between data points from different classes helps to develop a compact optimization model for classification [30]. $\mathbf{B} \in \{-1, 1\}^{n \times c}$ in the formulation is a constant matrix, and its element B_{ij} is defined as

$$B_{ij} = \begin{cases} +1, & Y_{ij} = 1 \\ -1, & Y_{ij} = 0. \end{cases} \quad (2)$$

B_{ij} denotes the dragging direction for elements in label matrix \mathbf{Y} . $\mathbf{M} \in \mathbb{R}^{n \times c}$ is a nonnegative matrix that records all ε s. The operator \otimes is the Hadamard product operator of matrices. Thus, $\mathbf{B} \otimes \mathbf{M}$ represents the dragging distance, and we have a new label matrix after the ε -dragging:

$$\mathbf{Y}' = \mathbf{Y} + \mathbf{B} \otimes \mathbf{M}. \quad (3)$$

With the least square loss function defined as

$$\Psi(\mathbf{W}) = \|\mathbf{X}\mathbf{W} - \mathbf{Y}\|_F^2, \quad (4)$$

we can attain our loss function $\Psi(\mathbf{W}, \mathbf{B}, \mathbf{M})$.

$$\Psi(\mathbf{W}, \mathbf{B}, \mathbf{M}) = \|\mathbf{X}\mathbf{W} - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}\|_F^2. \quad (5)$$

Next, regularization items used in the formulation are $l_{2,1}$ -norm and F -norm, and we take row-wise feature selection and view-wise feature selection into account.

$$\Phi(\mathbf{W}) = \|\mathbf{W}\|_{2,1} = \sum_{i=1}^d \sqrt{\sum_{j=1}^c w_{ij}^2}. \quad (6)$$

$$\Omega(\mathbf{W}) = \sum_{v=1}^m \|\mathbf{w}_v\|_F = \sum_{v=1}^m \sqrt{\sum_{i=1}^d \sum_{j=1}^c w_{ij}^2}. \quad (7)$$

$l_{2,1}$ -norm measures the distance of features as a whole and forces the weights of unimportant features to be assigned small values so that it can perform feature selection among all features. Similarly, F -norm measuring the distance between views forces the weights of unimportant views to be assigned small values [31]. The weight matrix \mathbf{W} is regulated by these penalty terms, and hierarchical feature selection is completed with row-wise and view-wise selection. $l_{2,1}$ -norm penalty corresponds to the low-level feature selection, and F -norm penalty corresponds to the high-level feature selection.

Therefore, the objective function of the HMVFS model is obtained and represented as (1). α and β are nonnegative constants that tune hierarchical feature selection. This model is also available with more than two views.

2.3. Optimization

In order to solve $l_{2,1}$ -norm minimization and F -norm minimization problems, the regularization terms $\|\mathbf{W}\|_{2,1}$ and $\sum_{v=1}^m \|\mathbf{W}_v\|_F$ need to be respectively relaxed by $Tr(\mathbf{W}^T\mathbf{C}\mathbf{W})$ and $Tr(\mathbf{W}^T\mathbf{D}\mathbf{W})$ [32]. The objective function is rewritten as

$$\begin{aligned} & \min_{\mathbf{W}, \mathbf{M}, \mathbf{C}, \mathbf{D}} \|\mathbf{X}\mathbf{W} - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}\|_F^2 + \alpha Tr(\mathbf{W}^T\mathbf{C}\mathbf{W}) + \beta Tr(\mathbf{W}^T\mathbf{D}\mathbf{W}), \\ & s.t. \quad C_{ii} = \frac{1}{2\|\mathbf{w}_i\|_2}, D_{jj} = \frac{1}{2\|\mathbf{W}_v\|_F}, \end{aligned} \tag{8}$$

where $\mathbf{C} \in \mathbb{R}^{d \times d}$ and $\mathbf{D} \in \mathbb{R}^{d \times d}$ are diagonal matrices and derived from \mathbf{W} . For D_{ii} , \mathbf{w}_i is the row vector of \mathbf{W}_v .

Though two more variables are introduced, we obtain a convex function, and we can solve the optimization problem iteratively. In each iteration, we update one variable while others are fixed, and all variables can be optimized in order. In view of \mathbf{C} and \mathbf{D} derived from \mathbf{W} , we fix \mathbf{M} and update \mathbf{W} at first. The derivative of (8) w.r.t. \mathbf{W} is calculated as

$$2\mathbf{X}^T(\mathbf{X}\mathbf{W} - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}) + 2\alpha\mathbf{C}\mathbf{W} + 2\beta\mathbf{D}\mathbf{W}. \tag{9}$$

Let (9) equal zero, then the updated \mathbf{W} can be obtained by solving the equation. If there are big-size data or high-dimensional data, the gradient descent method is recommended. Following that, \mathbf{C} and \mathbf{D} can be updated.

When it turns to \mathbf{M} , the optimization problem can be transformed from (8) to (10).

$$\begin{aligned} & \min_{\mathbf{M}} \|\mathbf{Z} - \mathbf{B} \otimes \mathbf{M}\|_F^2, \\ & s.t. \quad \mathbf{Z} = \mathbf{X}\mathbf{W} - \mathbf{Y}. \end{aligned} \tag{10}$$

According to the definition of F -norm, this problem can be decoupled into $n \times c$ subproblems [30] and represented as

$$\min_{M_{ij}} (Z_{ij} - B_{ij}M_{ij})^2. \tag{11}$$

With $B_{ij}^2 = 1$, (11) is equivalent to (12).

$$\min_{M_{ij}} (Z_{ij}B_{ij} - M_{ij})^2. \tag{12}$$

With the nonnegative constraint, M_{ij} is calculated as

$$M_{ij} = \max(Z_{ij}B_{ij}, 0). \tag{13}$$

Accordingly, \mathbf{M} can be updated as

$$\mathbf{M} = \max(\mathbf{Z} \otimes \mathbf{B}, 0). \tag{14}$$

Up to now, all variables are updated in the iteration and we present the optimization process in Algorithm 1.

After optimization, we obtain weight matrix \mathbf{W} learned across all views and then sort all features according to their importance. The importance is measured by the l_2 -norm value of each row vector of \mathbf{W} , $\|\mathbf{w}_i\|_2 (i = 1, 2, \dots, d)$. Feature selection can be completed with features ranked in descending order.

2.4. Convergence

In this subsection, we analyze the convergence of Algorithm 1. We need to guarantee the objective function decreases in each iteration of the optimization algorithm. The following lemma is used to verify its convergence.

Lemma 1. For any nonzero values $a, b \in \mathbb{R}$, the following inequality holds:

$$2ab \leq (a^2 + b^2) \Rightarrow a - \frac{a^2}{2b} \leq b - \frac{b^2}{2b}. \tag{15}$$

Theorem 1. The objective Function (1) monotonically decreases in the iteration of Algorithm 1.

Proof. According to Step 6 and Step 7 in Algorithm 1, we have \mathbf{W}_{t+1} and \mathbf{M}_{t+1} as follows:

$$\mathbf{W}_{t+1} \Leftarrow \min_{\mathbf{W}} \|\mathbf{X}\mathbf{W}_t - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}_t\|_F^2 + \alpha \text{Tr}(\mathbf{W}_t^T \mathbf{C}_t \mathbf{W}_t) + \beta \text{Tr}(\mathbf{W}_t^T \mathbf{D}_t \mathbf{W}_t), \tag{16}$$

$$\mathbf{M}_{t+1} \Leftarrow \min_{\mathbf{M}} \|\mathbf{X}\mathbf{W}_{t+1} - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}_t\|_F^2. \tag{17}$$

Firstly, according to (16) and (17), there is

$$\begin{aligned} & \|\mathbf{X}\mathbf{W}_{t+1} - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}_{t+1}\|_F^2 + \alpha \text{Tr}(\mathbf{W}_{t+1}^T \mathbf{C}_t \mathbf{W}_{t+1}) + \beta \text{Tr}(\mathbf{W}_{t+1}^T \mathbf{D}_t \mathbf{W}_{t+1}) \\ & \leq \|\mathbf{X}\mathbf{W}_{t+1} - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}_t\|_F^2 + \alpha \text{Tr}(\mathbf{W}_{t+1}^T \mathbf{C}_t \mathbf{W}_{t+1}) + \beta \text{Tr}(\mathbf{W}_{t+1}^T \mathbf{D}_t \mathbf{W}_{t+1}) \leq \|\mathbf{X}\mathbf{W}_t - \mathbf{Y} - \mathbf{B} \otimes \mathbf{M}_t\|_F^2 + \alpha \text{Tr}(\mathbf{W}_t^T \mathbf{C}_t \mathbf{W}_t) + \beta \text{Tr}(\mathbf{W}_t^T \mathbf{D}_t \mathbf{W}_t). \end{aligned} \tag{18}$$

Thus, according to the definition of \mathbf{C} , we have

$$\begin{aligned} \alpha \text{Tr}(\mathbf{W}_{t+1}^T \mathbf{C}_t \mathbf{W}_{t+1}) &= \sum_{i=1}^d \frac{\|(\mathbf{w}_i)_{t+1}\|_2^2}{2\|(\mathbf{w}_i)_t\|_2} \\ &= \alpha \sum_{i=1}^d \|(\mathbf{w}_i)_{t+1}\|_2 - \alpha \left(\sum_{i=1}^d \|(\mathbf{w}_i)_{t+1}\|_2 - \sum_{i=1}^d \frac{\|(\mathbf{w}_i)_{t+1}\|_2^2}{2\|(\mathbf{w}_i)_t\|_2} \right) = \alpha \Phi(\mathbf{W}_{t+1}) - \alpha \left(\sum_{i=1}^d \|(\mathbf{w}_i)_{t+1}\|_2 - \sum_{i=1}^d \frac{\|(\mathbf{w}_i)_{t+1}\|_2^2}{2\|(\mathbf{w}_i)_t\|_2} \right) \end{aligned} \tag{19}$$

We also perform the same transformation with $\text{Tr}(\mathbf{W}_{t+1}^T \mathbf{D}_t \mathbf{W}_{t+1})$, $\text{Tr}(\mathbf{W}_t^T \mathbf{C}_t \mathbf{W}_t)$ and $\text{Tr}(\mathbf{W}_t^T \mathbf{D}_t \mathbf{W}_t)$. We can rewrite (18) as

$$\begin{aligned} & \Psi(\mathbf{W}_{t+1}, \mathbf{M}_{t+1}) + \alpha \Phi(\mathbf{W}_{t+1}) + \beta \Omega(\mathbf{W}_{t+1}) - \alpha \left(\sum_{i=1}^d \|(\mathbf{w}_i)_{t+1}\|_2 - \sum_{i=1}^d \frac{\|(\mathbf{w}_i)_{t+1}\|_2^2}{2\|(\mathbf{w}_i)_t\|_2} \right) - \beta \left(\sum_v^m \|(\mathbf{w}_v)_{t+1}\|_F - \sum_v^m \frac{\|(\mathbf{w}_v)_{t+1}\|_F^2}{2\|(\mathbf{w}_v)_t\|_F} \right) \\ & \leq \Psi(\mathbf{W}_t, \mathbf{M}_t) + \alpha \Phi(\mathbf{W}_t) + \beta \Omega(\mathbf{W}_t) - \alpha \left(\sum_{i=1}^d \|(\mathbf{w}_i)_t\|_2 - \sum_{i=1}^d \frac{\|(\mathbf{w}_i)_t\|_2^2}{2\|(\mathbf{w}_i)_t\|_2} \right) - \beta \left(\sum_v^m \|(\mathbf{w}_v)_t\|_F - \sum_v^m \frac{\|(\mathbf{w}_v)_t\|_F^2}{2\|(\mathbf{w}_v)_t\|_F} \right). \end{aligned} \tag{20}$$

According to Lemma 1, we arrive at

$$\Psi(\mathbf{W}_{t+1}, \mathbf{M}_{t+1}) + \alpha \Phi(\mathbf{W}_{t+1}) + \beta \Omega(\mathbf{W}_{t+1}) \leq \Psi(\mathbf{W}_t, \mathbf{M}_t) + \alpha \Phi(\mathbf{W}_t) + \beta \Omega(\mathbf{W}_t). \tag{21}$$

Thus, Algorithm 1 decreases the optimization problem in (1) for each iteration so (1) will converge to its global optimum according to its convexity.

Algorithm 1 The optimization algorithm for (8)

Input: The feature matrix across all views, $\mathbf{X} \in \mathbb{R}^{n \times d}$; the label matrix, $\mathbf{Y} \in \{0,1\}^{n \times c}$; the parameters α and β
Output: The weight matrix across all views, $\mathbf{W} \in \mathbb{R}^{d \times c}$

- 1: Calculate \mathbf{B} from \mathbf{Y} via (2)
- 2: Initialize \mathbf{W}_0 and \mathbf{M}_0
- 3: Initialize $t = 0$
- 4: **Repeat**
- 5: Calculate \mathbf{C}_t and \mathbf{D}_t from \mathbf{W}_t
- 6: $\mathbf{W}_{t+1} = (\mathbf{X}^T \mathbf{X} + \alpha \mathbf{C}_t + \beta \mathbf{D}_t)^{-1} (\mathbf{X}^T \mathbf{Y} + \mathbf{X}^T \mathbf{B} \otimes \mathbf{M}_t)$
- 7: $\mathbf{M}_{t+1} = \max((\mathbf{X} \mathbf{W}_{t+1} - \mathbf{Y}) \otimes \mathbf{B})$
- 8: $t = t + 1$
- 9: Calculate residue via (1)
- 10: **Until** convergence or maximum iteration number achieved

3. Material and Characterization

3.1. Data Collection and Cleaning

In this study, the fault data were collected from an AC transmission network located in a coastal populous city in Guangdong Province, China. These faults occurred between 2016 and 2019, and the voltage levels varied from 110 to 500 kV. Fault signals were recorded by digital fault recorders (DFRs) installed on substations. The DFR equipment involves PMUs and computer systems to synchronize, store and display analog data for voltage and current signals. These signals can be remotely accessed through a communication network and provide offline data stored in common format for transient data exchange (COMTRADE). The sampling rate is 5 kHz in the dataset. Environmental information and other associated monitoring data were obtained through the inner maintenance system. A patrol report of manual inspection was attached to each fault, describing the inspection result and labeling its cause. The original dataset comprised 551 samples, and 288 of them remained after cleansing. The distribution of fault-cause categories is shown in Figure 1. Lightning, external force and object contact are the three dominant causes. External force refers to collision or damage due to human activity. Object contact is usually caused by floating objects in the air. These are typical causes in a densely populated city, causing more than 90% of known faults.

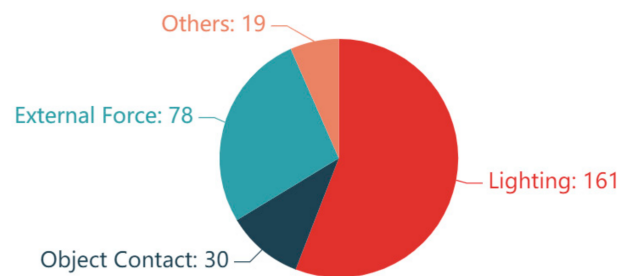


Figure 1. Distribution of transmission line fault cause after cleansing.

3.2. Waveform Characteristics

It is believed that the disturbance variation of electrical quantity after faults occurring contains important transient information for fault diagnosis [33]. The original waveform data are recorded in COMTRADE files with the sampling frequency of 5 kHz. The first step is to acquire fault segments and extract valid waveform segments without disturbance caused by tripping. In this paper, the beginning of valid segments is determined by inspection thresholds based on root mean squared (rms) current magnitude. dI is the difference between consecutive values.

$$dI \geq 0.15 \text{ pu or } I \geq 1.2 \text{ pu.} \quad (22)$$

The start thresholds are determined by inspection to make sure that fault measurements in this study are correctly captured. Since COMTRADE stores not only electrical

signals in analog channels but also tripping information in digital channels, one and a half cycles after tripping enabling signal is regarded as the end of the segment. In characterization, we extend previous research work on waveform characterization. The following waveform features are considered and extracted.

1. **Maximum Change of Sequence Components:** Instantaneous magnitude is calculated relative to prefault amplitude in order to be compatible with measurements from different voltage levels and operation conditions. Karenbauer transformation is used to obtain zero, positive and negative components of three-phase signals, denoted by s , $s = 0, 1, 2$.

$$V_{s(\max)} = \frac{\max(V_{s(\text{fault})})}{V_{s(\text{pre-fault})}}, s = 0, 2. \quad (23)$$

$$I_{s(\max)} = \frac{\max(I_{s(\text{fault})})}{I_{s(\text{pre-fault})}}, s = 0, 1, 2. \quad (24)$$

2. **Maximum Rate of Change of Sequence Components:**

$$\Delta V_{s(\max)} = \frac{\max(|\Delta V_s|)}{V_{s(\text{pre-fault})}}, s = 0, 1, 2. \quad (25)$$

$$\Delta I_{s(\max)} = \frac{\max(|\Delta I_s|)}{I_{s(\text{pre-fault})}}, s = 0, 1, 2. \quad (26)$$

3. **Sequence Component Values at t -cycle:** t is set to be 0, 0.5, 1 and 1.5. For instance, $t = 0.5$ means the measuring point is 1/2 cycle from the start.

$$V_{s(t)} = \frac{V_{s(t)}}{V_{s(\text{pre-fault})}}, s = 0, 1, 2. \quad (27)$$

$$I_{s(t)} = \frac{I_{s(t)}}{I_{s(\text{pre-fault})}}, s = 0, 1, 2. \quad (28)$$

4. **Custom Time Constant of Sequence Current:** Inspired by a linear time-invariant system, time content is introduced to reflect the dynamic response of the network [23]. Time content is the time required to rise from the zero point to $1/e$ of the maximum current. In this study, $1/e$ is replaced with a custom value, m . These features are denoted as $TC_{I_s(m)}$, $m = 0.1, 0.2, \dots, 0.9, 1$
5. **DC and Harmonic Content:** Hilbert–Huang transform is used to conduct spectrum analysis [17]. The harmonic content and DC content are calculated from the ratio of the specific component to the fundamental component. DC and harmonic content are denoted as Har_k , $k = 0, 3, 5, 7, 9, 11$
6. **Wavelet Energy and Energy Entropy:** Discrete wavelet transform is applied to decompose fault-phase current signals into three wavelet scales. Wavelet energy E and energy entropy S are calculated for each scale.

$$p_j = \frac{E_j}{\sum_j E_j} = \frac{\sum_j |C_j|^2}{\sum_j \sum_j |C_j|^2}, S_j = -p_j \log_2(p_j). \quad (29)$$

where C_j , E_j , p_j denote wavelet coefficient, wavelet energy and relative energy in scale j , $j = 1, 2, 3$.

- Maximum DC Current: Equation (30) is used to calculate the maximum DC current on three-phase signals. N_s is the number of data points in one cycle, and $n = 0$ means the triggering point.

$$I_{dc(max)} = \max(I_{dc_a}, I_{dc_b}, I_{dc_c}), I_{dc} = \frac{\sum_{n=1}^{N_s} i_n - \sum_{n \leq -N_s} i_n}{\max(I_{pre-fault})}. \tag{30}$$

- Time Domain Factors: Form factor, crest factor, skewness and kurtosis, denoted as t_1-t_4 , respectively, are introduced to reflect characteristics of waveform shape and the shock for fault-phase current signals. SD denotes their standard deviation.

$$t_1 = \frac{\sqrt{\frac{1}{N_s} \sum_{n=1}^{N_s} (i_n)^2}}{\frac{1}{N_s} \sum_{n=1}^{N_s} |i_n|}, t_2 = \frac{\max(i_n)}{\frac{1}{N_s} \sum_{n=1}^{N_s} |i_n|}, t_3 = \frac{\sum_{n=1}^{N_s} (i_n - \bar{i})^3}{SD^3 N_s}, t_4 = \frac{\sum_{n=1}^{N_s} (i_n - \bar{i})^4}{SD^4 N_s} \tag{31}$$

- Approximation Constants δ for Neural Waveform: In order to learn more from the front wave, the waveform of rms neutral voltage/current is approximated by (32), as introduced in [33].

$$f(t, \delta) = 1 - e^{-\delta t}, \tag{32}$$

where t is time step and δ is the approximation constant. Equation (32) estimates the closest value with regard to the actual waveform in per unit value.

- Fault Inception Phase Angle (FIPA): FIPA is calculated based on the trigger time after the last zero crossing point prior to fault happening.

All waveform features are listed in Table 2. Faulted phase features are included in the next subsection.

Table 2. Feature pools.

Pool Type	Feature	Total Number
Waveform	Maximum sequence voltage/current	5
	Maximum change of three-phase signals and sequence components	6
	Sequence component values	24
	Custom time constant of sequence current	30
	DC and harmonic content	6
	Wavelet energy and energy entropy	6
	Maximum DC current	1
	Form factor, crest factor, skewness and kurtosis	4
	Approximation constants	2
	FIPA	1
Contextual	Time stamp: season, day/night, month, hour	4
	Location: landform, zone	2
	Meteorological data: weather, temperature, humidity, rainfall, cloud cover, maximum wind speed, wind scale	7
	Protection data: reclosing, fault phase, fault duration, tripping time, breaker quenching time, reclosing time, number of triggering	7
	Others: voltage level, number of faults	2

3.3. Contextual Characteristics

Most monitoring technologies are developed for specified causes and work independently with interconnected data. In this study, due to data restriction, available nonwaveform data include time stamps, meteorological data, geographical data, protection data and query information. These informative values are preprocessed and integrated into the pool

of candidate contextual features, as shown in Table 2. Considering that there is no accurate discretization standard, we only discretize text data roughly if necessary. The time stamp information is discretized twice based on season and day/night as a contrast of months and daytime. As for dynamic records such as meteorological value, the records closest to the fault time are retained. Protection data are feedback information of protection devices after fault, usually obtained from the production management system. Although these collected data are related to fault events, they are not suitable for fault cause identification. These irrelevant features pose a great challenge in feature selection.

4. Experiments and Discussion

4.1. Experiment Setup

To validate the effectiveness and efficiency of HMFVS, we conducted comparison experiments using the mentioned field data previously. Three strategies for utilizing multiview data with feature selection were considered, namely single-view learning, feature concatenation after selection and feature selection after concatenation. The last two are the simplest early fusion methods. Single-view learning is represented via best single view (BSV) method, through which the most informative view achieves the best performance among views. As for the dataset in this paper, contextual features are more representative than hand-crafted waveform features. Feature concatenation after selection (FSFC) employs a feature selection technique separately and concatenates features selected from different views. Feature selection after concatenation (FCFS) concatenates original feature sets of two views and then performs feature selection. Adaptable feature selection methods listed in the next subsection are applied to select discriminative features.

The fault dataset was split into training data and testing data in a stratified fashion according to the ratio of 3:1. All samples were normalized by standard deviation after zero-mean standardization. Then, feature selection methods were used to seek the optimal feature combination using training sets and transform all samples for fault-cause classification. ML classifiers were utilized to finish the classification. In the presence of imbalanced data, criteria such as G-mean and accuracy were used to quantitatively assess classification performance. Since G-mean is a metric within biclass concepts, its microaverage was computed and adopted. The final results of each metric were calculated as the average of the 5 trials.

4.2. Comparison Feature Algorithms

As reviewed in [34], there are many feature selection methods. We conducted comparison experiments between our MVFS and several typical feature selection algorithms, namely Fisher score (F-Score), mutual information (MI), joint mutual information (JMI), joint mutual information maximization (JMIM), ReliefF, Hilbert–Schmidt independence criterion lasso (HSIC Lasso) [35] and recursive feature elimination (RFE). F-Score ranks features through variance similarity calculation, and the same rank can be obtained by analysis of variance (ANOVA). MI ranks features according to values of their mutual information with class labels. JMI and JMIM are developed from MI [36]. RFE ranks and discards features after training a certain kind of classifier. Starting from all features, the elimination process continues until the feature number or output error is settled to a minimum.

The above algorithms are developed for single-view learning and can be used in BSV, FCFS and FSFC directly. Except for RFE, all of them are filter feature selection approaches, as is HMFVS. Besides, the comparison algorithms designed for multiview learning are kernel canonical correlation analysis (KCCA) [24] and discriminant correlation analysis (DCA) [37]. These feature extraction approaches map multiview data into a common feature space so their results are attached to the comparison in FCFS. As for the proposed algorithm, there are two hyperparameters in HMFVS. In the experiments, these hyperparameters α and β were tuned ranging in $\{10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ through grid search on the training

sets. Moreover, experiments without any feature algorithm were conducted using BSV features and all features, tabbed as RAW_BSV and RAW.

4.3. Overall Classification Performance

In this subsection, we compare the mentioned dimension reduction approach on the basis of SVM to verify the effectiveness of multiview learning and HMVFS. Two concatenating rules were applied to FSFC. The first rule tries to keep 1:1 proportion of waveform and contextual features. There is one more contextual feature when the total number is odd. The second rule holds the same proportion of waveform and contextual features as that in HMVFS.

The results in terms of Gmean with different numbers of selected features are shown in Figure 2. By comparing single-view feature selection methods among strategies, we notice that most of them perform best in BSV rather than in FSFC and FCFS. Added fault features from the other view will even degrade their classification, and this indicates that simple concatenation cannot help conventional feature selection methods adapt to multiview classification. A similar conclusion is drawn in [23]. Thus, the introduction of MVL appears vital in particular. HMVFS has comprehensive advantages in the comparison of FSFC and FCFS and achieves the best performance compared with methods in BSV. HMVFS outperforms others in the middle of feature increasing, and its result with 14 selected features is the global or near-global optimum. When features from the other view increase, the performance is degraded to a certain extent, and then it rises to another peak. Most methods in BSV produce a zigzag rise curve and reach their best when almost all view features are selected. They are also inferior to HMVFS in FSFC and FCFS. ReliefF is the best competitor that achieves acceptable performance in different strategies. As for KCCA and DCA, their performance is low. Figure 2 illustrates that HMVFS is more capable of obtaining the best performance combining waveform and contextual features.

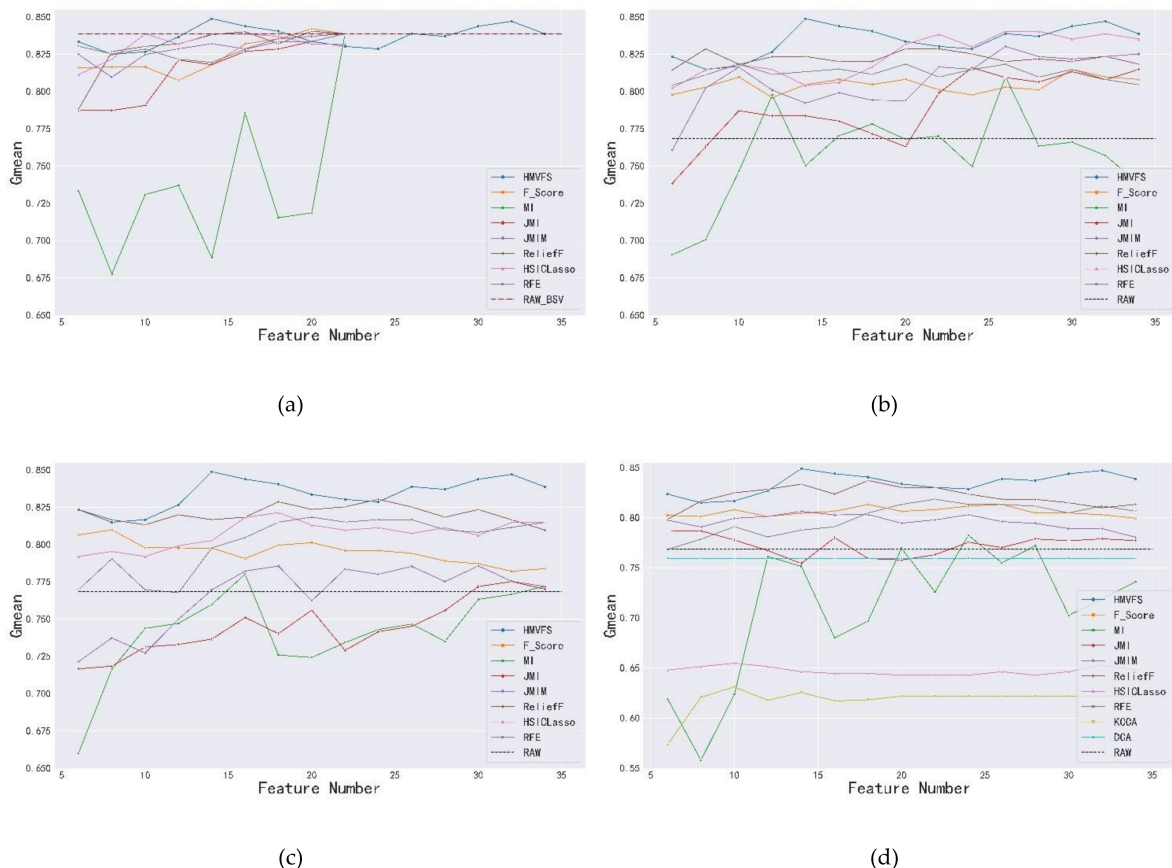


Figure 2. Classification comparison between HMVFS and other feature algorithms in strategies: (a) BSV; (b) FSFC_rule1; (c) FSFC_rule2; (d) FCFS.

Due to the limit of yield data condition and fault signature study, irrelevant and redundant features are introduced with increasing feature numbers. This problem is more prominent in the waveform view in both theoretical and experimental studies. The advantage of HMFVS is that it selects features with independent and complementary information of all views, while the single-view methods are easily affected by irrelevant features facing concatenated assembly or meeting the limitation of single-view features. As seen from Figure 2, concatenating and mapping fail to select or transform discriminative features with combined waveform and contextual features. There are two local optimums for HMFVS, and they are better than the performance of competitors, which demonstrates that HMFVS overcomes the negative effect of redundant features in multiview data.

4.4. Parameter Sensitivity

Determination of hyperparameters is an open problem for many algorithms. We conducted parameter sensitivity study by testing different settings of parameters α and β . Since these parameters help HMFVS perform hierarchical feature selection, it is clear that HMFVS will be sensitive to parameter change, and this study may reveal a hierarchical feature relationship. The candidate set was $\{10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ for each parameter. Classification performance and average running time are recorded and illustrated in Figure 3.

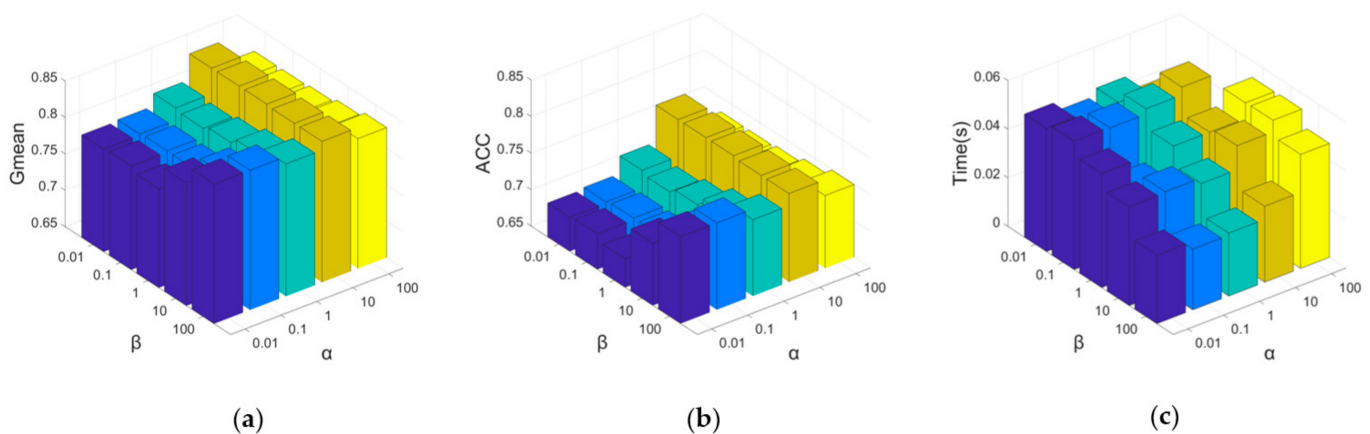


Figure 3. Performance variation of HMFVS with different values for the parameters α and β in terms of (a) Gmean; (b) ACC; (c) time.

It is observed that $\alpha = 10$ is beneficial to final selection and maintains relatively high classification performance, among which lower β has slight advantages. View importance is different in multiview learning. From the perspective of view importance, when only two views exist and one of them is generally better, acceptable performance can be achieved by one view, and additional features are expected for improvement. High-level feature selection is weak because the other view has relatively more redundant features and will be ignored with higher β . Meanwhile, appropriately higher α enhances low-level feature selection to exploit the most representative features from the unimportant view. Moreover, acceptable performance is achieved with $\alpha = 10^{-2}$, $\beta = 10^2$ and $\alpha = 10^{-1}$, $\beta = 10^2$. High-level selection is enhanced, and low-level selection is restrained, which results in limited performance approximating in single-view learning and short convergence time.

4.5. Comparison between ML Classifiers

In order to investigate the effect of classifiers and explore better identification accuracy, we employed different ML learners to complete fault-cause classification with HMFVS. Owing to space limitation and performance stability, F_Score and ReliefF were used for comparison. The typical individual classifiers CN2, LR, KNN, SVM and ANN, which have been proven effective in fault-cause identification studies, were tested, and the results

are presented in this subsection. Ensemble models promote fault-cause identification by combining individual learners [22], so we also explored the performance of various ensemble models, including random forest (RF), AdaBoosting, stacking ensemble and dynamic ensemble. META-DES, DES-Clustering and KNORA-U are dynamic ensemble techniques based on metalearning, clustering and k-nearest neighbors, respectively. Classification models were developed using Python machine learning library, scikit-learn and DESlib. Table 3 presents the best performance for each combination of feature selection methods and classifiers. Considering some data may be similar, AUC is introduced as a supplement criterion, which is derived from receiver operating characteristic (ROC) analysis and calculated as the area under the ROC curve.

Table 3. Best performance comparison with different ML classifiers.

Classifier	Feature Selection	Feature Number	Gmean	ACC	AUC
CN2	F_Score	39	0.707	0.581	0.834
	ReliefF	33	0.707	0.580	0.836
	HMVFS	28	0.730	0.612	0.841
LR	F_Score	16	0.833	0.756	0.889
	ReliefF	15	0.833	0.756	0.896
	HMVFS	33	0.831	0.752	0.896
KNN	F_Score	14	0.838	0.764	0.891
	ReliefF	11	0.835	0.760	0.895
	HMVFS	7	0.848	0.778	0.909
SVM	F_Score	18	0.812	0.728	0.908
	ReliefF	18	0.837	0.761	0.906
	HMVFS	14	0.849	0.779	0.921
ANN	F_Score	18	0.837	0.761	0.891
	ReliefF	15	0.850	0.780	0.911
	HMVFS	36	0.842	0.769	0.915
RF	F_Score	27	0.878	0.821	0.926
	ReliefF	12	0.876	0.819	0.935
	HMVFS	9	0.875	0.817	0.935
AdaBoost	F_Score	36	0.781	0.684	0.797
	ReliefF	19	0.777	0.679	0.830
	HMVFS	14	0.784	0.690	0.846
META-DES	F_Score	19	0.876	0.816	0.930
	ReliefF	11	0.872	0.812	0.928
	HMVFS	12	0.881	0.824	0.937
DES-Clustering	F_Score	32	0.872	0.812	0.916
	ReliefF	13	0.875	0.817	0.932
	HMVFS	10	0.882	0.827	0.945
KNORA-U	F_Score	15	0.872	0.812	0.926
	ReliefF	14	0.870	0.809	0.932
	HMVFS	12	0.884	0.829	0.942
Stacking	F_Score	16	0.880	0.824	0.930
	ReliefF	13	0.874	0.814	0.936
	HMVFS	11	0.886	0.831	0.939

As seen from the table, HMVFS outperforms F_Score and ReliefF except with LR and ANN. It is observed that HMVFS always takes fewer features to achieve the best performance in the remaining comparisons. In the group of RF, the best scores of F_Score, ReliefF and HMVFS are very close to each other because RF has the ability of variable selection. Thus the features that function in final classification are similar if selected feature subsets are large enough to contain valuable features. Except for mentioned learners, HMVFS has advantages in both score and feature number.

From the perspective of learners, the classification performance improves with the enhancement of model complexity. CN2 as a rule-based learner cannot cope with multiview features to achieve acceptable performance. Individual learners cannot achieve accuracies greater than 0.8, which are apparently inferior to most ensemble models. Among ensemble models, stacking ensemble realizes the best fault-cause identification in this study. The experimental results of ML classifiers indicate that HMFVS is more suitable for classifiers with high generalization and that ensemble models can bring significant improvement for fault-cause identification.

5. Conclusions

Associated multisource data for transmission line fault-cause diagnosis are divided and extracted as waveform and contextual features in this paper. MVL is introduced to appropriately combine these features for performance improvement. A novel hierarchical multiview feature selection method based on an ϵ -dragging technique and sparsity regularization is proposed to perform hierarchical feature selection with multiview data. The ϵ -dragging is applied in the loss function to enlarge sample distance between classes. $l_{2,1}$ -norm and F -norm conduct row-wise and view-level selection, respectively, which can be viewed as the low-level and high-level feature selection. We also develop the optimization algorithm and prove its convergence theoretically. The proposed HMFVS is evaluated by comparisons on yield data. The results reveal that HMFVS outperforms conventional feature selection methods in single-view and early fusion strategies. The further experiments concerning ML classifiers also demonstrate the superiority and effectiveness of the proposed method with high generalization learners. This study has shown the combined use of waveform and contextual features with HMFVS can cause significant improvement for fault-cause identification. In future work, more multiview data and further fault signature study are needed to refine the feature pools, and the performance of HMFVS is expected to be further improved.

Author Contributions: Conceptualization, S.J. and X.P.; methodology, H.Y. and C.S.L.; software, S.J.; experiment, validation and analysis, S.J. and H.Y.; investigation, S.J. and X.P.; resources, X.P.; data curation, S.J.; writing—original draft preparation, S.J. and X.P.; writing—review and editing, S.J., H.Y., C.S.L. and L.L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (61903091) and the Science and Technology Project of China Southern Power Grid Company Limited (031800KK52180074).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Ferreira, V.H.; Zanghi, R.; Fortes, M.Z.; Sotelo, G.G.; Silva, R.; Souza, J.; Guimarães, C.; Gomes, S., Jr. A survey on intelligent system application to fault diagnosis in electric power system transmission lines. *Electr. Power Syst. Res.* **2016**, *136*, 135–153.
2. Chen, Y.; Fink, O.; Sansavini, G. Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction. *IEEE Trans. Ind. Electron.* **2018**, *65*, 561–569. [[CrossRef](#)]
3. Minnaar, U.J.; Gaunt, C.T.; Nicolls, F. Characterisation of power system events on South African transmission power lines. *Electr. Power Syst. Res.* **2012**, *88*, 25–32. [[CrossRef](#)]
4. Cai, Y.; Chow, M. Cause-Effect modeling and spatial-temporal simulation of power distribution fault events. *IEEE Trans. Power Syst.* **2011**, *26*, 794–801. [[CrossRef](#)]
5. Gui, M.; Pahwa, A.; Das, S. Bayesian network model with Monte Carlo simulations for analysis of animal-related outages in overhead distribution systems. *IEEE Trans. Power Syst.* **2011**, *26*, 1618–1624. [[CrossRef](#)]
6. Núñez, V.B.; Meléndez, J.; Kulkarni, S.; Santoso, S. Feature analysis and automatic classification of short-circuit faults resulting from external causes. *Int. Trans. Power Syst.* **2013**, *23*, 510–525. [[CrossRef](#)]

7. Liang, Y.; Li, K.; Ma, Z.; Lee, W. Typical fault cause recognition of single-phase-to-ground fault for overhead Lines in nonsolidly earthed distribution networks. *IEEE Trans. Ind. Appl.* **2020**, *56*, 6298–6306. [[CrossRef](#)]
8. Xu, L.; Chow, M.; Taylor, L.S. Power distribution fault cause identification with imbalanced data using the data mining-based fuzzy classification E-algorithm. *IEEE Trans. Power Syst.* **2007**, *22*, 164–171. [[CrossRef](#)]
9. Xu, L.; Chow, M.; Timmis, J.; Taylor, L.S. Power distribution outage cause identification with imbalanced data using artificial immune recognition system (AIRS) algorithm. *IEEE Trans. Power Syst.* **2007**, *22*, 198–204. [[CrossRef](#)]
10. Xu, L.; Chow, M. A classification approach for power distribution systems fault cause identification. *IEEE Trans. Power Syst.* **2006**, *21*, 53–60.
11. Cai, Y.; Chow, M.; Lu, W.; Li, L. Statistical feature selection from massive data in distribution fault diagnosis. *IEEE Trans. Power Syst.* **2010**, *25*, 642–648. [[CrossRef](#)]
12. Chang, G.W.; Hong, Y.; Li, G. A hybrid intelligent approach for classification of incipient faults in transmission network. *IEEE Trans. Power Deliv.* **2019**, *34*, 1785–1794. [[CrossRef](#)]
13. Morales, J.; Orduña, E.A.; Rehtanz, C. Identification of lightning stroke due to shielding failure and backflashover for ultra-high-speed transmission line protection. *IEEE Trans. Power Deliv.* **2014**, *29*, 2008–2017. [[CrossRef](#)]
14. Jiang, X.; Stephen, B.; McArthur, S. Automated distribution network fault cause identification with advanced similarity metrics. *IEEE Trans. Power Deliv.* **2021**, *36*, 785–793. [[CrossRef](#)]
15. Liang, H.; Liu, Y.; Sheng, G.; Jiang, X. Fault-cause identification method based on adaptive deep belief network and time–frequency characteristics of travelling wave. *IET Gener. Transm. Distrib.* **2019**, *13*, 724–732. [[CrossRef](#)]
16. Tse, N.C.F.; Lai, L.L. Wavelet-based algorithm for signal analysis. *EURASIP J. Adv. Signal Process.* **2007**, *2007*, 1–10. [[CrossRef](#)]
17. Malik, H.; Sharma, R. Transmission line fault classification using modified fuzzy Q learning. *IET Gener. Transm. Distrib.* **2017**, *11*, 4041–4050. [[CrossRef](#)]
18. Tse, N.C.F.; Chan, J.Y.C.; Lau, W.H.; Poon, J.T.Y.; Lai, L.L. Real-time power-quality monitoring with hybrid sinusoidal and lifting wavelet compression algorithm. *IEEE Trans. Power Deliv.* **2012**, *27*, 1718–1726. [[CrossRef](#)]
19. Tse, N.C.F.; Chan, J.Y.C.; Lau, W.H.; Lai, L.L. Hybrid wavelet and Hilbert transform with frequency shifting decomposition for power quality analysis. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 3225–3233. [[CrossRef](#)]
20. Asman, S.H.; Aziz, N.; Amirulddin, U.; Kadir, M. Decision tree method for fault causes classification based on RMS-DWT analysis in 275 kV transmission lines network. *Appl. Sci.* **2021**, *11*, 4031–4051. [[CrossRef](#)]
21. Qin, X.; Wang, P.; Liu, Y.; Guo, L.; Sheng, G.; Jiang, X. Research on distribution network fault recognition method based on time-frequency characteristics of fault waveforms. *IEEE Access* **2018**, *6*, 7291–7300. [[CrossRef](#)]
22. Dehbozorgi, M.; Rastegar, M.; Dabbaghjamesh, M. Decision tree-based classifiers for root-cause detection of equipment-related distribution power system outages. *IET Gener. Transm. Distrib.* **2020**, *14*, 5809–5815. [[CrossRef](#)]
23. Minnaar, U.J.; Nicolls, F.; Gaunt, C. Automating transmission-line fault root cause analysis. *IEEE Trans. Power Deliv.* **2016**, *31*, 1692–1700. [[CrossRef](#)]
24. Shi, Y.; Ji, H. Kernel canonical correlation analysis for specific radar emitter identification. *Electron. Lett.* **2014**, *50*, 1318–1319. [[CrossRef](#)]
25. Ramachandram, D.; Taylor, G. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process. Mag.* **2017**, *34*, 96–108. [[CrossRef](#)]
26. Lai, C.S.; Yang, Y.; Pan, K.; Zhang, J.; Yuan, H.L.; Wing, W.; Gao, Y.; Zhao, Z.; Wang, T.; Shahidepour, M.; et al. Multi-view neural network ensemble for short and mid-term load forecasting. *IEEE Trans. Power Syst.* **2020**. [[CrossRef](#)]
27. Lai, C.S. Compression of power system signals with wavelets. In Proceedings of the 2014 International Conference on Wavelet Analysis and Pattern Recognition, Lanzhou, China, 13–16 July 2014. [[CrossRef](#)]
28. Lai, C.S. High impedance fault and heavy load under big data context. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2016. [[CrossRef](#)]
29. Xu, Z.; Yang, P.; Zhao, Z.; Lai, C.S.; Lai, L.L.; Wang, X. Fault diagnosis approach of main drive chain in wind turbine based on data fusion. *Appl. Sci.* **2021**, *11*, 5804. [[CrossRef](#)]
30. Xiang, S.; Nie, F.; Meng, G.; Pan, C.; Zhang, C. Discriminative least squares regression for multiclass classification and feature selection. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 1738–1754. [[CrossRef](#)] [[PubMed](#)]
31. Zhu, X.; Li, X.; Zhang, S. Block-row sparse multiview multilabel learning for image classification. *IEEE Trans. Cybern.* **2016**, *46*, 450–461. [[CrossRef](#)]
32. Zhang, Y.; Wu, J.; Cai, Z.; Yu, P.S. Multiview multilabel learning with sparse feature selection for image annotation. *IEEE Trans. Multimed.* **2020**, *22*, 2844–2857. [[CrossRef](#)]
33. Zin, A.; Karim, S. Protection system analysis using fault signatures in Malaysia. *Int. J. Electr. Power Energy Syst.* **2013**, *45*, 194–205. [[CrossRef](#)]
34. Lai, C.S.; Zhong, C.; Pan, K.; Ng, W.W.Y.; Lai, L.L. A deep learning based hybrid method for hourly solar radiation forecasting. *Expert. Syst. Appl.* **2021**, *177*, 114941. [[CrossRef](#)]
35. Yamada, M.; Jitkrittum, W.; Sigal, L.; Xing, E.; Sugiyama, M. High-dimensional feature selection by feature-wise Kernelized Lasso. *Neural Comput.* **2014**, *26*, 185–207. [[CrossRef](#)]

-
36. Bannasar, M.; Hicks, Y.; Setchi, R. Feature selection using Joint Mutual Information Maximisation. *Expert Syst. Appl.* **2015**, *42*, 8520–8532. [[CrossRef](#)]
 37. Haghghat, M.; Abdel-Mottaleb, M.; Alhalabi, W. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1984–1996. [[CrossRef](#)]