

Article

Forced Oscillation Detection Using Hybrid Knowledge–Deep Learning Features

Jiaxin Li ^{1,2} , Xiaomei Yang ^{1,2,*}  and Haoran Wu ^{1,2} 

¹ College of Electrical Engineering, Sichuan University, Chengdu 610065, China; leejiaxin@stu.scu.edu.cn (J.L.); hnanw0831@stu.scu.edu.cn (H.W.)

² Key Laboratory of Information and Automation Technology in Sichuan Province, Sichuan University, Chengdu 610065, China

* Correspondence: yangxiaomei@scu.edu.cn

Abstract

Accurate detection of forced oscillations is important for the stable operation of power systems. The method based on prior knowledge relies on manual feature extraction, which has limited ability to characterize non-stationary signals. While deep learning (DL) methods can automatically learn features, they may overlook the physical mechanisms of power systems, potentially leading to misjudgments. We propose a Hybrid Knowledge-DL network (HKD-SVM) that utilizes Support Vector Machine (SVM) as the classifier. In our method, Discrete Wavelet Transform (DWT) is used to represent the time–frequency structure of the input signals, and DL features are extracted by Convolutional Neural Network (CNN) from this time–frequency representation. These learned features are subsequently fused with prior knowledge features that carry explicit physical interpretations, thereby constructing a more discriminative feature representation space. Finally, SVM is adopted as the classifier, making the network well-suited for nonlinear, high-dimensional classification scenarios with limited training samples, which are common in power system applications. Experiments on both simulated and real-world phasor measurement unit (PMU) data demonstrate that HKD-SVM outperforms purely data-driven and purely knowledge-driven methods. The proposed method provides an effective solution for power system oscillation detection.

Keywords: forced oscillation detection; convolutional neural network; discrete wavelet transform; feature fusion

1. Introduction

With the expansion of interconnected power grids and the increasing penetration of distributed energy resources [1], the issue of low-frequency oscillations has become increasingly prominent. Such oscillations may trigger resonance effects, leading to power fluctuations, and may also induce cascading failures, threatening the safe and stable operation of power grids [2,3]. According to the generation mechanisms of low-frequency oscillations, they can be classified into two categories: one is natural oscillations caused by weak system damping, and the other is forced oscillations excited when external periodic disturbances are close to the natural modal frequency of the system [4,5]. Both types of oscillations exhibit highly similar constant-amplitude steady-state behaviors, as shown in Figure 1. However, their suppression methods differ significantly [3,6]. Therefore, accurately detecting the type of oscillation is a critical prerequisite for achieving effective suppression and ensuring the safe and stable operation of power grids.



Academic Editor: Andrea Gaetano Chiariello

Received: 30 June 2026

Revised: 31 July 2026

Accepted: 6 August 2026

Published: 7 August 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

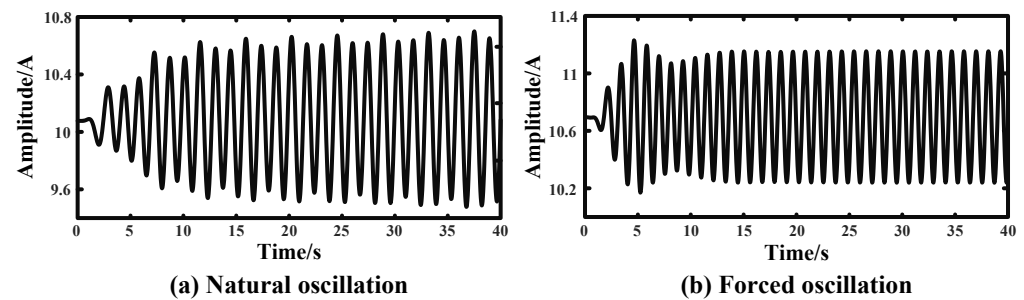


Figure 1. Current amplitude curves of different oscillation types. (a) Natural oscillation caused by weak system damping; (b) forced oscillation excited by external periodic disturbances.

Traditional forced oscillation detection methods primarily rely on signal processing techniques to extract time–frequency domain features of oscillation signals, combined with expert-defined discrimination rules for classification. These approaches are collectively referred to as knowledge-based forced oscillation detection methods. For feature extraction, time-domain characteristics typically include statistical metrics [7], curvature [8], envelope morphology [9], and phase-space trajectory radius [10]. In the frequency domain, Fourier transform and power spectral density analysis are commonly employed to obtain parameters such as oscillation frequency and slope [11–14]. Alternatively, researchers examine energy distribution [15,16], coherence [17], or spectral characteristics [18,19] across different frequency components for feature extraction. The advantages of such methods lie in their clear physical interpretations, mathematical foundations at each step, and high computational efficiency, which bypasses the reliance on extensive training samples. However, their detection performance heavily depends on the effectiveness of extracting robust discriminative features. Moreover, the feature sets are often designed for specific scenarios, making them difficult to adapt to the variable operating conditions of power systems. Additionally, these methods exhibit limited capability in characterizing strongly nonlinear and non-stationary signals. These limitations become particularly pronounced in modern power systems with a high penetration of renewable energy sources [20].

With the rapid development of artificial intelligence technology, data-driven Deep Learning (DL) has demonstrated significant application potential in the field of power system fault detection and diagnosis [21], with common models including Long Short-Term Memory (LSTM) network, Convolutional Neural Network (CNN), and Spiking Neural Network. Among these, CNN and its derivative models have received widespread attention in oscillation detection due to their effective feature extraction and spatial pattern recognition capabilities. For instance, some studies have utilized CNN to process Phasor Measurement Unit (PMU) data for extracting time–frequency images [22,23] and time-domain waveform features [24]; or introduced transfer learning to migrate CNN to different power systems [25]. In addition, Discrete Wavelet Transform (DWT) has been incorporated into detection frameworks owing to its excellent time–frequency decomposition capability, enabling models to extract time–frequency features more clearly and effectively. For instance, DWT has been combined with a dilated CNN and meta-learning for detection in few-shot scenarios [26], and embedded into a CNN-LSTM network with a transfer learning fine-tuning strategy to accurately identify forced oscillations in power systems [27]. Meanwhile, other studies have combined artificial neural networks with randomly scaled conjugate gradient algorithms [28], or constructed hybrid models of spiking recurrent neural networks and LSTM for detection [29].

However, purely data-driven approaches often face significant challenges in forced oscillation detection. On one hand, they operate as black boxes that tend to overlook essential domain-specific prior knowledge, leading to inferences that lack interpretability

and robust analytical foundations. On the other hand, real-world data are often limited and inherently imbalanced, which complicates model convergence and degrades classification performance towards the majority class. While recent advancements have introduced effective strategies to address these imbalances, such as weight optimization [30] and adaptive scaling [31], these approaches primarily focus on manipulating the data distribution rather than enhancing the underlying feature representations.

To further address signal complexity and improve generalization, researchers have explored innovative DL architectures and constraint-based modeling. For instance, physics-informed neural networks embed governing equations for stability prediction [32,33], while attention-augmented CNNs capture long-range temporal dependencies [34]. Furthermore, multi-scale time–frequency decomposition and hybrid networks have enhanced robustness under fluctuating conditions [35,36]. Despite these structural innovations, there remains a critical need to explicitly integrate domain knowledge at the feature extraction level to guide model learning.

To address the aforementioned limitations, we recognize a strong complementary relationship between knowledge-based methods and pure DL paradigms. While data-driven models excel at extracting high-dimensional abstract patterns, they struggle to incorporate expert domain knowledge. Conversely, classical statistical methods provide explicit engineering significance but lack the representational capacity for complex non-stationary dynamics. This dichotomy motivated our hybridization strategy, which uses prior knowledge to guide feature extraction and constructs a domain-relevant feature space. Driven by this design rationale, we propose the Hybrid Knowledge and DL with SVM classification (HKD-SVM) framework for forced oscillation detection. This architecture synergistically integrates domain-specific prior knowledge extracted from PMU data with DL features obtained through time–frequency decomposition and CNN processing. Specifically, DWT is utilized to generate time–frequency representations of signals, based on which a CNN adaptively learns DL features from the decomposed subbands. Subsequently, the knowledge-driven and data-driven features are fused, and SVM is employed to fully exploit its advantages in addressing nonlinear classification tasks. Therefore, the novelty of the proposed framework lies not in the simple integration of these algorithms, but in the complementary fusion mechanism that introduces prior knowledge boundaries and engineering constraints into deep representations. In summary, the main contributions of this paper are as follows:

- A hybrid knowledge–deep learning network is proposed to construct a discriminative feature space by fusing knowledge-driven features, which are rooted in domain-specific prior knowledge, with the extracted DL features. This method retains the explicit engineering significance of knowledge-driven features while enhancing the overall representation capacity through DL. It effectively alleviates the limitations of traditional methods in feature representation and compensates for the neglect of power system domain expertise in pure data-driven models.
- A DWT-CNN structure is designed to facilitate joint time–frequency feature extraction. This architecture leverages multi-scale decomposition to improve model sensitivity to transient dynamic behaviors, providing a more comprehensive representation of oscillation patterns than traditional single-domain feature extraction.
- A robust SVM-based classification strategy is implemented to enhance generalization. By utilizing SVM on the fused feature space, the framework mitigates the overfitting risks typical of deep networks in small-sample scenarios, ensuring reliable performance under diverse grid operating conditions.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 details the proposed method, covering feature extraction, network architecture, SVM classification, and training. Section 4 presents the experimental evaluation. Finally, Section 5 concludes the paper.

2. Related Works

2.1. DWT

DWT is a multi-resolution time–frequency analysis tool. It discretizes signals through scaling and translation factors, enabling the simultaneous characterization of local temporal features and spectral distribution properties of a signal [37]. In power system oscillation detection, DWT is commonly used to decompose the original signal $x(t)$ into an approximation component $A_l(t)$ and a series of detail components $D_l(t)$, expressed as:

$$x(t) = A_l(t) + \sum_{l=1}^L D_l(t), l = 0, 1, \dots, L, \quad (1)$$

where l denotes the decomposition level [38]. $A_l(t)$ corresponds to the low-frequency component, while $D_l(t)$ represents high-frequency fluctuations at different decomposition levels.

This approach facilitates the separation of oscillation modes in different frequency bands and provides a more domain-relevant time–frequency representation basis for subsequent feature extraction.

2.2. CNN

One-dimensional convolutional neural network (1DCNN, hereafter abbreviated as CNN for simplicity) is applicable to 1D time-series data like power system oscillation signals. It can directly extract deep abstract features from raw waveforms or pre-processed feature sequences [39], bypassing complex manual feature engineering. In its forward propagation process, the feature map y_l at the l -th layer can be expressed through the convolution operation as

$$y_l = \sigma(w_l * y_{l-1} + b_l), \quad (2)$$

where w_l represents the 1D convolutional kernel weights, b_l denotes the bias term, $*$ signifies the convolution operation, and σ is the activation function.

Furthermore, the CNN we employed possesses a more concise network architecture. This compact design allows it to achieve effective model training and classification tasks even under small sample conditions.

2.3. SVM

SVM, rather than classical Softmax, is utilized as a classifier to implement the detection of forced oscillation, fully leveraging the generalization performance of SVM in small-sample scenarios, high-dimensional feature spaces, and nonlinear classification tasks.

The core idea of SVM is to find a satisfactory hyperplane that maximizes the classification margin between two sample classes [40]. For linearly inseparable cases in the low-dimensional space, SVM introduces kernel functions to map the original features to a high-dimensional space, thereby transforming the problem into a linearly separable one. Meanwhile, slack variables $\xi_i \geq 0$ and a penalty parameter C are introduced, where C controls the model's tolerance for misclassified samples.

The determination of the optimal hyperplane is transformed into a convex quadratic programming problem:

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad \text{s.t.} \quad y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad (3)$$

where w is the normal vector, b is the bias term, $y_i \in \{-1, 1\}$ are binary labels, and $\phi(x_i)$ is a function that maps x_i to a high-dimensional feature space.

3. Proposed Method

This section elaborates on the proposed HKD-SVM network, as shown in Figure 2. The core procedures of this network are as follows: first, the extracted prior knowledge features are fused with the DL features obtained via the DWT-CNN. This fusion strategy addresses the limitations of prior-knowledge-based detection methods, which rely on handcrafted features with limited representational capacity, as well as the reliance of DL-based detection methods on large-scale labeled samples. Furthermore, owing to the embedding of DWT into CNN, DWT-CNN enables the effective extraction of time–frequency characteristics from oscillatory signals. Specifically, DWT is highly suitable for extracting the time–frequency characteristics of non-stationary oscillation signals. Subsequently, a CNN is employed, as it is capable of automatically learning deep abstract features from the DWT outputs. Finally, the selected fused features are fed into SVM for forced oscillation detection, thereby better handling problems of small-sample scenarios and nonlinear classification.

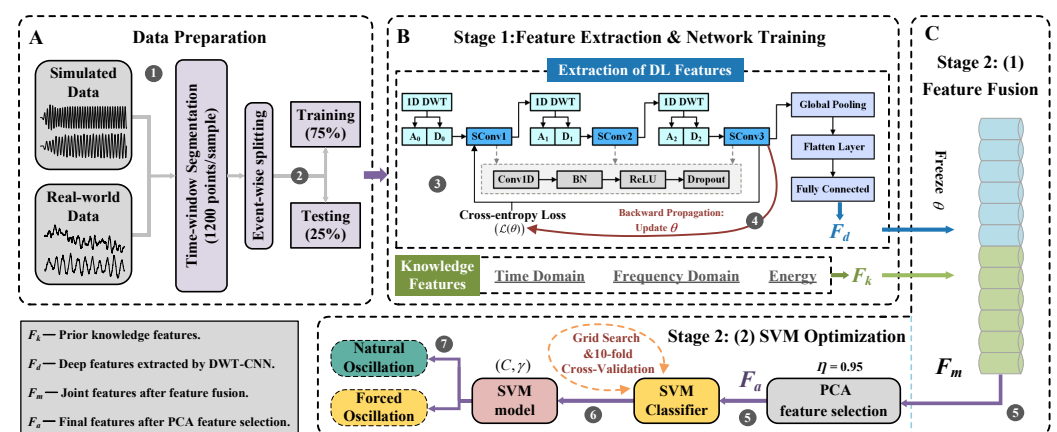


Figure 2. Flowchart of the complete methodology for the proposed HKD-SVM. The overall workflow is divided into three sequential modules: (A) Data Preparation; (B) Stage 1: Feature Extraction and Network Training, which encompasses the parallel extraction of DL features using the DWT-CNN and physical prior knowledge; and (C) Stage 2: Feature Fusion and Classification. The numbered markers (1–7) indicate the execution order of the algorithm.

3.1. Prior Knowledge Features Extraction

To fully leverage the statistical properties and dynamic behaviors of signals, we extracted prior knowledge features from the power system’s PMU data, including time-domain features, frequency-domain features, and energy features.

3.1.1. Definition of Time-Domain Features

To characterize the essential properties of forced oscillations and natural oscillations, we select the key time-domain features directly from PMU data, as shown in Table 1. Basic statistical features ($T_1 \sim T_4$) describe the distribution morphology of the signal $x(n)$ ($n = 0, 1, \dots, N - 1$), where N represents the signal length. Dynamic change features ($T_5 \sim T_7$) characterize the change rate and smoothness of the signal. Additionally, correlation features (T_8) reveal the strength of signal periodicity.

Table 1. Summary of Time-Domain Features.

Feature Name	Expression
Mean	$T_1 = \frac{1}{N} \sum_n x(n)$
Variance	$T_2 = \frac{1}{N} \sum_n (x(n) - T_1)^2$
Peak-to-Peak Value	$T_3 = \max(x(n)) - \min(x(n))$
Kurtosis	$T_4 = \frac{\sum_n (x(n) - T_1)^4}{NT_2^2}$
Mean Absolute Difference	$T_5 = \frac{\sum_n x(n) - x(n-1) }{N-1}$
Mobility	$T_6 = \sqrt{\frac{\sum_n [x(n) - x(n-1)]^2}{T_2}}$
Standard Deviation of Residual with $x_s(n) = \text{meanfilter}(x(n))$	$T_7 = \sqrt{\frac{1}{N} \sum_n (x(n) - x_s(n))^2}$
Autocorrelation Peak-to-Peak Value with $R_{xx}(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k)$	$T_8 = \max(R_{xx}(k)) - \min(R_{xx}(k))$

3.1.2. Definition of Frequency-Domain Features

Frequency domain features, obtained through Fourier transform and power spectral density analysis of PMU data, as shown in Table 2. Spectral center and distribution features (F_1, F_2) describe the dominant frequency position and bandwidth of the frequency distribution; meanwhile, spectral peakedness features (F_3, F_4) characterize the sharpness and central tendency of the spectral shape, and spectral entropy features (F_5, F_6) measure the randomness and distribution uniformity of spectral components.

Table 2. Summary of Frequency-Domain Features.

Feature Name	Expression
Mean Frequency	$F_1 = \frac{\sum_k X(k) \cdot f(k)}{\sum_k X(k) }$
Frequency Spread	$F_2 = \sqrt{\frac{\sum_k (f(k) - F_1)^2 \cdot X(k) }{\sum_k X(k) }}$
Frequency Peak Ratio	$F_3 = \frac{\max(X(k))}{\frac{1}{N} \sum_k X(k) }$
Frequency Domain Kurtosis	$F_4 = \frac{\frac{1}{N} \sum_k (X(k) - \bar{X})^4}{\left(\frac{1}{N} \sum_k (X(k) - \bar{X})^2\right)^2}$
Spectral Entropy	$F_5 = -\sum_k \frac{ X(k) }{\sum_k X(k) } \log_2 \frac{ X(k) }{\sum_k X(k) }$
Power Spectral Density Entropy	$F_6 = -\sum_k \frac{P(k)}{\sum_k P(k)} \log_2 \frac{P(k)}{\sum_k P(k)}$

Note: The discrete Fourier transform of $x(n)$ is expressed as $X(k)$, where k is the frequency index. $P(k)$ is the power spectral density, and $f(k)$ is the frequency value.

3.1.3. Definition of Energy Features

Energy features capture the intensity and fluctuation characteristics of oscillatory signals by analyzing the energy distribution and dynamic variations in the signal, as shown in Table 3. For example, total energy (E_1) directly reflects the overall oscillation intensity or power level of the signal. And energy entropy (E_2) measures the randomness or concentration degree of energy distribution over the time series. Meanwhile short-time energy standard deviation (E_3) analyzes the fluctuation of local energy through short-time Fourier transform.

Table 3. Summary of Energy Features.

Feature Name	Expression
Total Energy	$E_1 = \sum_{n=1}^N x(n)^2$
Energy Entropy	$E_2 = -\sum_n p_E(n) \cdot \log_2 p_E(n)$ with $p_E(n) = x(n)^2 / E_1$
Short-Time Energy Std.	$E_3 = \sqrt{\frac{1}{T} \sum_t (E_{\text{short}} - \bar{E}_{\text{short}})^2}$ with $E_{\text{short}}(t) = \sum_k X(k, t) ^2$ and $\bar{E}_{\text{short}} = \text{mean}(E_{\text{short}}(t))$.

Finally, by extracting the aforementioned time-domain, frequency-domain, and energy features from the PMU data, the prior knowledge features F_k , which characterize low-frequency oscillations, can be obtained in this study.

3.2. DL Features Extraction

To compensate for the limitations of prior knowledge features stemming from the limited understanding of low-frequency oscillations, a three-layer DWT-CNN network is constructed to extract DL features, as illustrated in Figure 3, where the tensor dimensions at each stage are explicitly marked. Specifically, within each DWT-CNN layer, a single-level 1D DWT is employed to decompose the signal $x(n)$ into a low-frequency approximation component A_l and a high-frequency detail component D_l , expressed as

$$(A_l, D_l) = \text{DWT}(F_l), l = 0, 1, 2, \quad (4)$$

where $F_0 \in \mathbb{R}^{1 \times S}$ denotes the original input signal with the signal length S , and $F_1 \in \mathbb{R}^{16 \times S/2}$, $F_2 \in \mathbb{R}^{32 \times S/4}$ are the intermediate features.

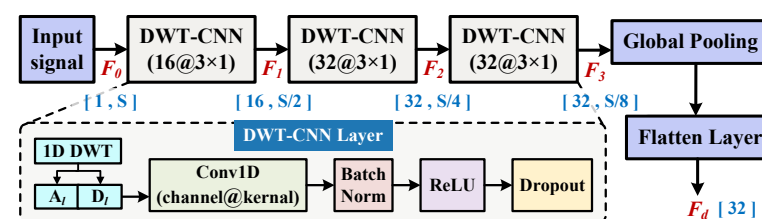


Figure 3. Detailed architecture of the proposed DWT-CNN feature extractor. Tensor dimensions are denoted as [Channel, Signal Length], and [Number of Features] for the final output F_d .

After concatenating (A_l, D_l) along the horizontal direction into a vector, a CNN operation is performed in the wavelet domain to extract the intermediate features F_{l+1} , as shown in the dashed box in Figure 3. The features F_{l+1} can be expressed as

$$F_{l+1} = \text{ReLu}(\text{BN}(\text{Conv}_1(\text{Concat}(A_l, D_l)))) , l = 0, 1, 2, \quad (5)$$

where $\text{ReLu}(\cdot)$ denotes the ReLu activation function, which accelerates the convergence speed of network training; $\text{BN}(\cdot)$ represents the batch normalization operation; $\text{Conv}_1(\cdot)$ refers to the 1D convolution operation.

Following three DWT-CNN layers, the DL features $F_d \in \mathbb{R}^{1 \times 32}$ are extracted from the features $F_3 \in \mathbb{R}^{32 \times 5/8}$ through global average pooling (GAP), flattening, and a fully connected layer (FC), expressed as

$$F_d = \text{FC}(\text{Flatten}(\text{GAP}(F_3))), \quad (6)$$

where $\text{Flatten}(\cdot)$ denotes the operation of reshaping the pooled features into a vector.

3.3. Classification with SVM

Utilizing the previously extracted knowledge features and DL features, we fuse them into the combined feature set F_m . Considering the potential redundancy within F_m , we apply Principal Component Analysis (PCA) [41] for feature selection.

The selected features F_a from F_m are fed into the SVM for oscillation classification, where the SVM model is derived by optimizing Equation (3). To construct nonlinear decision surfaces, the training data are projected into a high-dimensional space via the mapping function $\phi(\cdot)$ in (3). To avoid the high computational cost of calculating $\phi(x_i, x_j)$ in this high-dimensional space, a positive definite kernel function $\kappa(x_i, x_j)$ is introduced, defined as

$$\phi(x_i, x_j) = \kappa(x_i, x_j), \quad (7)$$

In this work, $\kappa(x_i, x_j)$ is selected as the radial basis function (RBF), expressed as

$$\kappa(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad (8)$$

where γ denotes the kernel width, which directly affects the classification performance of the SVM. The decision function for the nonlinear decision surfaces is then given by

$$f(x) = \text{sign}(\sum_{i \in S} \alpha_i y_i \kappa(x_i, x) + b) \quad (9)$$

where α_i are the Lagrange multiplier, S represents the index set of support vectors, and the bias term b is determined by the support vectors that satisfy $0 < \alpha_i < C$. If $f(x) > 0$, the sample is classified into the forced oscillation; otherwise, it is assigned to the natural oscillation.

3.4. Network Training

To train the DWT-CNN network, we utilize the cross-entropy function as the loss function $\mathcal{L}(\cdot)$, defined as

$$\mathcal{L}(\theta) = -\frac{1}{M} \sum_{i=1}^M \sum_{c=1}^2 (p_c(x_i) \log(q_c(\theta, x_i))), \quad (10)$$

where M denotes the number of training samples x_i , $p_c(x_i)$ and $q_c(\theta, x_i)$ are the true and predicted probabilities of sample x_i belonging to the c -th class, respectively; and θ represents the parameters of the DWT-CNN network, which are updated during the training process.

Considering the impact of the SVM hyperparameters C in (3) and γ in (8) on classification performance, these parameters are further optimized via a grid search strategy combined with stratified grouped cross-validation. The average classification loss is expressed as

$$\mathcal{L}_s(C, \gamma) = \frac{1}{V} \sum_{i=1}^V (1 - \frac{1}{2} \sum_{c=1}^2 \frac{TP_{c,i}}{TP_{c,i} + FN_{c,i}}), \quad (11)$$

where V denotes the number of folds in cross-validation, where $TP_{c,i}$ and $FN_{c,i}$ are the number of true positives and false negatives for the c -th class in the i -th validation subset.

Finally, the satisfied parameter pair (C^*, γ^*) that minimizes the cross-validation loss is selected. In this stage, the parameters of the trained DWT-CNN network are frozen. The complete training algorithm for the proposed classification task is presented in Algorithm 1.

Algorithm 1 The Training Process of HKD-SVM Network.

Input: PMU data F_0 with labels, batch size B , maximum training epochs E , initial learning rate r .

Stage 1: Feature Extraction

1. Extract the prior knowledge features F_k .
2. Initialize the DWT-CNN network parameters θ .
3. **for** $epoch = 1$ **to** E **do**
 - a. **Forward Propagation:** Extract DL features F_d , and compute the loss $\mathcal{L}(\theta)$ according to Equation (10).
 - b. **Backward Propagation:** Update parameters θ and adjust the learning rate r : $\theta \leftarrow \theta - r \cdot \nabla_{\theta} \mathcal{L}(\theta)$.
- end for**

Stage 2: SVM parameter optimization

1. Freeze θ , fuse F_d with F_k to obtain the selected features F_a , and perform PCA.
2. Search for the satisfied SVM hyperparameters (C^*, γ^*) .
3. Save the final SVM model.

Output: Detection results of SVM for oscillation.

4. Experimental Results and Discussion

This section evaluates the proposed method using simulated and real-world PMU data. We demonstrate its effectiveness through ablation studies and comparisons with baseline models, particularly in small-sample scenarios. Additionally, we analyze the model's robustness, computational efficiency, and the selection of key hyperparameters, such as the DWT wavelet basis and SVM settings.

4.1. Experimental Setup

4.1.1. Parameter Setting

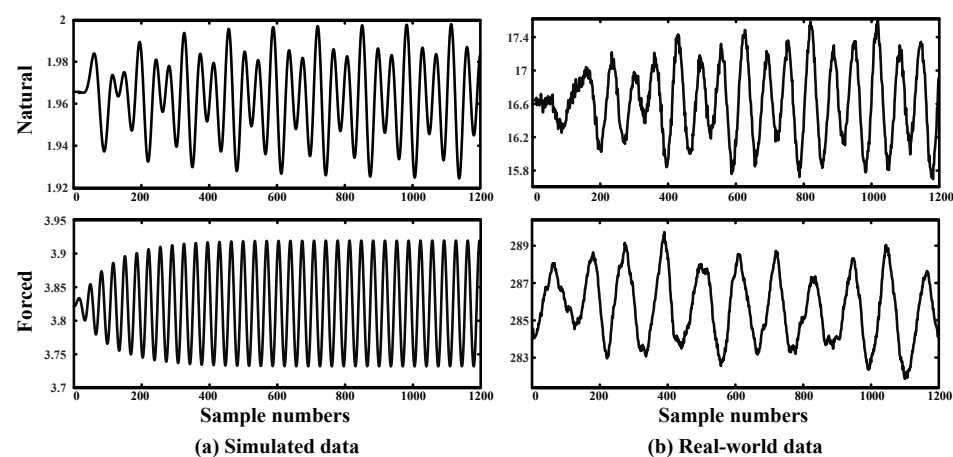
The software development environment for this experiment is the Python programming language, relying heavily on two core frameworks: PyTorch 2.8.0 and scikit-learn 1.6.1. The detailed experimental parameter settings are summarized in Table 4. The Daubechies 4 wavelet (db4) is selected as the wavelet basis for DWT. During the DL feature extraction process, the batch size is set to 256, and the model is trained for 50 epochs. AdamW is adopted as the optimizer, combined with a OneCycleLR learning rate scheduling strategy to enhance training efficiency. To mitigate overfitting, a progressive dropout strategy is implemented, utilizing a ratio of 0.1 at the initial layer to preserve fundamental signal features, and increasing to 0.3 in subsequent layers to regularize abstract representations when dealing with limited and noisy data. Prior to classification, PCA is employed for dimensionality reduction, with the cumulative variance contribution rate threshold (η) set to 0.95 to retain essential feature information while eliminating redundancy. In the SVM classification stage, the parameter pair (C, γ) is set to (100, 0.1) through a grid search with 10-fold cross-validation.

Table 4. Detailed experimental parameter settings.

Module	Parameter	Setting/Value
DWT-CNN Model	Batch Size	256
	Training Epochs	50
	Optimizer	AdamW
	Initial Learning Rate	5×10^{-4}
	Progressive Dropout Ratio	0.1 (Initial) $\rightarrow$ 0.3 (Final)
SVM Classifier	η of PCA	0.95
	C in (3)	100
	γ in (8)	0.1

4.1.2. Data Generation

The dataset used in this experiment is constructed from simulated data and real-world PMU data. As shown in Figure 4a, the simulated data are generated via TSAT from the WECC 179-bus power system model [42], using 40-s time-domain simulations with an integration step size of 0.25 cycles and an output sampling rate of 30 Hz. The real-world PMU data are collected from the ISO New England (ISO-NE) system [43], including 6 actual oscillation events. These events were primarily triggered by typical grid disturbances, such as equipment issues and generator anomalies within and outside the ISO-NE area. Representative oscillation signals are shown in Figure 4b. The PMU data records for each event last between 180 and 1200 s. This diversity in triggers, source locations, and durations provides a representative real-world basis for evaluating the proposed model. Since the oscillation characteristics are well reflected in the current signals, we utilize the current signals to construct the sample sets. Each sample is uniformly configured to contain 1200 sampling points via time-window segmentation, and the datasets are detailed in Table 5.

**Figure 4.** Current amplitude curves of natural and forced oscillations.**Table 5.** Overview and sample distribution of datasets utilized.

Category	Cases Quantity	Oscillation Type	Train/Test Samples	Total Signals
Simulated	9	Natural	3682/1052	4734
	18	Forced	7034/2434	9468
Real-World	6	Natural	385/110	495
		Forced	431/160	591

To strictly prevent data leakage, an event-wise splitting strategy is adopted, where the dataset is divided into training and test sets at a 3:1 ratio based on independent events. All data transformations, including PCA fitting and hyperparameter optimization, are performed exclusively on the training set. To ensure the reliability of the results, all experiments are conducted through 50 independent runs with consistent parameter settings. The randomness across these runs is attributed to the stochastic initialization of network weights and the mini-batch shuffling process during training, which allows for a robust assessment of model stability.

4.1.3. Evaluation Metrics

To comprehensively reflect the performance of the network, we employ Accuracy, Precision, Recall, and F1 Score, which are calculated as

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+FP+TN+FN} & \text{Recall} &= \frac{TP}{TP+FN} \\ \text{F1 Score} &= \frac{2 \cdot \text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}} & \text{Precision} &= \frac{TP}{TP+FP} \end{aligned} \quad (12)$$

where TP, FP, TN, and FN denote the numbers of true positives, false positives, true negatives, and false negatives.

4.2. Ablation Analysis

4.2.1. Contribution of Feature Extraction Modules

To verify the effectiveness of fusing DL features and knowledge features in the proposed HKD-SVM network, we construct oscillation detection models using only DL features or only knowledge-based features, namely DWT-CNN-SVM and Knowledge-SVM. Additionally, to independently evaluate the roles of DWT and CNN in the deep feature extractor, we construct variants of the proposed HKD-SVM with DWT or CNN removed, namely CNN-SVM and DWT-SVM. Thus, five models are used in total:

- HKD-SVM (Ours): Fused DL and knowledge features.
- DWT-CNN-SVM: Only DL features without knowledge ones.
- Knowledge-SVM: Only knowledge features without DL ones.
- CNN-SVM: DL features extracted via CNN without DWT.
- DWT-SVM: DL features extracted via DWT without CNN.

It should be noted that in the experiments, hyperparameter settings, as specified in Section 4.1.1, and data partitioning strategy were kept completely consistent. To ensure a fair comparison, only the feature extraction method was modified. The SVM hyperparameters for each model were re-tuned independently, using the same grid search strategy applied to HKD-SVM. This allows each model to be evaluated under comparable conditions.

Figure 5 shows the accuracy convergence curves of each method during the training process, where Knowledge-SVM does not have such a curve due to the absence of the DL process. HKD-SVM exhibits favorable convergence behavior and accuracy, reflecting the advantages in optimization efficiency and training stability brought by prior knowledge guidance and deep feature extraction via DWT-CNN. However, its convergence speed is initially slower than that of CNN-SVM, which may be attributed to the fact that DWT elevates feature dimensionality and complexity, requiring the model to undergo more iterations to learn the fusion mapping. Nevertheless, relying solely on deep temporal or convolutional learning without prior knowledge constraints makes it difficult to achieve robust and stable training performance.

The confusion matrices further validate the aforementioned conclusions, as shown in Figures 6 and 7. On real-world data, Knowledge-SVM exhibits a large number of unidirectional misclassifications, while DWT-CNN-SVM suffers from notable bidirectional

misclassifications. In contrast, HKD-SVM yields fewer cross-class misclassified samples, with its confusion matrix heavily dominated by the main diagonal, demonstrating effective inter-class discriminative capability. The statistical mean and standard deviation of the test results further validate this, as shown in Table 6.

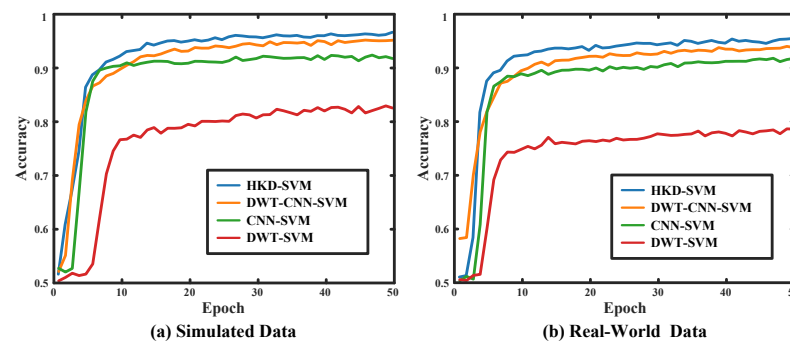


Figure 5. Accuracy curves of different feature extraction methods on simulated data and real-world data.

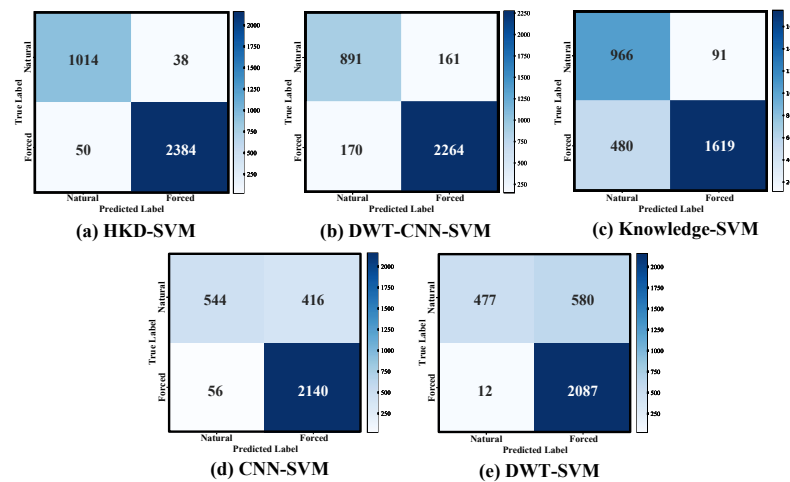


Figure 6. Confusion matrix of different feature extraction methods on simulated data.

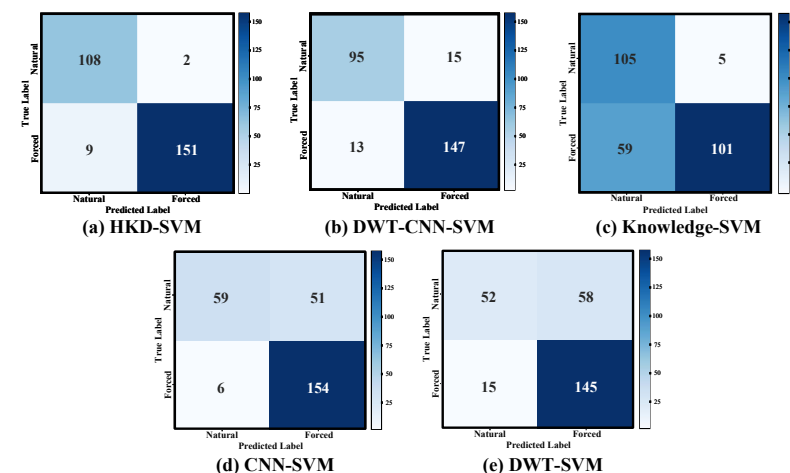


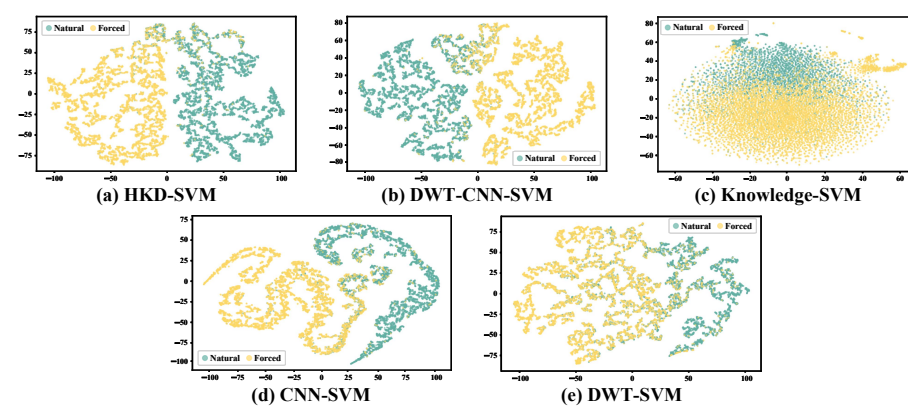
Figure 7. Confusion matrix of different feature extraction methods on real-world data.

In addition, we observed from Table 6 that the proposed method achieves favorable performance on both datasets. The performance degradation of the baseline methods is particularly pronounced on real-world data, indicating that HKD-SVM exhibits stronger adaptability to practical measurements. The integration of knowledge and DL features enables it to maintain stable and high accuracy even in complex data scenarios.

Table 6. Performance of different feature extraction methods on simulated and real-world data.

Method	Accuracy (%)		Precision (%)		Recall (%)		F1 Score (%)	
	Simulated	Real World	Simulated	Real World	Simulated	Real World	Simulated	Real World
DWT-CNN-SVM	94.45 ± 1.80	90.30 ± 0.54	95.35 ± 1.12	88.41 ± 0.75	92.29 ± 2.86	88.74 ± 0.83	93.53 ± 2.19	88.55 ± 0.62
Knowledge-SVM	81.29 ± 0.87	73.26 ± 0.20	80.35 ± 0.56	75.12 ± 0.07	83.90 ± 0.51	79.59 ± 0.11	80.53 ± 0.81	72.60 ± 0.18
CNN-SVM	84.35 ± 0.99	83.39 ± 0.04	84.74 ± 1.55	83.71 ± 0.12	81.86 ± 0.23	80.59 ± 0.38	83.27 ± 0.87	82.12 ± 0.14
DWT-SVM	81.23 ± 0.08	77.48 ± 0.36	87.56 ± 0.61	74.06 ± 0.66	72.36 ± 0.07	69.03 ± 0.19	74.71 ± 0.06	70.49 ± 0.28
HKD-SVM (Ours)	96.57 ± 0.78	95.50 ± 1.19	96.58 ± 0.77	95.92 ± 0.95	97.14 ± 0.33	95.59 ± 1.77	96.55 ± 0.81	95.58 ± 1.15

Additionally, as shown in Figure 8, we provide a 2D visualization of features derived from real-world data for different methods using t-SNE. It can be observed that the features extracted by HKD-SVM exhibit relatively compact intra-class aggregation and clear inter-class separation boundaries.

**Figure 8.** T-SNE distribution maps with two-dimensional features extracted from different feature extraction methods on real-world data.

4.2.2. Sensitivity Analysis of Network Architecture

To investigate the effect of the network architecture, we evaluated multiple depth and width configurations, as summarized in Table 7. The results indicate that the 1-layer configuration is insufficient for capturing complex oscillatory dynamics, leading to an accuracy drop of 10.71%. Conversely, widened 3-layer variants or deeper 5-layer models provide small gains in accuracy but incur disproportionate increases in computational costs. For instance, while the widened 3DWT-CNN [16-32-64] improves accuracy by less than 1%, it increases model parameters and FLOPs by approximately 30% and 47%, respectively, alongside a noticeable rise in inference latency. Similarly, deeper 5-layer models substantially increase the computational burden without corresponding accuracy benefits. Therefore, given the performance requirements of our task, the baseline 3-layer DWT-CNN provides a favorable balance between detection accuracy and computational efficiency for this specific application scenario.

Table 7. Architecture sensitivity analysis in terms of detection performance and computational efficiency.

Architecture	Accuracy (%)	FLOPs (M)	Latency (ms)	Params (M)	Δ Acc. (%)	Δ Lat. (%)
1DWT-CNN [16]	85.90	0.12	1.89	0.0144	−10.71	−21.96
3DWT-CNN [16-32-32] (Ours)	96.61	2.07	2.43	0.0243	-	-
3DWT-CNN [16-32-64]	97.53	3.05	3.08	0.0316	+0.92	+26.94
5DWT-CNN [16-32-64-64-128]	95.72	7.25	3.52	0.1079	−0.89	+44.95
5DWT-CNN [16-32-64-128-256]	96.36	15.78	4.22	0.2846	−0.25	+73.96

Note: The values in brackets represent the number of channels for each layer. Δ Lat. denotes the relative change in inference latency compared to our proposed method.

4.2.3. Impact of Prior Knowledge Feature Groups

To assess the specific contribution of individual knowledge feature groups, we performed a component-wise ablation study. In these experiments, the DL features were fused sequentially with only one category of prior knowledge features. The results summarized in Table 8 indicate that the full fusion model generally achieves superior performance compared to any single feature subset, suggesting that each physical domain provides complementary information for oscillation detection.

Table 8. Detection accuracy (%) of HKD-SVM with different knowledge feature subsets, including time-domain features, frequency-domain features and energy features.

Feature Group	Simulated Data	Real-World Data
Time-Domain	96.21 ± 0.62	93.59 ± 0.53
Frequency-Domain	96.14 ± 0.85	93.65 ± 1.08
Energy	95.11 ± 1.32	92.68 ± 0.79
Time + Freq + Energy	96.57 ± 0.78	95.50 ± 1.19

Furthermore, the feature importance analysis in Figure 9 reveals that features derived from domain-specific prior knowledge are critical to the classification decisions. Specifically, the Frequency Peak Ratio (F_3) and kurtosis (T_4) rank as the primary contributors, followed by Frequency Domain Kurtosis (F_4) and energy features. From an engineering perspective, F_3 captures the sharpness of the spectral peak, a characteristic indicator of forced oscillations excited by periodic disturbances. In the time domain, T_4 evaluates the impulsiveness and non-Gaussian nature of the signal, making it sensitive to the transient spikes and amplitude variations associated with grid anomalies. Additionally, energy features such as Energy Entropy (E_2) provide valuable supplementary information by quantifying energy concentration, which helps distinguish sustained forced oscillations from decaying natural modes.

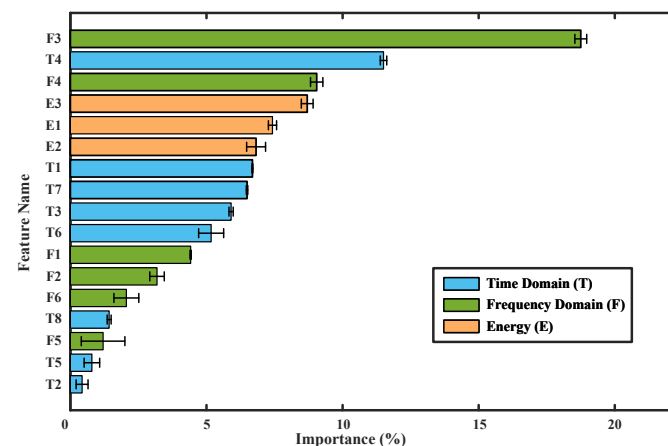


Figure 9. Feature importance analysis of the extracted prior knowledge features, where each feature's contribution is normalized to a percentage. Error bars represent the standard deviation across independent experimental runs. Features are grouped into time-domain (T), frequency-domain (F), and energy (E) categories.

4.2.4. Effectiveness of SVM

To verify the effectiveness of SVM as a classifier in the extracted feature space, we compared it against alternative classifiers, including Random Forest (HKD-RF), Softmax (HKD-Softmax), and XGBoost (HKD-XGBoost).

As shown in Table 9 and Figure 10, HKD-RF and HKD-XGBoost show a marginal performance gap compared to SVM on simulated data, but suffer from significant per-

formance degradation on real-world data. This is likely because tree-based models are more sensitive to noise and distribution shifts in real data, tending to overfit the training distribution. Although HKD-Softmax performs reasonably well on simulated data, it lacks the robust margin-maximization capability of SVM in high-dimensional feature spaces. As seen from the confusion matrix, the number of misclassified samples for HKD-Softmax is noticeably larger than that of HKD-SVM. This further confirms that SVM offers superior stability and generalization in nonlinear classification tasks when dealing with complex, non-ideal measurements.

Table 9. Performance of different classifiers on simulated and real-world data.

Method	Accuracy (%)		Precision (%)		Recall (%)		F1 Score (%)	
	Simulated	Real World	Simulated	Real World	Simulated	Real World	Simulated	Real World
HKD-RF	93.16 $\pm$ 1.12	89.09 $\pm$ 1.84	93.42 $\pm$ 0.91	89.46 $\pm$ 1.06	92.58 $\pm$ 1.52	88.86 $\pm$ 2.31	93.00 $\pm$ 1.21	89.16 $\pm$ 1.67
HKD-Softmax	91.76 $\pm$ 2.04	90.75 $\pm$ 0.87	92.41 $\pm$ 1.71	90.67 $\pm$ 0.90	90.51 $\pm$ 2.68	90.63 $\pm$ 0.90	91.45 $\pm$ 2.20	90.65 $\pm$ 0.92
HKD-XGBoost	93.97 $\pm$ 1.44	89.83 $\pm$ 0.79	94.18 $\pm$ 1.26	90.40 $\pm$ 0.57	93.52 $\pm$ 1.79	89.51 $\pm$ 1.16	93.85 $\pm$ 1.52	89.95 $\pm$ 0.72
HKD-SVM (Ours)	96.57 $\pm$ 0.78	95.50 $\pm$ 1.19	96.58 $\pm$ 0.77	95.92 $\pm$ 0.95	97.14 $\pm$ 0.33	95.59 $\pm$ 1.77	96.55 $\pm$ 0.81	95.58 $\pm$ 1.15

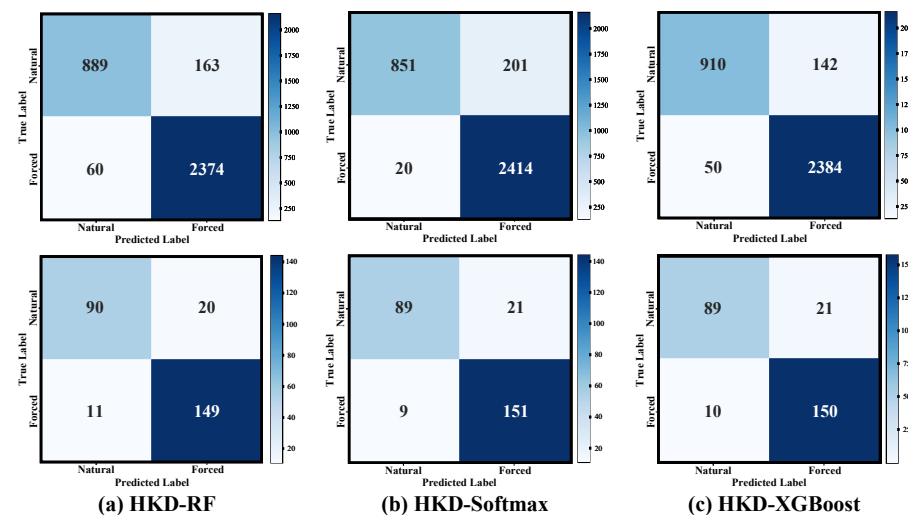


Figure 10. Confusion matrix of different classifiers. The top row illustrates the classification performance on simulated data, while the bottom row demonstrates the performance on real-world data.

4.3. Comparison Experiments

To further demonstrate the effectiveness of the proposed HKD-SVM framework, we comprehensively compared it with several recent methods. The comparative baseline models include: (1) SRNN-LSTM [29], a hybrid spiking recurrent neural network; (2) DWT-DiCNN-MAML [26], a discrete wavelet transform and dilated CNN network; (3) TCN-SVM [44], a temporal convolutional network approach optimized for transient stability; (4) CNN-Transformer [25], an advanced architecture utilizing transfer learning; and (5) LSTM. All models were evaluated under the identical event-wise data splitting strategy to ensure strict fairness.

As shown in Table 10, while most purely data-driven models achieve satisfactory performance on simulated data, they generally suffer a noticeable accuracy degradation when evaluated on real-world data. This degradation implies that these models are prone to overfitting the source distribution and lack sufficient engineering constraints to generalize in complex practical grids. In contrast, HKD-SVM consistently outperforms the comparative approaches. However, it is worth noting that HKD-SVM exhibits a marginally higher standard deviation on real-world data compared to certain methods, such as DWT-DiCNN-MAML, indicating a slight sensitivity to real-world noise variance across different folds.

Nevertheless, by integrating the prior knowledge, HKD-SVM establishes a robust physical boundary that effectively prevents large-scale misclassifications and ensures superior overall generalization.

Table 10. Performance of different detection approaches on simulated and real-world data.

Method	Accuracy (%)		Precision (%)		Recall (%)		F1 Score (%)	
	Simulated	Real-World	Simulated	Real-World	Simulated	Real-World	Simulated	Real-World
SRNN-LSTM [29]	83.37 ± 1.84	82.32 ± 3.55	85.02 ± 2.24	82.32 ± 3.99	79.14 ± 1.82	80.4 ± 2.46	81.98 ± 2.02	81.34 ± 3.14
DWT-DiCNN-MAML [26]	91.46 ± 0.34	91.12 ± 0.87	91.91 ± 0.24	91.09 ± 0.97	90.49 ± 0.50	90.80 ± 0.71	91.19 ± 0.37	90.95 ± 0.84
TCN-SVM [44]	83.10 ± 0.07	82.57 ± 0.06	85.07 ± 0.08	82.82 ± 0.10	78.34 ± 0.08	79.59 ± 0.01	81.57 ± 0.08	81.17 ± 0.05
CNN-Transformer [25]	93.50 ± 1.60	90.72 ± 1.03	94.06 ± 1.34	91.20 ± 0.68	92.45 ± 2.12	89.62 ± 1.65	93.25 ± 1.74	90.40 ± 1.17
LSTM-SVM	92.26 ± 1.92	89.78 ± 1.27	92.44 ± 1.86	90.21 ± 1.17	91.75 ± 2.22	89.54 ± 1.40	92.09 ± 2.04	89.87 ± 1.25
HKD-SVM (Ours)	96.57 ± 0.78	95.50 ± 1.19	96.58 ± 0.77	95.92 ± 0.95	97.14 ± 0.33	95.59 ± 1.77	96.55 ± 0.81	95.58 ± 1.15

Furthermore, the accuracy curves in Figure 11 reveal the distinct training dynamics of the comparative methods. While models like SRNN-LSTM and TCN-SVM experience a significant delay in initial learning (remaining stagnant for the first 10–15 epochs), and pure DL models frequently exhibit validation fluctuations, HKD-SVM converges rapidly to a stable, high-level plateau. Although its real-world accuracy curve displays microscopic local jitter at the plateau phase compared to its smoother simulated counterpart, this trade-off is acceptable. Ultimately, this confirms that prior-knowledge features can effectively narrow the optimization search space of neural networks, alleviating slow convergence in the initial training stage and improving overall training efficiency and stability.

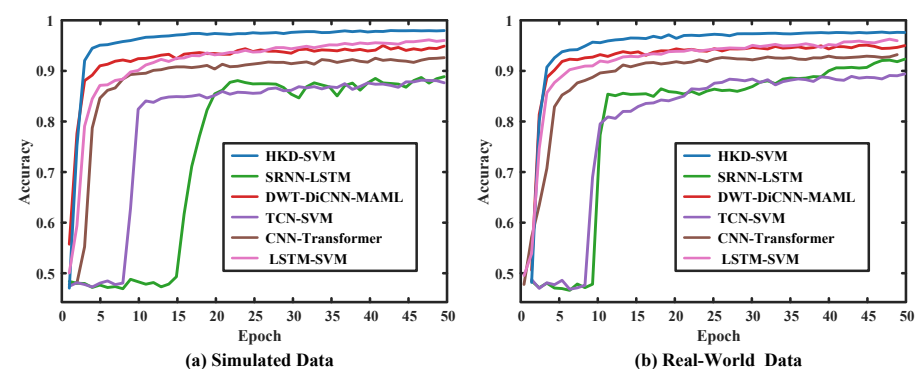


Figure 11. Accuracy curves of different detection approaches on simulated and real-world data.

4.4. Robustness and Efficiency Analysis

Having established the fundamental classification performance of the proposed HKD-SVM, its internal ablation variants, and external comparative baselines, this section comprehensively evaluates all these models from three advanced perspectives: statistical reliability, robustness under data scarcity, and computational efficiency.

4.4.1. Statistical Significance Analysis

To rigorously evaluate the statistical robustness of the performance improvements, we conducted a Wilcoxon signed-rank test [45] on the testing accuracy obtained from 50 independent experimental runs. The analysis was formulated under the alternative hypothesis that the proposed HKD-SVM consistently yields higher accuracy than the comparative models. Table 11 summarizes the corresponding statistical significance. The results consistently yield p -values less than 0.01 across all comparisons, labeled as H. This formal verification, supported by the statistical test, confirms that the superior performance of HKD-SVM is robust and statistically distinguishes our model from all considered baseline approaches.

Table 11. Statistical significance analysis (Wilcoxon signed-rank test) of Test Accuracy.

Comparison Model	Median Diff. (%)	Statistic (W)	p-Value	Significance
HKD-SVM vs. Ablation Study Variants				
DWT-CNN-SVM	4.53	1275.0	<0.01	H
Knowledge-SVM	22.29	1275.0	<0.01	H
CNN-SVM	11.19	1275.0	<0.01	H
DWT-SVM	17.56	1275.0	<0.01	H
HKD-SVM vs. Comparative Baselines				
SRNN-LSTM	12.60	1275.0	<0.01	H
DWT-DiCNN-MAML	3.69	1275.0	<0.01	H
TCN-SVM	12.26	1275.0	<0.01	H
CNN-Transformer	1.42	1099.0	<0.01	H
LSTM-SVM	5.15	1275.0	<0.01	H

Note: The Wilcoxon signed-rank test is conducted under the alternative hypothesis that the proposed HKD-SVM achieves a higher test accuracy than the comparative baselines based on 50 independent runs. Median Diff. (%) is calculated as $(\text{Accuracy}_{HKD} - \text{Accuracy}_{Baseline}) \times 100$. W denotes the Wilcoxon test statistic. H denotes Highly Significant, indicating $p < 0.01$.

4.4.2. Effectiveness in Small-Sample Scenarios

Considering the scarcity of labeled oscillation data in power systems, we validate the effectiveness of the proposed method under limited experimental data scenarios. This section designs two small-sample settings: utilizing only (a) 10% and (b) 5% of the available training samples. To fully reflect their inherent small-sample learning capabilities, no specific hyperparameter adjustments were made for the comparative methods under these constrained settings.

Figure 12 provides a visual comparison of the classification accuracy across three training sample ratios. It is observed that as the number of labeled samples decreases sharply, the performance of all methods degrades to varying degrees, indicating that data scarcity significantly limits their learning capabilities. In contrast, HKD-SVM exhibits a noticeably smaller performance decline. This demonstrates that integrating the prior knowledge can suppress overfitting to a certain extent and effectively exploit discriminative information from limited samples.

Furthermore, the training dynamics under limited-data scenarios are presented via the accuracy curves in Figure 13. Advanced data-driven architectures, such as CNN-Transformer and LSTM-SVM, exhibit noticeable validation fluctuations and instability when restricted to 5% or 10% of the training data. Meanwhile, complex temporal models like TCN-SVM and SRNN-LSTM suffer from slow initial convergence, remaining stagnant for up to 15–25 epochs before initiating any meaningful progress. It is worth acknowledging that methods explicitly designed for few-shot learning, such as DWT-DiCNN-MAML, exhibit commendable resilience and successfully avoid these prolonged initial delays. Nevertheless, the accuracy curve of HKD-SVM demonstrates highly stable training dynamics, rising rapidly and converging to a high-level plateau without severe oscillation. This confirms that integrating explicit prior knowledge provides a robust physical boundary that significantly reduces the neural network's search space, thereby avoiding initial learning delays, enhancing overall training stability, and mitigating the degradation typically observed in small-sample scenarios. Additionally, by incorporating domain-specific engineering signatures, the model becomes more sensitive to the distinct patterns of minority-class events, ensuring stable F1-score performance across skewed test sets even with limited data.

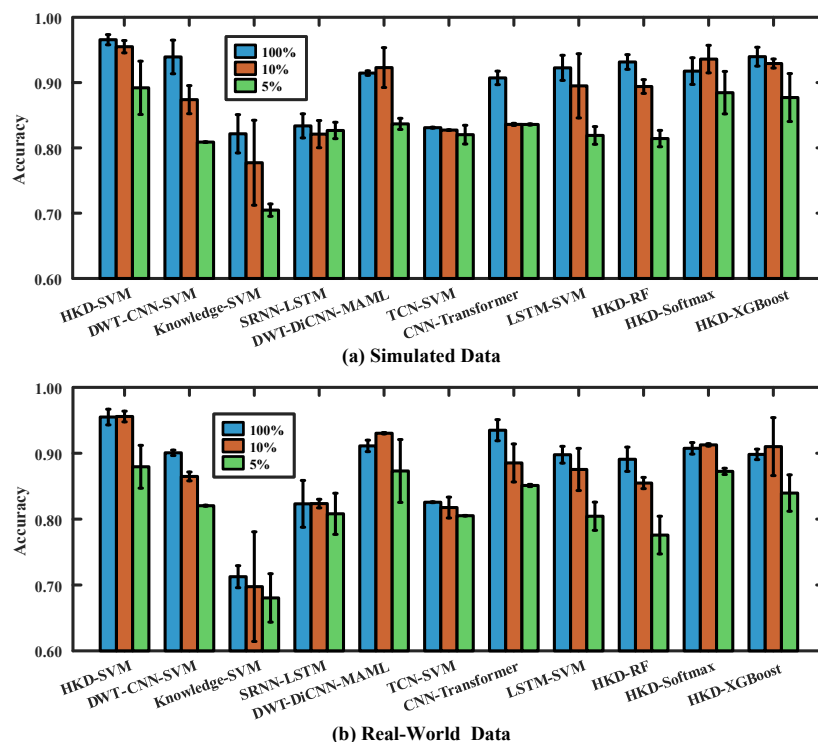


Figure 12. Classification accuracy of different methods under small-sample scenarios.

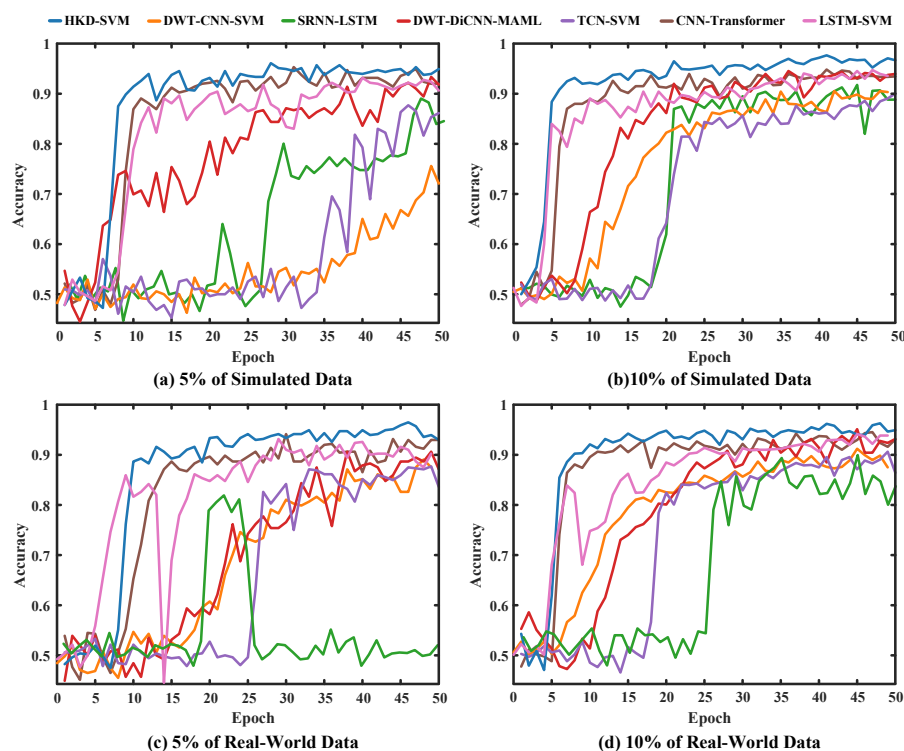


Figure 13. Training dynamics under limited-data constraints. The curves illustrate the accuracy convergence over 50 epochs for the proposed HKD-SVM and comparative baselines using 5% (a,c) and 10% (b,d) of the available simulated and real-world data.

4.4.3. Computational Efficiency Analysis

Considering that forced oscillation detection is fundamental to real-time situational awareness, we evaluated the computational efficiency of the HKD-SVM framework by comparing its computational complexity, training duration, and inference latency with

those of the baseline models. As summarized in Table 12, HKD-SVM requires a computational effort of 2.0704 MFLOPs per sample, which is lower than that of the CNN-SVM model. Although the total training time is longer than that of the HKD-Softmax, the inference latency of 2.46 ms per sample exhibits competitive efficiency. This performance suggests that the model is a viable candidate for high-speed power system monitoring, where decisions are often required within sub-second intervals. While these results validate the framework's real-time viability, further architectural refinements remain a potential direction for maintaining this level of efficiency under increasingly complex and resource-constrained conditions.

Table 12. Computational efficiency comparison of different detection models.

Model	FLOPs (M)	Training Time (s)	Latency (ms)
DWT-CNN-SVM	2.0580	428.56	3.22
Knowledge-SVM	0.0032	5.70	0.33
CNN-SVM	6.0241	1614.63	1.94
DWT-SVM	0.0225	1577.02	4.51
HKD-RF	2.0769	961.91	8.73
HKD-Softmax	2.0769	280.07	1.35
HKD-XGBoost	2.0769	982.23	4.60
SRNN-LSTM	8.2427	1157.86	62.95
DWT-DiCNN-MAML	3.0462	298.68	2.70
TCN-SVM	14.5989	393.59	3.10
CNN-Transformer	1.5670	601.33	1.90
LSTM-SVM	6.7973	1111.23	9.71
HKD-SVM (Ours)	2.0704	650.18	2.46

4.5. Selection of Key Parameters

The performance of the proposed network depends on the selection of key hyperparameters. In this section, parameter sensitivity experiments are conducted for three core components: wavelet basis, cumulative variance contribution rate threshold η of PCA, and penalty coefficient C and kernel coefficient γ of SVM.

4.5.1. Selection of Wavelet Basis

The basis function of DWT directly determines the signal decomposition characteristics across multiple scales. To investigate the most suitable wavelet basis for HKD-SVM, we compared eight commonly used wavelet bases on both simulated and real-world data: Daubechies (db2, db4, db8, db16), Symlets (sym4, sym8), and Coiflets (coif2, coif4). As observed in Figure 14a, db4 achieves high classification accuracy with low variance on both simulated and real-world data. Therefore, we ultimately select db4 as the basis function for DWT.

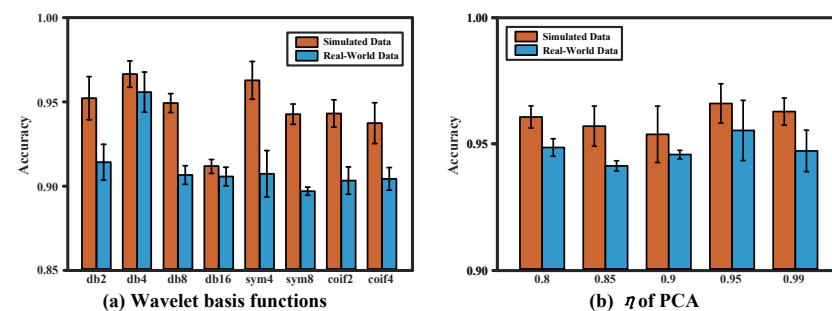


Figure 14. Parameter sensitivity analysis of the HKD-SVM framework on both datasets. The subplots evaluate the impact of (a) different wavelet basis functions, and (b) the cumulative variance contribution rate threshold (η) for PCA.

4.5.2. Necessity Analysis and Hyperparameter Selection

To validate the necessity of the PCA module, we first conducted an ablation study comparing the proposed HKD-SVM against a variant without PCA. As summarized in Table 13, the inclusion of PCA simultaneously improves detection accuracy and reduces computational costs across both simulated and real-world datasets. These results suggest that PCA serves as an effective pre-processing step to suppress redundant noise, which facilitates more stable classification performance.

Table 13. Ablation study of the PCA module on detection performance and computational efficiency.

Model	Data Type	Accuracy (%)	Training Time (s)	Latency (ms)
HKD-SVM w/o PCA	Simulated	95.22	768.94	3.60
	Real world	93.95	764.27	3.22
HKD-SVM (Ours)	Simulated	96.57	675.56	2.61
	Real world	95.50	650.18	2.46

Furthermore, we investigated the influence of the cumulative variance contribution rate η on the performance. The threshold η determines the balance between feature information retention and noise suppression. Five commonly used thresholds were selected for comparative experiments, and the results are presented in Figure 14b. Comparative experiments indicate that as η increases, the accuracy exhibits minor fluctuations. A low η threshold risks losing key discriminative information, while an excessively high η retains redundant dimensions that may hinder SVM generalization. Consequently, we set η to 0.95, which retains sufficient feature information while achieving stable performance. At this threshold, the feature space is reduced to 12 dimensions for simulated data and 17 dimensions for real-world data, reflecting the increased complexity of practical grid measurements.

4.5.3. Selection of SVM Hyperparameters

The penalty coefficient C and kernel coefficient γ of SVM are critical hyperparameters influencing the classification boundary and generalization ability. This section employs a grid search combined with 10-fold stratified cross-validation to systematically evaluate the performance across $C \in \{0.01, 0.1, 1, 10, 50, 100, 200, 500, 1000\}$ and $\gamma \in \{0.0001, 0.001, 0.01, 0.05, 0.1\}$. Through this systematic grid search and multiple experimental trials, a specific parameter combination was identified based on its superior stability, exhibiting the lowest standard deviation rather than merely focusing on absolute peak accuracy. As illustrated in Figure 15, this favorable combination occurs at $C = 100$ and $\gamma = 0.1$.

Figure 15b further validates that this configuration maintains high accuracy and stable performance on both simulated and real-world datasets. When γ is excessively small, the kernel function becomes overly smooth, which leads to underfitting and a decline in accuracy. Conversely, when γ is excessively large, the influence of support vectors becomes highly localized, causing performance degradation on real-world data due to overfitting. Similarly, an excessively small C provides insufficient penalty on training samples, resulting in an overly conservative boundary. Although a very large C may yield slight improvements on simulated data, it often shows signs of overfitting on real-world data, leading to a decline in generalization performance.

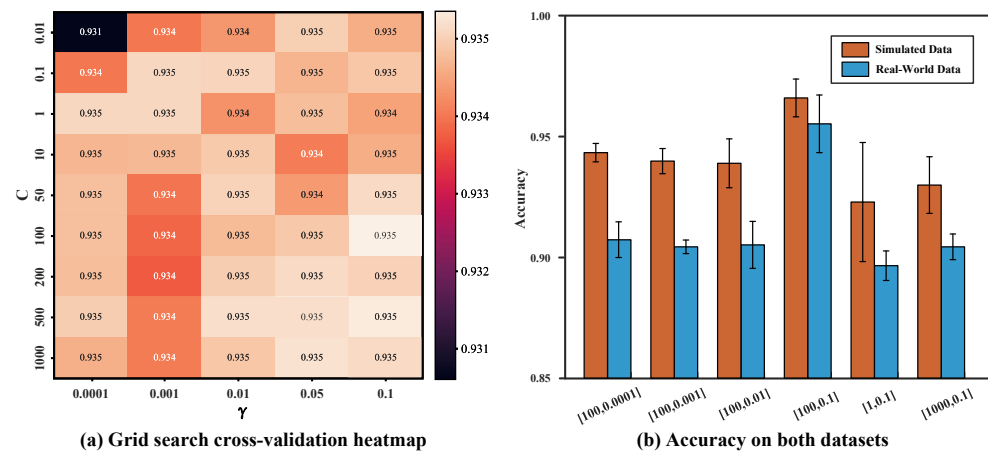


Figure 15. Optimization and sensitivity analysis of SVM hyperparameters. (a) Grid search cross-validation accuracy heatmap. (b) Classification accuracy on simulated and real-world datasets.

4.6. Discussion

4.6.1. Robustness on Different Noise Level

To evaluate the robustness of the proposed framework against measurement disturbances common in practical power systems, we assessed detection performance under additive white Gaussian noise, impulsive noise, and colored noise environments. The signal-to-noise ratio (SNR) levels were varied from 0 dB to 30 dB, with the detailed detection accuracies presented in Table 14. As anticipated, the detection accuracy of all models tends to decrease as the SNR drops. However, the HKD-SVM framework maintains more stable performance across the tested noise environments. When subject to the most severe low-SNR condition of 0 dB, the HKD-SVM exhibits a smaller performance decline compared to the baseline models. This resilience suggests that the integration of prior knowledge features contributes to a more robust decision boundary, which helps mitigate the sensitivity of purely data-driven feature extractors to non-ideal measurement conditions.

Table 14. Robustness comparison of detection accuracy across different noise types and SNR levels on simulated data.

Noise Type	Model	SNR			
		0	10	20	30
AWGN	DWT-CNN-SVM	73.17 ± 0.03	81.05 ± 7.21	87.09 ± 1.49	93.33 ± 0.56
	Knowledge-SVM	62.70 ± 2.34	70.89 ± 0.66	73.55 ± 1.56	79.70 ± 0.27
	HKD-RF	71.33 ± 2.00	74.25 ± 3.95	87.60 ± 2.82	93.04 ± 0.97
	HKD-Softmax	69.90 ± 0.75	71.27 ± 0.36	81.21 ± 3.39	89.52 ± 1.36
	HKD-XGBoost	72.83 ± 0.13	75.15 ± 1.06	88.64 ± 2.82	93.45 ± 1.57
	HKD-SVM (Ours)	78.81 ± 0.92	84.14 ± 4.27	89.27 ± 0.07	93.24 ± 0.26
Impulsive	DWT-CNN-SVM	73.29 ± 0.04	82.02 ± 1.84	92.73 ± 0.37	94.31 ± 0.17
	Knowledge-SVM	59.48 ± 1.91	73.56 ± 2.34	79.30 ± 0.29	80.82 ± 0.24
	HKD-RF	70.01 ± 2.61	80.67 ± 2.10	89.57 ± 0.52	92.33 ± 1.94
	HKD-Softmax	71.50 ± 1.45	77.30 ± 0.75	84.37 ± 4.63	90.30 ± 0.95
	HKD-XGBoost	71.76 ± 0.90	74.47 ± 3.13	92.33 ± 1.77	93.05 ± 1.39
	HKD-SVM (Ours)	73.00 ± 1.41	83.22 ± 0.47	92.42 ± 1.49	94.80 ± 0.89
Colored	DWT-CNN-SVM	73.17 ± 0.03	74.29 ± 0.16	86.21 ± 2.64	93.98 ± 0.30
	Knowledge-SVM	62.53 ± 0.14	71.68 ± 0.11	77.58 ± 2.42	80.19 ± 0.78
	HKD-RF	72.81 ± 0.04	79.21 ± 6.89	89.66 ± 1.92	93.14 ± 1.66
	HKD-Softmax	69.10 ± 1.76	71.64 ± 1.08	78.81 ± 5.43	90.82 ± 0.65
	HKD-XGBoost	72.91 ± 0.08	75.28 ± 5.61	91.23 ± 0.71	93.30 ± 1.20
	HKD-SVM (Ours)	73.45 ± 0.24	75.68 ± 0.67	90.63 ± 0.83	94.41 ± 1.23

4.6.2. Cross-Domain Generalization Analysis

To evaluate the generalization robustness of the proposed model against distribution shifts, a bidirectional cross-domain experiment was conducted, as shown in Table 15. The proposed HKD-SVM demonstrates a higher absolute accuracy in both transfer directions (Sim $\rightarrow$ Real and Real $\rightarrow$ Sim). While purely data-driven architectures tend to overfit when trained on limited real-world events, HKD-SVM leverages domain-specific prior knowledge to stabilize its decision boundaries. Furthermore, the purely knowledge-driven baseline (Knowledge-SVM) exhibited less performance degradation in the Real $\rightarrow$ Sim direction compared to Sim $\rightarrow$ Real. This suggests that features rooted in prior knowledge extracted from noisy real-world conditions generally transfer better to idealized simulated environments than the reverse.

Table 15. Comparison of detection accuracy (%) and performance degradation (Δ Acc.) across different approaches.

Method	Sim $\rightarrow$ Real		Real $\rightarrow$ Sim	
	Accuracy (%)	Δ Acc. (%)	Accuracy (%)	Δ Acc. (%)
DWT-CNN-SVM	88.41	−6.04	76.82	−13.48
Knowledge-SVM	60.31	−20.98	68.61	−4.65
HKD-RF	87.76	−5.40	71.77	−17.32
HKD-Softmax	89.15	−2.61	77.43	−13.32
HKD-XGBoost	87.58	−6.39	74.86	−14.98
SRNN-LSTM	76.35	−7.02	76.50	−5.82
DWT-DiCNN-MAML	88.13	−3.33	81.74	−9.39
TCN-SVM	81.32	−1.78	78.31	−4.26
CNN-Transformer	89.79	−3.71	78.02	−12.70
LSTM-SVM	82.33	−9.93	76.50	−13.28
HKD-SVM (Ours)	90.49	−6.08	83.50	−12.01

Nevertheless, certain limitations should be objectively noted. Despite yielding the best overall performance, HKD-SVM still experiences an inevitable performance degradation in the Real $\rightarrow$ Sim scenario. Although some baseline models display a smaller degradation (Δ Acc), this is largely attributable to their inherently lower in-domain baseline accuracy, which prevents them from capturing complex features in either domain. The observed degradation in HKD-SVM indicates that while the integration of the prior knowledge significantly reduces domain discrepancy, bridging the domain gap remains a challenge when real-world training samples are highly scarce.

5. Conclusions

The paper proposes a hybrid forced oscillation detection network, HKD-SVM, which integrates prior knowledge and DL networks. The synergy between knowledge-based features and DL-based feature extraction enhances explicit engineering significance and generalization capability, providing an effective solution for power system oscillation identification. Experimental results indicate that HKD-SVM consistently outperforms purely data-driven and purely knowledge-driven baselines on the current dataset, providing compelling preliminary evidence for its effectiveness in oscillation identification. For industrial deployment, integrating the proposed framework into wide-area measurement systems could provide system operators with a practical reference for rapid oscillation source identification, thereby assisting in timely decision-making for remedial control actions.

Future work will focus on three key areas. First, we intend to expand the dataset to include a wider range of grid topologies and disturbance types, further validating the

robustness of the proposed framework. Second, to accommodate time-critical industrial applications, we will explore model optimization and hardware acceleration techniques to enhance real-time deployment efficiency. Finally, we will investigate domain adaptation strategies to improve the model's resilience against severe measurement noise and varying operational conditions.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Wang, Y.; Guo, Y.; Ning, H.; Li, P.; Cen, B.; Zhao, H.; Zou, H. Research on Power Quality Control Methods for Active Distribution Networks with Large-Scale Renewable Energy Integration. *Processes* **2025**, *13*, 1469.
2. Wang, B.; Sun, K. Location Methods of Oscillation Sources in Power Systems: A Survey. *J. Mod. Power Syst. Clean Energy* **2017**, *5*, 151–159.
3. Feng, S.; Wu, X.; Jiang, P.; Xie, L.; Lei, J. Mitigation of Power System Forced Oscillations: An E-STATCOM Approach. *IEEE Access* **2018**, *6*, 31599–31608.
4. Trudnowski, D.J.; Guttromson, R. A Strategy for Forced Oscillation Suppression. *IEEE Trans. Power Syst.* **2020**, *35*, 4699–4708.
5. Ghorbaniparvar, M. Survey on Forced Oscillations in Power System. *J. Mod. Power Syst. Clean Energy* **2017**, *5*, 671–682.
6. Feng, S.; Jiang, P.; Wu, X. Suppression of Power System Forced Oscillations Based on PSS with Proportional-resonant Controller. *Int. Trans. Electr. Energy Syst.* **2017**, *27*, e2328.
7. Gonzalez-Jimenez, D.; Del-Olmo, J.; Poza, J.; Garramiola, F.; Madina, P. Data-Driven Low-Frequency Oscillation Event Detection Strategy for Railway Electrification Networks. *Sensors* **2023**, *23*, 254.
8. Liu, J.; Yao, W.; Wen, J.; He, H.; Zheng, X. Active Power Oscillation Property Classification of Electric Power Systems Based on SVM. *J. Appl. Math.* **2014**, *2014*, 218647.
9. Ye, H.; Liu, Y.; Zhang, P.; Du, Z. Analysis and Detection of Forced Oscillation in Power System. *IEEE Trans. Power Syst.* **2017**, *32*, 1149–1160.
10. Kim, B.; Park, J.; Cho, H.; Nam, S.; Lee, B. An Enhanced Geometric Featured Extraction Method for Real-Time Oscillation Detection in Power System. *IEEE Access* **2025**, *13*, 143223–143232.
11. Ghorbaniparvar, M.; Zhou, N.; Li, X.; Trudnowski, D.J.; Xie, R. A Forecasting-Residual Spectrum Analysis Method for Distinguishing Forced and Natural Oscillations. *IEEE Trans. Smart Grid* **2019**, *10*, 493–502.
12. Zhou, N.; Dagle, J. Initial Results in Using a Self-Coherence Method for Detecting Sustained Oscillations. *IEEE Trans. Power Syst.* **2015**, *30*, 522–530.
13. Khan, M.A.; Pierre, J.W. Detection of Periodic Forced Oscillations in Power Systems Using Multitaper Approach. *IEEE Trans. Power Syst.* **2019**, *34*, 1086–1094.
14. Yang, D.; Xing, L.; Wang, L.; Ye, H.; Liu, Y. Response Analysis and Type Discrimination of Power System Forced Oscillation. In Proceedings of the 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), Changsha, China, 26–29 November 2015; pp. 1346–1350.
15. Donnelly, M.; Trudnowski, D.; Colwell, J.; Pierre, J.; Dosiek, L. RMS-energy Filter Design for Real-Time Oscillation Detection. In Proceedings of the 2015 IEEE Power & Energy Society General Meeting, Denver, CO, USA, 26–30 July 2015; pp. 1–5.
16. Follum, J.; Holzer, J.; Etingov, P. A Statistics-Based Threshold for the RMS-energy Oscillation Detector. *Int. J. Electr. Power Energy Syst.* **2021**, *128*, 106685.
17. Zhou, N. A Cross-Coherence Method for Detecting Oscillations. *IEEE Trans. Power Syst.* **2016**, *31*, 623–631.

18. Agrawal, U.; Pierre, J.W. Detection of Periodic Forced Oscillations in Power Systems Incorporating Harmonic Information. *IEEE Trans. Power Syst.* **2019**, *34*, 782–790.
19. Follum, J.; Pierre, J.W. Detection of Periodic Forced Oscillations in Power Systems. *IEEE Trans. Power Syst.* **2016**, *31*, 2423–2433.
20. Alizadeh, M.; Wang, X.; Chen, L.; Nassif, A.; Mitra, B.; Feng, S.; Cheng, Y.; de Melo Vieira, J.C.; Badizi, E.M. Sustained Oscillations in Modern AC Distribution Systems with High DER Infiltration. *IEEE Trans. Power Syst.* **2025**, *40*, 244–255.
21. Lin, G.; Zhang, H.; Chen, L.; Chen, X. Deep Learning Based Fault Detection and Diagnosis Method for Power Systems. *Appl. Math. Nonlinear Sci.* **2025**, *10*, 20250200.
22. Dorado-Rojas, S.A.; Xu, S.; Vanfretti, L.; Ayachi, M.I.I.; Ahmed, S. ML-Based Edge Application for Detection of Forced Oscillations in Power Grids. In Proceedings of the 2022 IEEE Power & Energy Society General Meeting (PESGM), Denver, CO, USA, 17–21 July 2022; pp. 1–5.
23. Wang, T.; Chen, Q.; Lang, X.; Xie, L.; Li, P.; Su, H. Detection of Oscillations in Process Control Loops from Visual Image Space Using Deep Convolutional Networks. *IEEE/CAA J. Autom. Sin.* **2024**, *11*, 982–995.
24. Li, Q.; Xu, Z.; Qin, X.; Zhang, K.; Zhang, C.; Ye, S. A Convolution Neural Network Method for Power System Oscillation Type Identification. In Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 30 October–1 November 2020; pp. 499–504.
25. Yu, Z.; Shi, D.; Li, J.; Wang, Y.; Zhao, X.; Wang, Z.; Li, J. Using Transfer Learning to Distinguish between Natural and Forced Oscillations. In Proceedings of the 2020 IEEE Power & Energy Society General Meeting (PESGM), Montréal, QC, Canada, 2–6 August 2020; pp. 1–5.
26. Yang, X.; Huang, X.; Li, J.; Li, Z. Forced Oscillation Detection via a Discrete Wavelet Transform-Dilated Convolutional Neural Network with MAML. *Electr. Power Syst. Res.* **2026**, *251*, 112336.
27. Yang, X.; Shan, T.; Xiong, H.; Xiao, X.; Huang, X.; Wang, Y. Forced Oscillation Recognition With Interpretable Network Framework for Class Imbalance and Limited Real-World Data. *IEEE Trans. Smart Grid* **2026**, *17*, 780–793.
28. Singh, P.; Prakash, A.; Parida, S. Neural Network Based Pattern Recognition for Classification of the Forced and Natural Oscillation. *Electr. Power Syst. Res.* **2023**, *224*, 109706.
29. Yang, X.; Wang, J.; Huang, X.; Wang, Y.; Xiao, X. Forced Oscillation Detection via a Hybrid Network of a Spiking Recurrent Neural Network and LSTM. *Sensors* **2025**, *25*, 2607.
30. Khan, W.A. Balanced Weighted Extreme Learning Machine for Imbalance Learning of Credit Default Risk and Manufacturing Productivity. *Ann. Oper. Res.* **2025**, *348*, 833–861.
31. Öcalan, G.; Türkoğlu, İ. Data-Efficient Degradation Progression Modeling in Industrial Compressors via Baseline-Referenced Deep Feature Learning and Unsupervised Clustering. *Appl. Sci.* **2026**, *16*, 6895.
32. Guo, S.; Zhao, X.; Zhou, R.; Feng, Q.; Yu, J. Research on Artificial Intelligence Time Series Prediction Technology in Power System Stability Analysis. In Proceedings of the Tenth International Conference on Energy Systems, Electricity, and Power (ESEP 2025); SPIE: Bellingham, WA, USA, 2026; Volume 14179, p. 226.
33. Wang, J.; Dai, X.; Zhang, M.; You, K.; Liu, J. A Physics-Constrained Hybrid Deep Learning Model for State Prediction in Shipboard Power Systems. *Modelling* **2026**, *7*, 65.
34. Shao, H.; Henriques, R.; Morais, H.; Tedeschi, E. Attention-Enhanced Residual Networks for Real-Time Multi-Label Power Quality Disturbance Classification with Fast Iterative Filtering. *Appl. Energy* **2026**, *406*, 127233.
35. Altun, B.E.; Alpsalaz, F.; Uzel, H.; Türkay, Y. Explainable DL Based Classification for Power Quality Disturbances in Renewable-Energy-Integrated Distribution Networks. *IET Renew. Power Gener.* **2026**, *20*, e70269.
36. Xiao, F.; Li, R.; Wang, H.; Zheng, C.; Shang, Y.; Ai, Q. Distribution Network Fault Detection and Classification Using an Improved S-transform and a Modified Convolutional Neural Network. *IET Gener. Transm. Distrib.* **2026**, *20*, e70245.
37. Daubechies, I. The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Trans. Inf. Theory* **1990**, *36*, 961–1005.
38. Wei, X.; Wu, X.; Yoshimura, K.; Cheng, C.; Huang, H.; Ding, Z.; Song, Y. Climate-Informed Long-Term Forecasting of Wind and Photovoltaic Power Using a Hybrid DWT–BES–CNN–LSTM Model. *Energy* **2025**, *338*, 138677.
39. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444.
40. Vapnik, V., N. An Overview of Statistical Learning Theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999.
41. Berrich, Y.; Guennoun, Z. EEG-based Epilepsy Detection Using CNN-SVM and DNN-SVM with Feature Dimensionality Reduction by PCA. *Sci. Rep.* **2025**, *15*, 14313.
42. Maslennikov, S.; Wang, B.; Zhang, Q.; Ma, a.; Luo, a.; Sun, a.; Litvinov, E. A Test Cases Library for Methods Locating the Sources of Sustained Oscillations. In Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, USA, 17–21 July 2016; pp. 1–5.
43. Test Cases Library of Power System Sustained Oscillations. 2019 Available online: <https://web.eecs.utk.edu/~kaisun/Oscillation/actualcases.html> (accessed on 14 February 2024).

44. Massaoudi, M.; Zamzam, T.; Eddin, M.E.; Ghrayeb, A.; Abu-Rub, H.; Refaat, S.S. Fast Transient Stability Assessment of Power Systems Using Optimized Temporal Convolutional Networks. *IEEE Open J. Ind. Appl.* **2024**, *5*, 267–282.
45. Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.

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.