Multiscale Spatiotemporal Ensemble Learning for Imbalanced Chiller Fault Diagnosis Using Generative Adversarial Networks

1st Yuzhe Wang
College of Mechanical and Electronic Engineering
SEAC Key Laboratory of Big Data Applied Technology
Dalian Minzu University
Dalian, China
wangyz@dlnu.edu.cn

2nd Huaijun Guang
College of Computer Science and Engineering
SEAC Key Laboratory of Big Data Applied Technology
Dalian Minzu University
Dalian, China
15708541524@163.com

3rd Fangbo Fu
College of Computer Science and Engineering
SEAC Key Laboratory of Big Data Applied Technology
Dalian Minzu University
Dalian, China
1755429354@qq.com

4th Jinglong Liu
College of Mechanical and Electronic Engineering
SEAC Key Laboratory of Big Data Applied Technology
Dalian Minzu University
Dalian, China
1310959322@qq.com

5th Ran Li
College of Mechanical and Electronic Engineering
SEAC Key Laboratory of Big Data Applied Technology
Dalian Minzu University
Dalian, China
2408758073@qq.com

6th Xiaodong Duan
College of Computer Science and Engineering
SEAC Key Laboratory of Big Data Applied Technology
Dalian Minzu University
Dalian, China
dxd_dlnu@163.com

***Abstract*—Imbalanced class distributions and spatiotemporally coupled fault features in chiller operational data hinder reliable fault diagnosis. To address these challenges, this paper proposes a multi-scale spatiotemporal ensemble learning framework for imbalanced chiller fault diagnosis that integrates generative data augmentation with deep ensemble learning.**

The framework consists of three components. First, a CTGAN–Tomek hybrid sampling scheme uses a conditional tabular generative adversarial network (CTGAN) to generate minority-class samples and Tomek Links to remove majority-class boundary samples, thereby improving data diversity and inter-class separability. Second, an improved projection iterative optimization algorithm (IPIMO) uses macro-averaged F1 as the objective to jointly optimize key CTGAN hyperparameters, reducing reliance on empirical tuning. Third, a multi-scale spatiotemporal stacked ensemble network (MCTS-Net) combines multi-scale one-dimensional convolutional neural networks (1D-CNNs) and temporal convolutional networks (TCNs) to capture local abrupt patterns and long-range dependencies, while squeeze-and-excitation (SE) attention and stacking further enhance classification robustness.

Experiments on the ASHRAE RP-1043 dataset show that, under a simulated class imbalance ratio of about 10:1, the proposed CTGAN–Tomek–IPIMO–MCTS-Net outperforms competing methods in precision, recall, and macro-averaged F1 for multi-class fault diagnosis. Compared with the non-augmented baseline, the proposed framework improves macro-averaged F1 by about 7.3 percentage points, demonstrating its effectiveness for imbalanced chiller fault diagnosis.

***Index Terms*—Chiller fault diagnosis; Imbalanced data; Generative adversarial networks; Ensemble learning; Spatiotemporal deep learning**

I. INTRODUCTION

Buildings are major contributors to global energy consumption and carbon emissions, and recent studies show that buildings account for about 36% of global energy use, while HVAC systems account for around 40% of total building energy consumption [1], [2]. As key cooling equipment in commercial HVAC systems, chillers strongly affect overall energy efficiency and operational reliability. Recent experimental and diagnostic studies have shown that common chiller faults, such as refrigerant leakage, fouling, and flow degradation, can cause substantial energy waste and maintenance risks if not detected in time [3]. Accurate and timely fault detection and diagnosis (FDD) is therefore essential for reliable and low-carbon operation.

With the development of intelligent building operation, data-driven deep learning methods have been widely adopted for FDD. Convolutional neural networks (CNNs), recurrent models such as long short-term memory (LSTM) networks, and more recent Transformer and temporal convolutional network

(TCN) models can effectively extract nonlinear spatiotemporal features from monitoring data [4].

However, real chiller operation logs are inherently imbalanced: normal conditions dominate because faults are rare and usually repaired quickly, while fault data are costly to obtain and label.

To mitigate the shortage of fault samples, oversampling methods such as ADASYN [6], SMOTE [7], and generative adversarial networks (GANs) [8] have been introduced in imbalanced learning [5]. Nevertheless, traditional oversampling may increase class overlap in high-dimensional time series, GAN-based augmentation can suffer from mode collapse and weak physical consistency, and many diagnostic models still rely on single-scale features and separately tuned augmentation and classifier parameters.

Despite recent progress in data-driven chiller FDD, accurate diagnosis under severe class imbalance and complex operating conditions remains challenging. Therefore, the main contributions of this paper are summarized as follows:

1) *CTGAN–Tomek hybrid sampling*: We propose a CTGAN–Tomek scheme that combines a conditional tabular generative adversarial network (CTGAN) with Tomek Links to generate minority-class samples and remove majority-class boundary samples, thereby improving data diversity and inter-class separability for imbalanced chiller time-series data.
2) *IPIMO-based joint hyperparameter optimization*: We develop an improved projection iterative optimization algorithm (IPIMO) that uses macro-averaged F1 as the objective to jointly optimize key CTGAN hyperparameters, reducing reliance on manual and fragmented tuning.
3) *MCTS-Net multi-scale spatiotemporal ensemble network*: We develop a multi-scale spatiotemporal stacked ensemble network (MCTS-Net), in which multi-branch one-dimensional convolutional neural networks (1D-CNNs) and temporal convolutional networks (TCNs) jointly capture local abrupt patterns and long-range dependencies, while squeeze-and-excitation (SE) attention and stacking improve robustness and generalization. Experiments on the long-tailed ASHRAE RP-1043 chiller dataset show superior performance over baseline methods, especially for mild faults under severe class imbalance.

II. Related Work

A. Generative Adversarial Networks and Variants

Goodfellow et al. [8] proposed generative adversarial networks (GANs), which learn data distributions through adversarial training and are widely used for data augmentation. However, vanilla GANs often suffer from mode collapse and unstable training. WGAN [9] improves training stability by introducing the Wasserstein distance, while CGAN [10] incorporates class-conditional information for targeted generation. For industrial tabular data, CTGAN proposed by Xu et al. [11] better models mixed continuous and categorical variables and performs well on imbalanced tabular datasets. Recent studies have further combined GAN-based generation with sample-balancing and feature-screening mechanisms to enhance minority-class representation and suppress noisy samples in imbalanced fault diagnosis [12].

B. Tomek Links and Hybrid Sampling Strategies

Tomek Links is a classical undersampling method for refining decision boundaries in imbalanced data by removing ambiguous majority-class samples [13]. It identifies nearest-neighbor pairs from different classes and removes the majority-class sample in each pair, thereby reducing class overlap. Because of its boundary-cleaning effect, Tomek Links is often combined with oversampling methods such as SMOTE [7]. These hybrid strategies first expand minority classes and then remove noisy or overlapping samples, which can improve classification performance [14]. Related neighborhood-cleaning methods, including edited nearest neighbors (ENN), one-sided selection (OSS), and neighborhood cleaning rule (NCL), are also widely used for noise suppression in industrial datasets.

C. Intelligent Optimization Algorithms

Intelligent optimization algorithms are widely used for hyperparameter tuning and model training. Meta-heuristic methods such as particle swarm optimization (PSO) [15], genetic algorithms (GA), grey wolf optimizer (GWO), and differential evolution (DE) balance global exploration and local exploitation in high-dimensional non-convex spaces. When combined with generative models and deep classifiers, these methods can improve fault diagnosis on imbalanced industrial datasets, motivating more task-specific joint optimization schemes.

III. Method

A. Overview

The proposed fault diagnosis framework, illustrated in Fig. 1, consists of two stages. First, a CTGAN–Tomek hybrid sampling scheme is used to construct a balanced training set, while an improved projection iterative optimization algorithm (IPIMO) jointly tunes key hyperparameters. Second, the proposed deep diagnostic network MCTS-Net extracts multi-scale spatiotemporal features and employs a stacking ensemble for accurate multi-fault classification.

B. CTGAN–Tomek Hybrid Sampling Scheme

Minority fault classes in chiller FDD are often scarce and scattered, while sensor noise and varying operating conditions increase inter-class overlap. To avoid the boundary blurring caused by interpolation-based oversampling in high-dimensional spaces, we adopt a *generation + boundary refinement* strategy: (i) CTGAN learns class-conditional distributions and generates additional minority samples; and (ii) Tomek Links removes ambiguous majority-class samples near decision boundaries to improve class separability.

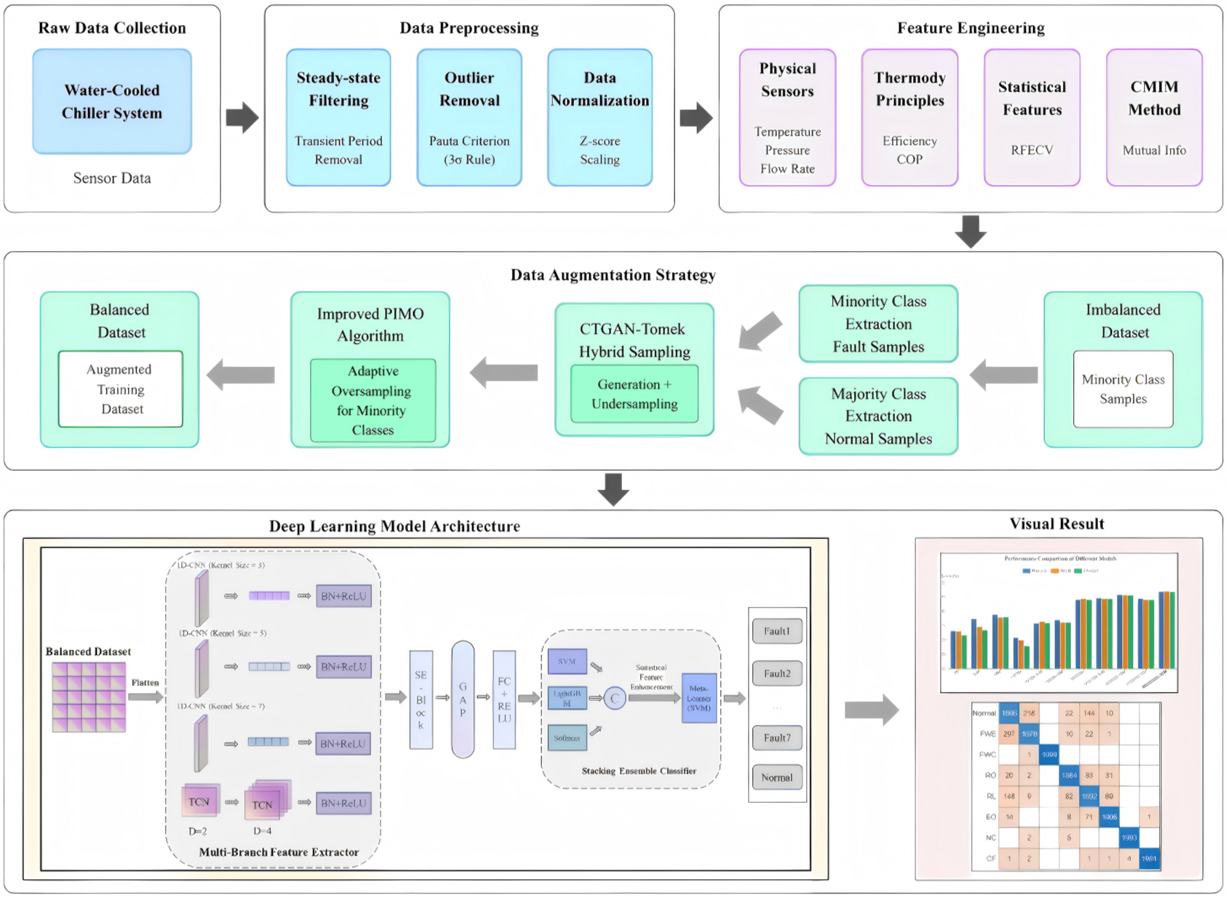

Fig. 1. Overall architecture of the proposed CTGAN–Tomek–IPIMO–MCTS-Net framework.

1) CTGAN-based conditional augmentation: CTGAN generates tabular samples conditioned on class labels, i.e., $\tilde{x} = G(z, c)$, where z is noise and c is the condition vector. The adversarial objective is

$$\begin{aligned}\min_G \max_D \mathcal{L}(D, G) = \mathbb{E}_{x\sim p_{\text{data}}}[D(x, c)] \\ - \mathbb{E}_{z\sim p_z,\, c\sim p_c}[D(G(z, c), c)] \\ - \lambda\, \mathbb{E}_{\hat{x}\sim p_{\hat{x}}}\left[(\|\nabla_{\hat{x}} D(\hat{x}, c)\|_2 - 1)^2\right],\end{aligned} \tag{1}$$

where $D(\cdot, c)$ is a class-conditional critic (without a sigmoid), $\hat{x} = \epsilon x + (1 - \epsilon)\tilde{x}$ with $\tilde{x} = G(z, c)$ and $\epsilon \sim \mathcal{U}(0, 1)$, and λ is the gradient-penalty coefficient. Following CTGAN [11], we adopt Wasserstein GAN with gradient penalty (WGAN-GP) [16] to stabilize training. For mixed-type features, CTGAN applies mode-specific normalization to multi-modal continuous variables using a Gaussian mixture model (GMM), and Gumbel–Softmax reparameterization to discrete variables to enable backpropagation through sampling [17].

2) Tomek Links boundary refinement: After generative augmentation, Tomek Links (TL) is used to refine the decision boundary [13]. Two samples (x_i, x_j) form a TL if they are mutual nearest neighbors and $y_i \neq y_j$. We remove the *majority-class* sample involved in any TL pair:

$$S' = S \setminus \{x_i \mid \exists x_j,\ \text{is\_TL}(x_i, x_j) \wedge y_i = y_{\text{maj}}\}, \tag{2}$$

Algorithm 1 Tomek Links Boundary Cleaning

Input: Dataset $S = \{(x_i, y_i)\}_{i=1}^{N}$, majority label y_{maj}

$T \leftarrow \emptyset$
for $i = 1$ to N **do**
  $j \leftarrow \text{NN}(x_i)$
  if $y_i \neq y_j$ **and** $\text{NN}(x_j) = i$ **then**
    $T \leftarrow T \cup \{i\}$
  end if
end for
$S' \leftarrow S \setminus \{(x_i, y_i) \mid i \in T \wedge y_i = y_{\text{maj}}\}$
return S'

where S and S' denote the original and cleaned datasets, respectively, and y_{maj} denotes the majority class.

C. Improved Projection Iterative Optimization (IPIMO)

To reduce the arbitrariness of manual hyperparameter selection for CTGAN and MCTS-Net, we propose an improved projection iterative optimization algorithm (IPIMO).

1) Objective and Decision Variables: IPIMO aims to maximize the macro-averaged F1 score on the validation set so that performance remains balanced across classes. The objective function is

$$\max_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}) = \frac{1}{C} \sum_{i=1}^{C} \frac{2P_i R_i}{P_i + R_i}, \tag{3}$$

Algorithm 2 IPIMO: Improved Projection Iterative Optimization Algorithm

Input: Population size N, maximum iterations $T_{\max}$, search space Ω

```
Initialization: Generate the initial population X_0 using a logistic chaotic map;
Set t ← 0;
while t < T_max do
    for each individual X_i in X_t do
        Evaluate fitness J(X_i) on the validation set;
    end for
    Mutation: Use Cauchy mutation in the first 50% of iterations and Gaussian mutation in the remaining 50%;
    Projection: If the population variance is below a threshold δ, perform a Nelder–Mead simplex projection search around the current best solution X_best [18];
    Update the global best solution X*_best;
    t ← t + 1;
end while
return Optimal hyperparameter vector θ* = X*_best
```

where C is the number of classes, and P_i and R_i are the precision and recall of class i, respectively. The decision vector $\boldsymbol{\theta}$ contains key hyperparameters of CTGAN and MCTS-Net, whose search ranges are listed in Table I.

CTGAN hyperparameters affect the quality of generated samples, whereas MCTS-Net hyperparameters determine model capacity on the resulting training set. To avoid mismatched augmentation and classifier settings, IPIMO jointly searches their hyperparameter space and maximizes the validation macro-F1 in (3).

TABLE I
HYPERPARAMETERS OPTIMIZED BY IPIMO AND THEIR SEARCH RANGES

Model	Hyperparameter	Search Range
CTGAN	Generator LR	$[1 \times 10^{-5}, 1 \times 10^{-3}]$
	Discriminator LR	$[1 \times 10^{-5}, 1 \times 10^{-3}]$
	Batch Size	$\{64, 128, 256, 512\}$
MCTS-Net	CNN Filters	$[16, 128]$
	TCN Dilation Limit	$\{2, 4, 8\}$
	Stacking Meta-Learner C	$[0.1, 10]$

2) Algorithm Details: IPIMO extends conventional projection iterative algorithms with logistic chaotic initialization and hybrid mutation strategies to alleviate premature convergence. The overall procedure is summarized in Algorithm 2.

D. MCTS-Net Deep Diagnostic Network

As shown in Fig. 2, MCTS-Net combines multi-branch deep feature extraction with stacking ensemble learning to capture multi-scale spatiotemporal patterns in chiller fault data. It contains three stages:

(1) Multi-branch feature extraction. To capture both short-term disturbances and long-term degradation trends, four parallel branches are used: three multi-scale 1D-CNN branches with different kernel sizes for local patterns, and one TCN branch for long-range temporal dependencies.

(2) Fusion and attention refinement. Features from all branches are concatenated and reweighted by a squeeze-and-excitation (SE) block to model inter-channel dependencies and suppress noisy channels. Global average pooling (GAP) then produces a compact embedding, followed by fully connected (FC) layers with rectified linear unit (ReLU) activation for feature transformation.

(3) Stacking ensemble classifier [19]. A two-level stacking scheme is adopted to improve generalization. At Level-0, support vector machine (SVM) [20], Light Gradient Boosting Machine (LightGBM) [21], and a Softmax classifier generate preliminary predictions. Their outputs are concatenated with simple statistical descriptors, and an SVM meta-learner at Level-1 produces the final fault label.

1) Network Configuration: MCTS-Net consists of three multi-scale 1D-CNN branches, a temporal convolutional network (TCN) branch [22], and a squeeze-and-excitation (SE) attention module [23]. Key layer settings are listed in Table II.

TABLE II
KEY LAYER CONFIGURATIONS OF MCTS-NET

Branch/Module	Kernel Size	Filters	Stride
1D-CNN ($k = 3$)	3	64	1
1D-CNN ($k = 5$)	5	64	1
1D-CNN ($k = 7$)	7	64	1
TCN	3 (dilation=1,2,4)	32	1
SE-Block	–	ratio=16	–

2) Stacking Training Strategy: The stacking module uses 5-fold cross-validation to reduce overfitting. Base learners generate out-of-fold predictions, which are concatenated with statistical enhancement features to train the SVM meta-learner. All deep components are optimized with Adam [24] using an initial learning rate of 0.001 and early stopping.

IV. EXPERIMENTS AND RESULTS

All experiments are conducted on a workstation running Windows 11 with an Intel Core i7 CPU and an NVIDIA RTX A4000 GPU.

A. Dataset Description

We use the ASHRAE RP-1043 dataset, a benchmark for centrifugal chiller fault diagnosis [29]. It contains seven fault types, each with four severity levels (SL1–SL4). Because early diagnosis is most important in practice, we focus on SL1 to evaluate the proposed method on subtle fault signatures.

B. Data Preprocessing

The raw RP-1043 data include transient operating periods and potential outliers. Before training, we apply three preprocessing steps:

1) *Steady-state filtering:* The chiller operates transiently during startup and shutdown (0–1000 s and 50,500–51,900 s). We therefore remove these intervals and retain only steady-state segments.

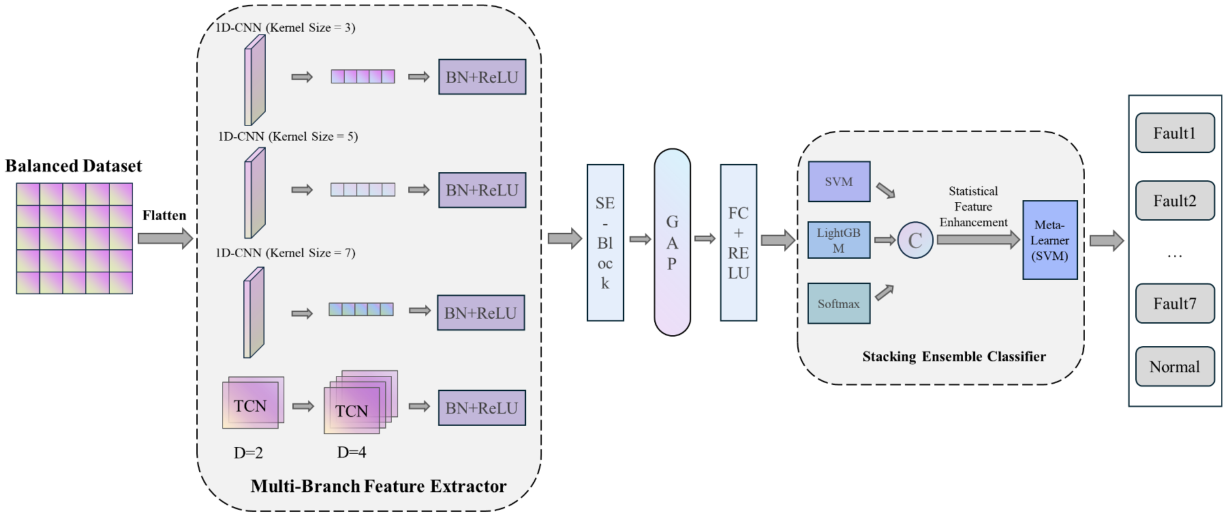

Fig. 2. Overview of MCTS-Net: multi-scale 1D-CNN ($k = 3, 5, 7$) + TCN, SE-based fusion, and a two-level stacking ensemble.

2) *Outlier removal:* We use the Pauta criterion [25] (the 3σ rule) to remove outliers. Any sample satisfying

$$|x - \mu| > 3\sigma, \tag{4}$$

is regarded as abnormal and discarded, where x is a sample value, and μ and σ are the mean and standard deviation of that feature.

3) *Feature scaling:* To eliminate the effects of different physical units and value ranges, all features are standardized by Z-score scaling:

$$x^* = \frac{x - \mu_X}{\sigma_X}, \tag{5}$$

where μ_X and σ_X denote the mean and standard deviation of feature X, respectively.

After preprocessing, the data are first split by operating run to prevent temporally adjacent samples from the same sequence from appearing in both the training and test sets. Stratified random sampling without replacement is then applied within each split to construct a long-tailed training set and a class-balanced test set. To reflect real operating conditions, the training set is built with an imbalance ratio of about 10:1 (normal vs. each fault class; Table IV), whereas the test set remains balanced for fair multi-class evaluation.

C. Feature Selection via RFECV–CMIM

To reduce computational cost and overfitting risk, we adopt a two-stage RFECV–CMIM feature selection strategy [26], [27]. CMIM first retains informative and less redundant features [26], and RFECV further removes low-contribution features through cross-validation [27]. The final RP-1043 feature set contains 12 key variables, listed in Table III.

Table IV summarizes the resulting dataset composition and label assignment.

D. Baseline Models on Imbalanced Data

Using the above dataset, we first evaluate the proposed MCTS-Net and 10 commonly used diagnostic models, including random forest (RF), k-nearest neighbors (KNN), and support vector machine (SVM), directly on the imbalanced training data. The comparative results are summarized in Table V.

TABLE III
KEY CHILLER VARIABLES SELECTED FROM THE RP-1043 DATASET

ID	Variable	Description
1	TCI	Chilled-water inlet temperature
2	TCO	Chilled-water outlet temperature
3	TWCO	Cooling-water outlet temperature
4	kW	Active power
5	FWC	Cooling-water flow rate
6	TCA	Compressor inlet air temperature
7	TRC	Compressor discharge temperature
8	PRC	Compressor discharge pressure
9	TRC_sub	Compressor discharge subcooling/superheat
10	P_lift	Pressure lift
11	PO_feed	Supply-side outlet pressure
12	PO_net	Network-side outlet pressure

TABLE IV
COMPOSITION AND LABEL DISTRIBUTION OF THE CONSTRUCTED DATASET

Fault	FWC	FWE	RL	RO	CF	EO	NC	Normal
Train size	100	100	100	100	100	100	100	1000
Test size	2000	2000	2000	2000	2000	2000	2000	2000
Label	1	2	3	4	5	6	7	8

As shown in Table V, all models are adversely affected by class imbalance. Even the best baseline, MCTS-Net, reaches an F1-score of only 86.93%, whereas the worst model achieves 67.93%. This indicates that sparse fault samples remain a major obstacle to learning discriminative fault representations.

E. Effectiveness of CTGAN–Tomek Data Augmentation

To alleviate this problem, we propose CTGAN–Tomek, which combines minority-sample generation with boundary refinement. To visualize its effect, features before and after augmentation are projected onto three principal components using PCA, as shown in Fig. 3.

Fig. 3 shows that CTGAN–Tomek produces denser minority-class distributions and clearer inter-class separation,

TABLE V
PERFORMANCE COMPARISON OF DIFFERENT MODELS ON IMBALANCED DATA

Model	Precision (%)	Recall (%)	F1-score (%)
RF	73.18	73.01	71.63
SVM	77.44	74.54	73.38
KNN	78.91	77.83	78.04
1DCNN	70.85	69.84	67.93
1DCNN-SVM	75.78	76.38	75.95
1DCNN-KNN	76.95	76.20	76.21
M1DCNN	84.13	84.40	84.16
M1DCNN-SVM	84.61	84.44	84.43
M1DCNN-KNN	85.80	85.74	85.72
M1DCNN-TCN	84.53	84.13	84.03
MCTS-Net	**86.95**	**86.99**	**86.93**

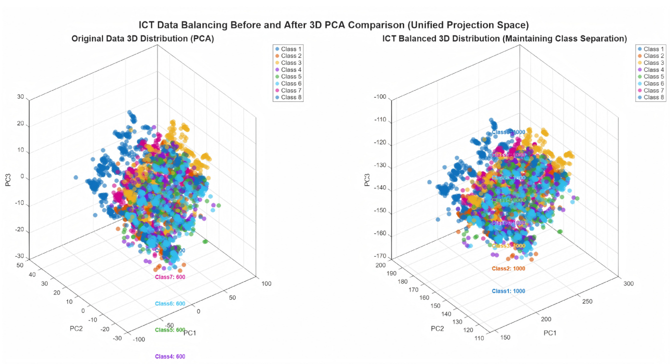

Fig. 3. Visualization of class distributions before and after CTGAN–Tomek augmentation (PCA projection).

indicating reduced overlap in the original feature space. We further compare it with representative oversampling methods under the same classifier settings; the results are reported in Table VI.

Table VI shows that CTGAN–Tomek consistently achieves the best or near-best performance across classifiers. Compared with ADASYN, SMOTE and its variants, CTGAN, and WGAN-GP, it yields higher accuracy, recall, and F1-score, confirming its effectiveness for high-dimensional imbalanced fault time series.

F. Ablation Study on Data Augmentation and Hyperparameter Optimization

To quantify the contributions of different components, we conduct an ablation study on MCTS-Net under the same data split and training settings, varying only the augmentation and hyperparameter optimization strategies. Five configurations are considered: (A) MCTS-Net trained directly on the imbalanced data; (B) CTGAN + MCTS-Net; (C) CTGAN + IPIMO + MCTS-Net; (D) CTGAN + Tomek + MCTS-Net; and (E) CTGAN + Tomek + IPIMO–MCTS-Net. The results are reported in Table VII.

Table VII shows that the baseline MCTS-Net achieves a macro-F1 of 86.93%. CTGAN alone brings almost no improvement, suggesting that naive synthetic augmentation is insufficient. Adding IPIMO increases macro-F1 to 89.80%, indicating that joint hyperparameter optimization helps reduce the mismatch between generated samples and the classifier.

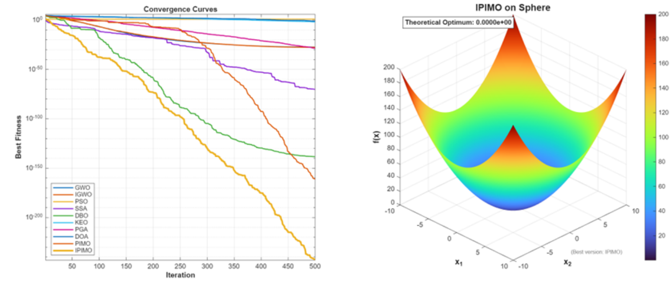

Fig. 4. Convergence comparison of IPIMO and competing optimization methods.

Incorporating Tomek Links further improves macro-F1 to 92.03%, validating the effectiveness of the “generation + boundary refinement” strategy. The full framework achieves 94.20% macro-F1, improving by 7.27 percentage points over the baseline.

G. Convergence Analysis of IPIMO

To evaluate search efficiency, Fig. 4 compares IPIMO with representative meta-heuristics, including particle swarm optimization (PSO) [15], the sparrow search algorithm (SSA) [28], and other common baselines.

As shown in Fig. 4, IPIMO converges faster and reaches a better final solution than the competing methods.

V. CONCLUSION

This paper presents an integrated fault diagnosis framework for HVAC chillers under severe class imbalance. By combining CTGAN–Tomek hybrid sampling, IPIMO-based hyperparameter optimization, and the MCTS-Net deep ensemble, the proposed method improves data balance, class separability, and diagnostic robustness under complex operating conditions.

The main contributions are:

1) *CTGAN–Tomek hybrid sampling:* A “generation + boundary refinement” strategy in which CTGAN generates minority-class samples and Tomek Links removes boundary-ambiguous majority samples, improving data diversity and inter-class separability.
2) *IPIMO optimization algorithm:* An improved projection iterative optimization method with chaotic initialization, hybrid mutation, and projection search for joint tuning of key hyperparameters in the augmentation and diagnosis framework.
3) *MCTS-Net diagnostic network:* A multi-scale spatiotemporal stacked ensemble network that captures local patterns and long-range dependencies, while stacking further enhances robustness and generalization.

Experiments on the *ASHRAE RP-1043* dataset show that the proposed *CTGAN–Tomek–IPIMO–MCTS-Net* consistently outperforms representative baselines under highly imbalanced conditions. In particular, the full framework improves macro-F1 from 86.93% to 94.20%. Future work will investigate cross-system generalization to other chiller types and HVAC equipment, as well as lightweight deployment through pruning and quantization for real-time applications.

TABLE VI
PERFORMANCE COMPARISON OF DIFFERENT OVERSAMPLING METHODS AND CLASSIFIERS

Model	**Ours (CTGAN–Tomek)**			ADASYN			SMOTE			SMOTE-ENN			SMOTE-Tomek			CTGAN			WGAN-GP		
	Acc.	**Rec.**	**F1**	Acc.	Rec.	F1	Acc.	Rec.	F1	Acc.	Rec.	F1	Acc.	Rec.	F1	Acc.	Rec.	F1	Acc.	Rec.	F1
RF	**90.09**	**90.18**	**90.06**	83.76	83.66	83.58	84.84	84.91	84.70	84.44	84.59	84.41	84.78	84.99	84.76	84.18	84.49	84.16	85.04	84.96	84.57
SVM	**90.60**	**90.65**	**90.47**	87.16	86.94	87.16	86.89	87.10	86.85	86.86	86.98	86.73	86.99	87.11	86.92	86.58	86.75	86.61	85.37	84.93	84.44
KNN	**90.59**	**90.48**	**90.51**	85.71	85.69	85.68	85.64	85.53	85.56	85.41	85.44	85.41	87.21	87.49	87.24	84.78	84.71	84.73	78.64	77.88	77.94
1DCNN	**85.67**	**85.73**	**85.65**	85.36	85.71	85.31	84.27	84.41	84.19	85.51	85.39	85.34	85.23	85.24	85.17	83.78	84.21	83.78	75.69	76.35	75.28
1DCNN-SVM	**90.08**	**89.84**	**89.85**	82.77	81.01	81.17	82.86	81.35	81.55	84.83	83.26	83.72	83.44	81.71	81.86	82.91	82.73	82.56	86.43	86.13	86.22
1DCNN-KNN	**89.41**	**89.27**	**89.31**	86.55	86.59	86.54	86.55	86.59	86.52	86.38	86.34	86.30	86.64	86.75	86.66	84.75	84.96	84.83	84.34	84.28	84.30
M1DCNN	**89.46**	**89.58**	**89.51**	86.81	86.98	86.87	86.98	86.97	86.92	86.79	86.96	86.82	87.01	87.20	87.08	86.61	86.84	86.68	85.61	85.76	85.64
M1DCNN-SVM	**91.41**	**91.42**	**91.41**	85.10	85.41	85.17	85.87	85.93	85.84	86.53	85.97	86.14	85.00	85.40	84.97	85.96	85.76	85.84	86.05	86.18	86.09
M1DCNN-KNN	**90.03**	**89.96**	**89.97**	87.17	87.26	87.20	87.45	87.46	87.43	87.36	87.44	87.37	87.50	87.62	87.51	86.72	86.87	86.76	86.73	86.59	86.65
M1DCNN-TCN	**88.19**	**88.03**	**88.05**	85.56	85.58	85.52	85.22	85.26	85.17	86.23	86.05	86.06	85.59	85.54	85.48	85.79	85.87	85.81	85.36	85.26	85.27
MCTS-Net	**92.07**	**92.01**	**92.03**	**87.18**	**87.42**	**87.24**	**87.68**	**87.83**	**87.72**	**87.54**	**87.74**	**87.58**	**87.80**	**87.92**	**87.80**	**86.87**	**86.98**	**86.90**	**87.51**	**87.37**	**87.40**

TABLE VII
ABLATION STUDY ON DATA AUGMENTATION AND HYPERPARAMETER OPTIMIZATION

Model	Accuracy (%)	Recall (%)	F1-score (%)
MCTS-Net	86.95	86.99	86.93
CTGAN + MCTS-Net	86.87	86.98	86.90
CTGAN + IPIMO + MCTS-Net	88.20	89.00	89.80
CTGAN + Tomek + MCTS-Net	92.07	92.01	92.03
CTGAN + Tomek + IPIMO–MCTS-Net	**92.80**	**93.50**	**94.20**

ACKNOWLEDGMENT

This work was supported by the National Key Research and Development Program of China under Grant No. 2018YFB1004402.

REFERENCES

[1] F. J. Zarco-Soto, I. M. Zarco-Soto, S. S. S. Ali, and P. J. Zarco-Periñán, "Energy consumption in buildings: A compilation of current studies," *Energy Reports*, vol. 13, pp. 1293–1307, 2025, doi: 10.1016/j.egyr.2024.12.069.

[2] S. L. Zhou, A. A. Shah, P. K. Leung, X. Zhu, and Q. Liao, "A comprehensive review of the applications of machine learning for HVAC," *DeCarbon*, vol. 2, art. no. 100023, 2023, doi: 10.1016/j.decarb.2023.100023.

[3] Z. Wang, P. Xia, S. Zhou, L. Wang, J. Guo, X. Li, J. Shi, and J. Chen, "Examining the impact of common faults on chiller performance through experimental investigation and parameter sensitivity analysis," *Energy and Buildings*, vol. 317, art. no. 114389, 2024, doi: 10.1016/j.enbuild.2024.114389.

[4] J. Bi, H. Wang, E. Yan, C. Wang, K. Yan, L. Jiang, and B. Yang, "AI in HVAC fault detection and diagnosis: A systematic review," *Energy Reviews*, vol. 2, art. no. 100071, 2024, doi: 10.1016/j.enrev.2024.100071.

[5] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.

[6] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2008, pp. 1322–1328, doi: 10.1109/IJCNN.2008.4633969.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: 10.1613/JAIR.953.

[8] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[9] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," arXiv preprint arXiv:1701.07875, 2017.

[10] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

[11] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling tabular data using conditional GAN," in *Advances in Neural Information Processing Systems*, 2019.

[12] H. Chen, J. Wei, H. Huang, L. Wen, Y. Yuan, and J. Wu, "Novel imbalanced fault diagnosis method based on generative adversarial networks with balancing serial CNN and Transformer (BCTGAN)," *Expert Systems with Applications*, vol. 258, art. no. 125171, 2024, doi: 10.1016/j.eswa.2024.125171.

[13] I. Tomek, "Two modifications of CNN," *IEEE Trans. Syst., Man, Cybern.*, vol. 6, no. 11, pp. 558–563, 1976, doi: 10.1109/TSMC.1976.4309452.

[14] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, 2004, doi: 10.1145/1007730.1007735.

[15] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, 1995, pp. 1942–1948.

[16] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Advances in Neural Information Processing Systems*, 2017, pp. 5767–5777.

[17] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with Gumbel-Softmax," arXiv preprint arXiv:1611.01144, 2017.

[18] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965, doi: 10.1093/comjnl/7.4.308.

[19] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[21] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017, pp. 3146–3154.

[22] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," arXiv preprint arXiv:1803.01271, 2018.

[23] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.

[25] F. Pukelsheim, "The three sigma rule," *The American Statistician*, vol. 48, no. 2, pp. 88–91, 1994, doi: 10.1080/00031305.1994.10476030.

[26] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, 2004.

[27] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 2002, doi: 10.1023/A:1012487302797.

[28] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow Search Algorithm," *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020, doi: 10.1080/21642583.2019.1708830.

[29] M. C. Comstock and J. E. Braun, *Fault Detection and Diagnostic (FDD) Requirements and Evaluation Tools for Chillers*, ASHRAE Research Project RP-1043, Final Report, Ray W. Herrick Laboratories, Purdue University, 1999.