

Article

Efficient Data Construction Method for Electromagnetic Fault Diagnosis of Radar Timing Control Boards

Pengfei Liu ^{1,2,*}, Ziyan Wang ², Hui Wang ³, Bo Jiang ² and Wenxian Yu ^{1,*}

¹ Shanghai Key Laboratory of Intelligent Sensing and Recognition, Shanghai Jiao Tong University, Shanghai 200240, China

² East-China Research Institute of Computer Technology, Shanghai 201800, China; wzian2@gmail.com (Z.W.); b26jiang@126.com (B.J.)

³ School of Electronic Information Engineering, Beihang University, Beijing 100191, China; wanghuihui@buaa.edu.cn

* Correspondence: debug_lpf@sjtu.edu.cn (P.L.); wxyu@sjtu.edu.cn (W.Y.)

Abstract

This work targets four core bottlenecks restricting electromagnetic susceptibility (EMS) fault diagnosis of digital radar boards, including cold-start modeling caused by scarce fault samples, low-efficiency scanning prone to omitting narrow sensitive frequencies, insufficient feature representation of electromagnetic signals, and severe overfitting induced by imbalanced data. An integrated EMS dataset construction framework is accordingly proposed for radar timing control boards based on closed-loop electromagnetic interference (EMI) injection and multidimensional key signal point (KSP) monitoring. The framework adopts an adaptive coarse-to-fine scanning scheme to rapidly lock sensitive bands and accurately resolve fault boundaries. It further incorporates the hybrid synthetic minority oversampling technique and random undersampling (SMOTE-RU) strategy and time–frequency feature filtering to balance sample distribution and remove redundant information, yielding standardized high-quality EMS feature datasets. We validate dataset quality on a suite of classic machine learning and deep learning models. Test results show that all baseline networks attain classification accuracy above 85.1%, while the convolutional block attention module (CBAM) attention-augmented model delivers the highest accuracy of 95.0%. Cross-board tests on dissimilar radar boards show that fine-tuning with 10% target samples drastically raises cross-domain accuracy, and moderate interference yields the best generalization above 91.2%. Incurring minimal retraining costs, this lightweight scheme achieves strong cross-hardware generalization to build EMS databases for radar timing boards and implement multiclass fault diagnosis, effectively solving the cold-start problem for new hardware with insufficient historical fault data.

Keywords: radar timing control board; key signal point (KSP); electromagnetic susceptibility (EMS); SMOTE-RU resampling; electromagnetic interference (EMI); fault diagnosis

Academic Editor: Ahmed Abu-Siada

Received: 1 June 2026

Revised: 3 July 2026

Accepted: 15 July 2026

Published: 18 July 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.

1. Introduction

1.1. Background

Modern complex battlefields feature continuous, ultra-wideband, highly transient electromagnetic environments filled with densely superimposed, dynamically varying

signals, making full physical reproduction of actual interference nearly impossible. As the core carrier of radar signal processing, timing control boards are particularly vulnerable to EMI. Their EMS testing faces three fundamental limitations that plague all radar digital circuits.

First, newly developed radar boards lack accumulated fault datasets or mature EMS prediction models, creating a cold-start dilemma that drastically slows the convergence of diagnostic models, as also reported in industrial equipment EMS research [1,2]. Second, radar hardware exhibits threshold-sensitive, frequency-selective EMS responses. Minor adjustments to interference frequency or power can trigger abrupt functional degradation, yet conventional pass/fail EMC tests cannot quantify gradual performance decay, leaving ambiguous critical susceptibility boundaries undetermined. Third, real battlefield interference is time-varying and multi-band coupled; traditional steady-state fixed-frequency testing cannot reproduce transient disturbance waveforms, thereby concealing faint non-stationary fault signatures under background noise.

Existing industrial threshold-based EMC qualification for FPGA/CPU timing boards only provides binary pass-or-fail outcomes and cannot record continuous equipment degradation under critical EMI excitation. Meanwhile, full-board electromagnetic simulation suffers from poor reliability and heavy modeling overhead, lacking accurate device parameters for complex electromagnetic operating conditions. These drawbacks further translate into two prominent practical engineering bottlenecks:

1. Data collection bottleneck: No unified rapid testing pipeline is available to efficiently extract sensitive frequency ranges and fault threshold boundaries for uncharacterized radar boards, resulting in a severe training data shortage.
2. Feature extraction bottleneck: Conventional standard test signals fail to reconstruct dynamically interleaved battlefield-conducted and near-field radiated interferences, generating mismatched datasets incapable of supporting high-precision intelligent fault diagnosis.

Accordingly, there is an urgent demand for a low-cost, efficient EMS data construction method capable of producing high-quality labeled fault samples without pretrained prior models, so as to advance active electromagnetic protection for radar electronic hardware.

1.2. Contributions of This Paper

Aiming at the aforementioned EMS testing and fault diagnosis limitations of radar digital boards, this work proposes a holistic EMS dataset construction and feature enhancement framework, as illustrated in Figure 1. Taking a radar timing control board as the verification carrier, this study systematically elaborates an end-to-end data generation pipeline covering interference injection to feature optimization. The core contributions are summarized as three items below:

1. An adaptive coarse-to-fine scanning strategy for EMS data acquisition is proposed. The two-stage scanning scheme eliminates the inefficiency of full-band blind searching: wide-step coarse scanning rapidly locates candidate sensitive bands, followed by high-resolution fine scanning around suspicious frequency points. This scheme cuts total test duration from 81 min to 43 min and achieves a fault frequency characterization accuracy of ± 0.5 MHz while fully capturing nonstationary electromagnetic responses at fault critical boundaries.
2. A multidimensional monitoring and closed-loop data enhancement system is established. Unlike traditional single-point sampling, multi-KSP channels are deployed on power rails, clock circuits and bus interfaces to realize full-dimensional electromagnetic state perception. The integrated SMOTE-RU [3] hybrid algorithm and tree-based feature screening mitigate severe sample imbalance and remove redundant

features, reducing the original feature dimension by approximately 47% and boosting the generalization performance of training datasets.

3. Comprehensive cross-model validation verifies the superiority of the proposed data framework. Multiple mainstream machine learning and deep learning backbones are trained on the constructed dataset. All baseline models obtain classification accuracy above 85.1%, and the CBAM-augmented attention model reaches a peak accuracy of 95.0%. The framework greatly lowers the time and computational overhead for EMS fault modeling of digital circuit boards.

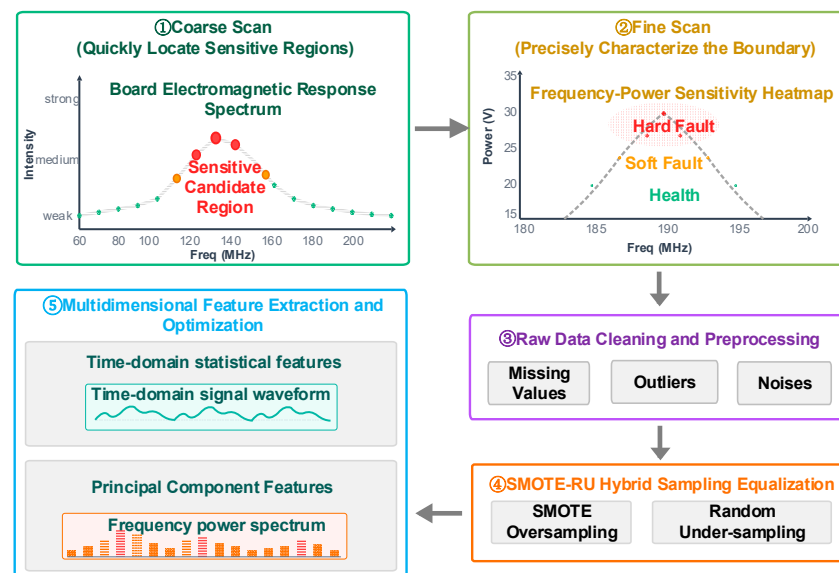


Figure 1. Framework of data construction and feature enhancement pipeline.

2. Related Work

2.1. Evolution of Board-Level Electromagnetic Fault Diagnosis

Research on electromagnetic fault diagnosis for digital circuit boards has gone through three developmental phases.

In the early stage, diagnosis depended on fixed threshold judgment based on single-voltage or clock signals. Such methods rely heavily on manual experience and fail to capture subtle transient signal distortions induced by time-varying EMI. Later, shallow machine learning techniques such as support vector machines (SVMs) and random forest (RF) were introduced for fault classification. These algorithms operate on handcrafted time–frequency features but cannot characterize long-range nonlinear correlations hidden in high-speed electromagnetic sequential data. With the boom of deep learning, scholars have adopted end-to-end temporal networks to automatically mine latent signal features. Even so, nearly all existing diagnosis pipelines ignore two core requirements of radar EMS tests: closed-loop data acquisition and multi-physical joint monitoring. They lack a complete optimized workflow for high-quality EMS dataset construction, which is the core focus of this work.

2.2. Limitations of State-of-the-Art Cross-Domain Diagnosis Algorithms

Three representative advanced intelligent diagnosis frameworks are reviewed, namely, time–frequency dual-domain prediction (TFDDP [4]), cross-sensor self-contrastive learning (CSSCL [5]) and graph contrastive domain adaptation (GCDA [6]), all of which show obvious deficiencies when applied to radar timing control boards.

The TFDDP method designs a lightweight dual-branch feature extraction structure, yet it does not support multi-channel collaborative collection of multi-physical signals or

adaptive screening of narrowband sensitive frequency bands, two key functions realized by our coarse-to-fine scanning and multi-KSP monitoring system.

The CSSCL algorithm improves feature discrimination via self-supervised learning, but its fault diagnosis accuracy drops sharply under the severe sample imbalance commonly seen in EMS injection tests. Our hybrid SMOTE-RU resampling module specially addresses this imbalance defect.

The GCDA network achieves cross-board generalization via graph convolution and domain adversarial alignment. Nevertheless, it cannot capture millisecond-level transient electromagnetic jumps, and it lacks a complete end-to-end closed-loop data generation pipeline, as proposed in this paper.

2.3. Deficiencies of Component-Level Electromagnetic Susceptibility Research

Most existing EMS mechanism analyses [7–9] only target discrete electronic components. Prior works have studied component susceptibility characteristics, reinforcement learning test strategies and physics-driven evaluation models while ignoring systematic testing and whole-board mechanism analysis.

In practical radar timing control hardware, power, clock and bus circuits are tightly coupled, generating complex inter-module electromagnetic crosstalk. Component injection experiments can only reflect the immunity of individual chips and cannot reconstruct real coupled interference environments on complete boards. The susceptibility rules summarized from single-device tests provide limited guidance for full-board fault diagnosis. At present, no research builds a closed-loop data pipeline with multi-KSP holographic monitoring, which is the core innovation of our proposed framework.

2.4. Comprehensive Review of Four Mainstream Intelligent Diagnosis Paradigms

Current intelligent EMS fault diagnosis methods can be divided into four categories: traditional machine learning, generic deep learning, time-series modeling and feature optimization algorithms. Each category is analyzed below, with a clear comparison against our integrated framework.

1. Traditional machine learning (SVM, XGBoost, LightGBM [10–12]).

These algorithms deliver interpretable baseline classification results, but they fully depend on artificially designed features and have weak fitting ability for complex nonlinear coupling among power ripple, clock jitter and bus errors. Their generalization capacity declines greatly when facing heterogeneous circuit boards. By contrast, our automatic time–frequency feature screening module avoids manual feature design and strengthens cross-board adaptability.

2. Convolution and attention-based deep learning [13–15].

CNNs and graph attention networks possess powerful fitting capability for complex signals. Relevant studies optimize CNNs for UAV links and design graph learning for mechanical equipment. However, these networks are customized for communication or rotating machinery, without targeted optimization for nonstationary, superimposed interference on radar PCBs. Our multidimensional KSP synchronous acquisition module specially extracts unique electromagnetic signatures of digital radar boards.

3. Time-series models (Bi-LSTM, Transformer [16–18]).

Such models excel at capturing long-term temporal correlations among sequential data. However, they require massive balanced labeled samples, which are hard to obtain in real EMS tests. In addition, they cannot accurately identify abrupt millisecond-scale electromagnetic state jumps. The adaptive coarse–fine scanning strategy in this paper automatically collects dense fault boundary samples to compensate for this shortage.

4. Feature optimization algorithms [19–21].

Existing feature enhancement and data augmentation schemes relieve sample scarcity and feature redundancy to some extent. Nevertheless, most of them are verified on optical imaging or analog sensor data, without joint time–frequency screening designed for imbalanced electromagnetic fault datasets. Our Pearson–tree combined feature filtering module fills this research gap.

Synthesizing the above literature analysis, four universal defects exist in existing EMS diagnosis technologies:

- (1) Few models consider complete circuit topological coupling, leading to low precision when solving full-board EMS critical boundaries;
- (2) Coupled features from power, timing and data channels are not fully excavated;
- (3) Algorithms lack robustness against wideband dynamically superimposed battlefield interference;
- (4) Cross-board transferable diagnosis schemes are rarely studied.

These unresolved bottlenecks motivate the integrated closed-loop EMS data construction framework proposed in this paper.

3. EMS Data Construction Method

This section presents an efficient framework for data construction, which comprises: a multi-channel synchronous perception hardware architecture design, a state-feedback adaptive EMS scanning strategy, and a closed-loop mechanism for data enhancement and feature mining. Through hardware perception, strategy scanning, and data preprocessing, this framework enables the generation of standardized electromagnetic feature datasets from physical digital boards.

In Figure 2, the overall dataset construction workflow consists of four sequential modules, namely, KSP multi-channel signal acquisition, adaptive electromagnetic scanning, SMOTE-RU imbalance sample processing and multi-domain feature optimization, as illustrated in the flowchart. All parameter configurations and invocation rules are detailed in the symbol list in Appendix A.1 and the pseudocode in Appendix A.2.

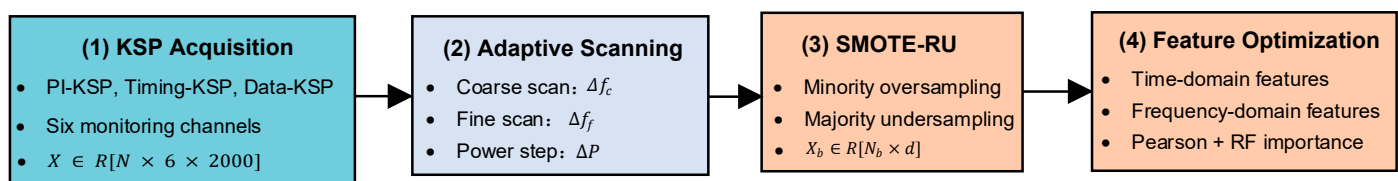


Figure 2. Complete EMS dataset construction workflow.

1. KSP Acquisition Module.

This module realizes synchronous collection of three types of key signal profiles via six independent monitoring channels. The raw collected signal tensor is defined as $X \in R[N \times 6 \times 2000]$, where N denotes the total number of sampling trials, 6 represents the number of monitoring channels, and 2000 stands for the sampling points per channel in a single measurement.

2. Adaptive Scanning Module.

An electromagnetic injection scanning strategy combining coarse and fine frequency searching is adopted to locate sensitive frequency points accurately. The coarse scanning frequency step is set as $\Delta f_c = 20$ MHz, while the fine scanning step is refined to $\Delta f_f = 1$ MHz for precise sensitive region localization. Meanwhile, the injection power step is fixed to $\Delta P = 1$ dBm to cover diverse EMI intensity levels.

3. SMOTE-RU Imbalance Balancing Module.

The hybrid SMOTE-RU resampling algorithm is utilized to mitigate the severe class imbalance of electromagnetic fault samples. It simultaneously executes minority class oversampling and majority class undersampling to reconstruct a balanced sample set, whose balanced data matrix is denoted as $X_b \in R[N_b \times d]$, where N_b is the quantity of balanced samples and d represents the initial feature dimension.

4. Multi-Domain Feature Optimization Module.

Time-domain statistical features and frequency-domain FFT transform features are extracted from the balanced multi-channel signals to form a raw multidimensional feature pool. Afterwards, Pearson correlation coefficient analysis and RF feature importance ranking are jointly deployed to filter redundant and irrelevant features, retaining only the physically discriminative feature subsets.

3.1. Multi-Channel Synchronous Perception Hardware Architecture

To deeply characterize the coupling evolution of external interference on the internal logic state of digital boards, this paper designs and implements a non-intrusive, high-fidelity electromagnetic monitoring framework. Starting from physical-layer signal integrity, this architecture ensures that the collected data highly restores the intrinsic sensitivity characteristics of the board.

3.1.1. Design of High-Fidelity Multi-Channel Acquisition Unit

To solve the problems of high-frequency digital signal distortion caused by probe loading effects and spatial radiation coupling noise in electromagnetic injection experiments, an efficient multi-channel acquisition unit is used. The acquisition unit integrates a four-channel high-speed analog acquisition architecture built around the zynq radio frequency system-on-chip (RFSoc). It supports a sampling rate of 2.5 GSa/s and 14-bit resolution. The measured signals are conditioned by fixed attenuation and anti-interference circuits and then digitized by analog-to-digital conversion modules. On-chip high-speed interconnect buses achieve multi-channel time synchronization, keeping the inter-channel timing error at the picosecond level. Digital data is transmitted to the main controller via a PCIe bus with direct memory access, which effectively eliminates timing jitter and data loss induced by intermediate buffers.

The multi-channel acquisition unit integrates four wideband sensing links and achieves nanosecond-level synchronous sampling via common-source clock triggering. It ensures precise time-axis alignment of power fluctuation, clock jitter, data bus error codes and other signals, which is essential for analyzing the instantaneous chain response mechanism of interference signals propagating from the power distribution network (PDN) to the logical timing links. Meanwhile, the acquisition links adopt a band-isolated shielding structure and a dynamic impedance compensation scheme, which greatly reduce the distributed capacitance and inductance of the sampling loop. This design guarantees distortion-free transmission of the amplitude–frequency characteristics of original EMS signals over a wide frequency range and finally acquires pure basic data that can truly reflect the logical state variations of the circuit board.

3.1.2. Multidimensional Configuration of KSP

To construct a holographic mapping model of the electromagnetic behavior of the board, this paper proposes a function-layered KSP configuration strategy.

Figure 3 visualizes the multi-layer KSP deployment layout designed for radar timing control hardware. Based on the signal transmission and data processing functional

partitions of the target board, all collected monitoring signals are categorized into three mutually complementary state vector branches as detailed below:

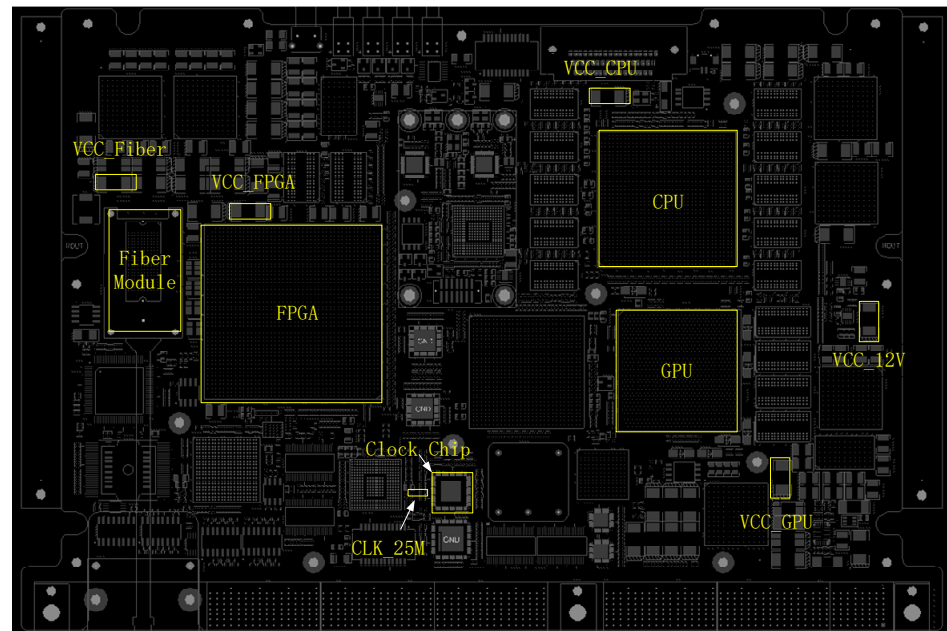


Figure 3. Multidimensional KSP configuration of the radar timing control board.

- (1) Power Integrity KSP (PI-KSP): Measurement nodes are laid out along the full power distribution chain, covering board-level LDO output terminals down to core power rails of internal chips. Since power delivery networks are most susceptible to conducted EMI, PI-KSP channels capture voltage drop events and high-frequency coupled noise triggered by interference pulses, enabling quantitative evaluation of power-domain electromagnetic susceptibility responses.
- (2) Timing Reference KSP (Timing-KSP): This monitoring category targets the system global clock source and frequency division logic. EMI interference distorts clock edge shapes and introduces cycle jitter, which further triggers misalignment of digital logic sequences. Sampled clock waveforms act as quantitative physical indicators that quantify timing margin degradation under electromagnetic excitation.
- (3) Data Interaction KSP (Data-KSP): This branch monitors high-speed differential lines, parallel buses and optical communication interfaces. Signal anomalies within data channels represent the final visible outcome of hardware performance degradation. Fault feature information extracted from these points maintains strong coupling correlations with power and timing distortions, forming a complete multidimensional feature closed loop for electromagnetic fault analysis.

Table 1 enumerates the six specific KSP monitoring points illustrated in Figure 3, alongside their hardware positioning descriptions.

Table 1. Detailed definitions of six configured KSP monitoring points.

KSP Category	KSP Label	Hardware Description
PI-KSP	VCC_12V	External 12 V Power Supply
PI-KSP	VCC_CPU	CPU Core Voltage, 0.8 V
PI-KSP	VCC_FPGA	FPGA Core Voltage, 1.0 V
PI-KSP	VCC_GPU	GPU Core Voltage, 0.85 V
Timing-KSP	CLK_25M	Clock Chip Reference Clock
Data-KSP	VCC_Fiber	Optical Module Voltage, 3.3 V

For power domain monitoring, the radar timing control board takes 12 V external voltage as the primary input, which is converted into multi-channel core supply voltages via multi-stage DC-DC and LDO circuits integrated within the AST4650 power management chip. The AST4650 is a dual-mode step-down regulator supporting either two 25 A independent channels or one single 50 A channel. It integrates functional modules for frequency synchronization, multi-phase parallel operation, burst power saving and sequential power rail tracking, plus an embedded thermal diode to track internal device temperature. To quantitatively characterize voltage sag, ripple amplification and high-frequency crosstalk induced by EMI on power distribution networks (PDNs), four PI-KSP nodes are arranged along critical power paths. These monitoring points cover both board-side input terminals and chip core voltage nodes to continuously record dynamic variations in voltage drop, ripple magnitude, transient response and PDN impedance metrics.

In terms of timing synchronization architecture, the hardware adopts TGEN63xxH low-jitter clock generators as the unified timing reference. Each chip embeds two configurable phase-locked loops (PLLs) and outputs multiple ultra-low-jitter clock signals, which expand timing design margins for high-performance radar computing systems compliant with PCIe standards. Under experimental test conditions, the 25 MHz external crystal reference is locked by the clock chip to generate 125 MHz and 100 MHz clock branches separately: the 125 MHz signal serves as the timing reference for FPGA internal logic and PCIe high-speed links, while the 100 MHz signal drives Ethernet and optical communication peripherals. The global 25 MHz reference clock CLK_25M is selected as the Timing-KSP channel to capture EMI-induced clock degradation phenomena, including edge distortion, period jitter and phase offset, and quantify corresponding parameter shifts.

For data interface monitoring, the radar timing control board equips Ethernet, DP display and fiber optical communication peripherals. The CPU unit connects GPU modules through PCIe buses to output DP video streams, while FPGA chips exchange high-speed serial data with optical transceivers over GTX links and communicate with the main CPU via PCIe interfaces. The VCC_Fiber node is configured as Data-KSP to track signal integrity (SI) metrics and bit error rate (BER) fluctuations. Signal integrity quantifies waveform distortion, jitter and amplitude attenuation of high-speed serial signals under external EMI.

By unifying synchronous sampling over power supply branches, global clock circuits and high-speed data interfaces, the proposed KSP layout realizes full-chain tracking of the typical electromagnetic fault evolution chain: power supply disturbance → timing characteristic degradation → functional failure of communication interfaces. This multi-domain monitoring scheme not only demonstrates reliable engineering practicability for radar electromagnetic test platforms but also constructs a clear physically interpretable mapping between hardware electromagnetic behaviors and extracted features, supporting subsequent deep learning-based fault modeling and causal correlation analysis.

3.2. State-Feedback Adaptive EMS Scanning Strategy

The hardware perception architecture realizes high-fidelity acquisition of electromagnetic signals of digital boards. On this basis, an efficient scanning strategy needs to be designed to accurately locate EMS frequency bands and fault boundaries. Therefore, this section proposes a state-feedback adaptive EMS scanning strategy. This strategy takes the key state response of the board as a feedback signal and dynamically adjusts the scanning resolution and injection parameters according to real-time observation results to efficiently characterize EMS frequency bands and fault boundaries.

The key parameters of the adaptive scanning framework are formulated based on accumulated engineering test experience and statistical characteristics of radar board EMS responses from massive pre-tests:

- (1) Frequency scanning parameters: A coarse scanning step of 20 MHz is adopted to rapidly traverse the full frequency range and capture potential sensitive intervals. Once suspected sensitive frequency regions are identified, the system automatically switches to fine scanning with a 1 MHz frequency step and a 1 dBm power gradient, enabling precise characterization of narrowband sensitive boundaries.
- (2) Signal anomaly discrimination thresholds: For multi-channel KSP monitoring signals, two quantitative criteria are defined: a peak waveform error threshold of 20% and a ripple coefficient threshold of 5%. Frequency points that exceed either threshold are automatically labeled as candidate sensitive positions for subsequent high-resolution fine scanning.

3.2.1. Fast Positioning of Sensitive Frequency Bands (Coarse Scan Stage)

In the initial stage of adaptive scanning, a large-step parameter configuration is used to quickly pre-scan the target frequency band and efficiently screen potential sensitive regions. Specifically, within the preset full frequency range (e.g., 100–400 MHz), fast scanning is performed with a frequency step of 20 MHz, while the injection power is kept within a safe and responsive working range.

During scanning, an online state-feedback discrimination mechanism is embedded into the test procedure. The system real-time acquires the raw response signals of each KSP and compares them with the baseline waveforms under normal operating conditions. A frequency point is automatically marked as an EMS candidate band when the peak waveform deviation exceeds 20% or the ripple coefficient is no less than 5% in any monitoring channel. This two-stage screening strategy enables the coarse-scan stage to rapidly narrow down the frequency search space without relying on prior EMS knowledge, providing reliable initial positioning for the sensitivity characteristic analysis of digital circuit boards.

The 20 MHz coarse frequency step is adopted for fast localization of potential sensitive bands, derived from massive radar board EMS tests. The electromagnetic susceptibility degradation of digital circuit boards exhibits continuous band-dependent features instead of isolated discrete single-frequency responses. Once any sensitive candidate frequency is detected, the system triggers a 1 MHz high-precision fine scan covering the ± 10 MHz adjacent interval. This two-stage nested scanning avoids missed detection of narrowband sensitive regions while guaranteeing overall test efficiency.

Sub-bands outside the ± 10 MHz range of sensitive frequency bands are exempted from fine scanning. Multiple preliminary coarse scanning experiments verify that neither obvious electromagnetic susceptibility responses nor effective fault excitation points can be observed in these two frequency intervals.

3.2.2. Fine Characterization of Fault Boundaries (Fine Scan Stage)

On the basis of the sensitive candidate frequency bands obtained in the coarse scan stage, a high-resolution fine scan test is further carried out to accurately characterize the performance degradation process of the board under electromagnetic stress. The specific strategy involves taking the candidate frequency point as the center, narrowing the frequency scanning step to 1 MHz within its ± 10 MHz range, and gradually adjusting the interference injection power with a gradient of 1 dBm, so as to construct a high-density test grid in the frequency–amplitude two-dimensional parameter space.

All test samples are automatically labeled as healthy, soft fault, or hard fault according to equipment functional integrity and fault recoverability. The detailed classification criteria are provided in Section 4.1.2.

Through the above fine-scan mechanism, the continuous performance degradation curve and critical fault boundary of the board's EMS can be obtained at a low test cost, providing high-quality data support for subsequent electromagnetic fault modeling,

health assessment, and prediction analysis. Experimental results show that the adaptive scanning strategy ensures the accuracy of sensitive boundary characterization while significantly improving the overall test efficiency, which is suitable for rapid electromagnetic characteristic evaluation scenarios of digital boards.

3.3. Closed-Loop Data Enhancement and Feature Mining

Under complex EMI environments, the original test data of digital boards generally suffer from problems such as unbalanced sample distribution, strong signal non-stationarity, and high feature redundancy, which directly affect the training stability and generalization performance of fault diagnosis models. Therefore, this paper systematically optimizes the original data from both the data layer and the feature layer: on the one hand, a hybrid sampling balancing strategy is introduced to alleviate the category bias caused by the difference in the number of samples between normal and fault states; on the other hand, aiming at the multi-scale characteristics of EMI signals in the time and frequency domains, a time–frequency multidimensional feature representation is constructed, and feature optimization is further realized combined with the feature importance evaluation mechanism. This method ensures sufficient expression of fault features while effectively reducing feature dimensions and computational complexity, laying a data and feature foundation for the efficient training and online real-time deployment of subsequent diagnostic models.

3.3.1. Hybrid Sampling Balancing Technique (SMOTE-RU)

During the experiment, due to noise in the acquisition channel, communication interruption, or unstable interference injection, the original data may contain missing values, outliers, and high-frequency noise signals. After completing the collection of coarse-scan and fine-scan experimental data, the original signal sequence is first cleaned, including missing value interpolation, outlier removal, filtering, and noise reduction, to ensure data quality.

To avoid information leakage caused by resampling the entire undivided dataset, we execute dataset partitioning as the first step in preprocessing. The whole original EMS dataset is split into training, validation and test sets according to a fixed proportion. The proposed SMOTE-RU hybrid balancing algorithm is only applied to the training set to generate synthetic minority fault samples. The validation set and test set always maintain the original real sample distribution without any artificial synthetic data, which guarantees objective and unbiased model evaluation results.

In the sample enhancement stage, targeted supplementation of samples is carried out according to the sample distribution ratio after coarse and fine scanning. In the experiment, the number of fault samples is significantly less than that of normal samples. Let the normal sample set and fault sample set be:

$$D_{\text{fault}} \ll D_{\text{nor}} \quad (1)$$

This class imbalance problem easily leads the model to bias toward the majority class, thereby reducing the ability to recognize fault samples. To this end, this paper adopts a hybrid sampling strategy named SMOTE-RU. In the SMOTE process, a neighbor sample is randomly selected from the neighbor set of the fault sample feature vector, the number of SMOTE neighbors $k = 5$, and synthetic samples are generated by linear interpolation:

$$x_{\text{new}} = x_i + \lambda(x_j - x_i), \lambda \sim U(0,1) \quad (2)$$

where x_i is the original fault sample, x_j is the neighbor sample of x_i , λ is a random number between 0 and 1, and $U(0,1)$ is a uniform distribution over (0,1). See Table A1 in Appendix A for the full definition of all symbols and hyperparameters. This process expands the distribution of minority classes in the feature space so that the synthetic samples

can cover the original fault feature region. Subsequently, a random undersampling strategy is adopted for the majority class samples:

$$D'_{nor} \subset D_{nor} \quad (3)$$

The majority class is undersampled to a 1:1 ratio. Finally, a class-balanced training set is constructed. This method avoids information redundancy while effectively enhancing the model's ability to learn rare fault features and improving the generalization performance of the diagnostic model. To intuitively demonstrate the effectiveness of the proposed hybrid balancing strategy, the feature distributions before and after SMOTE-RU balancing are visualized using t-distributed stochastic neighbor embedding (t-SNE), as discussed in Section 4.2.2.

3.3.2. Multidimensional Feature Optimization Mechanism

To characterize the nonstationary signal characteristics under complex EMI environments, this paper constructs a time–frequency multidimensional feature matrix.

1. Time-domain statistical feature extraction.

From the time-domain perspective, five statistical features, including mean, standard deviation, skewness, kurtosis, and energy ratio, are extracted to reflect the overall distribution characteristics of the signal.

2. Frequency-domain principal component feature extraction.

To capture the coupling relationship between the interference injection frequency and the resonance characteristics of the board, a fast Fourier transform (FFT) [22] is performed on the timing signal. By analyzing the spectrum amplitude, the main frequency components and their energy distribution are extracted as frequency-domain features. The frequency resolution of FFT frequency-domain analysis is 1 MHz.

3. Feature importance evaluation and screening.

To reduce the interference of redundant features on model training, a feature optimization mechanism combining correlation coefficients and tree models is adopted. First, the Pearson correlation coefficient [23] between feature f_i and category label y is calculated:

$$\rho_{f_i, y} = \frac{Cov(f_i, y)}{\sigma_{f_i} \sigma_y} \quad (4)$$

where $Cov(f_i, y)$ denotes the covariance between feature f_i and the class label y ; σ_{f_i} denotes the standard deviation of feature f_i across all samples; and σ_y denotes the standard deviation of the class label y . $\rho_{f_i, y}$ is used to measure the linear correlation between features and fault categories. See Table A1 in Appendix A for the full definition of all symbols and hyperparameters.

Meanwhile, a feature importance score based on tree models (such as random forest or gradient boosting tree) is introduced:

$$Importance(f_i) = \sum_{t=1}^T \Delta\{Gini_t(f_i)\} \quad (5)$$

where $\Delta\{Gini_t(f_i)\}$ represents the impurity reduction brought by feature f_i in the t -th decision tree. Through the joint integration of Pearson correlation coefficients and tree-model-derived feature importance metrics, only features with $|\rho| \geq 0.3$ and importance scores no less than 0.05 are preserved, and features with weak correlation or high redundancy are discarded.

This generates highly discriminative feature vectors and drastically cuts computation overhead for downstream electromagnetic fault diagnosis networks. We adopt uniform manifold approximation and projection (UMAP) visualization and Pearson correlation analysis to validate the feature optimization strategy, with detailed results analyzed in Section 4.2.2.

3.4. Complete EMS Data Construction Pipeline

Combining the multi-channel synchronous KSP acquisition hardware architecture (Section 3.1), adaptive coarse–fine frequency scanning scheme (Section 3.2), and SMOTE-RU resampling plus multidimensional time–frequency feature optimization module (Section 3.3), this section integrates all constituent parts into a unified end-to-end EMS data construction workflow. To facilitate full method reproduction and algorithm reproducibility, the complete executable pseudocode for the whole pipeline is attached in Appendix A.2. This pseudocode systematically encapsulates every core functional module introduced in Sections 3.1–3.3, covering hardware initialization, sensitive frequency band localization, sample balancing processing, and discriminative feature screening.

It is worth emphasizing that the dataset generated by this integrated workflow covers diverse equipment operating states instead of merely normal working conditions. During the full electromagnetic interference injection test, raw response waveforms are collected and hierarchically annotated according to hardware functional integrity and fault recoverability. In practical tests, the high-fidelity acquisition unit only provides four sampling channels. Therefore, at least two rounds of measurements are required to capture data from all six measurement points, followed by unified standardization processing of the data acquired in the two rounds.

The core design goal of the proposed integrated framework lies in generating standardized, fault-labeled, high-quality EMS datasets to break the cold-start dilemma of intelligent diagnosis for new radar boards. The established dataset possesses dual application values: it can be archived as a standardized electromagnetic susceptibility database for radar digital circuits and also acts as standardized input data to validate the classification accuracy and cross-board generalization capacity of various diagnostic models in the subsequent experimental section.

4. Experimental Verification and Result Analysis

4.1. Experimental Procedures

4.1.1. Definition of Variable Working Conditions

This subsection systematically defines all tunable experimental parameters supported by the closed-loop EMI injection and multi-channel synchronous acquisition platform. All adjustable working conditions are grouped into four categories: excitation parameters, KSP sampling schemes, tested hardware, and fault severity grades.

(1) EMI Excitation Parameters.

The RF signal source supports continuous tuning of interference frequency and power to build diverse excitation environments. The core test frequency band covers 100–400 MHz, with switchable scanning steps of 20 MHz (coarse scan) and 1 MHz (fine scan) for automatic screening of sensitive frequency points. Interference power ranges from 15 to 35 dBm with a 1 dBm step. Three power intervals are defined to correspond to distinct fault degrees: 15–22.3 dBm for mild soft faults, 22.3–27.8 dBm for moderate soft faults, and above 27.8 dBm for irreversible hard faults.

(2) KSP Sampling Configurations.

The multidimensional KSP monitoring system supports flexible channel combinations across PI-KSP, Timing-KSP and Data-KSP branches. Operators can activate one single branch, any two combined branches, or four KSP channels for synchronous data collection. The ablation experiments in Section 4.4 adopt multiple sampling combinations to quantify how monitoring channel layouts affect feature extraction performance.

(3) Tested Circuit Hardware.

Three heterogeneous radar boards from the same radar platform are selected for cross-domain verification: the source radar timing control board (original dataset carrier), target radar data processing board A with multi-channel high-speed data circuits, and target radar display control board B equipped with an exclusive clock and video bus topology. Obvious discrepancies in power supply, clock distribution and peripheral interfaces among the three boards create typical cross-domain test scenarios.

(4) Fault Severity Grading.

Samples are auto-labeled into three hierarchical states based on post-interference equipment recoverability and BER: healthy state: no observable voltage ripple or clock jitter drift; soft fault: recoverable signal distortion and communication errors, further divided into mild and moderate subclasses; and hard fault: irreversible timing chaos and permanent link breakdown.

Graded fault samples under a variable injection power supply diversified fault features to evaluate the model's classification ability. In brief, the four groups of adjustable parameters form the full variable space of the test platform. All comparative and ablation experiments in this work adopt different parameter combinations to comprehensively validate the stability and cross-hardware generalization of the proposed EMS data construction framework under complex practical test environments.

4.1.2. Standardized Classification Criteria for Health/Soft Fault/Hard Fault Labels

The three sample labels adopted in this paper are formulated by combining practical radar hardware EMC test experience and quantitative BER thresholds collected from the closed-loop electromagnetic injection experimental platform. We take the automatic recoverability of circuit functions after interference withdrawal as the core boundary to distinguish soft faults from hard faults and set fixed BER values as objective quantitative judgment indicators:

- (1) Health: BER remains 0 under all interference excitation levels; no obvious clock jitter, power supply ripple or data transmission distortion is observed on each KSP monitoring channel.
- (2) Soft Fault: $0 < \text{BER} \leq 10^{-3}$. Transient signal distortion and occasional communication errors occur during interference injection. After the EMI source is cut off, the BER recovers to 0 within 3 s, and no permanent damage to power, clock or data circuits is detected. Soft faults are further subdivided into mild ($0 < \text{BER} \leq 10^{-4}$) and moderate ($10^{-4} < \text{BER} \leq 10^{-3}$) categories during actual tests.
- (3) Hard Fault: BER exceeds 10^{-3} . Severe continuous transmission errors persist after removing interference, and circuit timing disorder/abnormal power ripple cannot be self-recovered within 10 s, representing irreversible electromagnetic damage to hardware modules.

4.1.3. Experimental Setup and Experimental Procedures

A radar timing control digital board is chosen to construct the EMS excitation test platform. Following the layout shown in Figure 3, six groups of KSPs are arranged around the power distribution circuits, clock traces and high-speed logic circuits along the board's critical signal paths. The system contains five voltage measurement channels and one reference clock channel to extract the board's electromagnetic signal characteristics. Detailed hardware descriptions of the six KSPs are listed in Table 1. The physical on-site photograph of the complete test setup is presented in Figure 4.

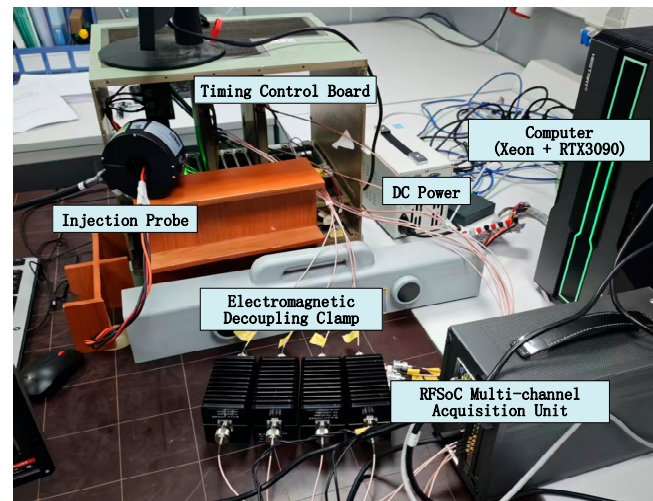


Figure 4. Physical photograph of the EMS excitation test environment.

A single-phase DC power supply supplies the test platform with a 12 V power line, a 3.3 V power line and two ground wires. A decoupling clamp is mounted at the power input port to attenuate external conducted noise. A ring-shaped injection probe is attached to the power cable connected to the timing control board. A power- and frequency-tunable signal generator produces interference waveforms, which are coupled into the target board through the power cable. The physical positions of the injection probe and decoupling clamp are kept identical for all test batches. The full wiring topology of the test platform is displayed in Figure 5.

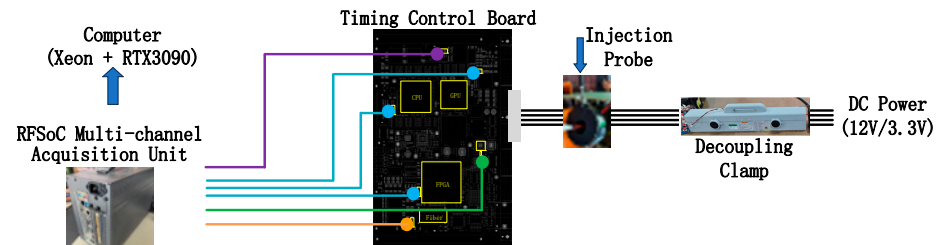


Figure 5. Wiring diagram of the EMS excitation test system.

In the above figure, the RFSoc multi-channel acquisition unit synchronously samples signals from four selected KSPs, which enables full capture of millisecond-level transient electromagnetic responses. In practical tests, four out of the six reserved acquisition channels are selected for sampling, and the captured data are streamed to a computer in real time for storage and postprocessing.

During the experiment, based on the conducted time-domain susceptibility test and the “5 + X” standard signal set [24] EMS test system, the interference characteristic data collection test is carried out on the radar digital module, and the set monitoring points are monitored to obtain the data changes at each point. The data samples are the electromagnetic characteristic parameter values recorded by six KSPs after the injection of the specified interference test, including frequency, voltage, and monitoring signal strength.

In the coarse scan stage, the scanning frequency range is 100–400 MHz, 16 frequency points are scanned, the frequency step is 20 MHz, the power is set to medium power 25 dBm, and the total test time is 8 min. The central frequency of the obtained sensitive frequency band is centered at 135 MHz, and the sensitive frequency band ranges from 110 MHz to 150 MHz. The 110–125 MHz and 145–150 MHz sub-bands are excluded from fine scanning.

In the fine scan stage, high-precision scanning is focused on the scanning frequency range of 125–145 MHz (± 10 MHz of the sensitive area), with a frequency step of 1 MHz, a power gradient of 1 dBm (15–35 dBm), $21 \times 21 = 441$ scanning points, and a test time of 35 min (total time consumed in all scanning stages is 43 min), focusing on depicting the fault boundary curve. Combined with the three-level fault status, the data are automatically labeled to determine the fault trigger critical frequency of 134.5 MHz (accuracy ± 0.5 MHz) and the fault power thresholds of 22.3 dBm (functional degradation) and 27.8 dBm (system failure).

During the experiment, the operating statuses of the internal FPGA clock, DP display, Ethernet port and Serial port of the board under different interference signals are recorded in real time. The electromagnetic fault statuses of the board are documented according to the severity of the observed susceptibility phenomena, as specified in Table 2:

Table 2. Classification of EMS phenomena of the timing control board.

EMS Level	Soft Fault *(Mild)	Soft Fault *(Moderate)	Hard Fault *
Clock (CLK)	Occasional transmission error	Functional abnormality; Recoverable jitter	Functional abnormality; Unrecoverable jitter
Display (DP)	Local water ripples	Entire screen water ripples	Screen distortion and blackout
Ethernet Port	Increased latency	Intermittent disconnection; Recoverable	Persistent disconnection; Unrecoverable
Serial Port	Increased latency	Intermittent disconnection; Recoverable	Persistent disconnection; Unrecoverable

*: All classification rules for electromagnetic susceptibility phenomena are formulated by integrating official EMC test specifications and long-term engineering experience of radar hardware measurements and calibrated based on quantitative bit error rate (BER) data acquired from the test platform.

At each test point, the high-fidelity multi-channel acquisition unit continuously samples at a rate of 4×10^{-10} s per point for a duration of 4×10^{-6} s, yielding 10,000 time-domain sample points per record. To ensure measurement stability, each test configuration is repeated 10 times, and the resulting waveforms are averaged before feature extraction. In total, 10,000 labeled data records are obtained across all scanned frequency–power combinations, each represented by a feature vector derived from the 10,000-point waveform.

A standardized sequential data processing pipeline is implemented to rule out information leakage. The full dataset with 10,000 samples is first divided into training, validation and test subsets at a 7:1:2 ratio using stratified sampling to keep a balanced fault category distribution in each partition. All data augmentation and feature processing operations, including SMOTE-RU hybrid balancing and the joint Pearson correlation and tree-based feature selection scheme, are exclusively applied to the training set. The validation set is used for model screening and hyperparameter optimization via 5-fold cross-validation, and the untouched test set is only utilized once to evaluate the final model performance after training convergence. No statistical information or synthetic samples derived from validation/test data participate in feature construction or model training.

This hierarchical structure enables data to be organized and accessed according to different interference frequencies and monitoring points. Meanwhile, the electromagnetic fault status information of the board provides additional context for the data, which can be used for subsequent analysis and modeling.

In the constructed EMS dataset, all samples are labeled according to the above three fault severity levels. The health category accounts for approximately 50% of the total samples, representing normal operating conditions without observable electromagnetic degradation. The soft fault class, which covers both mild and moderate soft fault

subcategories, accounts for around 20% of the total samples. Such samples correspond to recoverable functional anomalies induced by EMI, including transient communication interruptions, timing drift, and temporary signal distortion. The hard fault category accounts for approximately 30%, representing severe and unrecoverable failures, including persistent timing disorder, abnormal power ripple, or complete link failure. Therefore, the dataset contains not only normal operating samples but also fault-related samples with different severity levels and is used as the basis for both EMS database construction and subsequent fault identification experiments.

To provide an intuitive understanding of the signal characteristics within the constructed dataset, Figure 6 presents a representative three-panel visualization of a single data sample, using the VCC_FPGA monitoring channel under 135 MHz sinusoidal interference injection as an example.

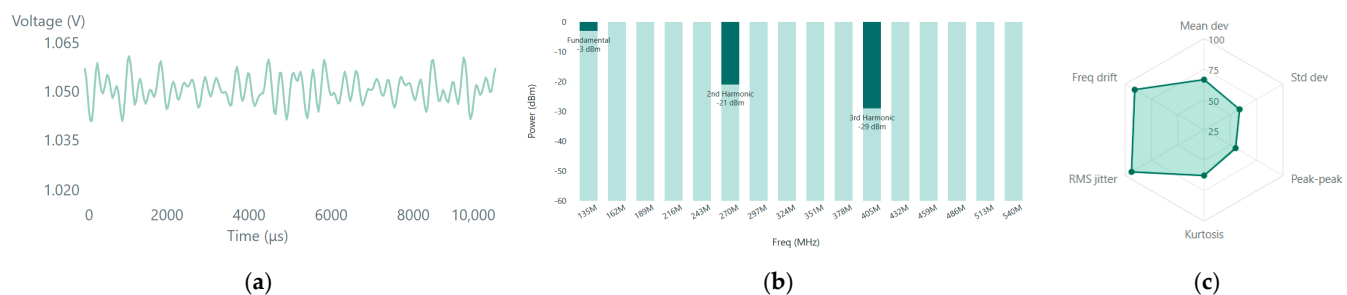


Figure 6. Three-panel visualization of a single data sample. (a) Time-domain voltage waveform; (b) FFT power spectrum; (c) statistical feature fingerprint radar chart.

The above figure shows the time-domain voltage waveform, the frequency-domain power spectrum and the statistical feature fingerprint as a radar chart, with six axes representing key statistical descriptors: mean deviation, standard deviation, peak-to-peak amplitude, kurtosis, RMS jitter, and dominant frequency drift.

Figure 6a displays the time-domain voltage waveform sampled at a 0.4 ns interval over a total duration of 4 μs, yielding 10,000 sampling points. The signal oscillates periodically around 1.000 V with prominent jitter and amplitude modulation, intuitively reflecting steady EMI superimposed on the clock power rail. Figure 6b presents the corresponding FFT power spectrum, which features a dominant peak at 135 MHz (−3 dBm) and a secondary second harmonic component at 270 MHz (−21 dBm). Spectral energy concentrates heavily on the injection frequency, enabling diagnostic models to efficiently extract frequency-dependent EMS signatures. Figure 6c illustrates the radar-style feature fingerprint for clock faults. It delivers distinctly elevated RMS jitter and frequency drift metrics, constructing a unique feature pattern easily separated from other fault modes, which validates the superiority of the multi-modal feature extraction scheme.

4.2. Validity Analysis

4.2.1. Dataset Evaluation Under Different Models

To comprehensively evaluate the effectiveness of the proposed dataset construction method, this paper selects a set of representative and gradient-different models as test objects, covering traditional machine learning, modern gradient boosting trees, classic deep sequence models, residual structures, and current mainstream advanced attention/self-attention mechanisms, as follows:

1. Traditional machine learning and modern tree models.

- SVM: The radial basis function (RBF) kernel is still used, and the input is the optimized frequency-domain features, serving as a classic linear/nonlinear separation baseline.
 - XGBoost: The default optimal parameters (tree depth 6, learning rate 0.1, subsample 0.8) are adopted, representing the strong generalization ability of gradient boosting trees.
 - LightGBM: Histogram acceleration + GOSS + EFB optimization, 500 trees, and leaf node split gain priority are used, reflecting the performance of modern efficient GBDT.
2. Mainstream deep sequence and convolutional models.
 - 1D-CNN (baseline) [25]: Multi-layer single-scale 1D convolution + global average pooling, without any attention or residual connections, used to verify the baseline effect of the pure convolution backbone.
 - Bi-LSTM: Bidirectional long short-term memory network (hidden layer dimension 256), capturing context-dependent timing dependencies.
 - ResNet-1D [26]: Based on a one-dimensional residual structure (BasicBlock $\times$ 3 groups), effectively alleviating the gradient disappearance problem of deep networks.
 3. Advanced attention and self-attention mechanism models.
 - CBAM-Net [27]: The convolutional block attention module with channel attention + spatial attention is embedded on the basis of ResNet-1D to verify the synergistic gain of dual attention.
 - ECA-Net [28]: Adopts efficient channel attention (ECA), with almost no additional parameter overhead, to achieve lightweight channel recalibration.
 - Transformer-1D [29]: Based on 1D convolutional position encoding + multi-head self-attention (four heads, hidden dimension 256), stacking three Transformer Encoder blocks to investigate the performance of pure self-attention in electromagnetic signal tasks.

Three core indicators in electromagnetic scenario classification tasks are used for comprehensive comparison: accuracy, macro-average F1 score, and area under the ROC curve (AUC). Table 3 presents the comparative classification performance, parameter counts and FLOPs of all tested models.

Table 3. Comparison of computational complexity and fault diagnosis performance among different models.

Model Category	Model Name	Parameters (M)	FLOPs (M)	ACC *(%)	F1 *(%)	AUC *(%)
Traditional ML	SVM (RBF)	-	-	85.1	82.7	86.0
	XGBoost	-	-	88.0	85.9	90.2
	LightGBM	-	-	89.3	86.7	91.0
Deep Learning Baseline	1D-CNN (baseline)	0.43	18.6	92.7	89.5	94.1
	Bi-LSTM	1.58	42.7	90.4	88.2	93.0
	ResNet-1D	0.92	31.5	92.6	90.4	94.4
Attention/Advanced Module	ECA-Net	0.95	32.1	94.3	91.5	96.3
	Transformer-1D	2.36	58.4	94.9	92.8	97.1
	CBAM-Net	1.08	35.9	95.0	91.7	96.0

*: ACC, accuracy; F1, F1-score; AUC, area under curve.

The test results show that even the basic SVM model can stably achieve an accuracy of about 85.1%, which again confirms that the proposed feature and data enhancement strategies significantly improve the interclass separability and intraclass compactness of

the original data. Convolutional models provide efficient local feature extraction, recurrent models capture temporal dependencies at the cost of higher sequential computation, and attention-enhanced models further improve diagnostic accuracy by emphasizing fault-sensitive channels and temporal regions. Among them, CBAM-Net achieves the best overall accuracy, while ECA-Net provides the most favorable trade-off between model complexity and diagnostic performance. The Transformer-1D model has the best performance, mainly due to the self-attention mechanism that can accurately capture the timing dependence of nonstationary electromagnetic signals. The SVM model has relatively low performance, mainly because the linear kernel has difficulty fitting the nonlinear characteristics of complex electromagnetic faults.

Figure 7 illustrates the convergence curves of six mainstream deep learning models (1D-CNN, Bi-LSTM, 1D-ResNet, CBAM-Net, ECA-Net, Transformer-1D) trained on the EMS feature dataset. All models converge stably within 80 epochs, with validation loss declining synchronously with training loss. This reveals the reasonable feature distribution of the dataset and the absence of severe overfitting. Attention-augmented models (CBAM, Transformer) maintain lower loss at late training stages and deliver substantially higher ACC and AUC than plain CNNs. CBAM-Net achieves the minimal final validation loss of 0.108.

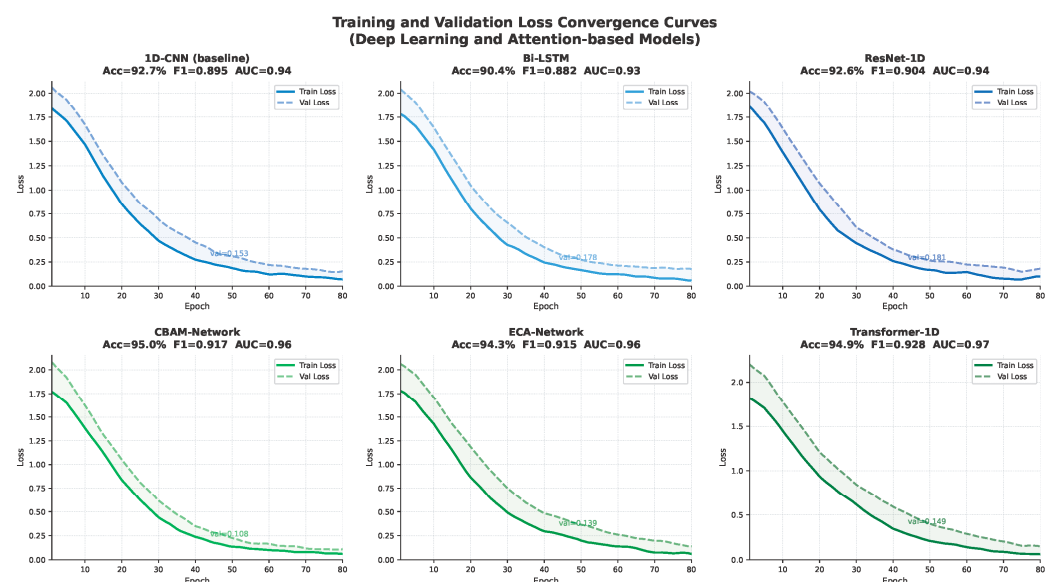


Figure 7. Training and validation loss convergence curves of deep learning and attention mechanism models.

Figure 8 shows receiver operating characteristic (ROC) curves comparing the overall classification capacity of traditional machine learning and advanced attention-based networks. All curves sit clearly above the random classification baseline. The SVM baseline only obtains an AUC of 0.86, while every attention-enhanced model delivers an AUC over 0.969; 1D-Transformer achieves the highest AUC value, proving that self-attention mechanisms excel at capturing long-term temporal correlations among nonstationary electromagnetic signals.

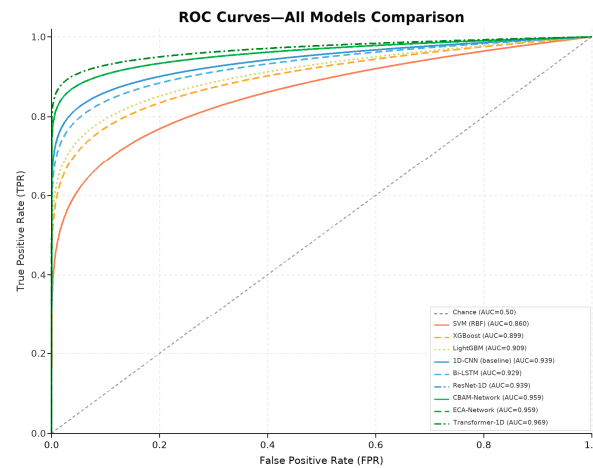


Figure 8. ROC curve comparison of the classification performance of different diagnostic models.

Figure 9 provides precision–recall (P-R) curves to evaluate model performance on imbalanced EMS samples processed by the SMOTE-RU hybrid resampling algorithm. The average precision under each P-R curve reflects the precision–recall balance of each classifier. Even classic gradient boosting models (XGBoost, LightGBM) achieve F1 scores above 0.85. Combined with the ROC results, it can be concluded that the proposed multi-dimensional feature optimization and sample balancing pipeline significantly strengthens the distinguishability of faint electromagnetic fault features.

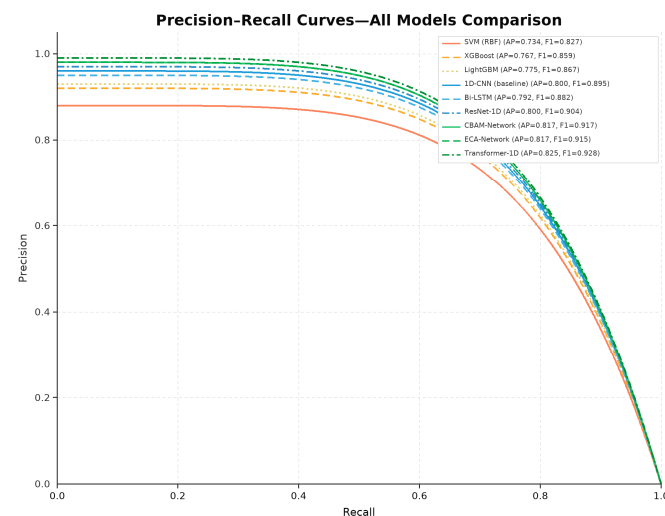


Figure 9. Precision–recall (P-R) curve comparison of the classification performance of all models.

Overall, all baseline models trained on the dataset generated by our framework reach classification accuracy above 85.1%, and the optimal attention-augmented model hits 95.0%. Consistent with the curve comparison results, attention architectures outperform conventional machine learning methods, and 1D-Transformer exhibits the strongest fitting capability for time-varying electromagnetic signals. The above results fully verify that our method effectively addresses the cold-start diagnosis challenge of newly designed radar circuit boards.

4.2.2. Feature Space Visualization Analysis

To further verify the effectiveness of the proposed SMOTE-RU balancing strategy and multidimensional feature optimization mechanism, t-SNE and UMAP are employed to project the high-dimensional EMS features into a two-dimensional space for visualization.

Figure 10 presents t-SNE and UMAP dimensionality reduction visualizations to separately verify the benefits of SMOTE-RU resampling and multidimensional feature optimization.

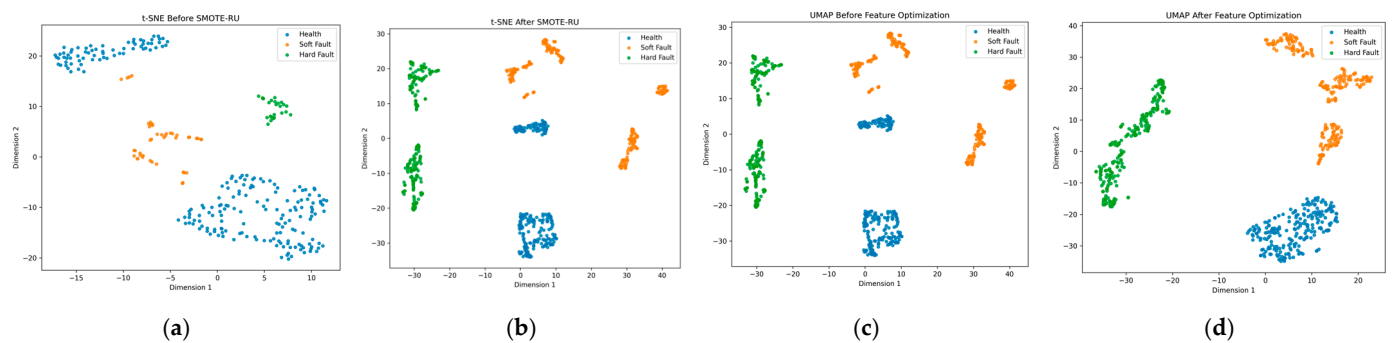


Figure 10. Feature space visualization before and after SMOTE-RU balancing and multidimensional feature optimization. (a) Original feature distribution before SMOTE-RU balancing. (b) Feature distribution after SMOTE-RU balancing. (c) UMAP visualization before feature optimization. (d) UMAP visualization after feature optimization.

Figure 10a illustrates the raw feature distribution without sample balancing: minority fault samples are extremely sparse and spatially separated from majority normal samples, which reflects severe native class imbalance. After applying the SMOTE-RU hybrid resampling strategy, Figure 10b shows densely supplemented minority fault clusters, while clear interclass boundaries remain intact. This confirms that the proposed balancing algorithm sufficiently enriches scarce fault samples without inducing overlapping feature regions across different categories. Figure 10c, d compare feature distributions before and after multidimensional time–frequency feature screening. By removing redundant feature dimensions, the optimized feature set achieves tighter intraclass aggregation and larger separable gaps between different fault categories.

Taken together, the two sets of visualization results jointly demonstrate two core improvements brought by our pipeline: first, SMOTE-RU effectively alleviates sample imbalance via synthetic minority sample generation; second, multidimensional feature optimization eliminates redundant information and enhances interclass separability. The combined two modules remarkably boost feature discriminability for electromagnetic fault classification tasks.

Figure 11 presents the Pearson correlation matrices before and after feature optimization. The Pearson correlation coefficient ranges from -1 to 1 , where values close to ± 1 denote strong linear correlation and values near 0 indicate weak correlation. As shown in Figure 11a, several statistical descriptors exhibit strong correlations, indicating the existence of redundant information within the original feature space. After the proposed feature optimization, as illustrated in Figure 11b, redundant correlations are significantly reduced while representative statistical features are retained. Consequently, the optimized feature set becomes more compact and informative, improving the robustness and computational efficiency of the subsequent fault diagnosis model.

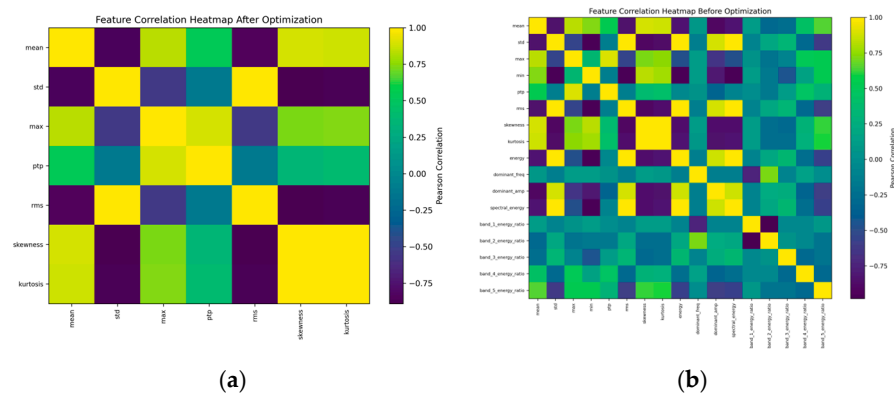


Figure 11. Pearson correlation matrices before and after feature optimization. (a) Correlation heatmap of original statistical features. (b) Correlation heatmap of optimized features via the proposed filtering method.

Overall, the visualization results presented in Figures 10 and 11 provide intuitive evidence that the proposed SMOTE-RU balancing and multidimensional feature optimization effectively suppress redundant information, enhance minority-class representation, and construct a more discriminative EMS feature space for fault diagnosis.

4.3. Computational Efficiency and Deployment Analysis

To evaluate the engineering applicability of the proposed data acquisition and preprocessing strategy, computational efficiency analyses are conducted from the perspectives of scanning time, feature optimization complexity, and diagnostic inference efficiency. The experiments are performed on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), an Intel Xeon CPU, and 64 GB RAM under the PyTorch 2.8.0 environment.

Table 4 summarizes the computational efficiency of different processing stages. Compared with traditional uniform equal-step scanning, the proposed coarse–fine adaptive scanning strategy reduces the total scanning time from 81 min to 43 min, achieving approximately 46.9% improvement in acquisition efficiency. This improvement is mainly attributed to the adaptive refinement mechanism, which focuses scanning resources on fault-sensitive frequency regions instead of uniformly traversing the entire spectrum.

Table 4. Efficiency metrics of each processing module.

Processing Stage	Time Consumption	Description
Uniform Equal-Step Scanning	81 min	Full-band fixed-step scanning
Adaptive Coarse–Fine Scanning	8 + 35 = 43 min	Adaptive fault-sensitive scanning
Raw Feature Extraction	100% Features	All original features retained
Feature Optimization	53% Features Retained	Pearson + tree model selection
Diagnostic Inference	6.1 ms/sample	Measured on RTX 3090 platform

In addition, the proposed Pearson correlation and tree-model-based feature optimization strategy effectively removes redundant and low-relevance features. After feature selection, only 53% of the original features are retained, resulting in an approximately 47.0% reduction in feature processing complexity while simultaneously improving classification accuracy. These results indicate that the optimized feature subset preserves the most discriminative information while suppressing feature redundancy and noise interference.

To further evaluate practical deployment capability, diagnostic inference latency is also measured using the collected dataset. The experimental results show that the average inference time is only 6.1 ms per sample on the RTX 3090 platform (as illustrated in Figure 4), satisfying the real-time requirements of online electromagnetic fault diagnosis applications.

Overall, the proposed data acquisition and preprocessing strategy achieves a favorable balance between computational efficiency, diagnostic accuracy, and practical deployment capability for large-scale radar timing signal analysis tasks.

4.4. Ablation Study

This section takes CBAM-Net as the unified baseline backbone and carries out two groups of ablation experiments in sequence. The first group ablates multi-channel KSP input configurations to clarify the value of multi-channel synchronous multi-domain sensing. The second group quantifies the independent performance gain brought by three core modules of the EMS data construction pipeline: adaptive coarse-to-fine scanning, SMOTE-RU hybrid resampling and Pearson-tree joint feature optimization.

4.4.1. Ablation on Multi-Channel KSP Input Configurations

To quantitatively analyze the independent contribution of each KSP branch to fault diagnosis accuracy, this paper designs six ablation control experiments with KSP measuring points added step by step. Limited by the quantity of acquisition channels, simultaneous sampling of all six KSP signals cannot be realized. As the primary power supply node of the circuit board, the 12 V power KSP usually presents the most obvious electromagnetic susceptibility features. Accordingly, the experiments start with VCC_12V, then sequentially incorporate the fiber channel and 25 MHz clock signal, followed by separate voltage measuring points of GPU, FPGA and CPU, forming six groups of tests in total. The complete experimental metrics are listed in Table 5.

Table 5. Ablation results of multi-channel KSP input combinations.

Experiment ID	VCC_CPU	VCC_FPGA	VCC_GPU	VCC_12V	VCC_Fiber	CLK_25M	ACC (%)	F1-Score (%)	AUC (%)
Test 1	X *	X	X	✓ *	X	X	88.3	87.1	90.4
Test 2	X	X	X	✓	✓	X	89.7	88.5	91.6
Test 3	X	X	X	✓	✓	✓	90.1	89.6	92.1
Test 4	X	X	✓	✓	✓	✓	92.5	91.6	94.4
Test 5	X	✓	X	✓	✓	✓	93.7	92.0	93.3
Test 6	✓	X	X	✓	✓	✓	94.1	92.7	96.1

*: ✓, KSP input used; X, KSP input unused.

The model's diagnostic performance improves continuously as more KSP measuring points are included in input branches. The baseline scheme that only adopts the single 12 V power measuring point (Test 1) achieves the lowest accuracy of merely 88.3%. After adding the fiber channel measuring point (Test 2) and the 25 MHz clock measuring point (Test 3) one after another, the accuracy rises to 89.7% and 90.1%, respectively. Further introducing the GPU voltage branch (Test 4) and the FPGA voltage branch (Test 5) pushes the accuracy to 92.5% and 93.7%. Test 6 adopts the configuration with VCC_12V, VCC_Fiber, CLK_25M and VCC_GPU activated, which achieves the optimal diagnostic performance with an accuracy of 94.1%, an F1-score of 92.7% and an AUC value of 96.1%. This result proves that the separate power domains of CPU, FPGA and GPU carry differentiated electromagnetic fault features.

The steady growth in diagnosis accuracy with the increasing number of measuring points reveals that multi-channel synchronous KSP sampling can fuse complementary electromagnetic features from power supply, optical communication and clock circuits. Monitoring with only a single branch cannot fully reflect the overall operating state of the circuit board, which further verifies that full-dimensional multi-physical KSP deployment is an essential condition for high-precision electromagnetic susceptibility fault diagnosis.

4.4.2. Ablation on Core Modules of EMS Data Construction Pipeline

Based on the high-quality dataset collected from all six KSPs described in Section 4.1.3, we further conduct ablation experiments on the three core modules of the proposed data generation pipeline. Three incomplete variants are designed for comparison: full-band equal-step scanning replacing adaptive coarse–fine scanning; dataset training without SMOTE-RU balancing; and raw feature input without Pearson–tree joint screening. The quantitative results are summarized in Table 6.

Table 6. Ablation results of core modules in the EMS data construction pipeline.

Method Variant	Adaptive Scanning	SMOTE-RU Balancing	Feature Optimization	ACC (%)	F1-Score (%)	AUC (%)
Equal-step Scanning	X *	✓ *	✓	93.1	92.4	95.0
Without SMOTE-RU	✓	X	✓	91.7	89.2	93.3
Without Feature Optimization	✓	✓	X	92.5	91.6	94.4
Full Proposed Pipeline	✓	✓	✓	95.0	91.7	96.0

*: ✓, functional module used; X, functional module unused.

Compared with the complete pipeline, each removed module brings clear performance degradation, and the specific quantitative gains are analyzed as follows:

Adaptive coarse-to-fine scanning: Replacing it with equal-step full-band scanning cuts ACC by 1.9% and AUC by 1.0%. The two-stage scanning accurately locks narrow sensitive bands and refines fault boundaries, bringing obvious performance improvement.

SMOTE-RU hybrid resampling: This module produces the most prominent promotion. Without balancing, the F1-score decreases from 91.7% to 89.2%, a drop of 2.5%. Severe sample imbalance suppresses the identification ability of minority fault samples, and the hybrid sampling strategy effectively alleviates this defect.

Pearson–tree joint feature optimization: Discarding the screening module reduces ACC by 2.5% and AUC by 1.6%. Redundant feature dimensions interfere with model training; the filtering step compresses feature space and simultaneously improves classification accuracy and generalization.

In conclusion, adaptive scanning, SMOTE-RU sample balancing and multidimensional feature optimization have complementary functions. The collaborative use of all three modules achieves the optimal fault diagnosis performance.

4.5. Cross-Board Validation

To verify the generalization performance of the proposed data construction framework, the full pipeline consisting of KSP configuration, adaptive coarse-to-fine frequency scanning, SMOTE-RU sample balancing and time–frequency feature optimization is implemented on two heterogeneous circuit boards of the same radar platform, namely, the display control board and data processing board. All boards share the unified EMI injection scheme and a 100–400 MHz test frequency band, yet they differ greatly in circuit architecture, functional logic and EMS response characteristics, which builds a convincing cross-board validation scenario.

Table 7 systematically compares the core hardware indicators of the source timing control board and two target boards. Although all boards adopt the same 3.3 V/12 V power supply and three-class KSP monitoring system, prominent structural gaps remain in core processors, clock frequency configurations and peripheral interface circuits. The source board equips the FPGA + CPU hybrid architecture with DP, Ethernet and optical fiber interfaces; Target Board A replaces the CPU with a DSP and introduces a dedicated RapidIO high-speed bus; and Target Board B adopts a CPU + GPU computing framework and rearranges the connection topology of display and communication interfaces.

Table 7. Hardware specifications and structural discrepancies between source and target radar boards.

Item	Source Board	Target Board A	Target Board B
Board Function	Radar timing control	Radar data processing	Radar display control
Main Processor	FPGA + CPU	FPGA + DSP	CPU + GPU
Power Input	3.3 V/12 V	3.3 V/12 V	3.3 V/12 V
Reference Clock	25/48/100 MHz	100/125 MHz	48/100 MHz
Key Interfaces	Ethernet, DP, optical fiber	RapidIO, optical fiber	Ethernet, DP
KSP Type	PI-KSP, Timing-KSP, Data-KSP	PI-KSP, Timing-KSP, Data-KSP	PI-KSP, Timing-KSP, Data-KSP
Differences	-	Different processor and interface	Different processor

Under consistent EMI excitation frequency and power, the above hardware discrepancies create unique electromagnetic coupling paths for each board, leading to completely different EMS fault response signals. Such obvious hardware heterogeneity raises strict requirements for the generalization capacity of diagnostic algorithms. Subsequent cross-board experimental results will prove that the multi-KSP acquisition and feature optimization framework proposed in this paper can extract universal electromagnetic fault features irrelevant to specific circuit structures. It can maintain stable high-precision diagnosis effects on hardware with disparate architectures, which fully verifies its remarkable cross-platform generalization and transfer learning performance.

For both target boards, EMS datasets are generated in strict accordance with the unified experimental criteria provided in Section 4.1. Interference injection power is split into three tiers bounded by 15 dBm, 22.3 dBm and 27.8 dBm, corresponding to mild soft faults, moderate soft faults and irreversible hard faults in sequence. Hierarchical contrast experiments are conducted to quantify the influence of interference power (fault excitation amplitude) on cross-board recognition performance. Two evaluation schemes are set to test the generalization capability of the constructed dataset:

- (1) Direct cross-board inference: The model trained on the source timing control board is directly deployed without any parameter adjustment;
- (2) Lightweight fine-tuning: The feature extraction backbone is frozen, and only the final classification layer is retrained using merely 10% of the labeled samples from each target board.

All cross-domain test results are aggregated in Table 8.

Table 8. Cross-board diagnostic accuracy under three interference power grades.

Target Circuit Board	Fault Severity (Injection Power)	Accuracy Without Fine-Tuning (%)	Accuracy After Fine-Tuning (10% Target Samples, %)
Data Processing Board	Mild Soft Fault	76.5	84.1
Data Processing Board	Moderate Soft Fault	86.9	91.8
Data Processing Board	Hard Fault	81.4	89.7
Display Control Board	Mild Soft Fault	78.2	85.7
Display Control Board	Moderate Soft Fault	88.1	91.2
Display Control Board	Hard Fault	83.6	90.5

The above table records the cross-board classification accuracy of two target boards under three interference power levels and compares the performance gap between direct inference and lightweight fine-tuning with only 10% of the target labeled data. The performance rules of three fault grades are summarized as follows: Mild soft faults (15–22.3 dBm): Weak interference only generates obscure fault features. Direct inference achieves only 76.5% and 78.2% accuracy on the two boards, respectively, with an accuracy gain of roughly 7–8 percentage points after fine-tuning. Low feature distinguishability makes mild soft faults the biggest obstacle to cross-domain diagnosis. Moderate soft faults (22.3–

27.8 dBm): Obvious signal distortion forms highly matched feature distributions between source and target boards. The direct inference accuracy exceeds 86.9%, and both boards obtain over 91% accuracy after fine-tuning, representing the optimal cross-board generalization interval. Hard faults (>27.8 dBm): Severe hardware damage brings salient fault modes, but inconsistent hardware failure mechanisms among different boards introduce obvious domain offsets. The untuned accuracy ranges from 81.4% to 83.6%, and the index climbs to nearly 90% after fine-tuning.

In general, cross-board diagnosis accuracy changes nonlinearly with injection power. It reaches the maximum value under moderate interference and decreases for mild and hard fault conditions. Mild soft faults become the main limiting factor of cross-domain identification. Even with only a small proportion of target samples for fine-tuning, the proposed data construction framework maintains favorable generalization on heterogeneous radar circuit hardware.

5. Conclusions

This paper proposes an efficient, rapid construction workflow for EMS datasets targeting radar timing control digital boards. The core technical contributions and experimental outcomes are concluded as follows:

1. A two-stage adaptive coarse–fine scanning strategy is developed specifically for radar digital board testing. Different from traditional full-band equal-step sweeping that suffers from low efficiency and missing narrowband sensitive frequency points, this scheme adopts a 20 MHz coarse step to quickly locate candidate frequency bands and a 1 MHz fine step to accurately depict fault boundaries. It shortens the overall test time from several hours to 43 min, with fault frequency positioning precision up to ± 0.5 MHz.
2. A unified multidimensional EMS data processing pipeline is constructed, consisting of multi-domain synchronous KSP sensing, SMOTE-RU hybrid resampling and time–frequency feature screening. This framework overcomes three typical defects of single-channel monitoring: incomplete signal acquisition, severe sample imbalance and redundant feature dimensions. It achieves full-spectrum holographic collection of electromagnetic signals, and the optimized feature subset halves the computational burden of subsequent diagnostic models.
3. Cross-hardware tests on two heterogeneous radar boards under three interference power levels demonstrate that lightweight fine-tuning using only 10% of the target samples significantly improves cross-domain classification accuracy. Moderate interference yields the best generalization performance, whereas mild soft faults with ambiguous features act as the main constraint for cross-board recognition. The diagnosis accuracy varies nonlinearly with excitation power, which validates the strong transferability of our method at a low retraining cost.
4. A hierarchically annotated EMS dataset covering multiple equipment degradation states is built. Unlike conventional single-criterion EMS data, the samples are labeled as mild soft fault, moderate soft fault and irreversible hard fault according to hardware damage severity. Such multi-grade labeled data can be archived as a standardized EMS database and also serve high-precision intelligent diagnosis under variable EMI intensities.

From an engineering perspective, the proposed integrated pipeline eliminates the cold-start barrier of EMS fault diagnosis for newly designed radar circuit boards. It can be directly applied to electromagnetic anomaly monitoring and active electromagnetic protection of brand-new digital hardware. In summary, this workflow offers a low-cost, standardized EMS testing solution for radar boards without sufficient historical fault samples.

Limitations of This Work:

1. Limited test environment: All cross-board experiments are conducted on a single radar platform with fixed sinusoidal interference and static KSP layouts. Its generalization ability on cross-system radar boards, public EMC benchmark datasets and diverse interference waveforms (pulse, swept, broadband) remains unvalidated.
2. Rigid monitoring setup: The KSP deployment scheme is fixed throughout the experiments, lacking self-adaptive adjustment logic to fit circuit boards with different topologies.

Future Research Directions:

1. Investigate feature extraction algorithms to capture inter-board electromagnetic coupling characteristics for multi-card collaborative radar systems.
2. Expand dataset coverage by incorporating complex excitation sources such as pulse and broadband interference to enrich fault diversity.
3. Design an online incremental data fusion algorithm to mitigate data distribution drift in real-time EMS fault diagnosis tasks.

Author Contributions: Conceptualization, P.L.; methodology, P.L.; validation, Z.W.; formal analysis, H.W.; investigation, P.L.; resources, B.J.; data curation, Z.W.; writing—original draft preparation, P.L.; writing—review and editing, P.L.; visualization, Z.W.; supervision, W.Y.; project administration, P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Acknowledgments: The authors would like to express their gratitude to Qiang Ren and Bing Li, from the EMC Laboratory of Beihang University, for their valuable help and support.

Conflicts of Interest: Pengfei Liu, Ziyang Wang and Bo Jiang are employees of the East-China Research Institute of Computer Technology. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EMS	Electromagnetic Susceptibility
EMC	Electromagnetic Compatibility
EMI	Electromagnetic Interference
CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
PCB	Printed Circuit Board
RFSoc	Radio Frequency System-on-Chip
KSP	Key Signal Point
PI	Power Integrity
Timing-KSP	Timing Reference KSP
Data-KSP	Data Interaction KSP
LDO	Low Dropout Regulator
t-SNE	t-Distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation and Projection
PLL	Phase Locked Loop
FFT	Fast Fourier Transform
SI	Signal Integrity
BER	Bit Error Rate

AUC	Area Under the ROC Curve
RBF	Radial Basis Function
SMOTE	Synthetic Minority Oversampling Technique
SMOTE-RU	Synthetic Minority Oversampling Technique and Random Undersampling

Appendix A

Appendix A.1. Unified Mathematical Symbols and Dataset Variable Notations

Unified notations for mathematical symbols and dataset variables are used in this manuscript:

Table A1. Unified notation for dataset- and sample-related variables.

Symbol	Definition	Unit/Value Range	Corresponding Equation & Section
D_{nor}	Set of normal operating samples	/	Equation (1), Section 3.3.1
D_{fault}	Set of electromagnetic fault samples	/	Equation (1), Section 3.3.1
D'_{nor}	Balanced normal sample subset after random undersampling	/	Equation (3), Section 3.3.1
x_i	Original feature vector of a single fault sample	Multidimensional feature	Equation (2), Section 3.3.1
x_j	Neighbor sample of in feature space	Multidimensional feature	Equation (2), Section 3.3.1
x_{new}	Synthetic fault sample generated by SMOTE interpolation	Multidimensional feature	Equation (2), Section 3.3.1
λ	Random interpolation weight for SMOTE	Uniform distribution	Equation (2), Section 3.3.1
y	Category label of sample (health/soft fault/hard fault)	Integer label	Equation (4), Section 3.3.2
f_i	Single-dimensional raw feature extracted from KSP signals	Scalar	Equation (4), Section 3.3.2

Appendix A.2. Pseudocode of the Proposed EMS Data Construction End-to-End Pipeline

To facilitate reproducibility, we present the complete pseudocode of the proposed EMS data construction pipeline as a unified end-to-end workflow.

Algorithm A1 Complete EMS Data Construction and Fault Diagnosis Pipeline

INPUT:

Frequency range $[f_{min}, f_{max}] = [100, 400]$ MHz;
 Power range $[P_{min}, P_{max}] = [15, 35]$ dBm; coarse step $\Delta f_c = 20$ Hz;
 Fine step $\Delta f_f = 1$ MHz; power gradient $\Delta P = 1$ dBm;
 SMOTE neighbors $k = 5$; correlation threshold $\rho = 0.3$;
 Feature importance threshold = 0.05.

OUTPUT:

Raw dataset D ; cleaned dataset D_{clean} ; balanced dataset $D_{balanced}$;
 Optimized dataset D_{opt} ; metrics $\{ACC, F1, AUC\}$.

// Phase 1: KSP Configuration and Hardware Initialization (Section 3.1)

1. Configure KSP monitoring: PI-KSP (VCC_12V, VCC_CPU, VCC_FPGA, VCC_GPU), Timing-KSP (CLK_25M), Data-KSP (VCC_Fiber);
2. Initialize RFSoc Acquisition Unit: 2.5 GSa/s sampling, 14-bit resolution, picosecond-level multi-channel synchronization;
3. Acquire normal-state baseline waveforms for all KSPs;

// Phase 2: Coarse Scan (Section 3.2.1)

4. Initialize candidate sensitive frequency set: $S_c \leftarrow \{\}$;

```

5. FOR  $f = f_{min}$  TO  $f_{max}$  WITH STEP  $\Delta f_c = 20$  MHz DO
6.   Inject interference at frequency  $f$ , power  $P = 25$  dBm;
7.   Collect KSP response: average over 10 repeated measurements  $\rightarrow W_f$ ;
8.   IF peak_error ( $W_f, W_{base}$ )  $\geq 20\%$  OR ripple_coeff ( $W_f$ )  $\geq 5\%$  THEN /
        $S_c \leftarrow S_c \cup \{f\}$ ;
9.   END IF
10. END FOR
// Phase 3: Fine Scan (Section 3.2.2)
11. Initialize raw dataset:  $D \leftarrow \{\}$ 
12. FOR EACH  $f_c$  IN  $S_c$  DO
13.   FOR  $f = (f_c - 10)$  TO  $(f_c + 10)$  MHz, STEP  $\Delta f_f = 1$  MHz DO
14.     FOR  $P = P_{min}$  TO  $P_{max}$ , STEP  $\Delta P = 1$  dBm DO
15.       Inject interference at ( $f, P$ ); collect  $W_{(f,P)}$ ;
16.       Monitor: BER, clock jitter, display/Ethernet/serial status;
17.       Assign 4-level label:
18.       Health / Mild Soft Fault ( $10^{-4} \geq \text{BER} > 0$ ) /
         Moderate Soft Fault ( $10^{-3} \geq \text{BER} > 10^{-4}$ ) / Hard Fault ( $\text{BER} > 10^{-3}$ );
19.        $D \leftarrow D \cup \{(W_{(f,P)}, \text{label})\}$ ;
20.     END FOR
21.   END FOR
22. END FOR
// Phase 4: Data Preprocessing (Section 3.3.1)
23. Clean  $D$ : missing value interpolation, outlier removal, noise filtering  $\rightarrow D_{clean}$ ;
// Phase 5: SMOTE-RU Hybrid Balancing (Section 3.3.1)
24. Stratified split  $D_{clean}$  into training set  $D_{train}$ , validation set  $D_{val}$ , /
    and test set  $D_{test}$ ;
25. Partition  $D_{train} \rightarrow D_{train\_nor}$  (majority),  $D_{train\_fault}$  (minority);
26. FOR EACH  $x_i$  IN  $D_{train\_fault}$  DO
27.   Find  $k = 5$  nearest neighbors  $\{x_j\}$  of  $x_i$  within  $D_{train\_fault}$ ;
28.   Select random  $x_j$ ; draw  $\lambda \sim \text{Uniform}(0,1)$ ;
29.    $x_{new} \leftarrow x_i + \lambda * (x_j - x_i)$ ;
30.    $D_{train\_fault} \leftarrow D_{train\_fault} \cup \{x_{new}\}$ ;
31. END FOR
32.  $D_{train\_nor}' \leftarrow \text{RandomSample}(D_{train\_nor})$  to obtain 1:1 ratio with  $D_{train\_fault}$ ;
33.  $D_{train\_balanced} \leftarrow D_{train\_fault} \cup D_{train\_nor}'$ ;
// Phase 6: Time-Frequency Feature Extraction (Section 3.3.2)
34. Extract time-frequency features from  $D_{train\_balanced} \rightarrow X_{train}$ ;
35. Extract the same features from  $D_{val} \rightarrow X_{val}$ ;
36. Extract the same features from  $D_{test} \rightarrow X_{test}$ ;
// Phase 7: Feature Importance Screening (Section 3.3.2)
37. Compute Pearson correlation  $\rho(f_i, y)$  using  $X_{train}$  only;
38. Train Random Forest on  $X_{train}$  only and compute Importance( $f_i$ );
39.  $F_{opt} \leftarrow \{f_i: |\rho(f_i, y)| \geq 0.3 \text{ AND } \text{Importance}(f_i) \geq 0.05\}$ ;
40. Build optimized datasets:
41.    $D_{train\_opt}$  using  $F_{opt}$ ,  $D_{val\_opt}$  using  $F_{opt}$ ,  $D_{test\_opt}$  using  $F_{opt}$ .

```

References

1. Grbovic, M.; Li, W.; Subrahmanya, N.A.; Usadi, A.K.; Vucetic, S. Cold Start Approach for Data-Driven Fault Detection. *IEEE Trans. Ind. Inform.* **2013**, *9*, 2264–2273. <http://dx.doi.org/10.1109/TII.2012.2231870>.
2. Wang, J.; Zhang, L.; Gao, R.X.; Vogl, A.W. Deep learning for smart manufacturing: Methods and applications. *J. Manuf. Syst.* **2018**, *48*, 144–156. <https://doi.org/10.1016/j.jmsy.2018.01.003>.
3. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–356. <https://doi.org/10.1613/jair.953>.
4. Xu, Y.; He, D.Q.; Sun, H.M.; Jin, Z.Z.; Zhao, M. Self-Supervised Learning for Train Bearing Fault Diagnosis Based on Time-Frequency Dual Domain Prediction. *Struct. Health Monit.* **2026**, In press. <https://doi.org/10.1177/14759217251405584>.
5. Xu, Y.; He, D.; Luo, X.; Jin, Z.; Sun, H. CSSCL: A cross-sensor self-contrastive learning for nondestructive fault detection of heavy locomotive motors in label-scarce scenarios. *Nondestruct. Test. Eval.* **2026**, 1–34, In press. <https://doi.org/10.1080/10589759.2026.2671370>.
6. Dai, C.Y.; He, D.Q.; Jin, Z.Z.; Zhang, X.W.; Chen, G.; Wu, J.X.; Zhao, J.Y.; Xu, Y. Digital Twin-Assisted Graph Contrastive Domain Adaptation for Small-Sample Bearing Fault Diagnosis. *Struct. Health Monit.* **2026**, In press. <https://doi.org/10.1177/14759217261440798>.
7. Du, C.; Xia, D.; Huang, Q.; Mao, C.; Cui, Z.; Fang, W.; Nie, X. Research on Electromagnetic Susceptibility of Electronic Modules in Component-Level HEMP PCI Test. *Energies* **2022**, *15*, 1409. <https://doi.org/10.3390/en15041409>.
8. Chen, J.; Cheng, L.; Zhang, J.; Li, W. Determining the highly susceptible waveform for electromagnetic susceptibility testing via reinforcement learning. *IEEE Trans. Electromagn. Compat.* **2024**, *66*, 1988–1998. <https://doi.org/10.1109/TEM.2024.3431570>.
9. Lu, W.; Duan, J.; Cheng, L.; Zhang, Y. Electromagnetic interference effect assessment under measuring testability limitation based on physics-informed neural network and Gaussian process regression. *IEEE Trans. Power Electron.* **2024**, *39*, 12413–12423. <https://doi.org/10.1109/TPEL.2024.3401706>.
10. Xue, H.; Yang, Q.; Chen, S. SVM: Support vector machines. In *The Top Ten Algorithms in Data Mining*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2009; pp. 51–74.
11. Zivkovic, M.; Bacanin, N.; Antonijevic, M.; Nikolic, B.; Kvascev, G.; Marjanovic, M.; Savanovic, N. Correction: Zivkovic et al. Hybrid CNN and XGBoost Model Tuned by Modified Arithmetic Optimization Algorithm for COVID-19 Early Diagnostics from X-Ray Images. *Electronics* **2022**, *11*, 3798. *Electronics* **2026**, *15*, 727. <https://doi.org/10.3390/electronics15040727>.
12. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
13. Xu, T.; Wang, Y.; Zhang, D.; Chen, Y. Prediction on EMS of UAV's data link based on SSA-optimized dual-channel CNN. *IEEE Trans. Electromagn. Compat.* **2022**, *64*, 1346–1356. <https://doi.org/10.1109/TEM.2022.3174635>.
14. Duan, H.; Chen, G.; Yu, Y.; Du, C.; Bao, Z.; Ma, D. DyGAT-FTNet: A Dynamic Graph Attention Network for Multi-Sensor Fault Diagnosis and Time-Frequency Data Fusion. *Sensors* **2025**, *25*, 810. <https://doi.org/10.3390/s25030810>.
15. Wang, X.; Jiang, H.K.; Dong, Y.T.; Mu, M.Z. Spatial-channel collaborative multi-scale graph interaction deep transfer learning for unsupervised rotating machinery fault diagnosis. *Eng. Appl. Artif. Intell.* **2026**, *176*, 114691. [https://doi.org/10.1016/j-engappai.2026.114691](https://doi.org/10.1016/j.engappai.2026.114691).
16. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. <https://doi.org/10.1109/78.650093>.
17. Meliboev, A.; Alikhanov, J.; Kim, W. Performance Evaluation of Deep Learning Based Network Intrusion Detection System across Multiple Balanced and Imbalanced Datasets. *Electronics* **2022**, *11*, 515. <https://doi.org/10.3390/electronics11040515>.
18. Yan, H.; Li, Y.; Zheng, J.; Zhang, Y. A novel method for susceptibility threshold prediction in communication systems via multitask learning. *IEEE Trans. Electromagn. Compat.* **2024**, *66*, 2102–2110. <https://doi.org/10.1109/TEM.2024.3472031>.
19. Jia, Z.; Chen, G.; Yang, S.; Chen, Y.; Weng, Y.; Su, D. Data-Driven Electromagnetic Susceptibility Modeling Method for Analog Sensors. *IEEE Sens. J.* **2024**, *24*, 12560–12569. <https://doi.org/10.1109/JSEN.2024.3370565>.
20. Xu, T.; Chen, Y.; Zhao, M.; Zhang, D. Adaptive EMS test design method on UAV data link based on Bayesian optimization. *IEEE Trans. Electromagn. Compat.* **2023**, *65*, 716–722. <https://doi.org/10.1109/TEM.2023.3261879>.
21. Yan, J.K.; Wang, Q.; Cheng, Y.; Su, Z.Y.; Zhang, F.; Zhong, M.L.; Liu, L.; Jin, B.; Zhang, W.H. Optimized single-image super-resolution reconstruction: A multimodal approach based on reversible guidance and cyclical knowledge distillation. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108496. <https://doi.org/10.1016/j.engappai.2024.108496>.
22. Cooley, J.W.; Lewis, P.A.W.; Welch, P.D. The fast Fourier transform and its applications. *IEEE Trans. Educ.* **2007**, *12*, 27–34. <https://doi.org/10.1109/TE.1969.4320436>.

23. Hastie, T.; Tibshirani, R.; Friedman, J.H. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2009.
24. Li, L.; Dai, F.; Hu, R.; Li, B.; Cai, S. Research Progress and Prospects of Electromagnetic Susceptibility Testing and Cognition Technology. *Saf. EMC* **2025**, *2*, 9–17, 34.
25. Tang, W.; Long, G.; Liu, L.; Zhou, T.; Jiang, J.; Blumenstein, M. Omni-Scale CNNs: A simple and effective kernel size configuration for time series classification. In Proceedings of the 10th International Conference on Learning Representations (ICLR), Virtual, 25–29 April 2022.
26. Wang, Z.; Yan, W.; Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. In *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 1578–1585.
27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018*; Springer: Cham, Switzerland, 2018; pp. 3–19.
28. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 11534–11542. <https://doi.org/10.1109/CVPR42600.2020.01155>.
29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008. <https://dl.acm.org/doi/10.5555/3295222.3295349>.

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.