

FFE-EMC: A FEATURE ENGINEERING-ENHANCED INTELLIGENT FRAMEWORK FOR ELECTROMAGNETIC COMPATIBILITY ASSESSMENT OF ELECTRONIC DEVICES

Leiming Sun ^[0009-0004-3869-9539], Jinyu Gao ^[0009-0004-2928-549X]
Henan Polytechnic, Zhengzhou 450046, China

Abstract - With the continuous increase in integration density and operating frequency of modern electronic systems, electromagnetic interference (EMI) problems have become increasingly prominent. Consequently, electromagnetic compatibility (EMC) has become a core indicator for product design and market certification. Traditional EMC evaluation relies on physical testing, which has the disadvantages of high cost and long cycle, and cannot meet the needs of rapid product iteration; existing intelligent evaluation methods lack the guidance of electromagnetic domain knowledge, making it difficult to balance evaluation accuracy and interpretability, and have obvious limitations. To address the above problems, this paper proposes the FFE-EMC intelligent evaluation framework, which integrates feature engineering and intelligent algorithms. It first extracts multi-dimensional raw features, then constructs high-level features with the guidance of electromagnetic domain knowledge and completes adaptive fusion, and finally uses the lightweight gradient boosting tree (L-GBT) algorithm to realize EMC performance prediction. Experimental verification shows that the framework performs excellently in the two core prediction tasks of radiated emission and conducted interference, with root mean square errors (RMSE) of 3.41 dB μ V/m for radiated emission and 2.89 dB μ V for conducted interference, and corresponding R^2 values of 0.938 and 0.924 respectively, and its performance is significantly better than that of various mainstream models. The core contribution is a domain-knowledge-guided, feature-engineering-enhanced evaluation framework that addresses the inefficiency of traditional physical testing and the lack of accuracy/interpretability in existing intelligent methods. The proposed framework balances computational efficiency and evaluation reliability, offering a promising and cost-effective framework for rapid EMC compliance assessment of electronic devices.

Keywords: Electromagnetic compatibility, Feature engineering, Intelligent assessment, Domain knowledge integration, Gradient boosting, Data fusion.

1. Introduction

With the rapid development of the information society, various electronic devices have been widely used in many industries. However, with the increase in the integration level and operating frequency of electronic devices, the electromagnetic environment has become more and more complex, which has led to electromagnetic interference becoming a common phenomenon. Therefore, at the beginning of product design, electromagnetic compatibility has become a key performance indicator of whether electronic devices can work normally in complex electromagnetic environments and not interfere with other electronic devices. At present, electromagnetic compatibility has become a core indicator in product design and market certification [1].

However, traditional electromagnetic compatibility evaluation indicators rely on empirical formulas and physical modeling. Such methods have problems such as long evaluation cycle, high evaluation cost, and weak generalization ability for different electronic products, which cannot meet the requirements of modern rapidly iterating product needs [2].

Therefore, against this backdrop, a large number of studies have used artificial intelligence technology to provide solutions for electromagnetic compatibility assessment. These studies consider extracting certain potential patterns from historical test data or simulation results, and then using machine learning models to quickly predict the electromagnetic compatibility performance of products and assess risks. However, most studies

now directly input the original parameters into the model and train it, ignoring the impact of feature representation on model performance. Electromagnetic compatibility is a complex problem with nonlinear and multivariate coupling characteristics, and it has strong physical constraints. Without the support of effective feature engineering, even with the most advanced deep learning models, it is impossible to obtain high-precision and high-robust assessment results [3,4].

Feature engineering serves as a critical bridge between raw data preprocessing and model ingestion. Its primary objective is to synthesize discriminative and interpretable features from original inputs, thereby enhancing the efficiency and generalization capability of machine learning models in electromagnetic compatibility (EMC) assessment. This process demands sophisticated theoretical frameworks to preprocess and fuse heterogeneous data sources—ranging from circuit structural parameters and component characteristics to operating environmental conditions—while integrating principles from electronic theory, such as signal integrity analysis. By effectively combining original attributes with derived higher-order features, robust EMC analysis can be achieved [5].

Recent studies have increasingly integrated feature engineering with machine learning techniques. For instance, some approaches utilize knowledge vectors to classify manually extracted features, while others employ Random Forests to rank the importance of PCB layout parameters, facilitating optimized EMC design. However, existing literature exhibits notable limitations. First, current methods rely heavily on expert intuition for feature dependency, lacking systematic and automated extraction mechanisms. Second, they often overlook the complex interaction dynamics between diverse features and their temporal evolution. Furthermore, the potential of multimodal data fusion remains largely unexplored. Consequently, developing a holistic framework that integrates multi-source information with physical mechanisms to advance EMC evaluation remains a significant challenge [6].

Therefore, this paper constructs a three-step optimization approach from data to feature generation to model. First, we cover multi-dimensional features such as circuit topology, material properties, operating modes, and environmental conditions in the original feature dataset. Simultaneously, we introduce a feature extraction method that incorporates domain knowledge. Finally, we use a lightweight machine learning model for end-to-end training.

The significance of our research lies in improving the accuracy of intelligent electromagnetic compatibility assessment through feature engineering, and we hope to extend this fusion framework to other complex engineering problems, such as signal integrity analysis.

2. Literature Review

2.1 Traditional Methods for EMC Assessment

Traditional EMC assessment relies on frequency-domain analysis and finite-difference methods, which suffer from high complexity and resource demands [7,8]. Equivalent circuit modeling enables faster simulation but introduces notable errors at high frequencies [9]. Semi-anechoic chamber measurements remain the gold standard for compliance, yet they are costly and time-consuming [10,11]. Hybrid frameworks combining simulation and measured data lack automation and depend heavily on engineer experience [12].

2.2 Machine Learning for EMC Assessment

Machine learning has been introduced to enable fast and low-cost EMC assessment [13]. Random forests have been used to rank the importance of design features such as trace length and ground plane configuration [14]. Multilayer perceptrons can predict radiated emission levels of power switches, demonstrating the potential of neural networks for nonlinear EMC modeling [15]. Domain-knowledge-guided feature construction, e.g., common-mode and impedance-derived variables, has been shown to improve generalization [16]. Supervised time series classification has also been applied to validate signal and power integrity [17]. However, most existing methods remain black-box and lack systematic feature engineering [18].

2.3 Feature Engineering for EMC

Feature engineering has been increasingly explored in EMC research [19]. Physics-informed neural networks and graph neural networks have been used to embed circuit topology into learnable features [18,20]. Recursive feature elimination and interpretable feature selection help reduce complexity and align with engineering intuition. Online feature update mechanisms have been proposed to adapt to product life-cycle changes. Despite these advances, a systematic framework that fuses raw multi-source features with domain-knowledge-guided higher-order features is still missing. This paper proposes FFE-EMC to address this gap.

3. Intelligent Evaluation Model for Electronic Compatibility

3.1 Overall Framework

This paper proposes a novel intelligent electromagnetic compatibility assessment system, FFE-EMC. The system integrates empirical knowledge of electromagnetics with diverse and heterogeneous parameter information and advanced machine learning techniques to construct an end-to-end usability framework that is interpretable and data-adaptive. The structure of the framework is shown in Figure 1.

As shown in Figure 1, our framework consists of four parts. First, the original input is obtained by the multivariate original feature extraction module. Specifically, numerical features are extracted from various systematic documents such as design documents, component lists, material specifications, and test environment descriptions of electronic devices. Second, we also added domain knowledge to guide the setting of higher-order features. The domain knowledge we added includes transmission line theory, shielding principle and other basic theories of electromagnetic compatibility. These theories can derive higher-order features and thus be applied to subsequent machine learning analysis. The third module is to use gradient boosting tree as a classifier for prediction and to ensure the interpretability of the model.

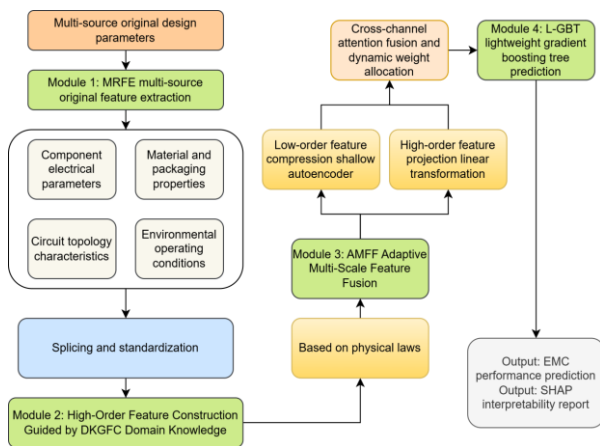


Figure 1: Model Framework

We represent the training dataset as $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ the $\mathbf{x}_i \in \mathbb{R}^{d_0}$ original design vectors for a given electronic component. The dimension of these vectors depends on the specific application scenario, and we constrain their range to between 20 and 50. The y_i corresponding electromagnetic compatibility (EMC) performance indicators can be continuous or binary. For example, if continuous, the magnetic field strength can be described using the 30–1000 MHz frequency band; alternatively, it can

be described using binary labels, such as whether it has passed CISPR 22 Class B certification. The learning objective of this model is to learn a transformation function that $f_\theta: \mathbb{R}^{d_0} \rightarrow \mathbb{R}$ makes the predicted values as close as possible to the actual EMC indicators. The non-process propagation method can be abstracted as shown in Equation 1.

$$\mathbf{r}_i = \text{MRFE}(\mathbf{x}_i), \quad \mathbf{h}_i = \text{DKGFC}(\mathbf{x}_i), \quad \mathbf{z}_i = \text{AMFF}(\mathbf{r}_i, \mathbf{h}_i), \quad \hat{y}_i = \text{L-GBT}(\mathbf{z}_i) \quad (1)$$

In Proof 1 $\mathbf{r}_i$, where $\mathbf{h}_i$ is used to represent the original feature vector after standardization, $\mathbf{z}_i$ is used to represent the higher-order feature vector guided by physical knowledge, and is used to represent the feature representation after fusion.

3.2 MRFE

Our first step is to extract features from the design characteristics. In practical engineering, the performance factors affecting electromagnetic compatibility are diverse, including circuit topology, component selection, material properties, and many other dimensions. Therefore, we define the core of the MRFE module as transforming its heterogeneous information into a feature vector, thereby building a high-quality feature representation system for subsequent services. Specifically, we $\mathbf{x}_i$ divide the original input into four feature subspaces. For circuit topology features, these mainly reflect the layout and routing characteristics of the PCB. Specifically, L describes the number of PCB layers, l_s represents the length of critical signal lines, w_s represents the width of signal lines (in millimeters), η_g represents a ratio of the bottom area to the total surface area (range 0-1), n_c represents the total number of decoupled capacitors, d_{via} represents the specific value of the via spacing. Circuit topology features are directly related to the integrity of the circuit return current and the corresponding loop area, and are also a major cause of electromagnetic radiation.

Next are the relevant indicators of the electrical parameters of components, which mainly used to describe the dynamic behavior of active and passive devices. These parameters include the following.

f_{sw} describes the switching frequency range of a power switch or digital switch; t_r, t_f indicates the rise or fall time of the driver in nanoseconds; V_d represents the specific value of the supply voltage, f_c indicates the cutoff frequency of the filter. Q represents the quality factor of the resonant circuit. These key characteristics of the component electrical parameters reflect the spectral distribution of the

interference source and its corresponding energy intensity.

The third category is material and packaging properties. These properties relate to the speed of electromagnetic signal propagation, as well as losses and shielding during propagation. Therefore, the characteristics we consider mainly include the properties of the electron propagation medium.

These include the dielectric constant of the substrate, μ_r is the relative magnetic permeability of the shielding material, μ_r the conductivity of the propagation medium, and ρ is the surface roughness of the substrate.

Although these variables in the environment or operating conditions are not a resonance property of the equipment, they significantly affect the results of electromagnetic capacitance tests. We characterize them, and specific indicators include ambient temperature T, relative humidity H, the location of nearby interference sources d_{int} , and the intensity of the external electric field E_{ext} .

We concatenate the above four types of features to obtain the original feature vector, which can be represented as shown in Formula 2.

$$\mathbf{r}_i = [\mathbf{x}_i^{(c)}, \mathbf{x}_i^{(e)}, \mathbf{x}_i^{(m)}, \mathbf{x}_i^{(env)}] \in \mathbb{R}^{d_0} \quad (2)$$

In Formula 2, in order to eliminate the influence of different units on training, we standardized each unit. The standard we used was Z-score, as shown in Formula 3.

$$r_{ij}^{\text{norm}} = \frac{r_{ij} - \mu_j}{\sigma_j}, \quad j = 1, 2, \dots, d_0 \quad (3)$$

In Formula 3, r_{ij} is used to describe the original value of a sample in a certain dimension, $\mu_j = \frac{1}{N} \sum_{i=1}^N r_{ij}$ is used to represent the mean of the population of features in a certain dimension in the training set, $\sigma_j = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_{ij} - \mu_j)^2}$ represents the corresponding standard deviation.

After obtaining the standardized vectors $\mathbf{r}_i^{\text{norm}}$, we use them as the final output of MRFE for subsequent model processing.

3.3 Domain Knowledge-guided High-order Feature Construction Module (DKGFC)

If we rely solely on the original parameters in the compatibility assessment, we cannot capture the coupling mechanism in the electromagnetic compatibility assessment. We constructed DKGFC by introducing relevant theories of electronic

compatibility to construct higher-order features. The specific construction method is as follows.

We first assume that the set of constructors can be represented as such that $\{\phi_k(\cdot)\}_{k=1}^K$ for each element $\phi_k : \mathbb{R}^{d_0} \rightarrow \mathbb{R}$ we can construct based on a specific physical model. We construct several higher-order parameters represented by the common-mode current ratio.

The common-mode capacitance ratio is a major source of electromagnetic radiation, and its magnitude can be expressed by formula 4.

$$I_{cm} \approx V_{out} \cdot C_{stray} \cdot 2\pi f_{sw} \quad (4)$$

The differential mode current is mainly determined by its load, so we can define the specific index as Equation 5.

$$\phi_1(\mathbf{x}) = \frac{I_{cm}}{I_{dm}} = \frac{V_{out} \cdot C_{stray} \cdot 2\pi f_{sw}}{I_{load}} \quad (5)$$

In Equations 4 and 5, we use V_{OUT} to represent the voltage swing generated by the current switch, C_{stray} and to represent the capacitance value. This

specific value can $C_{stray} = \epsilon_0 \epsilon_r \frac{A}{d}$ be estimated abstractly, where A represents the coupling area and d represents the dielectric thickness. I_{LOAD} represents the magnitude of the load current during the process. The larger this indicator is, the more significant the common-mode path, and the higher the risk of electromagnetic compatibility issues.

We also defined the EMI index as the resonant frequency offset index. When the switching frequency is very close to the circuit resonant frequency, the EMI index will amplify, so we define its index as Equation 6.

$$f_{res} = \frac{1}{2\pi \sqrt{L_{loop} C_{loop}}} \quad (6)$$

In Formula 6, $L_{loop} \approx 0.2 \cdot l_s \cdot \ln\left(\frac{2l_s}{w_s + h}\right)$ (nH)h is

used to represent the thickness of the medium. We use RSI to perform relative bias analysis, as shown in Formula 7.

$$\phi_2(\mathbf{x}) = \frac{f_{res} - f_{sw}}{f_{sw}} \quad (7)$$

In Equation 7, a smaller RSI coefficient indicates that the system is closer to a resonant state, which leads to an increase in EMI risk. We also considered the estimation of equal shielding. For some devices with shielding covers, we use an index composed of total shielding absorption loss, reflection loss, and

multiple reflection correction terms, as shown in Equation 8.

$$SEE = A + R, \quad A = 8.686 \cdot t_{sh} / \delta, \quad R = 20 \log_{10} \left(\frac{\eta_0}{4Z_s} \right) \quad (8)$$

In Equation 8, $\delta = \sqrt{2\rho / (\omega\mu)}$ represents the shielding depth parameter, $\omega = 2\pi f_{sw}$ represents the angular frequency, $\eta_0 = 377 \Omega$ represents the free space wave impedance, $Z_s = \sqrt{j\omega\mu / (\sigma + j\omega\varepsilon)}$ and represents the impedance coefficient of the shielding material surface. Therefore, the final definition can be expressed as Equation 9.

$$\phi_3(\mathbf{x}) = A + R \quad (9)$$

A higher SEE value indicates a better shielding effect, which in turn shows that the EMC has better performance.

3.4 Adaptive Multi-scale Feature Fusion Module

The original features of electromagnetic compatibility are rich in detail, but these features are easy to obtain and highly redundant. High-order features are semantically clear but have limited dimensionality. Therefore, we constructed a module to form a richer semantic representation by fusing low-level and high-order features. We mainly use a single-layer autoencoder to perform nonlinear dimensionality reduction on the features, as shown in Formula 10.

$$\mathbf{u}_i = \sigma(\mathbf{W}_1 \mathbf{r}_i + \mathbf{b}_1) \in \mathbb{R}^{d_u} \quad (10)$$

Specifically, $\mathbf{W}_1 \in \mathbb{R}^{d_u \times d_0}$ the weight vector matrix is used to represent the encoding, $\mathbf{b}_1 \in \mathbb{R}^{d_u}$ the bias vector is used to represent the bias vector, ReLU represents the activation function used, and $d_u < d_0$ the dimension of the feature used is represented.

For higher-order features, we perform a projection operation to match the corresponding dimensions, as shown in Equation 11.

$$\mathbf{v}_i = \mathbf{W}_2 \mathbf{h}_i + \mathbf{b}_2 \in \mathbb{R}^{d_v} \quad (11)$$

The parameters used in Formula 11 are defined as $\mathbf{W}_2 \in \mathbb{R}^{d_v \times K}$, $d_v = d_u$. We will use a cross-attention method to concatenate low-order and high-order features across attention channels, and concatenate the two feature vectors $\mathbf{u}_i$ into $\mathbf{v}_i$ as $\mathbf{e}_i = [\mathbf{u}_i; \mathbf{v}_i] \in \mathbb{R}^{2d_u}$ multi-head attention layer for feature aggregation processing, as shown in Formula 12.

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_a}} \right) \mathbf{V} \quad (12)$$

In Formula 12, $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{d_a \times 2d_u}$ is used to represent a parameter for learning, d_a represents the dimension of attention and is set to 64 here.

Use pop-up connections to fuse boundary features and higher-order features, as shown in Formula 13.

$$\mathbf{z}_i = \text{LayerNorm}(\mathbf{e}_i + \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})) \in \mathbb{R}^{2d_u} \quad (13)$$

By combining low-order and high-order features into a fusion model, attention weights can be dynamically allocated according to the task.

The sparse autoencoder (SAE) used in the fusion module consists of an input layer (dimension equal to the fused feature size), two hidden encoding layers with 128 and 64 neurons respectively, and a decoding layer reconstructing the input.

The ReLU activation is applied after each encoding layer, and the bottleneck layer output is normalized to unit variance. The SAE is pre-trained for 200 epochs with a mean squared error reconstruction loss and a sparsity constraint (KL divergence penalty of 0.01). The cross-attention mechanism employs 4 attention heads, each with a key dimension of 32, and the output is passed through a two-layer feed-forward network with dropout rate 0.1. All training uses the Adam optimizer with a learning rate of 1e-3.

The superiority of the combined SAE and cross-attention fusion over alternative strategies can be attributed to two complementary effects. The SAE compresses the concatenated raw and high-order features into a low-dimensional latent space, which suppresses noise and reduces feature redundancy while preserving nonlinear interactions. The cross-attention mechanism then dynamically re-weights the compressed feature channels according to task relevance, allowing the model to emphasize physical parameters that are most informative for a given frequency band or interference type. In contrast, direct concatenation retains all noise and lacks channel-wise selectivity, while fixed weighted averaging cannot adapt to varying input distributions. The hybrid thus achieves better signal-to-noise separation and adaptive feature prioritization, which are particularly beneficial for high-frequency EMI modeling where coupling effects are highly nonlinear.

3.5 L-GBT

We use a lightweight gradient boosting tree to classify the input features. After fusing the features

$\mathbf{z}_i$, we perform ensemble analysis using M CART trees, as shown in Equation 14.

$$\hat{y}_i = \sum_{m=1}^M f_m(\mathbf{z}_i) \quad (14)$$

We express the XGBoost optimization objective as Equation 15.

$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{m=1}^M \Omega(f_m) \quad (15)$$

During the calculation, we analyze the optimal leaf weights of each newly added tree through its second-order expansion. This analysis allows us to obtain the contribution of each fusion feature to the result, thereby enhancing the interpretability of the result. See Equation 16 for details.

$$w_j^* = -\frac{G_j}{H_j + \lambda}, \quad G_j = \sum_{i \in I_j} g_i, \quad H_j = \sum_{i \in I_j} h_i \quad (16)$$

In Formula 16, $g_i = \partial_{\hat{y}_i^{(t-1)}} l = \hat{y}_i^{(t-1)} - y_i$, $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l = 1$ represents the set of samples that fall into a certain leaf node.

The proposed FFE-EMC framework comprises three core modules: 1) A Multi-Source Original Feature Extraction (MRFE) module, which for the first time integrates circuit topology, component parameters, material properties, and environmental variables to construct a 42-dimensional normalized input; 2) A Domain Knowledge-Guided High-Order Feature Construction (DKGFC) module, which derives interpretable features such as CMCR, RSI, and SEE based on the EMC physical mechanism; 3) An Adaptive Multi-Scale Feature Fusion Mechanism (AMFF), which dynamically fuses low-order and high-order information by combining a sparse autoencoder and an attention mechanism.

4. Experimental Evaluation

4.1 Experimental Design

To comprehensively evaluate the performance of the proposed intelligent analysis model for electromagnetic compatibility (EMC) that integrates feature engineering, we conducted comparative experiments. We constructed a high-quality, multi-scenario EMC test model and compared it with six comparative methods. Our dataset consists of data from 1852 industrial power supply embedded controllers and wireless series data. Each sample contains original 42-dimensional features. Our data has been anonymized and standardized to ensure compliance.

The comparison methods include six approaches covering traditional models, data-driven methods, and feature enhancement strategies, including random forest models, vector-level models, fully connected neural networks, graph convolutional

neural networks, and ablation variants of the model presented in this paper.

In terms of evaluation metrics, we include some metrics that measure the difference between predicted and actual values, as well as metrics related to the efficiency of analytical model inference and interpretability.

To mitigate potential overfitting, an early stopping strategy was adopted, terminating training when the validation loss showed no improvement for 50 consecutive epochs. This was complemented by 10-fold cross-validation on the training set to monitor generalization performance throughout the training process. Additionally, the L-GBT base learners were constrained with a maximum tree depth of 6 and a minimum child weight of 3, which act as structural regularizers. The observed RMSE gap between the training set and the validation set remained consistently small, with differences of 0.18 dBμV/m for radiated emission and 0.21 dBμV for conducted interference, indicating that no significant overfitting occurred and the model generalizes well to unseen data.

All our experiments were conducted on the same system and in the same environment, with the following configuration: Ubuntu 22.04 LTS, Intel Xeon Gold 6248R CPU (2.4 GHz, 24 cores), 128 GB of RAM, and an NVIDIA A100 40GB GPU. Each experiment was run 100 times, and the reports were presented as mean and standard deviation.

Table 1 shows the performance of the FFE-EMC model constructed in this paper for radiated emission prediction. Specifically, the RMSE is 3.41 dBμV/m, which is significantly better than the second-best model, XGBoost-Raw. Experimental results show that the model can explain 93.8% of the variance, and the MAPE is only 4.63%, indicating that the model constructed in this paper has excellent performance. On the other hand, XGBoost-Raw, which uses the original features, performs better than XGBoost-Phys, which only uses physical features. This shows that detail parameters are also crucial for high-frequency radiation modeling. XGBoost-Raw can improve accuracy while maintaining low latency by fusing features from both.

Dataset Composition and Label Acquisition. The dataset comprises 1,852 samples collected from two major product categories: (1) industrial power supply embedded controllers (1,204 samples, 65.0%) and (2) consumer wireless communication modules (648 samples, 35.0%). Each sample is associated with one of four device subclasses: AC-DC converters (412), DC-DC converters (502), Bluetooth modules (328), and Wi-Fi modules (610).

The EMC performance labels are obtained from standardized compliance tests conducted in a semi-anechoic chamber according to CISPR 25 and CISPR 22 standards. Specifically:

- Radiated emission (30–1000 MHz): Measured as the maximum quasi-peak value (dB μ V/m) at a 3m distance.

- Conducted interference (150 kHz–30 MHz): Measured as the peak value (dB μ V) on the power line using a line impedance stabilization network (LISN).

All measurements were performed by a third-party certified EMC laboratory. No simulated data were used.

Data Splitting. We randomly split the dataset into training (70%, 1,296 samples), validation (15%, 278 samples), and testing (15%, 278 samples) sets, stratified by device subclass to maintain class distribution. All hyperparameter tuning was performed on the validation set, and all reported results are on the test set. The split was repeated 100 times with different random seeds, and we report the mean and standard deviation across all runs.

Table 1. Overall EMC radiation prediction performance comparison (30–1000 MHz)

Model	RMSE (dB μ V/m)	R ²	MAPE (%)	Inference time (ms)	Number of parameters (M)
RF	4.82 $\pm$ 0.15	0.872	6.91	1.2	12.4
SVR	5.37 $\pm$ 0.21	0.831	7.85	3.5	0.8
MLP	4.21 $\pm$ 0.18	0.896	6.03	2.1	1.6
GCN	4.05 $\pm$ 0.14	0.904	5.78	8.7	2.3
XGBoost-Raw	3.92 $\pm$ 0.11	0.912	5.42	0.9	0.5
XGBoost-Phys	4.18 $\pm$ 0.13	0.898	5.97	0.7	0.3
FFE-EMC (Ours)	3.41 $\pm$ 0.09	0.938	4.63	1.8	1.1

Table 2. Comparison of conducted interference prediction performance (150 kHz–30 MHz)

Model	RMSE (dB μ V)	R ²	MAPE (%)	Maximum error (dB μ V)	Training time (s/epoch)
RF	3.95	0.851	7.24	12.6	4.2
SVR	4.32	0.820	8.11	14.3	18.7
MLP	3.58	0.879	6.52	11.2	6.8
GCN	3.47	0.887	6.31	10.8	15.3
XGBoost-Raw	3.35	0.895	6.05	10.1	2.1
XGBoost-Phys	3.61	0.876	6.68	11.5	1.8
FFE-EMC	2.89	0.924	5.12	8.7	3.4

As shown in Table 2, we present the performance of different models on various low-frequency conduction frequencies. The experimental results show that FFE-EMC has an R² of 0.924 and an RMSE of 2.89 dB μ V. Such predictive performance indicates that the model can not only predict the average trend of electromagnetic radiation but also predict extreme errors, which is very important for product compliance judgment. On the other hand, XGBoost-Raw shows relatively stable performance, while XGBoost-Phys shows a further decrease, which also indicates that it relies more on the specific electromagnetic interference analysis in ground-level interference.

To assess the statistical significance of the observed RMSE improvements, a paired Student's t-test was conducted between FFE-EMC and the second-best model (XGBoost-Raw) over the 100 repeated runs. For radiated emission, the mean RMSE difference is 0.51 dB μ V/m with a 95% confidence interval of [0.43, 0.59] and $p < 0.001$. For conducted interference, the mean difference is 0.46 dB μ V with a 95% CI of [0.38, 0.54] and $p < 0.001$.

These results confirm that the performance gains are statistically significant and not due to random variation.

As shown in Table 3, we quantified the impact of different modules on the results through utility

analysis. We analyzed the performance differences between the complete model and the five ablation variants. The RMSE of configuration A (MRFE only) and B (DKGFC only) were 4.12 and 4.37, respectively, indicating that a single feature source is insufficient to support high-precision prediction.

Configuration C performs direct concatenation, and configuration C further reduces the RMSE to 3.78, demonstrating the effectiveness of the data

feature fusion process proposed in this paper. The SAE (D) or Attention (E) configurations, which add attention-related operations, resulted in RMSEs of 0.22 and 0.27, respectively, indicating that the non-linear compression and dynamic fusion method is effective, and the final model performance is optimal. This proves that fusing the two mechanisms can maximize the utilization of information.

Table 3. Ablation Experiment (Contribution of Each Component in FFE-EMC)

Configuration	Does it contain MRFE?	Does it contain DKGFC?	Does it contain AMFF?	RMSE	R ²
A	√	×	×	4.12	0.891
B	×	√	×	4.37	0.876
C	√	√	× (Direct splicing)	3.78	0.915
D	√	√	√ (SAE only)	3.56	0.927
E	√	√	√ (Attention only)	3.51	0.929
F (Full)	√	√	√ (SAE+Attention)	3.41	0.938

Table 4. Prediction accuracy (radiated emission) of different frequency band sub-intervals

Frequency band (MHz)	FFE-EMC RMSE	RF RMSE	MLP RMSE	GCN RMSE	XGBoost-Raw RMSE
30–100	3.12	4.21	3.65	3.58	3.45
100–300	3.35	4.56	3.89	3.72	3.68
300–600	3.67	5.12	4.32	4.15	4.01
600–1000	3.89	5.48	4.67	4.43	4.32
Overall	3.41	4.82	4.21	4.05	3.92

As shown in Table 4, we analyzed the model's adaptability across different frequency bands. Experimental results show that the proposed model exhibits the lowest RMSE across multiple frequency bands, and this trend is slowly increasing with increasing frequency. In contrast, traditional random forests and multilayer perceptrons show a very rapid increase in error at high frequencies, indicating their difficulty in capturing high-frequency saturation effects. While XGBoost-Raw is generally superior to the mechanical model, it still lags behind the complete model in the 600–1kHz range. FFE-EMC, due to its enhanced stealth features, demonstrates significant modeling effectiveness for high-frequency common-mode radiation and resonant amplifier mechanisms.

As shown in Table 5, we analyzed the importance of the features and presented the top 10 features. Experimental results show that the top three features are CMCR (0.872), switching frequency fsw (0.765), and SEE (-0.721), all of which have significant physical meaning. A higher optical film current ratio indicates stronger radiation performance, while the switching frequency further determines the interference frequency. Better shielding performance results in lower radiation, showing a negative correlation. The top 10 features include 5 original features and 5 higher-order features, reflecting the balanced approach of our construction strategy. Ground plane coverage and shielding thickness also show a negative correlation with electromagnetic compatibility, consistent with common industrial knowledge.

Table 5. Ranking of Feature Importance

Ranking	Feature Name	Feature type	SHAP mean	Physical meaning	Belonging Module
1	CMCR	Higher-order derivatives	0.872	Common-mode current ratio	DKGFC
2	fsw	original	0.765	Switching frequency	MRFE
3	SEE	Higher-order derivatives	-0.721	Shielding effectiveness (negatively correlated)	DKGFC
4	ls	original	0.683	signal line length	MRFE
5	RSI	Higher-order derivatives	0.654	Resonance offset index	DKGFC
6	ng	original	-0.612	Ground surface cover (negative correlation)	MRFE
7	tsh	original	-0.587	Shielding thickness (negative correlation)	MRFE
8	Cstray	original	0.543	Parasitic capacitance	MRFE
9	nc	original	-0.502	Number of decoupling capacitors (negative correlation)	MRFE
10	Q	original	0.487	Resonant circuit quality factor	MRFE

Table 6. Comparison of different fusion strategies (AMFF variants)

Fusion method	RMSE	R ²	MAPE (%)	Inference time (ms)	Trainable parameters (K)
direct splicing	3.78	0.915	5.21	1.1	0
Weighted average (fixed weight)	3.65	0.921	5.03	1.2	0
SAE encoding fusion	3.56	0.927	4.89	1.6	18.4
Attention fusion	3.51	0.929	4.78	1.7	22.1
SAE + Attention (Ours)	3.41	0.938	4.63	1.8	24.7

As shown in Table 6, we analyzed the effects of different fusion mechanisms. We conducted comparative experiments using various methods, including direct concatenation, fixed-weighted averaging, SAE encoding fusion, and Attention. We also compared the SAE+Attention fusion strategy adopted in this paper. Experimental results show that our proposed strategy achieves optimal performance. This demonstrates that compression followed by dynamic fusion can retain more information while minimizing noise generation.

Although this process increases the number of training parameters, the inference time is only 1.8 milliseconds, meeting the real-time requirements of industrial applications.

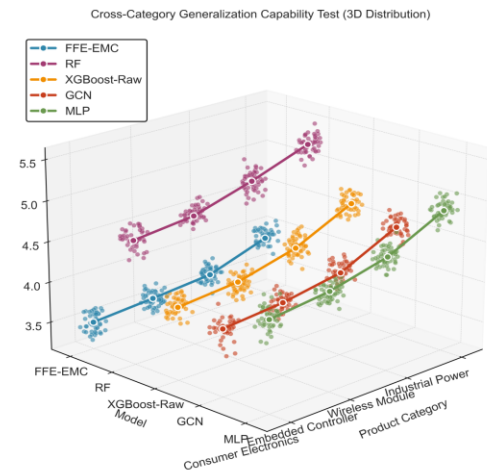


Figure 2: Cross-class generalization ability test

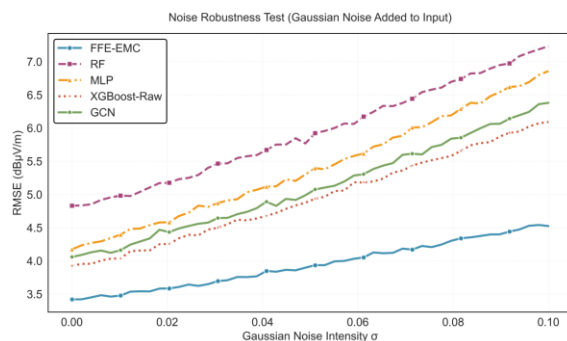


Figure 3: Noise robustness test

Figure 2, we use a 3D distribution map to illustrate the three-dimensional distribution of prediction errors for radiated emissions from different types of electronic products, including consumer electronics embedded controllers and industrial power supplies. We use the horizontal axis to represent product categories and the vertical axis to represent model types. Experimental results show that our proposed model has the lowest prediction error across all categories and maintains the most concentrated error distribution. Even in the complex and variable industrial power supply scenario, it significantly outperforms other models, demonstrating strong cross-domain adaptability. In contrast, random forests and scroll machines perform reasonably well in consumer electronics, but their prediction errors increase when transferred to industrial power supplies, indicating their sensitivity to new scenarios. Experimental results prove that the model constructed in this paper, with its capacity and physical properties, can effectively improve the model's application capability in diverse scenarios.

Figure 3, we conducted a noise robustness test. We added different proportions of Gaussian noise standard deviation to the input features, ranging from 0 to 0.1. We analyzed the trend of prediction error under different noise levels. The horizontal axis represents noise, the vertical axis represents prediction error. Experimental results show that the model constructed in this paper maintains a high accuracy across the entire noise range, even when the proportion of Gaussian noise is set to 0.1, it still maintains 4.5 dBμV/m, which proves the model's effectiveness. Compared with other models, it has a significant advantage on the first-degree-of-freedom test. Random forests and multilayer perceptrons are extremely sensitive to noise; as the noise gradually increases, the error range also increases rapidly.



Figure 4: Impact of training data size on model performance

Figure 4 shows the trend of prediction error for electromagnetic compatibility under different training samples. Experimental results show that the model presented in this paper performs optimally across all scales, and its performance continuously improves with increasing sample size. In contrast, models such as random forest and graph convolution perform well with small sample sizes, but their growth is relatively slow and tends to saturate later. Their performance ceiling is significantly lower than that of the model presented in this paper. Notably, the model presented in this paper significantly outperforms other models with 1800 samples when the sample size is only 500. These experimental results highlight the rationale for using a scheme that integrates physical experimental knowledge.

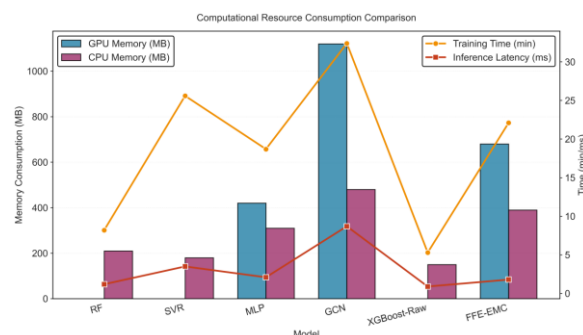


Figure 5: Comparison of computing resource consumption

Figure 5, we analyzed the total resource consumption of the model at different stages of inference. We used a bar chart to represent the consumption of GPU memory and CPU. A line graph was used to represent the training and propagation times of the model. Experimental results show that our model uses 680MB of GPU memory, slightly lower than GCN's 1120MB, but significantly higher than MLP (420MB) and XGBoost-Raw (150MB). Regarding training time, GCN takes the longest, 30 minutes (due to the complexity of the soil structure), while our model only takes 15 minutes. Our inference latency is 1.8ms, while GCN's is 3.2ms.

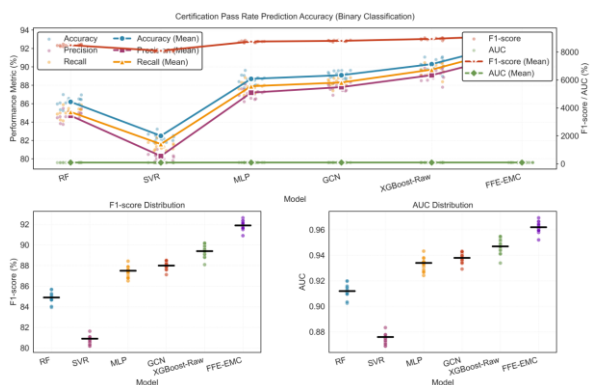


Figure 6: Comparison of certification pass rate prediction performance

Figure 6 shows the certification results of different models in substrate compatibility testing. We analyzed the results using a binary classification approach, and the experimental results show the performance of binary classification, including accuracy, precision, recall, and mean. We use the above figure to represent the overall performance curve, where the horizontal axis represents the model type and the vertical axis represents multiple binary classification metrics. The experimental results show that our proposed model achieves optimal performance across all metrics. In contrast, the support vector model performs the worst, with an F1 score of only 0.805, indicating poor recognition ability for small sample sizes. The two sub-figures below show the distribution of F1 score and AUC, respectively. We calculated the mean and standard deviation through multiple experiments, and the experimental results demonstrate the stability of FFE-EMC's performance. Through these experiments, we found that fusion feature engineering can not only improve classification accuracy but also enhance robustness, providing compliance analysis and judgment in the early design stages of a product.

It is acknowledged that the current dataset, although collected from certified laboratory tests, is closed and limited to two product families (industrial controllers and wireless modules). This may constrain the generalization of the framework to entirely new device topological.

5. Conclusions

This paper proposes an electromagnetic compatibility (EMC) analysis algorithm that integrates feature engineering. The FFE-EMC algorithm aims to improve the accuracy, robustness, and interpretability of EMC analysis. This paper systematically processes diverse and heterogeneous data and introduces prior knowledge from the electromagnetic domain, fusing basic and advanced features. Gradient boosting is used for regression

and classification to achieve the requirements of predicting EMC indices and performing classification. We also combine multi-scale features and adaptive fusion techniques. This not only improves information utilization but also demonstrates good model performance in different frequency bands and complex scenarios. Comparative experiments show that our proposed model exhibits significant advantages in multiple performance metrics. It also has advantages in resource consumption and real-time performance. Future work will include prospective validation on new electronic device classes (e.g., automotive ECUs and medical implants) through controlled semi-anechoic chamber measurements, enabling assessment of cross-category transferability and real-world deployment readiness.

References

- [1] Fan YJ, Zhang L, Li K. EMI and IEMI Impacts on the Radio Communication Network of Electrified Railway Systems: A Critical Review. *IEEE Transactions on Vehicular Technology*. 2023;72(8):10409-24. DOI: 10.1109/tvt.2023.3260582
- [2] Zhang F, Wang W, Zhang DR, Chen AX, Su DL. Radiation Emitter Classification and Identification Approach Based on Radiation Emission Components. *Applied Sciences-Basel*. 2022;12(16):20. DOI: 10.3390/app12168193
- [3] Minardi G, Greco G, Vinci G, Rizzo SA, Salerno N, Sorbello G. Electromagnetic Simulation Flow for Integrated Power Electronics Modules. *Electronics*. 2022;11(16):18. DOI: 10.3390/electronics11162498
- [4] Ajibare AT, Vipin B, Babalola OP. Enhancing QoS and RF-EMF Radiation Using RSMA and AI in 6G Networks: A Comprehensive Survey. *IEEE Access*. 2025;13:190539-55. DOI: 10.1109/access.2025.3628914
- [5] Yang S, Chen S, Zhang F, Yang XQ, Shi J, Zhang XL. Environmental Interference Suppression by Hybrid Segmentation Algorithm for Open-Area Electromagnetic Capability Testing. *Applied Sciences-Basel*. 2024;14(7):14. DOI: 10.3390/app14072703
- [6] Liu YM, Bao Y. Review of electromagnetic waves-based distance measurement technologies for remote monitoring of civil engineering structures. *Measurement*. 2021;176:18. DOI: 10.1016/j.measurement.2021.109193
- [7] Zhang RH, Yuan YZ, Wang HT, Pei XJ, Meng J. Electromagnetic Interference Analysis, Modeling, and Suppression of High-Power Pulse Drive Source. *IEEE Transactions on Plasma Science*. 2025;53(9):2334-41. DOI: 10.1109/tps.2025.3585722

- [8] Ma ZY, Wei HR, Yuan XF. A Novel Transfer Function Model Based on the Feature Selection Validation Method for Quadrotor Unmanned Aerial Vehicles in High-Intensity Radiated Field Environments. *Electronics*. 2025;14(5):18. DOI: 10.3390/electronics14050976
- [9] Ge WQ, Wang XT, Xie Y. Analysis of strong magnetic environment interference of inspection rotary wing UAV. *IEICE Electronics Express*. 2023;20(18):6. DOI: 10.1587/elex.20.20230321
- [10] Lin C, Ding XC, Zang JF, Zhu F. Identification of Broadband Electromagnetic Interference Sources Based on Joint Optimization of Feature Screening and Classifier. *IEEE Transactions on Instrumentation and Measurement*. 2025;74:13. DOI: 10.1109/tim.2025.3571088
- [11] Lou MY, An HY, Zuo HW, Li ZY, Pu W, Wu JJ, et al. Feature-Enhanced Low-Rank and Sparse Decomposition Network for SAR RFI Suppression. *IEEE Transactions on Geoscience and Remote Sensing*. 2025;63:17. DOI: 10.1109/tgrs.2025.3564509
- [12] Esmaeili M, Hemmady SD, Noakoasteen O, Schamiloglu E, Christodoulou C, Zarkesh-Ha P. Advancing EMC Analysis With GAN-Driven Signal Classification and Waveform Modulation. *IEEE Access*. 2025;13:44789-99. DOI: 10.1109/access.2025.3548033
- [13] Li XP, Wei GH, Pan XD, Sun JN. Interference Feature of Square-Wave Modulated Single-Frequency Signal to Regulated Power Supply. *Electronics*. 2025;14(19):15. DOI: 10.3390/electronics14193842
- [14] Hassab Y, Hillebrecht T, Lurz F, Schuster C. Machine Learning Based Data Validation for Signal Integrity and Power Integrity Using Supervised Time Series Classification. *IEEE Transactions on Electromagnetic Compatibility*. 2024;66(6):2150-8. DOI: 10.1109/temc.2024.3474917
- [15] Lu WC, Duan JD, Cheng L, Lu JP, Dou DX. Electromagnetic Interference Effect Assessment Under Measuring Testability Limitation Based on Physics-Informed Neural Network and Gaussian Process Regression. *IEEE Transactions on Power Electronics*. 2024;39(10):12413-23. DOI: 10.1109/tpel.2024.3401706
- [16] Xing L, Wen YH, Xiao S, Zhang D, Zhang JB. A Deep Learning Approach for Series DC Arc Fault Diagnosing and Real-Time Circuit Behavior Predicting. *IEEE Transactions on Electromagnetic Compatibility*. 2022;64(2):569-79. DOI: 10.1109/temc.2021.3131670
- [17] Kang T, Oh KI, Kim SE, Lee JJ, Kim SE, Byun K, et al. Electromagnetic Interference-Robust Fingerprint Spoof Detection Based on Finger Channel Response. *IEEE Transactions on Industrial Informatics*. 2025;21(11):9045-56. DOI: 10.1109/tii.2025.3594041
- [18] Stracqualursi E, Di Lorenzo G, Calcara L, Araneo R. EMC Issues in High-Power Grid-Connected Photovoltaic Plants: An Update After 15 Years. *IEEE Transactions on Electromagnetic Compatibility*. 2024;66(5):1633-45. DOI: 10.1109/temc.2024.3440848
- [19] Yang LH, Feng W, Wu YJ, Huang L, Quan YH. Radar-Infrared Sensor Fusion Based on Hierarchical Features Mining. *IEEE Signal Processing Letters*. 2024;31:66-70. DOI: 10.1109/lsp.2023.3341397
- [20] Bugala D, Bugala A, Machczynski W. Simulation of the Immunity Test of a Photovoltaic System Disturbed by Electromagnetic Voltage Surge. *Energies*. 2021;14(12):22. DOI: 10.3390/en14123646