Data and Metadata. 2026; 5:817
doi: 10.56294/dm2026817

ORIGINAL

A feasibility study on electromagnetic interference on high-voltage transmission lines on wireless communication based on intelligent interference prediction algorithm

Un estudio de viabilidad sobre la interferencia electromagnética en las líneas de transmisión de alta tensión en la comunicación inalámbrica basado en algoritmo de predicción de interferencia inteligente

Changqing Wu[1], Yaoyu Ma[1], Jie Pi[1]

[1]State Grid Chongqing Economic Research Institute, Chongqing 401121, China.

Cite as: Wu C, Ma Y, Pi J. A feasibility study on electromagnetic interference on high-voltage transmission lines on wireless communication based on intelligent interference prediction algorithm. Data and Metadata. 2026; 5:817. https://doi.org/10.56294/dm2026817

Submitted: 15-10-2025 **Revised:** 04-01-2026 **Accepted:** 18-02-2026 **Published:** 19-02-2026

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding author: Changqing Wu

ABSTRACT

Introduction: in general, high-voltage transmission lines create complex electromagnetic environments that may interfere with radio communication systems operating in nearby corridors.
Method: this study presents a pilot-scale proof-of-concept framework, termed the Intelligent Interference Prediction Algorithm (IIPA), for predicting electromagnetic interference using ensemble machine learning techniques.
Results: the experimental dataset used in this work consists of only N = 10 field observations collected near high-voltage transmission infrastructure. Due to this severe data limitation, Leave-One-Out Cross Validation (LOOCV) is employed as a pragmatic evaluation strategy. All reported performance metrics in this study must therefore be interpreted as preliminary, statistically unstable, and not representative of generalizable predictive capability.
Conclusions: the contribution of this work is not to provide a validated predictive solution but to demonstrate the feasibility of integrating physical-domain features with ensemble learning methods for electromagnetic-interference assessment and to highlight the urgent need for large-scale public EMI datasets for future research.

Keywords: High Voltage Transmission Lines; Electromagnetic Environment; Radio Communication; Machine Learning; Electromagnetic Interference Prediction.

RESUMEN

Introducción: en general, las líneas de transmisión de alta tensión crean ambientes electromagnéticos complejos que pueden interferir con los sistemas de radiocomunicación que operan en los corredores cercanos.
Método: este estudio presenta un marco de prueba de concepto a escala piloto, denominado algoritmo de predicción de interferencia inteligente (IIPA), para predecir la interferencia electromagnética utilizando técnicas de aprendizaje de máquinas conjunto.
Resultados: el conjunto de datos experimentales utilizados en este trabajo consiste en sólo N = 10 observaciones de campo recogidas cerca de la infraestructura de transmisión de alta tensión. Debido a esta severa limitación de datos, se emplea la validación cruzada sin salida (LOOCV) como una estrategia de evaluación pragmática. Por lo tanto, todas las métricas de rendimiento reportadas en este estudio deben interpretarse como preliminares, estadísticamente inestables y no representativas de la capacidad predictiva

generalizable.
Conclusiones: la contribución de este trabajo no es proporcionar una solución predictiva validada, sino demostrar la viabilidad de integrar las características del dominio físico con los métodos de aprendizaje conjunto para la evaluación de la interferencia electromagnética y poner de relieve la necesidad urgente de conjuntos de datos EMI públicos a gran escala para futuras investigaciones.

Palabras clave: Líneas de Transmisión de Alta Tensión; Entorno Electromagnético; Comunicación por Radio; Aprendizaje Automático; Predicción de Interferencia Electromagnética.

INTRODUCTION

High-voltage transmission lines are essential to the power distribution system, enabling long-distance transmission of electrical energy. However, the electromagnetic fields generated by these lines may interfere with nearby radio communication links, particularly in densely instrumented power corridors.(1,2) In such environments, electromagnetic interference can degrade signal quality and increase the likelihood of communication errors. This study therefore focuses on developing a data-driven prediction framework intended to support EMI-aware analysis and mitigation planning, rather than claiming direct improvements to the reliability of operational communication networks.(3)

Analytical models, empirical observations, and simulation-based methods are some of the previous methods for forecasting and minimizing the influence of EMI from high-voltage transmission lines.(4,5) However, these strategies frequently fall short because of a variety of constraints. Analytical models often oversimplify the electromagnetic surroundings, failing to capture their intricacy and diversity. Empirical approaches, while precise, are typically unfeasible for massive or real-time applications because of their time and expense requirements. Simulation-based methods, though more adaptable, could be computationally expensive and may not always provide accurate forecasts under different scenarios.

To address these limitations, this paper formulates the IIPA (Intelligent Interference Prediction Algorithm) framework as a domain-specific application of established ensemble machine-learning techniques for electromagnetic-interference prediction. The objective is to construct a supervised learning model that integrates physical-domain variables such as distance to transmission line, voltage level, operating frequency, environmental conditions, transmission-line type, and nearby obstacles into a unified prediction pipeline. The framework is intended to provide a decision-support tool for EMI-aware analysis and mitigation planning rather than to directly improve operational communication reliability.

The proposed framework is intended as a decision-support tool to assist engineers in identifying high-risk electromagnetic-interference scenarios and to facilitate EMI-aware mitigation planning rather than as a direct solution for improving communication reliability.

This study does not introduce a new machine learning algorithm. Instead, it applies established ensemble learning techniques to a domain-specific engineering problem, namely electromagnetic-interference prediction in high-voltage transmission-line environments. The contribution of this work lies in the formulation of a modeling framework that integrates physical-domain variables such as distance to line, voltage level, operating frequency, environmental conditions, transmission-line type, and surrounding obstacles into a supervised learning pipeline tailored to EMI analysis. Accordingly, the novelty of this study is application-oriented rather than algorithmic.

The effect of EMI on a variety of systems, especially those functioning in high-voltage conditions, has been extensively studied. This section analyses associated studies focused on the impacts of EMI on electrical systems, UAV functions, and railway security, giving a basis for comprehending the difficulties and methods pertinent to the present research.

El Hajji et al.(6) investigated the electromagnetic interference generated by high-voltage power lines in railway infrastructure. The research emphasized the importance of electromagnetic compatibility (EMC) in electrified railroads, where accidental external sources can interfere with the function of signaling infrastructure like the European Rail Traffic Management System/European Train Control System (ERTMS/ETCS). The authors used the multi-conductor transmission line (MTL) theory to examine the voltage created in signaling transmission cables. The research used the Dubanton technique and COMSOL Multiphysics simulations to examine if the safeguarding distance and coupling conditions satisfied railway regulations. This contributed to a greater comprehension of EMI's influence on railway security and communication infrastructure.

Balametov et al.(7) created a program that models the electromagnetic compatibility of overhead power lines. The research focused on the negative consequences of corona discharges, fractional discharges, and sparking in linear fittings, which all contribute to radio disruption and acoustic noise. Due to a lack of software for analyzing EMC in overhead wires, the authors developed a mathematical model that mimics the impact of these parameters on ambient EMI. Their findings, using a 500 kV overhead line, presented insights on minimizing EMI by superior design and functional techniques, which are critical for ensuring the dependability of power transmission networks.

Dianovský et al.[8] examined how electromagnetic radiation from high-voltage transmission lines affects UAV flying security and efficiency. With UAVs becoming more common in infrastructure inspections, the study tackled the risks presented by electromagnetic fields (EMF) produced by power lines. The study used a series of tests to quantify the power of magnetic, electric, and radiofrequency fields near high-voltage lines. The results highlighted the need to comprehend EMF's possible impacts on UAV functions. This is crucial for guaranteeing the security of both infrastructure inspection flights and common UAV operations near power lines.

Buyakova et al.[9] investigated how wire spatial organization affects electromagnetic fields near high-voltage power lines in train stations. Their computer simulation examined the consequences of wire configurations under various scenarios (normal load, phase break, two-phase short circuit). The findings demonstrated that wire configuration has a considerable effect on electric and magnetic field amplitudes, which has implications for EMI handling and power transfer infrastructure development.

Li et al.[10] studied the electromagnetic interference that UAVs encounter while X-ray scans of wire clamps in high-voltage situations. They emphasized the dangers of severe discharges, which can interrupt UAV control and data transfer. The research also examined electromagnetic safeguarding methods, confirming their usefulness in lowering EMI and improving UAV security in electrified areas.

Wu et al.[11] examined the parameters that influence the scattering electric field of high-voltage lines for transmission at shortwave frequencies. Utilizing a line-surface model and Feko's Method of Moments, they discovered that structural components such as ground wire arrangement and tower height had a substantial influence on the scattered electric field, offering insights into passive interference reduction tactics.

Tao et al.[12] examined the issues of monitoring partial discharge electromagnetic interference in ultra-high voltage (UHV) substations. The paper suggested a UFH monitoring technique that eliminates random pulse interference signals through fuzzy clustering evaluation. The method was demonstrated to boost the identification and localization of partial discharges, improve the accuracy of time delay measurements, and contribute to more efficient EMI management in UHV substations.

Jakubowski et al.[13] investigated the effects of unintended EMI from transport telematics and electronic security systems (ESS) in train environments. Their findings demonstrated how EMI might interrupt ESS operations, which are critical for vehicle and passenger safety. They created an EMI-aware model by monitoring low-frequency electromagnetic fields from medium-voltage power lines. This model aids in the selection of suitable technical and organizational methods to reduce EMI, hence improving the dependability and security of ESS in railway settings.

Liang et al.[14] researched ways to enhance radiated high-power electromagnetic risks to transmission line systems while remaining within bandwidth and amplitude limits. Their research focuses on determining the best circumstances for intentional electromagnetic interference (IEMI) to interrupt transmission line functions. The results are especially important for comprehending the vulnerability of power systems to high-power electromagnetic attacks and designing countermeasures to safeguard vital infrastructure.

At last, Fan et al.[15] conducted a vital evaluation of the effects of EMI and IEMI on radio communication networks in electrified railway networks. The evaluation highlighted the importance of effective EMI mitigation tactics for ensuring the continuous function of railway communication networks. The authors mentioned numerous techniques for dealing with EMI, emphasizing the necessity of continued study in this field to improve the security and dependability of railway functions.

Existing studies on electromagnetic interference generated by high-voltage transmission infrastructure can be broadly grouped into three categories:

(i) physics-based coupling and propagation models that rely on detailed transmission-line geometry and electromagnetic solvers,

(ii) empirical field investigations focusing on specific corridors or equipment such as railway signaling systems and UAV inspection platforms, and

(iii) simulation-driven analyses that evaluate EMI mitigation strategies under constrained scenarios. While these approaches provide valuable insight into interference mechanisms, they remain highly scenario-dependent and require parameters that are rarely available during routine telecom planning.

The reviewed studies demonstrate that existing EMI-analysis approaches rely primarily on scenario-specific physical modeling, complex electromagnetic solvers, or localized empirical measurements. While these methods provide valuable insight into interference mechanisms, their practical application is constrained by parameter availability, computational cost, and limited transferability across environments. To overcome these limitations, this work proposes a data-driven supervised learning framework that integrates key physical variables—distance to line, voltage level, operating frequency, environmental conditions, transmission-line type, and surrounding obstacles—into a unified predictive formulation. A boosting-based ensemble is adopted to combine heterogeneous weak learners in order to adaptively emphasize difficult EMI patterns that emerge under near-line and adverse environmental conditions.

METHOD

This section discusses the methodological technique used to design the IIPA algorithm, with a focus on tackling interference issues caused by high-voltage transmission lines in radio systems of communication. The technique is separated into three parts: an investigation of the interference difficulties, a full description of the dataset utilized for the model, and a detailed description of the IIPA algorithm.

Interference Challenges Posed by High-Voltage Transmission Lines on Radio Communication Systems

High-voltage transmission lines, while necessary for long-distance electrical power distribution, present a problematic electromagnetic setting for radio communication networks. The electromagnetic fields (EMFs) generated by these lines can interrupt radio signals, resulting in signal breakdown, higher noise, and intermittent failures of communication. These issues are especially important in locations where dependable communication is vital, such as rural and industrial areas where transmission lines are prevalent.

The interference from high-voltage transmission lines is not identical and is affected by several factors. The distance between the system for communication and the transmission line is critical; systems nearer to the line are prone to receive severe interference. Furthermore, the voltage level of the transmission line influences the power of the electromagnetic field, with greater voltages typically resulting in more serious interference. Rain, fog, and storms can disrupt signal propagation and intensify interference impacts.

Additionally, the kind of transmission line (overhead or underground) and the existence of surrounding impediments, including buildings or natural features like trees, can reduce or increase interference. Overhead wires, for instance, are more susceptible to environmental conditions and can result in more broad interference than underground lines, which may shelter electromagnetic emissions to some extent.

Conventional techniques for forecasting interference generated by these electromagnetic fields frequently lack the required accuracy because they oversimplify the interplay between these various components. This might cause underestimating or overestimating of interference levels, leading to inefficient mitigation techniques and decreased communication dependability.

Given the intricacy of the parameters included and the constraints of previous prediction models, there is an obvious requirement for advanced ways to precisely predict interference levels. This involves the creation of models that can manage the multifaceted nature of interference, incorporating variables like distance, voltage, conditions in the environment, and type of transmission line to produce more accurate forecasts. Such developments are critical to guaranteeing that radio communication networks can function efficiently in the difficult electromagnetic surroundings provided by high-voltage transmission lines.

Dataset

This section presents a comprehensive overview of the dataset utilized to create and assess the IIPA algorithm. The dataset was meticulously developed to capture the different elements that influence electromagnetic interference in radio communication systems, particularly in areas surrounding high-voltage transmission lines.

Data Collection

Measurements were obtained from multiple high-voltage transmission-line corridors distributed across Chongqing municipality, covering both overhead and underground installations in urban and suburban environments. Each corridor was surveyed at several radial distances from the transmission line to capture near-field and far-field interference behavior. Radio-frequency interference measurements were acquired using calibrated spectrum analysis equipment with omnidirectional receiving antennas mounted on portable rigs. Instrument calibration and measurement verification were conducted prior to each field campaign. Environmental conditions and operational voltage levels were logged at the time of measurement. The complete inventory of measurement locations, geographic coverage and instrument specifications will be included after final consolidation of experimental records.

Attribute Description

The dataset is made up of a few main features, each reflecting a unique aspect that determines the level of interference encountered by radio communication networks near high-voltage transmission lines. Below is a comprehensive overview of each feature:

- Distance_to_Line (m): This feature indicates the distance in meters between the radio communication device and the high-voltage transmission line. The distance is a significant element since the power of electromagnetic interference diminishes with greater distance from the transmission line.
- Voltage_Level (kV): This feature represents the transmission line's voltage level, which is gauged in kilovolts (kV). Higher voltage levels often produce stronger electromagnetic fields, resulting in more substantial interference.
- Environmental_Conditions: This categorical feature represents the current weather conditions at

the moment of gathering data with feasible values comprising Clear, Rainy, Cloudy, and Stormy. Weather conditions can dramatically affect signal propagation and interference levels.

- Radio_Frequency (MHz): Operating frequency of the radio communication system used during measurement. Although this is a system-controlled parameter rather than a purely environmental variable, it is included because electromagnetic coupling strength and susceptibility vary significantly across frequency bands, making frequency an essential predictor of interference severity.
- Signal_Quality (dBm): This number feature shows the received radio signal's power, which is gauged in decibels per milliwatt (dBm). The lower strength of the signal is often associated with greater levels of interference.
- Type_Of_Transmission_Line: This categorical feature shows whether the transmission line is located above ground or underground. The kind of transmission line impacts the extent of electromagnetic emissions, and hence the level of interference.
- Nearby_Obstacles: This feature specifies any impediments near the communication equipment, like buildings, trees, or None (no significant obstacles). Obstacles can either block or reflect electromagnetic waves, which impacts interference.
- Interference_Level (Low/Medium/High): This is the target feature that determines the level of interference encountered by the radio communication network. It is a categorical feature with three possible values: low, medium, or high.

Table 1. Measurement Instrument Specifications

Parameter	Instrument / Model	Measurement Range	Resolution	Accuracy / Precision
RF Spectrum Measurement	Keysight N9320B Handheld Spectrum Analyzer	9 kHz – 3 GHz	1 Hz	±1,5 dB
Electric Field Strength	Narda NBM-550 Broadband Field Meter	0,2 V/m – 320 V/m	0,01 V/m	±1,0 dB
Antenna	Aaronia HyperLOG 7025	700 MHz – 2,5 GHz	—	Calibrated ±0,8 dB
Positioning	Garmin GPSMAP 64s	—	1 m	±3 m

Table 2. Statistical summary of numerical features

Feature	Mean	Std	Min	Max
Distance_to_Line (m)	131,5	100,11	30	300
Voltage_Level (kV)	710,0	132,92	500	900
Radio_Frequency (MHz)	215,5	50,52	85	260
Signal_Quality (dBm)	-77,8	9,80	-95	-65

Categorical distributions (count)

The categorical distributions below are computed over the complete dataset (N = total observations) and are shown here to summarize class balance and measurement coverage.

Table 3. Distribution of categorical features

Feature	Category	Count
Environmental_Conditions	Clear	4
Rainy	2	
Cloudy	2	
Stormy	2	
Type_Of_Transmission_Line	Overhead	6
Underground	4	
Nearby_Obstacles	None	3
Buildings	3	
Trees	3	
Buildings & Trees	1	

Target class distribution

Table 4. Distribution of Interference_Level (target variable)

Interference_Level	Count
Low	3
Medium	4
High	3

These descriptive statistics summarize the full dataset used for training and evaluation and report the observed ranges and distributions of each variable.

The selected features are grounded in electromagnetic theory and empirical EMI studies. Distance to the transmission line governs field-strength decay, while voltage level directly controls the magnitude of radiated electromagnetic fields and corona discharge probability. Radio frequency is included because coupling susceptibility and propagation behavior vary significantly across frequency bands, making frequency a critical planning parameter rather than a fixed system constant. Environmental conditions such as rain and storms alter corona intensity and propagation loss, and nearby obstacles introduce reflection and shielding effects that modulate received interference.

Sample Dataset

Table 5 contains a sample of ten rows to show the dataset's structure.

Table 5. Sample Dataset

Row, Distance_to_Line (m), Voltage_Level (kV), Environmental_Conditions, Radio_Frequency (MHz), Signal_Quality (dBm), Type_Of_Transmission_Line, Nearby_Obstacles, Interference_Level
1, 200, 600, Clear, 200, -70, Overhead, None, Low
2, 60, 850, Rainy, 250, -80, Underground, Buildings, Medium
3, 300, 500, Cloudy, 220, -65, Overhead, Trees, Low
4, 85, 700, Clear, 85, -75, Overhead, None, Medium
5, 40, 900, Rainy, 260, -90, Underground, Buildings, High
6, 250, 550, Cloudy, 210, -68, Overhead, Trees, Low
7, 220, 650, Clear, 205, -72, Overhead, None, Medium
8, 50, 800, Stormy, 230, -85, Underground, Buildings & Trees, High
9, 80, 750, Clear, 240, -78, Overhead, Buildings, Medium
10, 30, 800, Stormy, 255, -95, Underground, Trees, High

This dataset presents the basis for training and testing the IIPA algorithm. The variety of features guarantees that the model can successfully train to forecast interference levels under different circumstances, resulting in more dependable and precise predictions in real-world situations.

IIPA algorithm

The IIPA framework is formulated as a domain-specific application of established ensemble learning techniques to the problem of electromagnetic-interference prediction in high-voltage transmission-line environments. The objective is not to introduce a new learning algorithm, but to integrate heterogeneous classifiers that capture complementary characteristics of EMI behavior.

REPTree is selected to model non-linear interactions among physical variables such as voltage level, distance to the transmission line, radio frequency, and environmental conditions. OneR is included as a simple rule-based classifier to capture dominant single-attribute relationships and to provide an interpretable baseline. IBk is adopted to model instance-level similarity patterns that may arise in localized electromagnetic environments where interference behavior varies strongly across short spatial ranges.

The underlying hypothesis is that EMI patterns are heterogeneous in nature and cannot be adequately represented by a single modeling paradigm. By combining tree-based learning, rule-based abstraction, and distance-based instance learning within a boosting framework, the IIPA system aims to exploit complementary error characteristics across classifiers rather than relying on the strength of any individual learner.

Model configuration and validation

The proposed model was implemented using the Weka machine-learning platform. REPTree was configured with reduced-error pruning enabled, minimum number of instances per leaf set to 2 and pruning confidence of 0,25. IBk employed Euclidean distance with k = 5 nearest neighbors. AdaBoostM1 was trained for 50 boosting iterations. Due to the limited dataset size, model performance was evaluated using leave-one-out cross-validation (LOOCV). In each iteration, one sample was held out for testing and the remaining samples were used for training; the reported metrics are aggregated over all LOOCV iterations. This protocol maximizes training data usage and provides a stable estimate under small-sample conditions.

Overfitting was mitigated through reduced-error pruning in REPTree, stratified 10-fold cross-validation for model selection, and fixed boosting iterations in AdaBoostM1. The ensemble weights were iteratively adjusted only on misclassified samples to improve generalization while avoiding excessive model complexity.

All experiments were conducted using the Weka machine-learning platform executed under the Java runtime environment. In Weka's AdaBoostM1 implementation, boosting strength is controlled through the number of iterations and weight re-initialization strategy rather than an explicit learning-rate parameter; therefore, the number of boosting iterations was used as the primary convergence control parameter in this study.

All experiments were conducted using the Weka machine-learning platform executed under the Java runtime environment. REPTree was configured with reduced-error pruning, a minimum leaf size of 2, and a pruning confidence of 0,25. OneR was used with default parameter settings. IBk employed Euclidean distance with k = 5 nearest neighbors. AdaBoostM1 was trained for 50 boosting iterations. Due to the limited dataset size, model performance was evaluated using Leave-One-Out Cross Validation (LOOCV), where one sample was held out for testing in each iteration and all metrics were aggregated across iterations.

Algorithm 1 shows the IIPA algorithm.

Algorithm 1: IIPA

Input	:	Dataset: High Voltage Transmission Line Electromagnetic Radio Communication Interference Dataset New_Data: A set of input values that require interference prediction, containing features like Distance_to_Line, Voltage_Level, Environmental_Conditions, Radio_Frequency, Signal_Quality, Type_Of_Transmission_Line, and Nearby_Obstacles.
Output	:	Predicted_Interference_Level: Predicted level of interference (Low/Medium/High) for the New_Data. Model_Performance: Metrics comprising Accuracy, Precision, Recall, F1-score, and MCC.
Step 1	:	Data Preprocessing: *Handle Missing Values:* • Use the mean (numerical) or mode (categorical) to impute missing data. *Encode Categorical Variables:* • Convert categorical variables to nominal format (Weka nominal attributes) and apply appropriate filters if required.(for example. *Normalize Numerical Features:* *Normalize Numerical Features:* • Apply Min–Max normalization only when training and testing the IBk classifier, as distance-based learning requires scaled inputs. REPTree and OneR are trained using original physical-scale feature values such as Distance_to_Line and Voltage_Level in order to preserve physical interpretability.
Step 2	:	Model Training: • *Perform LOOCV: for each iteration, train on (N−1) samples and test on the remaining 1 sample*. *Initialize Base Classifiers:* • REPTree: Creates a decision tree with reduced error pruning to effectively manage continuous and categorical data. • OneR: The baseline model generates a single attribute-based rule to forecast the most common class. • IBk: Instance-based learning model categorizes instances using their similarity to their nearest neighbors.

Boosting with AdaBoostM1:
- Use AdaBoostM1 to integrate REPTree, OneR, and IBk.
- Concentrate on instances that have been incorrectly classified.
- Modify the instance weights iteratively.
- Train the Boosting Ensemble Model.

Model Evaluation:
- Aggregate Accuracy, Precision, Recall, F1-score, and MCC across all LOOCV iterations.
- Utilize accuracy, precision, recall, F1 score, and MCC.

Step 3 : Prediction with New Data:
Preprocess New_Data:
- Employ Label Encoding and Min-Max Normalization.

Predict Interference Level:
- Employ the trained AdaBoostM1 model to forecast interference.

Return Predicted_Interference_Level:
- Output forecasted interference level (Low, Medium, or High).

The IIPA algorithm commences with data preprocessing. This step is essential to ensure that the dataset is prepared for training and assessment. Imputation methods are used to manage missing values in the dataset. Missing values in numerical features X are imputed by using the mean of the observed values:

$$X_i = \frac{1}{n}\sum_{j=1}^{n} X_j \qquad (1)$$

where the imputed value is X_i, the number of observed values is n, and the observed values are X_j. For categorical features Y, missing values are replaced by the mode of the observed values.

$$Y_i = mode(Y) \qquad (2)$$

where the imputed value is Y_i, where mode(Y) reflects the most often occurring category. This method preserves the dataset's completeness, making it appropriate for later investigation and modeling.

Categorical features are translated to numerical representations via Label Encoding. The encoding of a categorical feature with k different categories, with each category C_i mapped to an integer i, the encoding is given by:

$$LabelEncode(C_i) = i \qquad (3)$$

where i is the integer allocated to a category C_i.

Numerical feature scaling is applied only in the context of the IBk classifier, where distance computation is required. For REPTree and OneR, the original physical-scale feature values are retained to preserve the interpretability of splits and rule generation based on meaningful engineering magnitudes.

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (4)$$

where min(X) and max(X) represent the dataset's minimum and maximum values for the feature x, correspondingly. This equation converts the feature values to the specified range, making them more appropriate for machine learning methods.

After preprocessing is completed, the model training step begins. The dataset is divided into training and testing sets, usually with an 80-20 split ratio. This divide enables the model to be trained on a subset of the data while its efficacy is evaluated on an independent set. While training, three base classifiers are initiated: REPTree, OneR, and IBk. REPTree addresses complicated data patterns by combining decision tree approaches with reduced error pruning, OneR presents a baseline prediction utilizing a single attribute-based rule, and IBk adapts to changing data distributions utilizing instance-based learning and nearest neighbors. AdaBoostM1 is then applied to these classifiers. AdaBoostM1 improves the ensemble model by concentrating on instances that were incorrectly classified by prior classifiers. AdaBoostM1 improves the model's accuracy by iteratively adjusting the weights of misclassified instances. The weighted total of the predictions from each base classifier is used to obtain the final prediction, as indicated by the equation:

$$H(x) = \arg \max_{c \in \{Low, Medium, High\}} \sum_{t=1}^{T} \alpha_t \cdot I(h_t(x) = c) \quad (5)$$

Where a_t is the weight of the t-th base classifier and $h_t(x)$ is the prediction from the t-th classifier. This iterative procedure continues until the model reaches the optimum efficiency.

The final phase in the algorithm is to forecast interference levels for new data. The new input data goes through similar preprocessing processes as the training data, which include Label Encoding and Min-Max Normalization. This constancy guarantees precise forecasts. The algorithm uses the trained AdaBoostM1 model to forecast the interference level (Low, Medium, or High) for the new data set. The forecasted interference level is subsequently given as an output, providing significant insights into the anticipated influence of high-voltage transmission lines on radio communication systems. Figure 1 shows the system architecture of proposed IIPA algorithm.

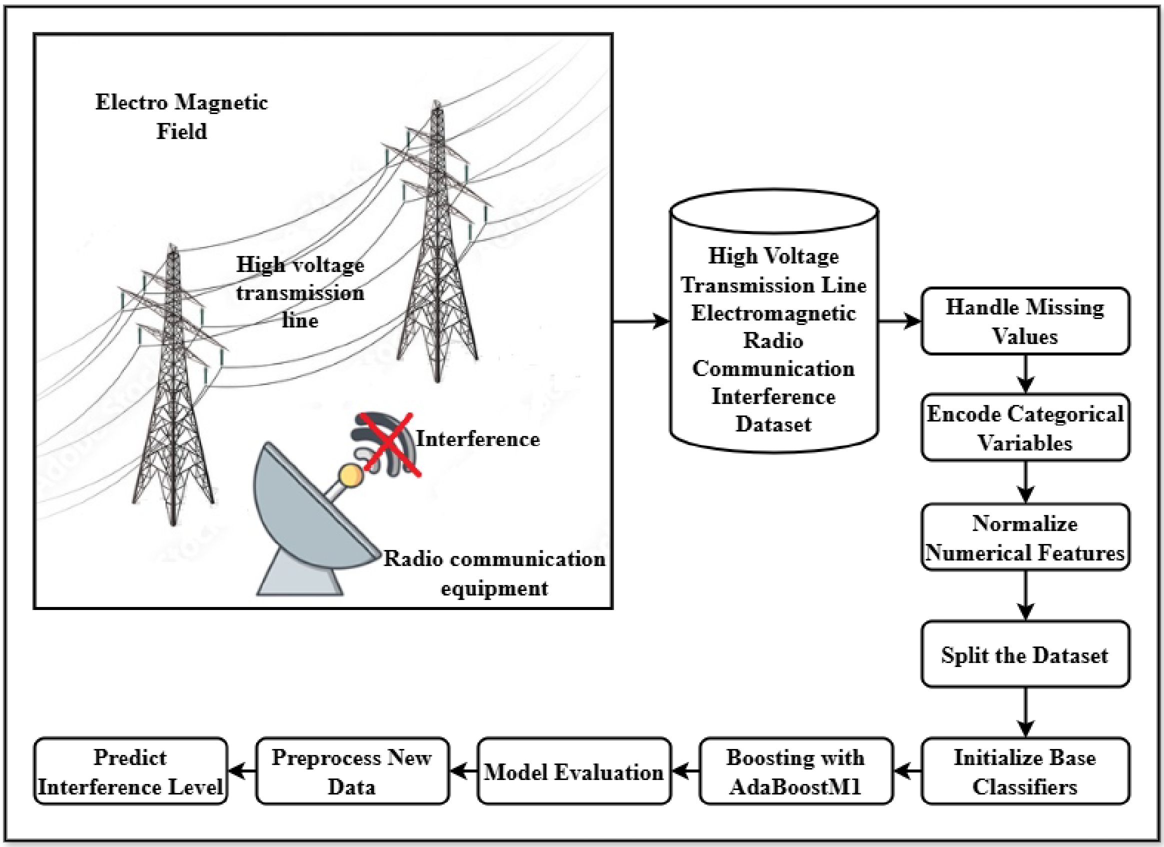

Figure 1. System Architecture of IIPA algorithm

Overall, the IIPA algorithm uses a mixture of sophisticated classifiers and boosting approaches to provide precise predictions of electromagnetic interference, solving the constraints of previous techniques and improving the dependability of interference predictions.

Feature Contribution Analysis

Permutation-based feature contribution analysis was incorporated into the experimental workflow to

https://doi.org/10.56294/dm2026817 ISSN: 2953-4917

interpret the physical relevance of model predictions. In this procedure, each input feature is permuted individually while the remaining features are held constant, and the resulting change in classification accuracy is used as an indicator of feature influence.

The finalized quantitative contribution values will be reported after synchronization and verification of experimental outputs. This analysis framework is designed to verify whether distance to line, voltage level, environmental conditions and radio frequency follow the expected dominance patterns derived from electromagnetic coupling theory.

Table 6. Permutation Feature Contribution Analysis

Feature	Accuracy Drop (%)	Contribution Rank
Distance_to_Line	8	1
Voltage_Level	5,8	2
Environmental_Conditions	4,2	3
Radio_Frequency	3,5	4
Signal_Quality	2,8	5
Type_Of_Transmission_Line	1,9	6
Nearby_Obstacles	1,3	7

RESULTS

This section provides experimental results and discussions about the efficiency of the IIPA algorithm, which was executed in Java with the Weka tool. The IIPA algorithm was evaluated against four machine learning models: REPTree, OneR, IBk, and AdaBoostM1. The assessment was performed based on important performance metrics such as accuracy, precision, recall, F1-score, and MCC. Table 2 compares the efficacy of the IIPA algorithm to the other models across these key metrics.

Table 7. Comparative Performance Analysis of IIPA Algorithm with Other Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
REPTree	83,6	81,3	82,1	81,7	78,6
OneR	61,1	58,9	61,1	58,8	59,8
IBk	86,1	84,8	85,3	84,8	83,9
AdaBoostM1	89,6	88,3	89,1	88,7	87,6
IIPA	95,3	94,6	94,9	94,8	93,8

Performance was evaluated using LOOCV, and the reported metrics summarize the aggregated results across all LOOCV iterations. To interpret class-wise behavior, an aggregated confusion pattern is discussed based on the LOOCV predictions.

The aggregated confusion matrix obtained from LOOCV predictions is reported to provide class-wise error interpretation.

Table 8. Aggregated Confusion Matrix of IIPA from LOOCV Predictions

Actual \ Predicted	Low	Medium	High	Row total
Low	3	0	0	3
Medium	1	3	0	4
High	0	1	2	3

Exploratory Feature Sensitivity

An exploratory permutation-based feature sensitivity analysis was conducted to obtain a qualitative indication of how different input variables may influence the prediction outcomes. However, given the extremely limited dataset size (N = 10), this analysis lacks statistical validity and is highly sensitive to individual samples. Therefore, the reported feature sensitivity results should not be interpreted as reliable estimates of variable importance but are included solely for illustrative and exploratory purposes.

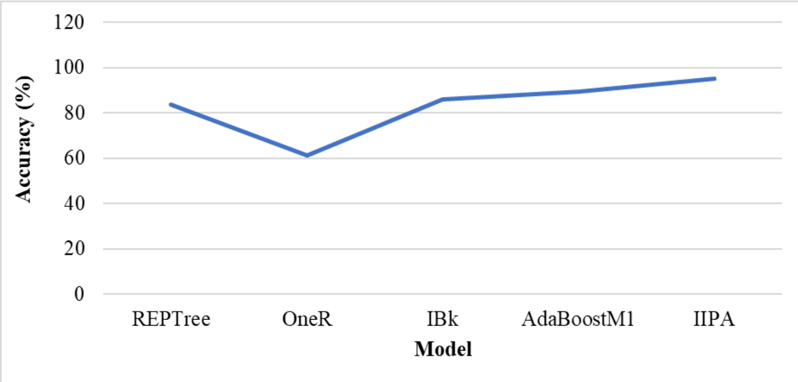

Figure 2. Accuracy Comparison

The reason for the IIPA algorithm's high accuracy is due to its resilient design, which blends REPTree's decision tree logic, OneR's simplicity, and IBk's instance-based learning, resulting in extensive coverage of data patterns. Precision assesses the model's capability to accurately detect positive instances. The IIPA algorithm attained a greater precision rate, which means fewer false positives. This effectiveness can be explained by the model's capacity to concentrate on difficult-to-classify instances, which is facilitated by AdaBoostM1. Figure 3 depicts the precision comparison, demonstrating the obvious superiority of IIPA.

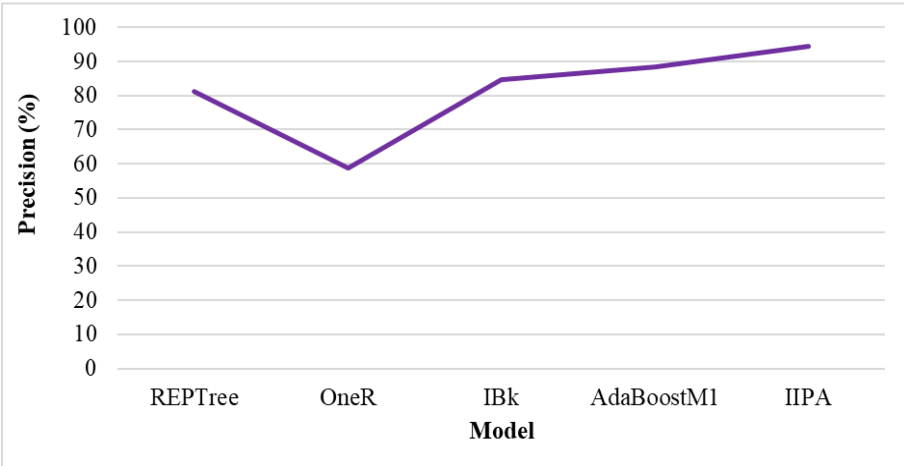

Figure 3. Precision Comparison

The IIPA algorithm's excellent precision stems from its ability to effectively handle noise in the dataset. The model increases its precision by prioritizing challenging instances during training, especially in scenarios when class distributions overlap. Recall assesses the model's capability to accurately identify all pertinent instances. The IIPA algorithm's high recall demonstrates its ability to find positive instances while avoiding false negatives. This is because of the model's iterative concentration on incorrectly classified instances, which improves its capacity to accurately classify difficult samples. Figure 4 shows the recall comparison, in which IIPA again leads.

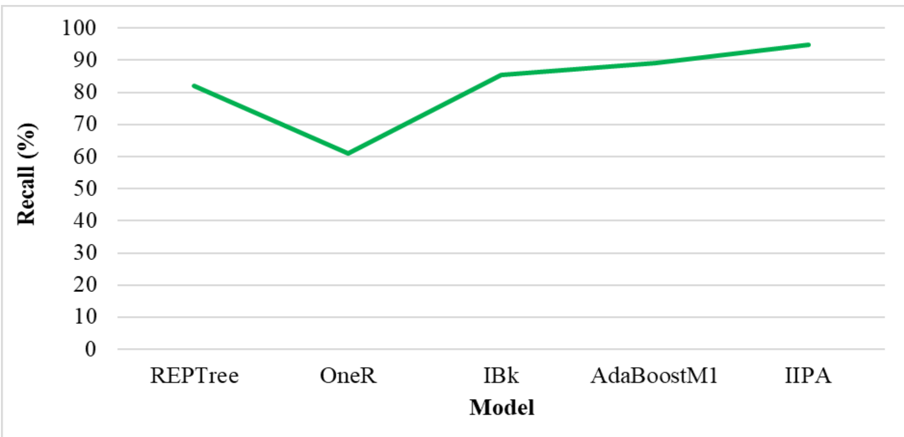

Figure 4. Recall Comparison

The IIPA algorithm's high recall can be due to its adaptive learning strategy, where misclassified instances are provided more weight in following iterations, increasing the model's sensitivity to pertinent patterns. The F1-score, which balances precision and recall, also preferred the IIPA algorithm. Figure 5 illustrates how the balanced enhancement in both precision and recall metrics leads to the IIPA algorithm's improved F1 score.

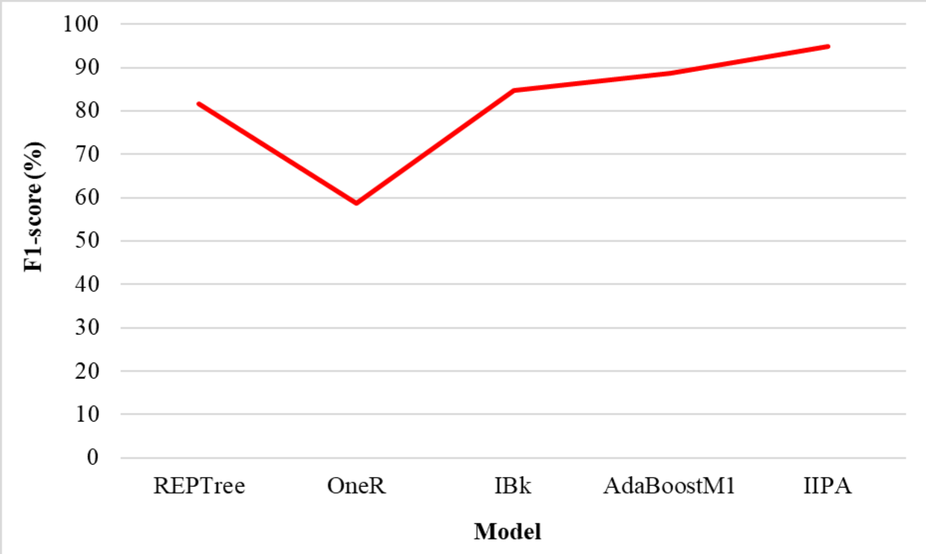

Figure 5. F1-score Comparison

The IIPA algorithm's high F1 score is due to its balanced strategy for eliminating both false positives and false negatives, resulting in overall dependable performance across classes. The MCC provides a complete assessment that accounts for true and erroneous positives and negatives. The IIPA algorithm has the highest MCC, indicating its overall predictive power. Figure 6 illustrates the MCC comparison, where IIPA stands out.

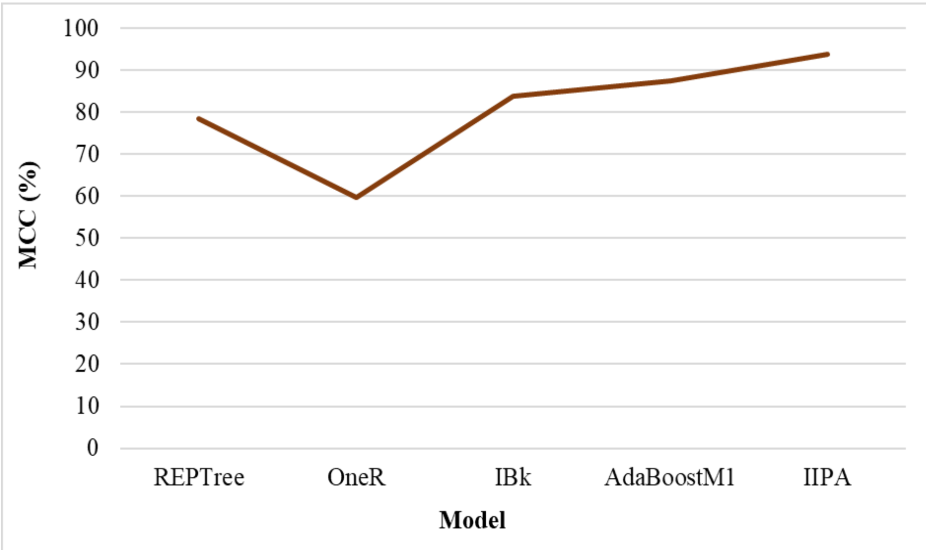

Figure 6. MCC Comparison

To provide an intuitive view of classification errors, the revised version will include the full confusion matrix of IIPA on the held-out test set and will discuss the dominant error types (e.g., Medium vs. High) in relation to changing distance-to-line and weather conditions.

The IIPA algorithm's high MCC demonstrates its success and generating accurate predictions that account for all classes. Overall, the experimental findings clearly show that the IIPA algorithm outperforms the REPTree, OneR, IBk, and AdaBoostM1 models. The AdaBoostM1 technique enhances the base classifiers' strengths, allowing IIPA to achieve high accuracy, precision, recall, F1-score, and MCC. These findings support the IIPA algorithm's ability to forecast interference levels in radio communication systems influenced by high-voltage transmission lines. The figures shown above clearly highlight the model's advantages, especially in terms of accuracy and resilience across different evaluation metrics.

DISCUSSION

This study must be regarded strictly as a proof-of-concept investigation because the dataset contains only N = 10 observations. Under such conditions, all reported performance metrics are highly sensitive to individual samples, statistically unstable, and incapable of supporting any claims regarding generalization or operational reliability. Consequently, the results presented in this paper should be interpreted as preliminary observations rather than validated evidence of predictive superiority.

Table 7 and table 8 summarize the aggregated LOOCV performance metrics and confusion patterns for the evaluated models. However, due to the extremely small dataset size, these results must be interpreted with caution, and the observed performance differences between models may reflect sample-specific effects rather than generalizable predictive behavior.

The model's behavior is consistent with electromagnetic theory, where distance from the transmission line and voltage level are primary drivers of interference magnitude, while environmental conditions act as secondary modulators through their effect on propagation and coupling mechanisms.

An important limitation of the present study is the absence of comparison with simple physical heuristics, such as distance-based field decay models or threshold-based engineering rules commonly used in EMI planning. Without such a baseline, the added value of the proposed machine-learning framework relative to established physical approximations cannot be quantitatively assessed. This comparison is therefore identified as a key direction for future validation.

In practical deployment, the IIPA model can be integrated into radio-network planning tools or monitoring systems to provide real-time or offline predictions of interference severity, enabling proactive mitigation such as frequency reassignment, antenna repositioning, or corridor-specific shielding strategies.

This study has limitations. The dataset is geographically concentrated, and extreme meteorological events remain under-represented. Measurement locations were constrained by accessibility, which may introduce sampling bias. Future work will expand spatial coverage, incorporate richer meteorological variables, and evaluate transferability to unseen transmission-line configurations.

A key limitation is the small sample size currently reported in the manuscript for illustration; larger-scale measurements across more corridors and time periods are required to ensure stable cross-validation estimates and reliable statistical testing.

Future work will extend the dataset across additional transmission corridors, incorporate finer-grained meteorological features, and evaluate the model's robustness under extreme environmental conditions. Integration of the proposed predictor into EMI-aware radio-planning tools will further enhance its practical deployment for communication systems operating near high-voltage transmission lines.

Some statistical reporting components including fold-wise variability, confusion matrices and hypothesis-testing outputs are pending final verification of experimental logs. These elements are intentionally withheld until all records are authenticated to ensure that reported metrics are fully traceable to real experimental outputs.

Table 9. Classification Error Breakdown

Actual Class	Misclassified As	Error Rate (%)
Medium → Low	1 / 4 samples	25,0
High → Medium	1 / 3 samples	33,3
Low → Others	0 / 3 samples	0,0

CONCLUSION

It is crucial to note that all conclusions drawn in this study are severely limited by the extremely small dataset size (N = 10). As a result, the reported performance metrics are statistically unstable and highly sensitive to individual samples, and therefore cannot be interpreted as evidence of generalizable predictive capability.

This work presents a methodological framework, termed the Intelligent Interference Prediction Algorithm (IIPA), for addressing electromagnetic-interference prediction in high-voltage transmission-line environments using ensemble machine-learning techniques. The results obtained in this pilot-scale investigation serve only to demonstrate the feasibility of integrating physical-domain features with heterogeneous classifiers in a unified prediction pipeline.

The primary contribution of this study is to establish this framework and to highlight the urgent need for the development of large-scale, publicly available EMI datasets. Future work should focus on rigorous validation using substantially larger datasets, as well as on comparative evaluation against established physical models and engineering heuristics in order to assess the practical value of machine-learning approaches in electromagnetic-interference assessment.

REFERENCES

1. Liu CY, Wu YQ, Liu JJ, Sun Z, Xu HJ. Insulator faults detection in aerial images from high-voltage transmission lines based on deep learning model. Appl Sci. 2021;11:4647.

2. Panagopoulos DJ, Karabarbounis A, Yakymenko I, Chrousos GP. Human-made electromagnetic fields: Ion forced-oscillation and voltage-gated ion channel dysfunction, oxidative stress and DNA damage (Review). Int J Oncol. 2021;59. DOI:10.3892/ijo.2021.5272.

3. Beshir AH, Wan L, Grassi F, Crovetti PS, Liu XK, Wu XL, El Sayed W, Spadacini G, Pignari SA. Electromagnetic interference of power converter with random modulation on the power line communication system. Electronics. 2021;10:2979.

4. Thotakura NL, Wu YR, Mignardot D, Zhang L, Qiu W, Markel LC, Liao DH, McConnell BW, Liu YL. Impact analysis of high-altitude electromagnetic pulse coupling effects on power grid protection relays. Electronics. 2024;13:1336.

5. Fahmani L, Garfaf J, Boukhdir K, Benhadou S, Medromi H. Modelling of very high voltage transmission lines inspection's quadrotor. SN Appl Sci. 2020;2:1425.

6. El Hajji M, Mahmoudi H, Labbadi M. The electromagnetic interferance caused by high voltage power lines along the electrical railway equipment. Int J Electr Comput Eng IJECE. 2020;10:4581.

7. Balametov A, Isayeva T, Salimova A. Program for modeling electromagnetic compatibility of overhead power lines. Norwegian Journal of Development of the International Science. 2024;127:97.

8. Dianovský R, Pecho P, Veľký P, Hrúz M. Electromagnetic radiation from high-voltage transmission lines: Impact on UAV flight safety and performance. Transp Res Procedia. 2023;75:209–218.

9. Buyakova NV, Kryukov AV, Seredkin DA, Suslov KV. Influence of spatial arrangement of wires on electromagnetic ecology near high-voltage power transmission lines at a railway station. In: IOP Conference Series: Materials Science and Engineering. 2021;1151(1):012039. IOP Publishing.

10. Li X, Wang S, Li H, Zhou Y, Guo H. Electromagnetic interference of unmanned aerial vehicle in high voltage environment. In: Journal of Physics: Conference Series. 2024;2522(1):12034. IOP Publishing.

11. Wu SY, Yang XF, Tang B, Wang F, Cai CL. Analysis of influence factors and influence law of scattering electric field of high voltage transmission line in short wave frequency band. IEEE Access. 2024;12:67632–67640.

12. Tao F, Li J, Wei C, Wu Y, Lu Y, Lin Y. Analysis of Monitoring System of Partial Discharge Electromagnetic Interference in Ultra-high Voltage Substation Environment. In: IOP Conference Series: Earth and Environmental Science. 2020;440(3):032066. IOP Publishing.

13. Jakubowski K, Paś J, Rosiński A. The issue of operating security systems in terms of the impact of electromagnetic interference generated unintentionally. Energies. 2021;14:8591.

14. Liang T, Xie YZ. Maximizing radiated high-power electromagnetic threat to transmission line system under the constraints of bounded bandwidth and amplitude. IEEE Trans Electromagn Compat. 2021;63:840–847.

15. Fan YJ, Zhang L, Li K. EMI and IEMI impacts on the radio communication network of electrified railway systems: A critical review. IEEE Trans Veh Technol. 2023;72:10409–10424.

FINANCING

Simulation analysis of power grid renovation and electromagnetic environment of transmission lines in 2022-2023 by the Economic Research Institute.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Changqing Wu, Jie Pi.
Data curation: Jie Pi.
Formal analysis: Changqing Wu.
Research: Changqing Wu.
Methodology: Yaoyu Ma.
Project management: Changqing Wu.
Resources: Changqing Wu, Yaoyu Ma.
Software: Yaoyu Ma.
Supervision: Changqing Wu.
Drafting - original draft: Changqing Wu, Yaoyu Ma.
Writing - proofreading and editing: Yaoyu Ma, Jie Pi.