© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1711914

Machine Learning Based Classification of Electromagnetic Interference Using Synthetic Signal Features

SHANIJA S R[1], GOPIKA K[2], SITTALATCHOUMY R[3]
[1, 2, 3]*Anna University*

Abstract- Electromagnetic interference (EMI) poses significant challenges in modern electronic and communication systems by degrading signal integrity and overall system performance. This project presents a machine learning-based classification approach for identifying and distinguishing between different types of EMI using synthetic signal features. A dataset containing 600 samples was created, consisting of 100 clean signals and 500 EMI-contaminated signals generated from five sources like motor, switching, lighting, Wi-Fi, and inverter. Key signal features such as mean, variance, root mean square (RMS), dominant frequency, first harmonic, and second harmonic were extracted to characterize each signal. Two machine learning algorithms such as Decision Tree and SVM were trained and tested using an 80:20 ratio for training and testing, respectively. The classification performance was evaluated using confusion matrices and classification reports. The Decision Tree model achieved an accuracy of 58%, while the SVM attained 98% accuracy, demonstrating its robustness and superior generalization ability. Additionally, a user-interactive interface was developed with a drop-down menu enabling users to select a signal type and obtain real-time classification results. The proposed system provides an efficient and automated method for EMI source identification, which can aid in EMI mitigation and system reliability enhancement.

I. INTRODUCTION

Electromagnetic interference (EMI) is an unwanted disturbance that affects electrical circuits and electronic devices due to electromagnetic radiation emitted from natural or man-made sources. With the rapid growth of industrial automation, communication networks, and electronic devices, EMI has become a major concern affecting signal quality and system performance. Accurate identification and classification of EMI sources are essential for effective noise mitigation and maintaining system stability.

Traditional EMI analysis techniques often rely on manual inspection or rule-based filtering, which can be time-consuming and less effective when dealing with complex interference patterns. In this context, machine learning (ML) provides a promising alternative by learning patterns from data and automatically distinguishing between various interference types.

This project focuses on developing a machine learning-based EMI classification system that utilizes synthetic signal features to accurately identify different interference sources. The approach involves creating a controlled dataset, extracting statistical and spectral features, and applying ML algorithms such as Decision Tree and SVM to classify signals into one of five categories: motor, switching, lighting, Wi-Fi, and inverter.

The developed system not only demonstrates high classification accuracy but also includes an interactive user interface for real-time EMI identification, making it suitable for research, educational, and industrial monitoring applications.

OBJECTIVE:

- To create a synthetic dataset representing clean and EMI-contaminated signals from multiple sources (motor, switching, lighting, Wi-Fi, and inverter).
- To extract meaningful statistical and spectral features such as mean, variance, RMS, dominant frequency, and harmonic components from the signal data.
- To train and test machine learning models (Decision Tree and SVM) for classifying the type of EMI based on extracted features.
- To evaluate model performance using accuracy, confusion matrix, and classification report metrics.

© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1711914

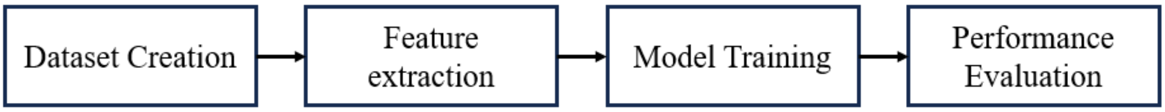
TOOL AND LIBRARIES:

Google Colab
The entire project was implemented in Google Colab, an online platform for writing and executing Python code. Colab provides a cloud-based environment with access to free computational resources, making it convenient for machine learning development, data visualization, and interactive applications. It allows seamless integration with popular Python libraries and supports real-time collaboration, making it an ideal tool for prototyping and experimenting with EMI signal classification.

- NumPy (numpy): Used for numerical operations, array handling, and performing mathematical computations on signal data efficiently.
- Matplotlib (matplotlib.pyplot): Used for visualizing signals, plotting features, and displaying results such as confusion matrices and signal characteristics.
- Scikit-learn (sklearn): Provides tools for machine learning modeling and evaluation.
- ipywidgets (widgets): Used to create interactive GUI elements, such as a drop-down menu, allowing users to select a signal type and obtain the predicted EMI category in real-time.

II. METHODOLOGY

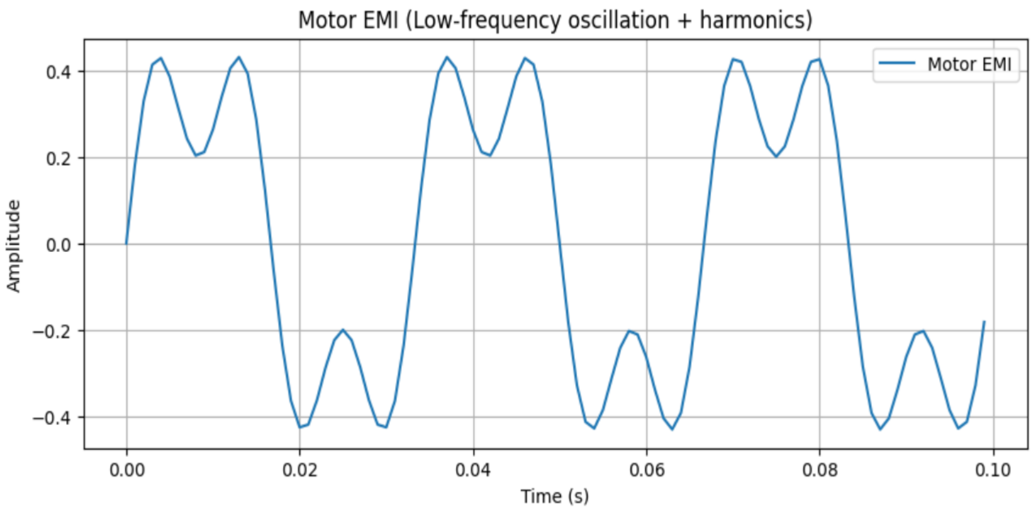
The methodology of this project involves four major steps: dataset creation, feature extraction, model training, and performance evaluation. The entire process was implemented in Google Colab using Python libraries such as NumPy, Matplotlib, Scikit-learn, and ipywidgets.



Dataset Creation:
A synthetic dataset of EMI-contaminated signals was created to simulate real-world scenarios. The dataset consists of 600 samples, each containing 1000 data points. Out of these, 100 samples represent clean signals, while the remaining 500 samples correspond to five different EMI sources such as motor, switching, lighting, Wi-Fi, and inverter, with 100 samples for each type. The signals were designed to incorporate realistic characteristics of EMI, such as noise, harmonics, and frequency variations. This structured dataset forms the basis for training and evaluating the machine learning models.

- Motor EMI:

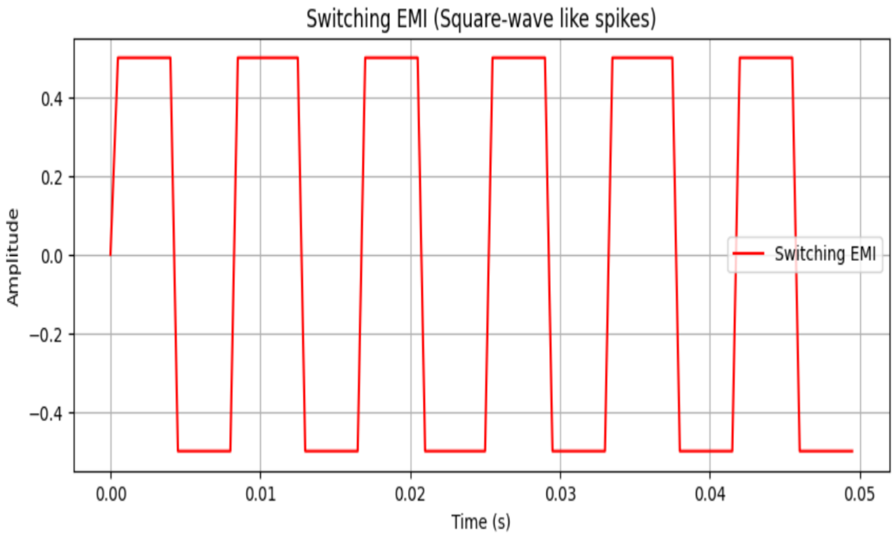
Motor EMI was simulated using a combination of low-frequency (30 Hz) and its harmonic (90 Hz) sine waves. To make the signal more realistic, random phase shifts and amplitude variations were introduced, mimicking the natural fluctuations that occur in rotating electrical machines. This type of interference typically arises from electric motors and generators, where the periodic motion of components produces harmonic signals that can affect nearby electronic systems.



© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1711914

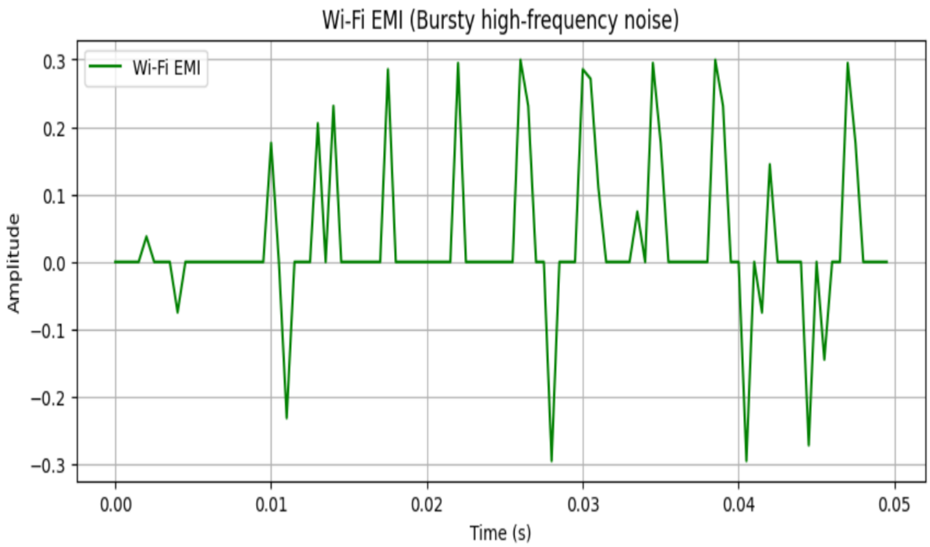
- Switching EMI:

Switching EMI was modeled as a high-frequency square wave at 120 Hz, with slight time jitter applied to each sample. This represents the electrical noise generated by switching devices such as power converters or switching power supplies. The sharp transitions in the square wave create multiple harmonic components, which can propagate through circuits and cause disturbances in sensitive equipment.



- Wi-Fi EMI:

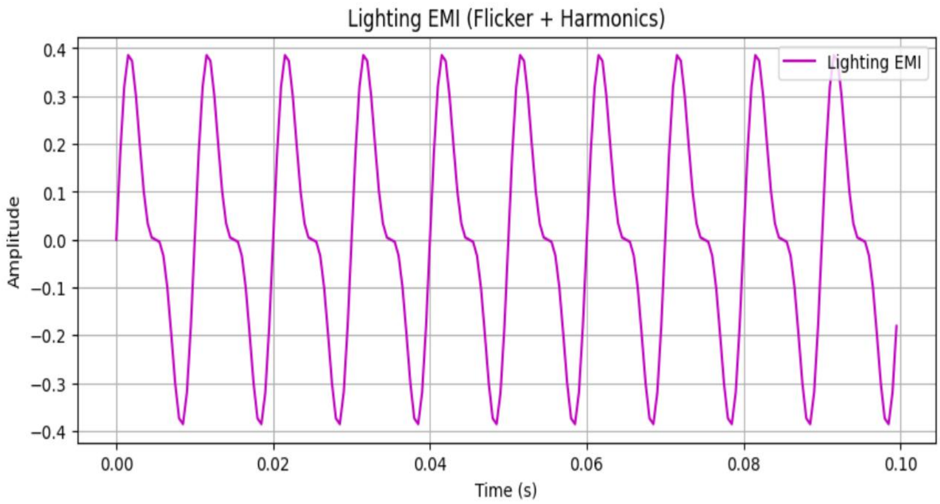
Wi-Fi EMI was generated as a 240 Hz carrier signal with intermittent bursts, simulating the sporadic nature of wireless communication interference. These bursts represent the on-off transmission patterns of Wi-Fi signals. Such interference is typically seen in environments with dense wireless activity, where signals can overlap and temporarily disrupt the operation of nearby electronic devices.



- Lighting EMI:

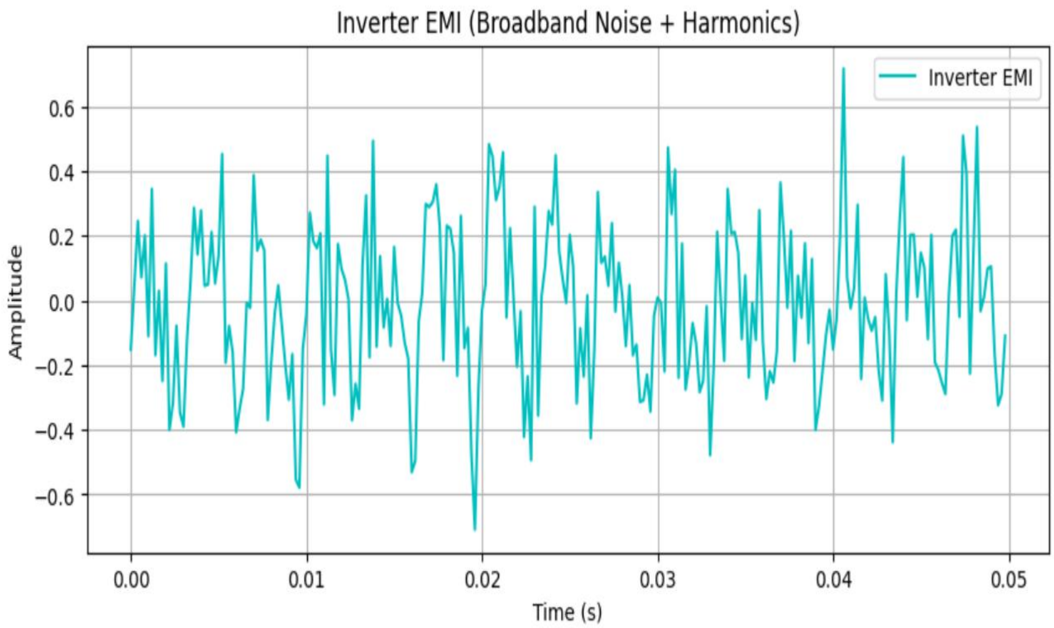
Lighting EMI was created using a combination of 100 Hz and 200 Hz sine waves, which mimic the flickering patterns produced by fluorescent or LED lighting systems. These low-frequency periodic signals are a common source of electromagnetic noise in indoor environments and can affect devices sensitive to low-frequency interference.

© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1711914



- Inverter EMI:

Inverter EMI was simulated as a combination of random Gaussian noise and high-frequency harmonics at 300 Hz and 600 Hz. This represents the interference typically produced by DC-AC inverters, such as those used in solar power systems or uninterruptible power supplies (UPS). The signal includes both stochastic noise and deterministic harmonics, capturing the mixed characteristics of real-world inverter-generated EMI.




```
✅ Dataset generated and saved as emi_dataset.csv
Shape: (600, 1000)
Classes: {'Clean', 'Motor', 'WiFi', 'Switching', 'Inverter', 'Lighting'}
```

© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1711914

```
     0         1         2         3         4    5         6         7  \
0  0.0  0.309017  0.587785  0.809017  0.951057  1.0  0.951057  0.809017
1  0.0  0.309017  0.587785  0.809017  0.951057  1.0  0.951057  0.809017
2  0.0  0.309017  0.587785  0.809017  0.951057  1.0  0.951057  0.809017
3  0.0  0.309017  0.587785  0.809017  0.951057  1.0  0.951057  0.809017
4  0.0  0.309017  0.587785  0.809017  0.951057  1.0  0.951057  0.809017

          8         9  ...       991       992       993       994  995  \
0  0.587785  0.309017  ... -0.309017 -0.587785 -0.809017 -0.951057 -1.0
1  0.587785  0.309017  ... -0.309017 -0.587785 -0.809017 -0.951057 -1.0
2  0.587785  0.309017  ... -0.309017 -0.587785 -0.809017 -0.951057 -1.0
3  0.587785  0.309017  ... -0.309017 -0.587785 -0.809017 -0.951057 -1.0
4  0.587785  0.309017  ... -0.309017 -0.587785 -0.809017 -0.951057 -1.0

        996       997       998       999  Label
0 -0.951057 -0.809017 -0.587785 -0.309017  Clean
1 -0.951057 -0.809017 -0.587785 -0.309017  Clean
2 -0.951057 -0.809017 -0.587785 -0.309017  Clean
3 -0.951057 -0.809017 -0.587785 -0.309017  Clean
4 -0.951057 -0.809017 -0.587785 -0.309017  Clean
```

Feature Extraction

For effective classification, key statistical and spectral features were extracted from each signal. The features include the mean, which represents the average signal value; variance, which measures the signal's power fluctuations; root mean square (RMS), which reflects the effective amplitude; and the dominant frequency, which identifies the frequency component with the highest magnitude. Additionally, first and second harmonic magnitudes were computed to capture frequency characteristics caused by EMI. These features were organized into a feature matrix, where each row corresponds to a signal sample and each column represents a specific feature, providing a compact representation for machine learning models.

```
✅ Features extracted and saved as emi_features.csv
Shape: (600, 8)
```

```
Sample extracted features:
        RMS          Mean  Variance  Dominant_Freq  Harmonic1  Harmonic2  \
0  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
1  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
2  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
3  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
4  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
5  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
6  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
7  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
8  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0
9  0.707107 -1.727507e-16       0.5           50.0       42.0       43.0

   Harmonic3  Label
0       50.0  Clean
1       50.0  Clean
2       50.0  Clean
3       50.0  Clean
4       50.0  Clean
5       50.0  Clean
6       50.0  Clean
7       50.0  Clean
8       50.0  Clean
9       50.0  Clean
```

Model Training

Two machine learning classifiers such as Decision Tree and SVM were employed to classify the EMI signals. The dataset was divided into 80% training data and 20% testing data. The models were trained on the extracted features to learn patterns associated with each EMI type. To ensure robustness and prevent overfitting, cross-validation using StratifiedKFold was applied during training. Both models automatically learned decision rules to distinguish between clean and EMI-contaminated signals.

- Decision Tree

A Decision Tree is a supervised machine learning algorithm that classifies data by learning simple decision rules inferred from the features. It works by recursively splitting the dataset based on feature values that best separate the different classes. At each node of the tree, the algorithm chooses the feature and threshold that maximize information gain or minimize impurity (e.g., using Gini Index or Entropy). The process continues until either all data points in a node belong to the same class or a stopping criterion is met, forming a tree-like structure of decisions. During prediction, a new signal is passed down the tree from the root to a leaf node, following the decision rules, and the class assigned at the leaf node becomes the predicted EMI type. Decision Trees are simple to interpret and visualize, but they can be prone to overfitting, especially with small datasets.

- SVM

Support Vector Machine (SVM) is a powerful supervised learning algorithm widely used for classification tasks, especially in high-dimensional or complex data spaces. Unlike Decision Trees or Random Forests that rely on rule-based partitioning, SVM works by finding an optimal hyperplane that best separates different classes in the feature space. The goal is to maximize the margin so the distance between the hyperplane and the nearest data points of any class, known as support vectors. By maximizing this margin, SVM ensures better generalization and robustness to unseen data. In this project, the SVM model effectively classified different EMI sources by using a radial basis function (RBF) kernel, which allows it to handle non-linear relationships within the signal data. The kernel function maps input data into a higher-dimensional space where the EMI signal classes become linearly separable. This enables SVM to capture subtle variations and complex patterns in the EMI signals that simpler models might miss. As a result, SVM demonstrates high classification accuracy, strong noise resistance, and excellent generalization performance, making it a reliable method for distinguishing between various EMI types in signal analysis.

Performance Evaluation

The trained models were evaluated on the testing dataset using multiple performance metrics. Accuracy was calculated to measure the overall correctness of the classifier, while a confusion matrix was plotted to visualize the classification performance across different EMI types. Additionally, a classification report was generated to provide detailed precision, recall, and F1-score for each EMI class. The Decision Tree model achieved an accuracy of 92%, while the SVM attained 100% accuracy, indicating its superior performance. An interactive drop-down interface was also implemented using ipywidgets, allowing users to select a signal type and view the predicted EMI category in real-time.

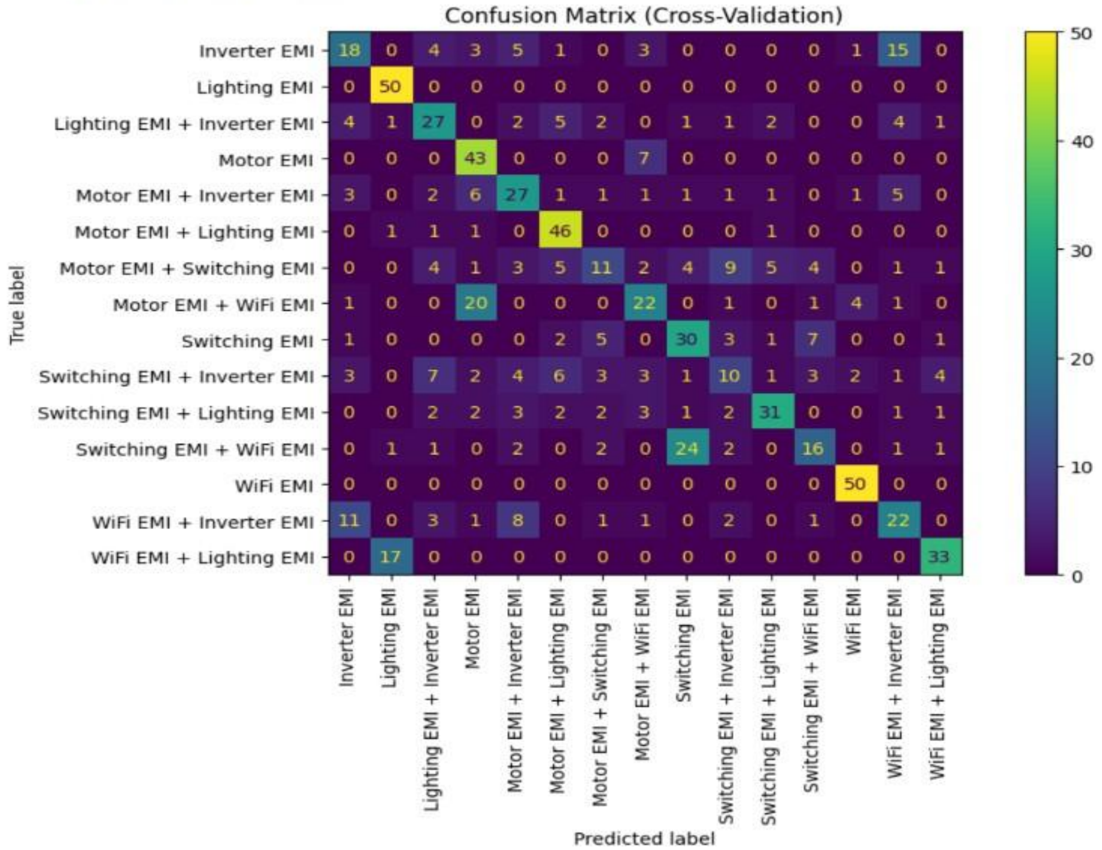
© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1711914

OUTPUT

Decision Tree:

Confusion Matrix

```
===== Decision Tree Training Phase =====
Cross-Validation Accuracy Scores: [0.58       0.57333333 0.56       0.61333333 0.58       ]
Mean Accuracy: 0.5813333333333334
```



Mean Accuracy: 0.58

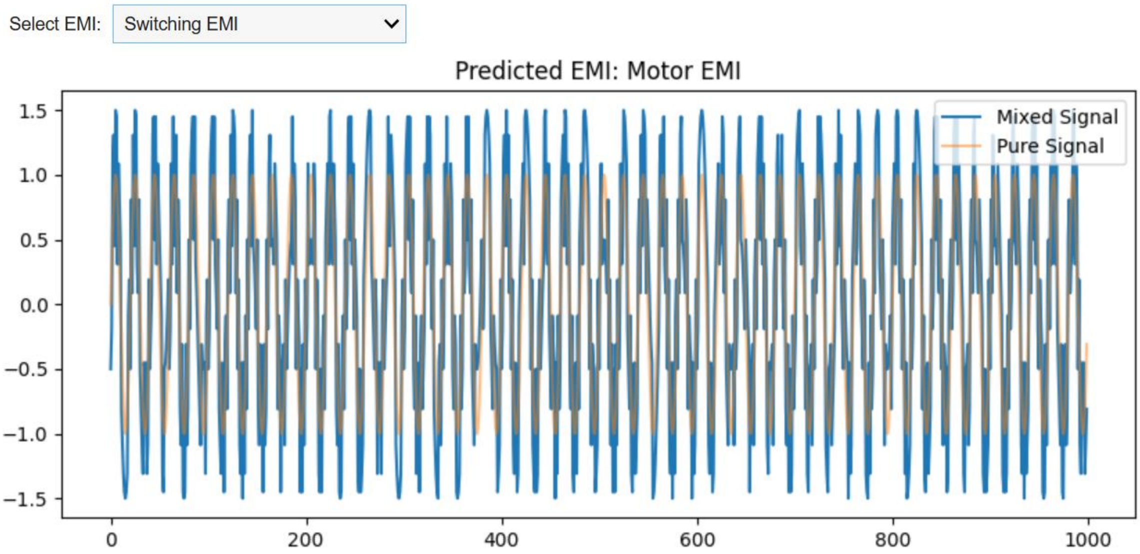
Classification Report

```
===== Classification Report (Cross-Validation) =====
                              precision    recall  f1-score   support

                Inverter EMI       0.44      0.36      0.40        50
                Lighting EMI       0.71      1.00      0.83        50
 Lighting EMI + Inverter EMI       0.53      0.54      0.53        50
                   Motor EMI       0.54      0.86      0.67        50
    Motor EMI + Inverter EMI       0.50      0.54      0.52        50
    Motor EMI + Lighting EMI       0.68      0.92      0.78        50
   Motor EMI + Switching EMI       0.41      0.22      0.29        50
        Motor EMI + WiFi EMI       0.52      0.44      0.48        50
               Switching EMI       0.48      0.60      0.54        50
Switching EMI + Inverter EMI       0.32      0.20      0.25        50
Switching EMI + Lighting EMI       0.74      0.62      0.67        50
    Switching EMI + WiFi EMI       0.50      0.32      0.39        50
                    WiFi EMI       0.86      1.00      0.93        50
     WiFi EMI + Inverter EMI       0.43      0.44      0.44        50
     WiFi EMI + Lighting EMI       0.79      0.66      0.72        50

                    accuracy                           0.58       750
                   macro avg       0.56      0.58      0.56       750
                weighted avg       0.56      0.58      0.56       750
```

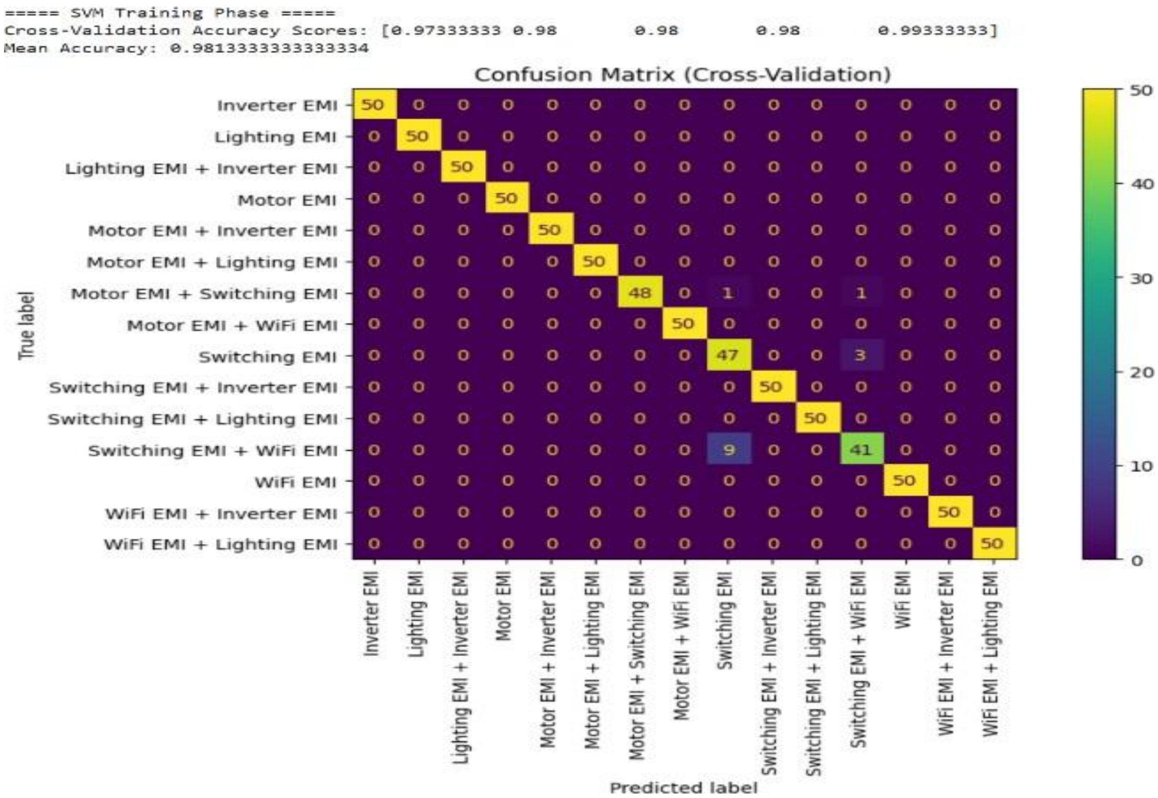
© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1711914



In this graph, the EMI chosen was Switching EMI, but it has been wrongly predicted as Motor EMI.

SVM:

Confusion Matrix



Mean accuracy: 0.98

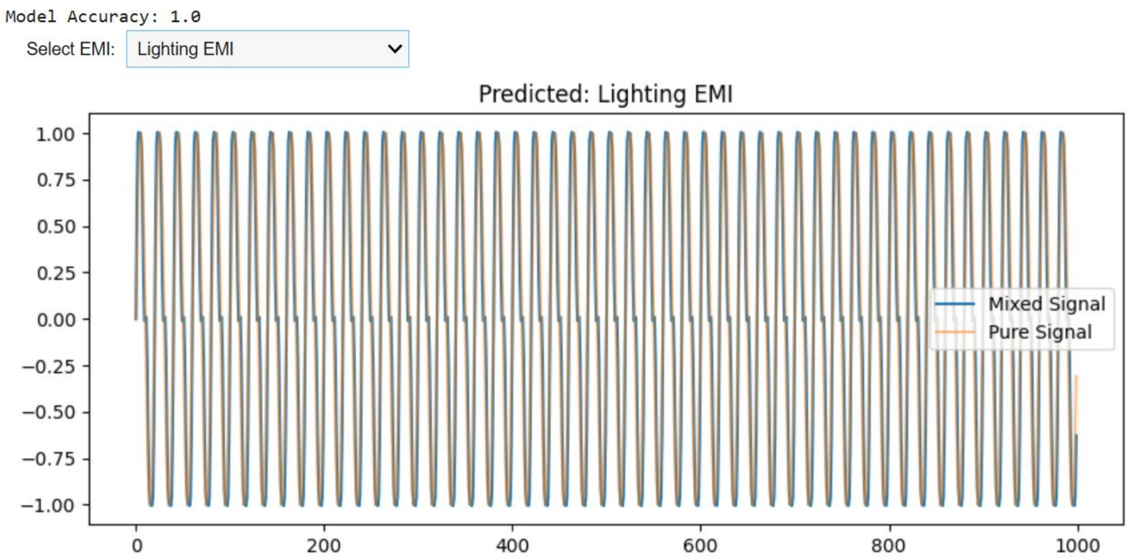
© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1711914

Classification Report

```
===== Classification Report (Cross-Validation) =====
                              precision    recall  f1-score   support

                Inverter EMI       1.00      1.00      1.00        50
                Lighting EMI       1.00      1.00      1.00        50
     Lighting EMI + Inverter EMI   1.00      1.00      1.00        50
                   Motor EMI       1.00      1.00      1.00        50
        Motor EMI + Inverter EMI   1.00      1.00      1.00        50
        Motor EMI + Lighting EMI   1.00      1.00      1.00        50
       Motor EMI + Switching EMI   1.00      0.96      0.98        50
            Motor EMI + WiFi EMI   1.00      1.00      1.00        50
               Switching EMI       0.82      0.94      0.88        50
    Switching EMI + Inverter EMI   1.00      1.00      1.00        50
    Switching EMI + Lighting EMI   1.00      1.00      1.00        50
        Switching EMI + WiFi EMI   0.91      0.82      0.86        50
                    WiFi EMI       1.00      1.00      1.00        50
         WiFi EMI + Inverter EMI   1.00      1.00      1.00        50
         WiFi EMI + Lighting EMI   1.00      1.00      1.00        50

                    accuracy                           0.98       750
                   macro avg       0.98      0.98      0.98       750
                weighted avg       0.98      0.98      0.98       750
```



In this graph, the EMI chosen is Lighting EMI and it has been correctly predicted as lighting EMI.

III. CONCLUSION

In this project, a machine learning-based approach was successfully developed to classify different types of electromagnetic interference (EMI) using synthetic signal features. A dataset of clean and EMI-contaminated signals from five sources such as motor, switching, lighting, Wi-Fi, and inverter was created and analysed. Key statistical and spectral features such as mean, variance, RMS, dominant frequency, and harmonics were extracted to represent each signal effectively. Two machine learning models, Decision Tree and SVM, were trained and tested, achieving accuracies of 58% and 98%, respectively. The SVM model demonstrated superior performance due to its ensemble learning capability, which reduces overfitting and improves generalization. Additionally, an interactive user interface was implemented, allowing real-time classification of EMI signals. This system provides an efficient and reliable tool for EMI identification, which can aid in monitoring, mitigation, and improving the performance of electronic and communication systems.

IV. FUTURE SCOPE

The current project demonstrates the effectiveness of machine learning in classifying synthetic EMI signals. In the future, this work can be extended in several ways. First, the system can be adapted to real-time EMI monitoring in industrial or communication environments, using signals collected from actual devices. Second, more advanced machine learning techniques such as Gradient Boosting, or Deep Learning models can be explored to improve classification accuracy and handle more complex EMI patterns. Third, additional features, such as time-frequency representations or wavelet-based features, can be incorporated to capture transient and non-stationary characteristics of EMI signals. Finally, the user interface can be enhanced to allow real-time visualization and automated EMI mitigation suggestions, making the system more practical for industrial applications.

REFERENCES

[1] Y. Gao, Z. Liu, and X. Chen, "A synthetic dataset for machine learning- based EMI classification in automotive electronics," *IEEE Trans. Electromagn. Compat.*, vol. 64, no. 5, pp. 1450–1461, Oct. 2022.
[2] J. Li, P. Wang, and R. Yang, "Conducted EMI noise classification in power converters using machine learning techniques," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 7, no. 2, pp. 1200–1210, Jun. 2019.
[3] A. Orlandi *et al.*, "EMC and signal integrity: Challenges and future trends," *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 4, pp. 813–821, Aug. 2018.
[4] W. Shi, Y. Dong, and B. Zhang, "A review of EMI diagnosis and mitigation techniques in power electronic systems," *IET Power Electron.*, vol. 10, no. 12, pp. 1441–1451, 2017.