International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 12, No. 3, March, 2025, pp. 900-908
Journal homepage: http://iieta.org/journals/mmep

Ensemble Learning Techniques for Improved Electromagnetic Interference Prediction in LED Driver Circuits

Mohammad Yanuar Hariyawan[1*], Mochamad Nizar Palefi Ma'ady[2], Helmy Widyantara[3], Devie Rosa Anamisa[4], Haykal Azrel Putra Sugijanto[1], Nathanael Tjahyadi[1]

[1] Department of Computer Engineering, Telkom University, Surabaya 60231, Indonesia
[2] Department of Information System, Telkom University, Surabaya 60231, Indonesia
[3] Department of Information Technology, Telkom University, Surabaya 60231, Indonesia
[4] Department of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan 69162, Indonesia

Corresponding Author Email: myanuar@telkomuniversity.ac.id

https://doi.org/10.18280/mmep.120316

Received: 24 December 2024
Revised: 14 February 2025
Accepted: 20 February 2025
Available online: 31 March 2025

Keywords:
EMI, LED driver, ensemble learning, spread spectrum, machine learning prediction

ABSTRACT

This study evaluates ensemble learning methods, including Random Forest (RF), XGBoost, and LightGBM, utilizing bagging and stacking techniques to predict electromagnetic interference (EMI) levels. The dataset includes various signal types, such as LorenzModif, Gaussian, Noise, Ampalt, Triangle, Square, and Sine, which are applied to modify the LED driver's switching frequency as part of the spread-spectrum techniques aimed at reducing EMI. Parameters such as frequency and amplitude were also considered to ensure a comprehensive analysis of EMI behaviour. During preprocessing, the data were first processed using one-hot encoding and feature scaling before being analyzed using machine learning models. The models were evaluated based on mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R^2). The results showed that Stacking achieved the highest R^2 value of 0.7347, slightly outperforming RF (0.7327), indicating better predictive accuracy. Statistical analysis revealed significant differences between Stacking and models such as XGBoost and LightGBM but not between stacking and RFs. Data visualization using residual plots and scatter diagrams further reinforces the superiority of stacking and RFs in EMI prediction. These findings suggest that ensemble learning can effectively enhance EMI prediction in LED drivers, ultimately supporting the design and manufacturing processes.

1. INTRODUCTION

EMI presents a significant challenge in LED driver operation owing to the rapid voltage and current fluctuations during switching. If not correctly managed, EMI can disrupt LED systems and other sensitive electronics, sometimes leading to failures in critical industries like forestry [1, 2]. To ensure stable operation, LED drivers must comply with electromagnetic compatibility (EMC) standards and implement effective mitigation strategies [3]. The severity of EMI depends on various factors, including interference power, frequency, induced power, and input resistance [4]. Addressing these challenges requires robust control strategies, microgrids (MGs), and energy storage systems (ESSs) to maintain stability [5].

Additionally, advanced monitoring techniques help detect vulnerabilities in power distribution networks (PDNs) before EMI-related failures occur, ensuring long-term reliability [6]. As smart grids become more complex, with thousands of interconnected devices, the risk of EMI disruptions increases. This growing complexity makes predictive models more essential than ever for ensuring stability and reliability [7].

Despite advancements in EMI filters and modulation techniques [2, 8-10], predicting EMI in LED drivers remains a significant challenge. Traditional approaches, such as EMI filters [11, 12] and chaotic signal modulation [13, 14], often struggle with the complexities of modern LED circuits. These methods require extensive parameter tuning and frequently fail to adapt to real-world conditions. While some signal modulation techniques can mitigate specific EMI frequencies, they tend to be less effective against broadband interference, resulting in inconsistent performance. More advanced computational techniques provide better accuracy but face difficulties handling high-dimensional nonlinear data and complex circuit interactions [15]. These limitations highlight the need for a more dynamic and adaptable approach to EMI prediction in LED drivers.

Machine learning, particularly ensemble learning techniques, offers a promising alternative for improving EMI prediction. Unlike single-model approaches, ensemble methods combine multiple models to enhance accuracy, generalization, and robustness. Techniques such as Bagging, Boosting, and Stacking help reduce overfitting, minimize variance, and capture complex EMI patterns [16]. RF enhances robustness by averaging multiple decision trees, reducing prediction noise. XGBoost efficiently manages feature interactions and minimizes bias using gradient boosting, while LightGBM optimizes speed and computational efficiency, making it ideal for large-scale datasets. By leveraging these techniques, machine learning can provide a deeper and more adaptable understanding of EMI behaviour in LED driver circuits.

This study evaluates the effectiveness of RF, XGBoost, and LightGBM in predicting EMI levels in LED driver circuits. The models were trained using data from various operating conditions, considering key system parameters. Performance was assessed using metrics such as prediction accuracy and stability, ensuring the models are reliable and practical for real-world applications. By applying these advanced machine learning techniques, this research aims to develop a more accurate, adaptable, and efficient EMI prediction framework, assisting engineers in designing better LED drivers while ensuring compliance with EMI regulations.

2. METHODOLOGY

2.1 Data collection and preprocessing

This study's data came from the conducted emission (CE) test generated by the LM3409 LED buck topology evaluation board. This study aims to measure the level of EMI generated by LED drivers under various operating conditions, precisely when the switching frequency of the driver is modified through signal injection with different characteristics. The experimental setup can be seen in Figure 1 and involves some of the following equipment:

- LED LM3409 evaluation board: This evaluation board was used as the testbed for measuring EMI under various signal injection conditions.
- Spectrum analyzer (SA): This instrument measured the EMI levels emitted by the driver during the test.
- Arbitrary function generator (AFG): This instrument injects periodic and non-periodic signals into the system.
- Line impedance stabilization network (LISN): Ensures LED drivers' precise and stable connection to the test equipment.
- PC: A computer is used to generate non-periodic signals.

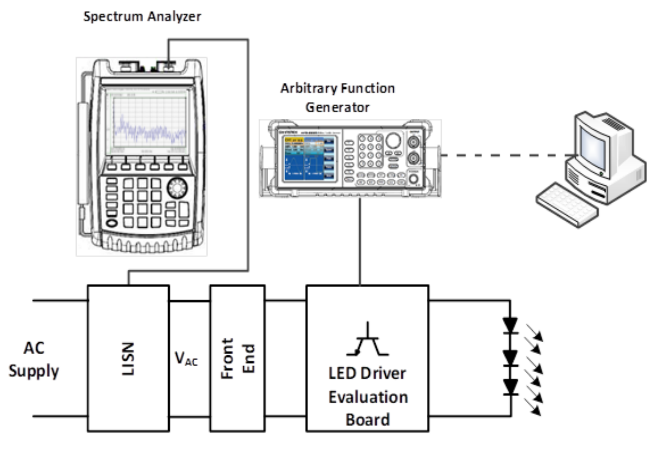


Figure 1. Conducted emission measurement setup

The data set was obtained from an experiment when the LED driver modified its switching signal by injecting various signals. The injected signals include periodic signals (Triangle, Square, and Sine) and non-periodic signals (Gaussian, Noise, LorenzModif, and Ampalt), which simulate highly unpredictable EMI to evaluate EMI behaviour comprehensively.

In particular, non-periodic signals are important in applying spread spectrum techniques. By injecting these signals into the LED driver, the signal energy is dispersed over a broader range of frequencies, not concentrated at a specific frequency. This dispersion helps reduce peak interference and EMI more effectively, making the spread spectrum a valuable strategy for improving EMC. By combining periodic-based non-periodic signals and spread spectrum, this study aims to predict the EMI level when switching frequency modifications are performed on LED driver circuits in various switching signal characteristics, periodic or non-periodic.

2.1.1 Features and data overview

Each sheet in the dataset contains one of the signal types with the following key features:

- Frequency (Hz): The frequency of the injected signal plays a crucial role in the intensity of the EMI generated.
- Signal characteristics (dBµV): The amplitude of the signal, measured in microvolt decibels (dBµV), reflects how strong the injected signal is.
- EMI (dBµV): The level of EMI generated, which is the primary target variable in this study, indicates the level of interference measured in microvolt decibels.

This dataset consists of 1,203 data points for each signal type. With eight different signal categories, the total data collected reached 9,624 points. This comprehensive data reflects the behaviour of EMI at switching frequencies that vary in the range of 150 kHz to 30 MHz, allowing for an in-depth analysis of the impact of signal characteristics on EMI levels.

The data is organized in separate sheets based on signal type to make analysis easier. This approach allows for a clearer understanding of how variations in signal characteristics affect EMI in LED systems. Each sheet records the relationship between signal frequency, signal strength, and EMI generated at various points during testing, providing more detailed insight into EMI patterns and trends in the system.

2.1.2 Data collection process

This data set was obtained from a series of controlled experiments in which the switching frequency of the LED driver was modified using different types of signals. The signals used include periodic signals, such as sine, square, and triangle, and non-periodic signals, including Gauss, Noise, LorenzModif, and Ampalt. This experiment is carried out through the following stages:

1) Signal generation: The periodic signal generated directly by the AFG is first injected into the driver to observe the basic EMI level. Later, non-periodic signals were introduced, such as LorenzModif signals, which are simulated before being fed into the system via the AFG to represent real-world interference better.
2) Measurement: The CE spectrum is measured to determine the EMI generated by the LED driver by both types of signals.

This experimental setup was created so that the data collected covered a wide range of possible EMI scenarios, and the results could be more relevant and valuable for real-world systems.

2.1.3 Preprocessing and data integrity

Several preprocessing steps are performed to prepare the data before analysis, including:

- Normalization: All features, such as frequency, signal characteristics, and EMI levels, are normalized so that the model can learn somewhat without being affected by differences in scale or data size.
- Outlier detection: Identify and remove outliers in the data so they do not affect the accuracy of the model's predictions.
- Feature engineering: Additional features like the signal-to-noise ratio are calculated to capture the more complex relationship between signal characteristics and EMI levels.

By implementing these preprocessing steps, the collected data becomes cleaner and more structured, making it more ready for use in machine learning models. This process ensures that the model can predict the EMI in an LED driver system more accurately and reliably, even under various operational conditions.

2.2 Model training

In this study, the initial model was trained using the default settings for each algorithm, including RF, K-Nearest Neighbor (K-NN), and Support Vector Regressor (SVR). This approach provided an initial overview of the underlying performance of each model.

To improve prediction accuracy, we perform hyperparameter tuning to find the best combination that maximizes model performance. This process involves cross-validation, grid search, and manual adjustment, ensuring the model optimally predicts EMI in the LED driver system. This process involves cross-validation, grid search, and manual adjustment, ensuring the model optimizes predicting EMI on the LED driver system.

2.2.1 Hyperparameter tuning process

Each model has hyperparameters adjusted based on its role in predicting EMI.

A). RF

The RF model is adjusted to various hyperparameters to improve the performance. One of the key parameters, n_estimators (the number of trees in the model), was initially set at 100 and then tested in the range of 50-200 to find the best balance between accuracy and computational efficiency. This range was chosen because the model was not too simple and was not heavy to run.

In addition, the max_depth (maximum depth of the tree) is set between 5 and 30 to control the tree's growth and prevent overfitting. With this approach, the tree can grow deep enough to capture relevant patterns without memorizing the training data.

Several other key parameters were also adjusted to optimize the model. min_samples_split, which determines the minimum number of samples needed to divide a node, was tested in the range of 2-10 to find the most appropriate level of granularity. Meanwhile, min_samples_leaf, which controls the minimum number of samples in each leaf node, varies between one to five.

This adjustment ensures that each node has sufficient samples to generate more generalized and accurate predictions, reducing the risk of overfitting without sacrificing the model performance.

B). K-NN

Some hyperparameters were adjusted to maximize model performance. The first is n_neighbors, which determines the number of neighbours used in the prediction process. The values of this parameter were tested in the range of 3 to 15 to determine the optimal balance between accuracy and the model's ability to recognize patterns in general. By adjusting these parameters, we can further understand how the number of neighbours selected affects the model's ability to classify the data more accurately.

In addition to the n_neighbors, several other important parameters were tested to improve the model's performance. One is weights, tested with two schemes: 'uniform', where each neighbour has the same influence, and 'distance', where closer neighbours have more weight than farther ones. This test aims to determine the most effective weighting method for handling data.

In addition, the distance metric used to measure the proximity between data points was also adjusted by comparing 'minkowski' and 'Euclidean' to evaluate how this choice affected the accuracy and performance of the model. To determine the most optimal combination of parameters, we tested the K-NN model using a grid search with cross-validation to determine the best configuration.

C). SVR

The performance of the SVR model was improved by adjusting several key hyperparameters. One of the parameters tested was C, which controls the balance between flexibility in customizing the training data and the model's simplicity. The C value was tested in the range of 0.1 to 10 to determine the optimal point that prevents underfitting and overfitting.

In addition, the epsilon parameter, which determines the error tolerance, was tested in the range of 0.01 to 0.5. This test aims to understand the extent to which the model can ignore minor variations in the data without overreacting to noise.

The selection of kernel functions is also explored to find the best approach for capturing complex patterns in the data. Some kernels tested include 'rbf', 'linear', and 'poly', which are analyzed to evaluate their impact on model performance.

A manual grid search was performed using cross-validation to obtain the optimal combination of parameters. This approach allows the systematic exploration of various configurations so that the model can achieve the best balance between accuracy and generalization when making predictions.

2.2.2 Ensemble learning tuning

In addition to individual models, ensemble learning techniques, particularly Bagging and Stacking, have also been applied and adapted. This method combines several basic models to improve overall prediction performance and accuracy.

A). Bagging (Bootstrap Aggregating)

The ensemble learning approach in this study applies Bootstrap Aggregating (Bagging) to improve the stability and accuracy of predictions by combining several models. This technique trains various models using different subsets of data, thereby improving generalization capabilities while reducing the risk of overfitting [17].

Bagging reduces the variance in the model, in which some of the same base models are trained on different parts of the training data. The final result is then obtained by averaging the output of all models.

This study applied bagging to the RF and K-NN models to evaluate how these techniques can improve prediction performance. The mathematical formulation of the Bagging method is described by Eq. (1).

$$\hat{y}_{bagging} = \frac{1}{B}\sum_{b=1}^{B}\hat{y}_b \tag{1}$$

where,

$\hat{y}$ is the output of the b-th base model's prediction.

B is the number of base models (or estimators).

The bagging regressor was implemented with a SVR as the base model, and 50 base models were used.

B). Stacking

Stacking, also known as stacked generalizations, is an ensemble learning technique that combines predictions from several basic models with the help of meta-learner models. This approach aims to improve the accuracy of predictions and strengthen the model's ability to generalize new data [18, 19]. This approach usually consists of two layers. In the first layer, some basic models produce the initial prediction, whereas in the second layer, meta-models learn from the prediction to produce a more accurate final result [20]. The stacking process typically involves cross-validation of the training data, where some basic models are trained, and initial predictions are generated. The predictions from these models are then used as an additional feature to train meta-learners, who combine information to improve the accuracy of the final prediction [18]. The essential models used in this study were RF, K-NN, and SVR. Linear regression was chosen as the meta-learner.

Predictions from each of the basic models ($\hat{y}_1$, $\hat{y}_2$,…… $\hat{y}_k$) are combined and used as input for the meta-learner, who then learns it to produce the final prediction. This process is formulated using Eq. (2).

$$\hat{y}_{stacking} = w_1\hat{y}_1 + w_2\hat{y}_2 + \cdots + w_k\hat{y}_k \tag{2}$$

where,

$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_k$ are the predictions from the base model.

$w_1, w_2, \ldots w_k$ does the meta-learner learn the weights.

The linear regression meta-learner determines the ideal blend of predictions from the base models to reduce the discrepancy between the actual and forecasted EMI values.

2.2.3 Model selection criteria

Once the hyperparameters were tuned, we evaluated each model based on several criteria to ensure it could be generalized effectively to new, unseen data. These criteria included the following:

(1) Cross-validation performance: Cross-validation was applied to check how well the model generalized across different subsets of the training data, ensuring that it did not overfit any specific portion of the dataset.
(2) Evaluation metrics: The performance of each model was assessed using:
    - MAE: To quantify the average magnitude of the prediction errors.
    - MSE: Measure the average squared differences between predicted and actual values.
    - R^2: Assess the proportion of variance in the EMI levels explained by the model.

The best hyperparameter set for each model was chosen based on the combination that achieved the lowest MSE, highest R^2, and shortest training time, thereby ensuring that the model provided the most accurate predictions without unnecessary complexity.

2.2.4 Model comparison and selection

The performance of each model was compared based on evaluation metrics (MAE, MSE, and R^2). The model with the highest R^2, lowest MAE, and lowest MSE was considered the best-performing model. The stability and generalization abilities were assessed by testing the models on the testing dataset. A comparative evaluation was conducted on the following machine-learning models: RF, K-NN, SVR, bagging (SVR), and stacking.

2.2.5 Model visualization

To further evaluate model performance, the following visualizations were employed:

- Actual vs. predicted plots: Scatter plots of actual versus predicted EMI values for each model to assess the model's fit visually.
- Learning curves: Plots showing the training and testing errors over time to understand how the model learns and generalizes.
- Residual plots: Plots showing the spread of residuals (errors between actual and forecasted values) used to assess the bias and variance in the model predictions.

These visualizations provide critical insights into the strengths and weaknesses of each model, enabling a more thorough comparison and discussion of its performance.

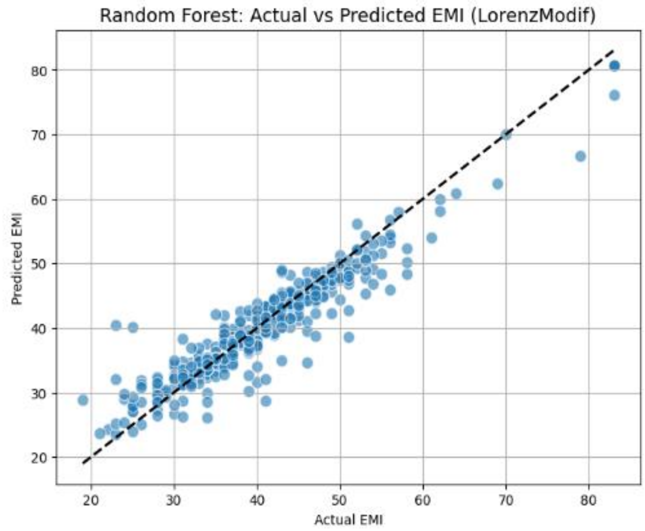
3. RESULT

The effectiveness of various ensemble learning models in predicting EMI levels was assessed using three key performance metrics: MAE, MSE, and R^2. These metrics comprehensively evaluate each model's accuracy, reliability, and generalization ability in handling EMI data. The results, summarized in Table 1, reveal that stacking and RF (bagging) outperform the other models, exhibiting the lowest MAE and MSE values and the highest R^2 scores (0.732741 and 0.732725, respectively). These findings indicate that both models are well-suited for EMI prediction, capturing complex relationships within the data with minimal bias and improved generalizability.

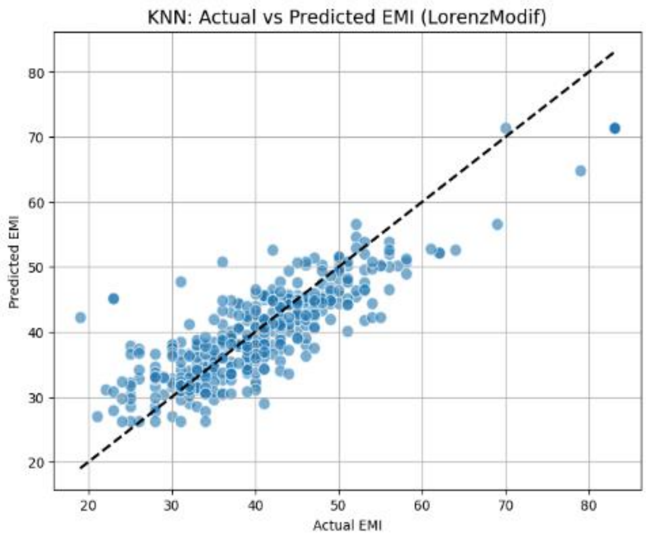
Conversely, K-NN and SVR demonstrated significantly weaker performance, with R^2 values of 0.273503 and 0.248959, respectively. The high prediction errors suggest that these models struggle to capture EMI's nonlinear and high-dimensional nature, particularly in complex signal types such as LorenzModif. Furthermore, the bagging (SVR) model does not substantially improve the standalone SVR, as its R^2 score (0.249134) remains nearly identical, highlighting its limited effectiveness in EMI prediction. Although XGBoost and LightGBM have been recognized for their robustness in other domains, such as financial forecasting, their application to EMI prediction may require extensive hyperparameter tuning and feature engineering, as indicated by previously reported R^2 values of 0.626928 and 0.636082.

Table 1. Model evaluation results (MAE, MSE, R^2)

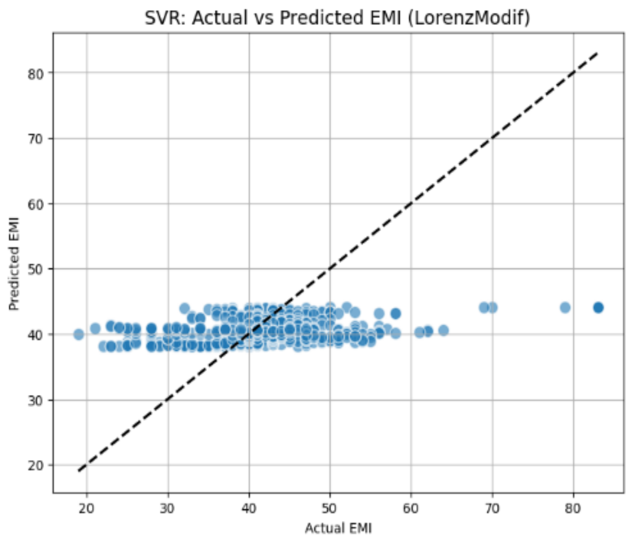
Model	MAE	MSE	R^2
RF (Bagging)	3.981386	27.701451	0.732725
KNN	6.728349	75.297072	0.273503
SVR	6.788039	77.840978	0.248959
Bagging (SVR)	6.789676	77.822789	0.249134
Stacking	3.970231	27.699791	0.732741



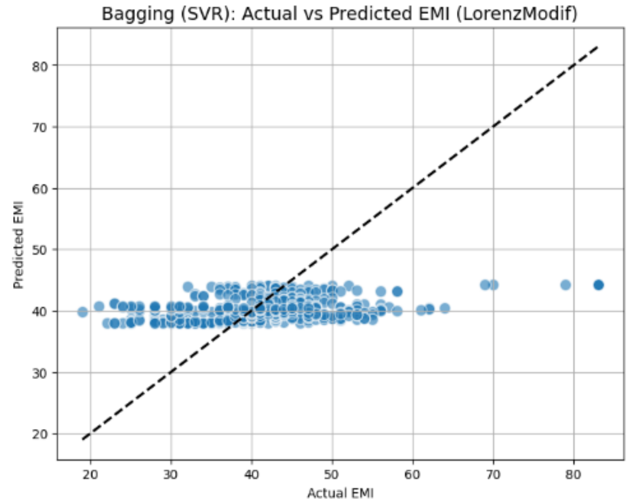
(a)



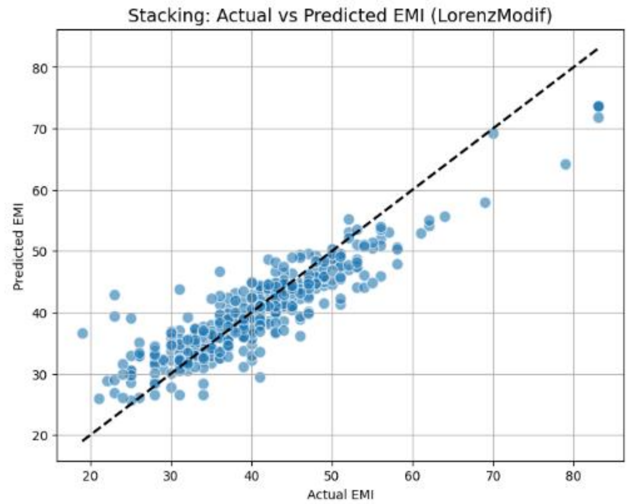
(b)



(c)



(d)



(e)

Figure 2. Scatter plots for actual vs. predicted EMI

Figure 2 illustrates scatter plots comparing actual vs. predicted EMI values across different models to validate these findings further. The RF (bagging) and stacking models exhibit tighter prediction groupings around the equality line (y = x), reinforcing their predictive superiority. In contrast, K-NN demonstrates a broader dispersion, suggesting higher prediction variance and lower accuracy. Moreover, Figure 3 presents EMI value vs. frequency line plots, providing insight into how each model captures frequency-dependent EMI variations. Here, RF (bagging) and stacking align closely with actual EMI trends, whereas bagging (K-NN) exhibits more pronounced deviations across specific frequency regions, further indicating prediction inconsistencies.

While these visual analyses highlight the strengths and weaknesses of each model, a residual analysis is necessary to confirm error distribution and potential biases. Residual plots can provide deeper insights into model limitations, particularly in handling nonlinear EMI data. The results suggest that stacking and RF (bagging) are the most reliable models for EMI estimation, while K-NN and SVR require significant modifications or alternative approaches to achieve competitive performance. Given the critical role of EMI prediction in ensuring EMC and reducing interference risks, future research should focus on enhancing model interpretability, refining feature extraction techniques, and integrating hybrid learning frameworks to optimize prediction accuracy further.

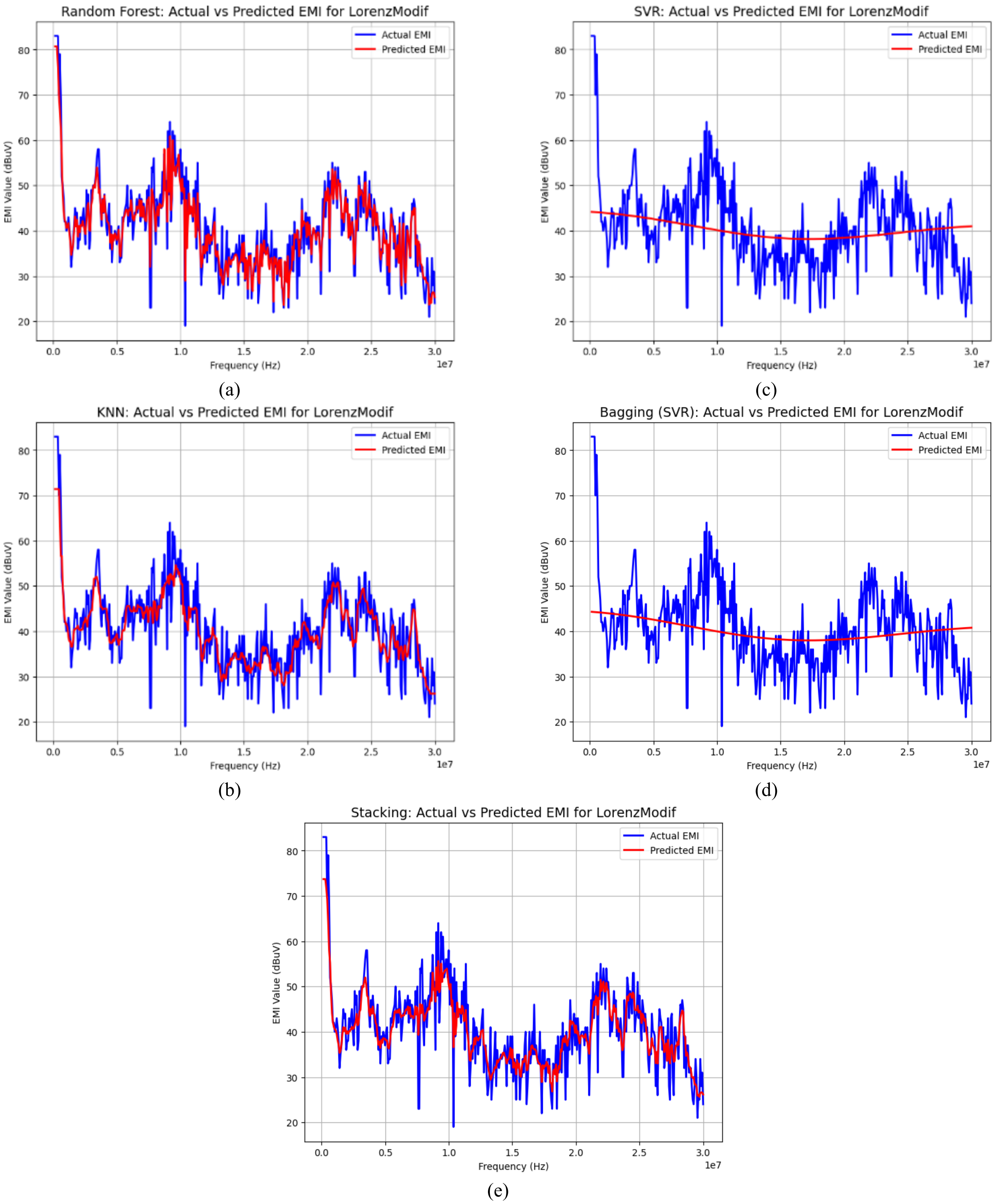


Figure 3. EMI value vs. frequency line plots

4. DISCUSSION

4.1 Interpretation of results

In this study, RF and Stacking showed outstanding performance, with high R^2 values (0.732725 for RF and 0.732741 for stacking), proving their effectiveness in predicting EMI levels. Both models can handle complex, high-dimensional, and noisy EMI data, with RF particularly excelling in modelling nonlinear relationships, so it works well on a wide range of signal types, including LorenzModif and other spread spectrum signals.

However, the advantages of RF and Stacking are not always applicable to all machine-learning applications. Although these models have shown strong performance in various domains [21-24], other studies have shown that alternative methods, such as Rotation Forest–RF [25] and XGBoost [26], could be more effective depending on the data type, signal characteristics, and problem-specific configuration. Therefore, while RF and Stacking hold promise for EMI prediction in LED drivers, their ability to generalize EMI conditions in the more complex real world still needs further validation. Further studies are needed to evaluate the model's effectiveness of the model in various operational environments and system configurations to ensure its resilience to diverse real-world scenarios.

4.2 Ensemble learning: bagging and stacking

Bagging and Stacking use an ensemble learning approach but with different success rates. Bagging (SVR) is expected to reduce variance and improve model performance, but the results do not significantly improve over the standard SVR. In contrast, Stacking, which incorporates several basic models such as RF, KNN, and SVR, provides a more significant improvement, achieving an R^2 value of 0.732741. These results show that Stacking is more effective in recognizing various patterns in data, producing more accurate and reliable predictions than individual models.

4.3 Impact of spread spectrum techniques in EMI mitigation

Spread spectrum techniques help improve the accuracy of EMI prediction by spreading the signal energy over a broader range of frequencies, making it easier to analyze using models such as RF and Stacking. Both models demonstrate high R^2 values and a reliable ability to track EMI on various signals, including spread spectrum signals [22, 27, 28]. This performance improvement allows the model to recognize EMI patterns more effectively and capture details that other methods may miss. However, validation under real-world conditions is still required to ensure the model can handle a wide range of environmental variables and operational conditions accurately.

Although it often outperforms individual models in various prediction tasks, Stacking still has room for further optimization in EMI prediction, as indicated by recent research [29-31]. Innovations such as gradient boosting RF (GBRF) [27] and double RF (DRF) [32] have opened up opportunities to improve model performance, particularly in processing spread spectrum signals more efficiently.

4.4 Implications for the LED driver industry

Integrating an ensemble learning model into a real-time EMI prediction system for LED drivers presents several challenges. Models such as RF (Bagging) and Stacking require significant computing power, slowing the inference process and making them less suitable for resource-constrained environments. To overcome this challenge, researchers should explore lighter alternatives, such as RF optimized with more efficient feature selection or deep learning architectures designed for low-latency applications, to ensure real-time performance.

In addition, real-time EMI prediction must be able to handle high-frequency data streams (150 kHz to 30 MHz), which presents challenges in acquisition, preprocessing, and inference that exceed the capabilities of standard microcontrollers. Real-world EMI conditions are also dynamic and affected by temperature fluctuations, power changes, and circuit ageing, so adaptive learning strategies, such as online learning and gradual updates, are required to maintain model accuracy in the long term.

In addition to technical challenges, compliance with EMI regulations (e.g., CISPR 15, FCC Part 15, IEC 61000) is also an important factor in the practical implementation of the model. Cloud-based solutions offer scalability but can increase operational costs, making embedded AI models a more efficient and cost-effective alternative to real-time EMI mitigation.

4.5 Limitations

While ensemble learning models have great potential for predicting EMI in LED driver circuits, some limitations need to be considered to be more generalized and ready to be applied in the real world. One of the main challenges is data representation, as the study uses only limited types of EMI signals, such asLorenzModif, Gaussian, Noise, Triangle, and Square signals. Real-world EMI patterns can be significantly more complex and influenced by environmental noise, circuit configuration changes, and load fluctuations, which are not fully covered in the current dataset. Future studies should aim to improve the durability and accuracy of the model by expanding the scope of EMI signals, testing models on various LED driver topologies, and leveraging data from live testing data under operational conditions. With this step, the model can become more reliable, accurate, and ready to be implemented for EMI monitoring and mitigation in next-generation LED drivers.

Each model has limitations that can affect the accuracy of the predictions. For example, RF (bagging) is difficult to extrapolate, making it less effective in recognizing EMI patterns with never-encountered frequencies. The Stacking Ensemble, while powerful, relies heavily on the quality of its base model. If a suboptimal model, such as a K-NN in high-dimensional space, is included, the overall performance can deteriorate. In addition, both models are susceptible to overfitting; they may work well on current datasets but have not been tested under various EMI conditions. Future research can enhance the model's reliability by applying more extensive regularization, data augmentation, and cross-validation, making it more stable, adaptive, and capable of handling diverse EMI scenarios.

Upgrade-based models such as XGBoost and LightGBM have not been optimized in this context. With lower R^2 values 0.63 and 0.64, and higher MAE 4.51 and 4.69. These results indicate that hyperparameter tuning has not been fully optimized for EMI data. Therefore, more sophisticated hyperparameter optimization techniques, such as Bayesian optimization or genetic algorithms, are required to improve the model's accuracy in predicting EMI more effectively.

However, one of the main challenges in Stacking Ensembles is the considerable computing power requirement. Although this model can provide high accuracy, its heavy processing makes it less ideal for real-time EMI monitoring, especially for LED drivers with limited computing resources. Future research should explore more efficient alternatives to enhance practicality, such as optimizing RF with feature selection or using deep learning architectures explicitly designed for low-latency real-time applications. By addressing these challenges, EMI prediction models will become more accurate, efficient, scalable, and ready to be implemented in the industry, particularly for next-generation LED drivers.

4.6 Future research

For future research looking to expand the scope of these studies, it is important to consider more variations in signal types and environmental factors that can affect EMI prediction. Currently, research has used signals such as LorenzModif and Gaussian signals, but there is still much potential to explore the impact of high-frequency signals, irregular switching patterns, and noise from the real environment on the accuracy of EMI prediction. Understanding how these factors affect the model performance is crucial for improving the resilience and reliability of EMI prediction systems.

In addition, applying ensemble learning models in real-world LED driver design still requires further development. Future research could focus on integrating these models into design tools or real-time monitoring systems to detect and mitigate EMI during device operation. Developing a dynamic EMI mitigation system capable of adapting to environmental changes and the operational conditions of LED drivers will be a significant step forward in real-time EMI management.

However, adaptive learning strategies, such as online learning or gradual updates, can help the model evolve by learning from the latest EMI patterns. Using this approach, models can remain accurate and relevant even as environmental conditions change or LED driver configurations evolve, making them more flexible and reliable when deployed in the real world.

However, the real-time implementation remains a significant challenge. Computing efficiency is critical in direct EMI monitoring; therefore, future research needs to develop lighter models, utilize edge computing, and apply hardware acceleration techniques such as TPUs or FPGAs to make the inference process faster without sacrificing accuracy. In addition, integrating AI-based EMI prediction into embedded systems or Edge AI solutions can improve scalability and accelerate the deployment of these models in real-world LED driver systems. This move not only helps to optimize the performance of devices but also encourages innovation in more sustainable energy efficiency in various industries.

5. CONCLUSION

This study presents a novel approach to predicting EMI in LED drivers using spread spectrum techniques to modify their switching mode. The ensemble learning methods-the RF and Stacking-achieved promising accuracy, with R^2 values of 0.7327 and 0.7347, respectively. The results highlight the potential of artificial intelligence to address challenges related to EMC. This approach enhances LED driver design, ensures compliance with EMI regulations, and improves the reliability of the lighting system. In addition, it lays the foundation for more effective EMI mitigation strategies and further research under diverse system conditions. Future exploration of advanced ensemble learning techniques could improve EMI prediction accuracy and contribute to developing industry standards.

ACKNOWLEDGMENT

This project was supported by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (No.: 043/SP2H/RT-MONO/LL4/2024) and Telkom University-LLDIKTI IV (No.: 0667/E5/AL.04/2024).

REFERENCES

[1] Qu, Y., Shu, W., Chang, J.S. (2018). A low-EMI, high-reliability PWM-based dual-phase LED driver for automotive lighting. IEEE Journal of Emerging and Selected Topics in Power Electronics, 6(3): 1179-1189. https://doi.org/10.1109/JESTPE.2018.2812902

[2] Araújo, C.M. (2018). Characterization of electromagnetic interference conducted in DC-DC buck converter LED Light in accordance with CISPR 25-Class 5. In 2018 13th IEEE International Conference on Industry Applications (INDUSCON), Sao Paulo, Brazil, pp. 1259-1265. https://doi.org/10.1109/INDUSCON.2018.8627158

[3] Zeghoudi, A., Bendaoud, A., Canale, L., Tilmatine, A., Slimani, H. (2021). Common mode and differential mode noise of AC/DC LED Driver. In 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Bari, Italy, pp. 1-6. https://doi.org/10.1109/EEEIC/ICPSEurope51590.2021.9584616

[4] Natarajan, S., Babu, T.S., Balasubramanian, K., Subramaniam, U., Almakhles, D.J. (2019). A state-of-the-Art review on conducted electromagnetic interference in non-Isolated DC to DC converters. IEEE Access, 8: 2564-2577. https://doi.org/10.1109/ACCESS.2019.2961954

[5] Hu, H., Yu, S.S., Trinh, H. (2024). A review of uncertainties in power systems-Modeling, impact, and mitigation. Designs, 8(1): 10. https://doi.org/10.3390/designs8010010

[6] Bandara, S., Rajeev, P., Gad, E. (2024). A review on condition assessment technologies for power distribution network infrastructure. Structure and Infrastructure Engineering, 20(11): 1834-1851. https://doi.org/10.1080/15732479.2023.2177680

[7] Ottenburger, S.S. (2021). Criticality-Based designs of power distribution systems: Metrics for identifying urban resilient smart grids. In Information Technology Applications for Crisis Response and Management. IGI Global, pp. 150-175. https://doi.org/10.4018/978-1-7998-7210-8.ch008

[8] Sreekumar, V. (2019). A review of the EMI mitigation techniques in power converters. In Proceedings of The International Conference on Systems, Energy & Environment (ICSEE). https://doi.org/10.2139/ssrn.3448322

[9] Zhou, Y. (2021). Prediction and analysis of conduction electromagnetic interference in communication power. Journal of Nanoelectronics and Optoelectronics, 16(12): 1892-1896. https://doi.org/10.1166/jno.2021.3155

[10] Najjar, M., Kouchaki, A., Nymand, M. (2020). Inductor size evaluation of an electromagnetic interference filter for a two-level power factor correction rectifier using

different modulation techniques. In 2020 22nd European Conference on Power Electronics and Applications (EPE'20 ECCE Europe). Lyon, France, pp. P-1-P.9. https://doi.org/10.23919/EPE20ECCEEurope43536.2020.9215713

[11] Liu, Y., Mei, Z., Jiang, S., Liang, W. (2020). Conducted common-mode electromagnetic interference suppression in the AC and DC sides of a grid-connected inverter. IET Power Electronics, 13(13): 2926-2934. https://doi.org/10.1049/iet-pel.2019.1087

[12] Kumar, M.S., Rani, A.J. (2018). Reduction of conducted electromagnetic interference by using filters. Computers & Electrical Engineering, 72: 169-178. https://doi.org/10.1016/j.compeleceng.2018.09.002

[13] Tsymbal, O., Lebedev, D. (2024). Improving EMI performance in automotive data transmission systems using spread spectrum modulation. Computer-Integrated Technologies: Education, Science, Production, 56: 59-70. https://doi.org/10.36910/6775-2524-0560-2024-56-07

[14] Yanuar, M.H., Hidayat, R., Firmansyah, E. (2016). An experimental study of conducted EMI mitigation on the LED driver using spread spectrum technique. International Journal of Electronics and Telecommunications, 62(3): 283-288. http://doi.org/10.1515/eletel-2016-0041

[15] Duan, Z., Wen, X. (2020). A new analytical conducted EMI prediction method for SiC motor drive systems. Etransportation, 3: 100047. https://doi.org/10.1016/j.etran.2020.100047

[16] Bahl, N., Bansal, A. (2019). Balancing performance measures in classification using ensemble learning methods. In Business Information Systems: 22nd International Conference, BIS 2019, Seville, Spain, pp. 311-324. https://doi.org/10.1007/978-3-030-20482-2_25

[17] Lee, T.H., Ullah, A., Wang, R. (2020). Bootstrap aggregating and random forest. Macroeconomic Forecasting in the Era of Big Data: Theory and Practice, pp. 389-429. https://doi.org/10.1007/978-3-030-31150-6_13

[18] Nukui, T., Onogi, A. (2023). An R package for ensemble learning stacking. Bioinformatics Advances, 3(1): vbad139. https://doi.org/10.1093/bioadv/vbad139

[19] Tyralis, H., Papacharalampous, G., Burnetas, A., Langousis, A. (2019). Hydrological post-Processing using stacked generalization of quantile regression algorithms: Large-scale application over CONUS. Journal of Hydrology, 577: 123957. https://doi.org/10.1016/j.jhydrol.2019.123957

[20] Lee, D.G., Ahn, K.H. (2021). A stacking ensemble model for hydrological post-Processing to improve streamflow forecasts at medium-Range timescales over South Korea. Journal of Hydrology, 600: 126681. https://doi.org/10.1016/j.jhydrol.2021.126681

[21] Li, Y., Chen, W. (2020). A comparative performance assessment of ensemble learning for credit scoring. Mathematics, 8(10): 1756. https://doi.org/10.3390/math8101756

[22] Li, Y., Li, G., Wang, K., Wang, Z., Chen, Y. (2024). Forest fire risk prediction based on stacking ensemble learning for Yunnan Province of China. Fire, 7(1): 13. https://doi.org/10.3390/fire7010013

[23] Ilia, I., Tsangaratos, P., Tzampoglou, P., Chen, W., Hong, H. (2022). Flash flood susceptibility mapping using stacking ensemble machine learning models. Geocarto International, 37(27): 15010-15036. https://doi.org/10.1080/10106049.2022.2093990

[24] Wang, C., Xu, X., Zhang, Y., Cao, Z., Ullah, I., Zhang, Z., Miao, M. (2024). A stacking ensemble learning model combining a crop simulation model with machine learning to improve the dry matter yield estimation of greenhouse pakchoi. Agronomy, 14: 1789. https://doi.org/10.3390/agronomy14081789

[25] Nhu, V.H., Shirzadi, A., Shahabi, H., Chen, W., Clague, J.J., Geertsema, M., Jaafari, A., Avand, M., Miraki, S., Asl, D.T., Pham, B.T., Ahmad, B.B., Lee, S. (2020). Shallow landslide susceptibility mapping by random forest base classifier and its ensembles in a semi-arid region of Iran. Forests, 11(4): 421. https://doi.org/10.3390/f11040421

[26] Arora, H.C., Bhushan, B., Kumar, A., Kumar, P., Hadzima-Nyarko, M., Radu, D., Cazacu, C.E., Kapoor, N.R. (2024). Ensemble learning based compressive strength prediction of concrete structures through real-time non-destructive testing. Scientific Reports, 14(1): 1824. https://doi.org/10.1038/s41598-024-52046-y

[27] Zhang, Z., Zhu, X., Liu, D. (2022). Model of gradient boosting random forest prediction. In 2022 IEEE International Conference on Networking, Sensing and Control (ICNSC), Shanghai, China, pp. 1-6. https://doi.org/10.1109/ICNSC55942.2022.10004112

[28] Mallick, J., Talukdar, S., Ahmed, M. (2022). Combining high resolution input and stacking ensemble machine learning algorithms for developing robust groundwater potentiality models in Bisha watershed, Saudi Arabia. Applied Water Science, 12(4): 77. https://doi.org/10.1007/s13201-022-01599-2

[29] Syahid, N.F., Weerapreeyakul, N., Srisongkram, T. (2023). StackBRAF: A large-scale stacking ensemble learning for BRAF affinity prediction. ACS Omega, 8(23): 20881-20891. https://doi.org/10.1021/acsomega.3c01641

[30] He, Y., Xiao, J., An, X., Cao, C., Xiao, J. (2022). Short-Term power load probability density forecasting based on GLRQ-stacking ensemble learning method. International Journal of Electrical Power & Energy Systems, 142: 108243. https://doi.org/10.1016/j.ijepes.2022.108243

[31] Zandi, O., Zahraie, B., Nasseri, M., Behrangi, A. (2022). Stacking machine learning models versus a locally weighted linear model to generate high-resolution monthly precipitation over a topographically complex area. Atmospheric Research, 272: 106159. https://doi.org/10.1016/j.atmosres.2022.106159

[32] Han, S., Kim, H., Lee, Y.S. (2020). Double random forest. Machine Learning, 109: 1569-1586. https://doi.org/10.1007/s10994-020-05889-1