

RESEARCH ARTICLE

Automated Prognostics and Diagnostics of Railway Tram Noises Using Machine Learning

JUNHUI HUANG, HAO LIU, WENYAN XI, AND SAKDIRAT KAEWUNRUEN^{ID}

Department of Civil Engineering, School of Engineering, University of Birmingham, B15 2TT Birmingham, U.K.

Corresponding author: Sakdirat Kaewunruen (s.kaewunruen@bham.ac.uk)

This work was supported in part by European Commission for H2020-Marie Skłodowska-Curie Actions (MSCA)-Research and Innovation Staff Exchange (RIS) through the Rail Infrastructure Systems Engineering Network (RIS-EN) under Project 691135; in part by H2020 Shift2Rail under Project 730849; in part by Long-Range Mannix (LORAM); in part by Brazilian Railway Authority; in part by China Academy of Railway Sciences (CARS); in part by Network Rail; and in part by the Rail Safety and Standard Board (RSSB), U.K. The APC has been kindly sponsored by the University of Birmingham Library's Open Access Fund.

ABSTRACT Railway noise, stemming from various sources such as wheel/rail interactions, locomotives, and track machinery, affects both human health and the environment. This study explores the application of machine learning (ML) models to quantify tram noise at sharp curves, considering variables such as weather conditions, train speed, crowd levels, and running directions. Data collection is carried out on a tram line in Birmingham, using an iPhone 11 to record acoustic data at a sample rate of 48 kHz. The noise is categorized into impact noise, rolling noise, flanging noise, and squeal noise based on frequency and power spectrum characteristics. Random Forests (RF) and Extreme Gradient Boosting (XGBoost) are employed to predict the root mean square (R.M.S) values of each type of noise. Results indicate that XGBoost outperformed RF with an R^2 up to 0.96 during k-fold cross-validation. This model provides a robust tool for railway operators to optimize noise control measures and contributes to improved compliance with environmental regulations and a better quality of life for communities near rail tracks.

INDEX TERMS Railway noise, machine learning, noise quantification, environmental factors, random forests, XGBoost.

I. INTRODUCTION

Railway noise comes from various sources, including wheel/rail noise, locomotives, warning signals, bridges, freight vehicles, flange squeal on tight curves, marshalling yards, maintenance machines, and track machinery horns [1]. Thompson and Joines detailed the primary sources of railway noise, focusing on wheel/rail interactions [2]. They categorized this noise into three types: rolling noise, impact noise, and squeal noise. Rolling noise, common at conventional and high speeds, results from wheel-rail interface roughness. Impact noise, a severe form of rolling noise, occurs at surface discontinuities like rail joints and welds. Squeal noise arises from frictional instability on sharp curves.

Investigating railway noise is very important due to its wide-ranging effects on health, and the environment.

The associate editor coordinating the review of this manuscript and approving it for publication was Mohsin Jamil^{ID}.

Long-term exposure to railway noise can cause health issues like sleep problems, heart diseases, and stress, which leads to a lower quality of life for people living near tracks. While prior research has focused on wheel-rail interactions and noise types such as rolling, impact, and squeal noise, there are still some critical factors that remain underexplored. Fields and Walk identified noise as a major environmental nuisance [3], while Mohler highlighted its disruptive effect on communication [4]. Öhrström and Skånberg noted that vibration increases noise annoyance [3], and Sorensen et al. found a link between railway noise above 60 dB and higher hypertension risk [5]. Grubliauskas et al. confirmed that excessive railway noise disrupts residential life, particularly affecting night-time comfort and health [6]. Environmentally, railway noise affects both people and wildlife by disrupting natural habitats [7]. Railway operators need to meet strict noise regulations to avoid legal issues and maintain good community relations. Therefore, continuous monitoring of

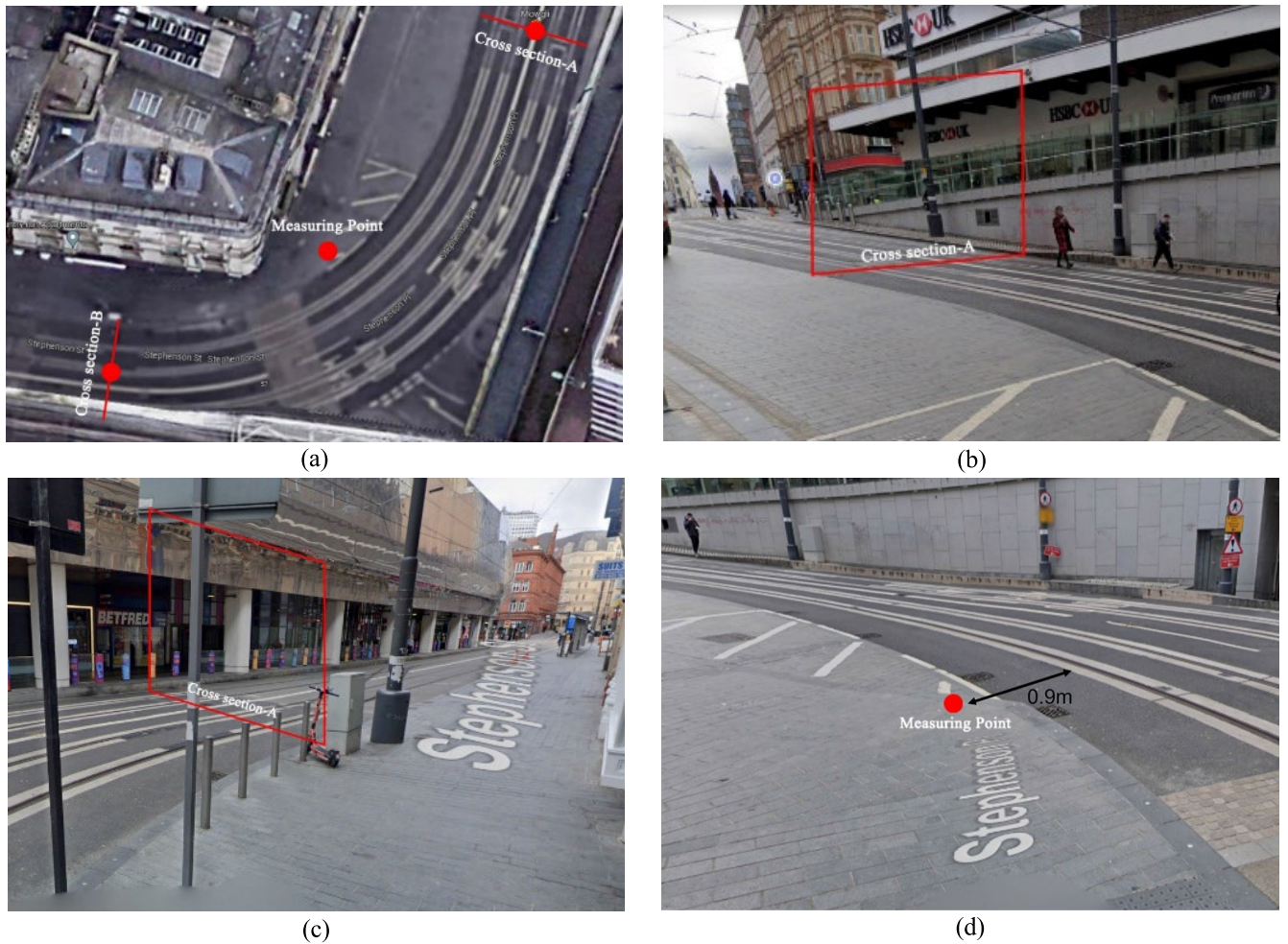


FIGURE 1. The tramline section for data collection (a) overview of the whole section (b) starting point A (c) starting point B (d) measuring point where is fixed during every measurement.

the noise is important to ensure sustainable and resilient transportation systems.

The first theoretical model of rolling noise was found to be produced by Remington [8], [9], [10], [11] based on irregularities of the wheel/rail and was later extended by Tompson [12]. Subsequently, the model was funded by the European Rail Research Institute resulting in a computer program Track-Wheel Interaction Noise Software [13] validated by extensive full-scale experiments [14], and predicted noise with high precision (within 2 dB) for various wheel and track designs. Subsequent studies have shown that optimizing wheel shape can reduce railway noise by up to 5 dB [15] and factors such as rubber material parameters, the number of rubber blocks, and the gap-rubber ratio can affect noise levels [16]. In addition, Andrés et al. [17] focused on optimizing track design parameters, revealing that an optimal combination can reduce sound radiation by up to 7.4 dB, with rail pad stiffness identified as the most influential factor. Sheng et al. [18] investigated the railway noise of high-speed trains running on a non-ballasted slab track. Their findings emphasized that train speed dramatically affects

railway noise, with both rolling noise and aerodynamic noise varying with speed [18], [19]. Numerous other factors impacting wheel/rail noise have also been studied, including locomotives' propulsion systems and static and dynamic loads [20].

While a wide range of factors, primarily from track and rail characteristics, have been extensively evaluated in terms of railway noise, the effects of other factors, such as weather conditions, remain unexplored. Climate change has significant implications for the evaluation of railway noise due to its effects on weather patterns and extreme weather events. According to Palin et al. [21] climate change is expected to increase the frequency and intensity of weather hazards, posing substantial risks to railway infrastructure resilience and performance. Key impacts include temperature fluctuations, greater precipitation, and the occurrence of extreme weather events, all of which can influence the structural and acoustic properties of railway materials, thereby affecting noise levels. This study proposes the use of machine learning (ML) models to quantify the correlation between railway noise and variables such as weather conditions

(temperature, humidity, pressure, wind, wind speed, weather condition, dew point, precipitation), train speed, crowd levels, and running directions (which affect the slope of the tracks). By incorporating these variables, the study aims to provide in-depth insights into existing models like TWINS, reducing predictive errors and extending the model’s applicability to a broader range of wheel and track designs. Furthermore, the proposed model can predict R.M.S levels for various types of railway noise based on the available features.

Section II details the data collection procedures. Section III provides methodologies for four types of noise extraction and ML models. Section IV elaborates on results and discussions while Section 5 presents conclusions, limitations and future directions of the study.

II. METHODOLOGIES

This section includes methodologies for data collection, data analysis, and machine learning models.

A. DATA COLLECTION

This section outlines the procedure for collecting data to quantify the correlation between railway noise and various factors at a sharp curve in Birmingham New Street, as shown in FIGURE 1(a). The tram line operates on a 1435 mm standard gauge track and is electrified with a 750 V DC overhead line. The running vehicle is Urbos 3 which can reach up to 70 km/h [22].

The data collection is conducted using an iPhone 11 equipped with the ‘Motiv Audio’ app, which allows for audio recording at a sampling rate of 48 kHz. This high sampling rate ensures the capture of high-quality acoustic data. Key attributes recorded include weather conditions, crowd levels, acoustic data, direction, and type of noise. This setup enables an analysis of the impact of these factors on railway noise.

FIGURE 1(b) and (c) define two starting points to show the direction of the train running. Here, if a tram is from ‘starting point A’, the direction is considered to be positive and the direction is negative for the tram from ‘starting point B’. The phone is placed at the measuring point and the subject starts recording when the train enters one of the starting points and ends recording when the tram passes another starting point.

The measuring point is for the pedestrian lane which is accessible, and the space is far enough away from the running tram to keep the subject safe. The chosen measuring point is situated next to the railway, which is ideal for capturing the sound emanating directly from the trains.

The available weather conditions include temperature, condition, wind, wind speed, humidity, precipitation, and dew point. The schedule for data collection is made to cover different weather conditions and various times of the data to capture different traffic densities. The ‘starting point A’ is close to a tram stop which allows the data collection to cover the accelerated period while the tram needs to decelerate to enter the stop. Meanwhile, Open areas like this are suitable

for studying the impact of weather on noise levels as there is likely less obstruction for wind, rain, and other weather conditions to affect the noise. This collected point is close to a busy station which introduces variability in crowdedness. It is worth noting that the tram concerning the slope can also be observed. If the tracks are on an inline or decline at this location, it can affect both the noise and speed of the trains.

Concurrently, it is important to log relevant metadata for each recording, including the date, time, weather conditions, and perceived train crowdedness. Weather conditions are from the built-in app on the phone. It is worth mentioning that the rule to categorize the crowd level adheres to the standards outlined by the United Kingdom Rail Safety and Standards Boards (UKRSSB) and can be found in FIGURE 2.

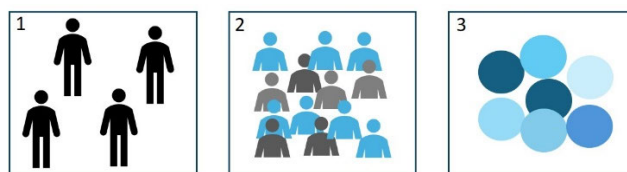


FIGURE 2. Visual illustration of crowd level in a public transport.

The crowd level is defined following the level of visibility of the passenger’s body. The whole-body is visible implying that a tram is labelled as ‘uncrowded’. Only the head and shoulders are visible leading to a ‘crowded’ while only the head can be seen resulting in an ‘overcrowded’.

The tram speed can be an important factor affecting the curve noise. Higher speeds cause greater dynamic forces and further lead to elevated noise levels [23], [24]. The ‘speed’ feature in this study is determined by the equation of velocity.

$$v = \frac{s}{t} \tag{1}$$

where:

- v represents the velocity.
- s defines the distance between two cross section points.
- t shows the duration of the recording.

Besides, any unusual events or disturbances during the recording that might affect the data quality are logged.

B. DATA PREPROCESSING

FIGURE 3 overviews the data pre-processing life cycle. The recording is conducted from point ‘b’ to point ‘a’ along the curve in FIGURE 3. The collected audio data then undergoes Fast Fourier Transform (FFT) analysis and filtering to isolate and identify different types of noise. The specific conditions are the frequency and energy signature for each different type of noise.

The tram noise is filtered based on Table 1 specifying the sound level and frequency bands that different types of noises vary before the label showing the type of noise is tagged to each component. The integrity of the label is confirmed by experts listening to the filtered samples. After the label integrity is ascertained, the sound pressure and label of each component are extracted. It is worth mentioning

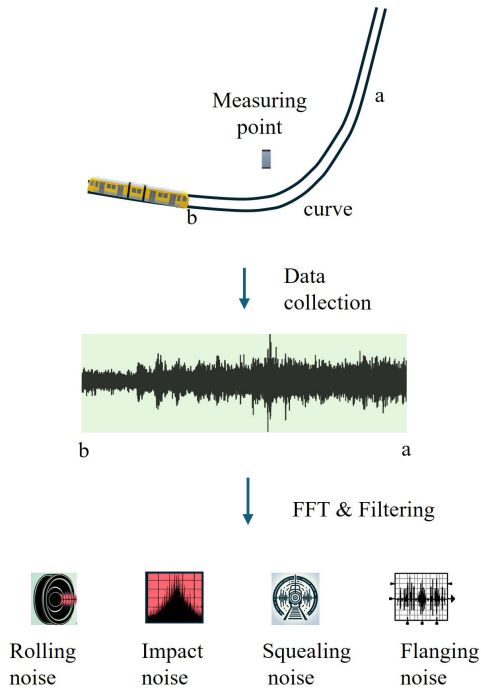


FIGURE 3. Data flowing.

TABLE 1. Conditions for filtering different types of noise.

Noise	Frequency (Hz)	Power spectrum (dB)
Impact noise	20 – 100	90
Rolling noise	80 – 200	90
Squeal noise	5000 – 8000	80
Flanging noise	2000 - 5000	90

that an additional condition is applied to the squeal noise. As shown in FIGURE 4, the occurrence of peaks among the fluctuating wave is another prerequisite for the squealing noise. To separate impact noise and rolling noise, the feature of the peak is also considered since impact noise is subject to transient peak due to the track irregularities while rolling noise is flatter due to the track and rail roughness.

C. MACHINE LEARNING MODELS

Multiple tree-based machine learning models such as Random Forests (RF) and Extreme Gradient Boosting (XGBoost) are available in this study to capture non-linear relationships and interactions between features. Multiple models are used for model comparison and robustness evaluation.

The tree-based model is selected as it has been widely proven that tabular datasets are more compatible with tree-based models. Friedman [25] showed that tree-based models like XGBoost and RF outperform deep learning models on medium-sized tabular datasets (~10K samples) with an extensive benchmark over 45 different tabular datasets conducted.

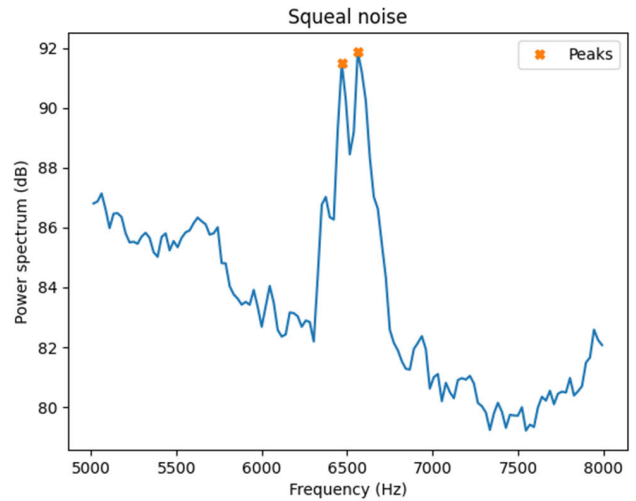


FIGURE 4. Squeal noise.

One of the main constraints to applying AI models in industries is interpretability. The tree-based models provide a visual and interpretable representation of the decision-making process, which can be valuable for understanding how the model is making predictions. This benefit can fill the gap that makes the potential users more confident to use the model, especially for areas where understanding the decision process is crucial for validation and trust [26].

1) RANDOM FORESTS

Since its development by Breiman in 2001 [27], RF has achieved great success. This ensemble model comprises multiple decision tree estimators (as can be seen in FIGURE 5 and is used for both classification and regression tasks. Each decision tree is trained on a sub-dataset obtained through bootstrap sampling [28], introducing variety and reducing overfitting. For classification problems, the final prediction is determined by voting on the outcomes of all trees, while for regression problems, the predictions are averaged.

RF is less prone to overfitting because each tree is trained on a different subset of data. This issue is further mitigated by adjusting hyperparameters such as the maximum depth of trees or the minimum number of samples required for leaf nodes. The nature of the model also enhances its robustness to noisy data by smoothing the impact of individual data points that may not represent the main pattern. Another advantageous feature of RF is the implicit feature selection. By randomly selecting subsets of data to train each tree instead of using all attributes, spurring to identify the most informative features. Therefore, this reduces the influence of less important features on the final predictions and minimizes the need for extensive feature selection.

2) XGBoost

Tree boosting is a highly effective and widely applied technique in ML. Chen and Guestrin [29] introduced XGBoost,

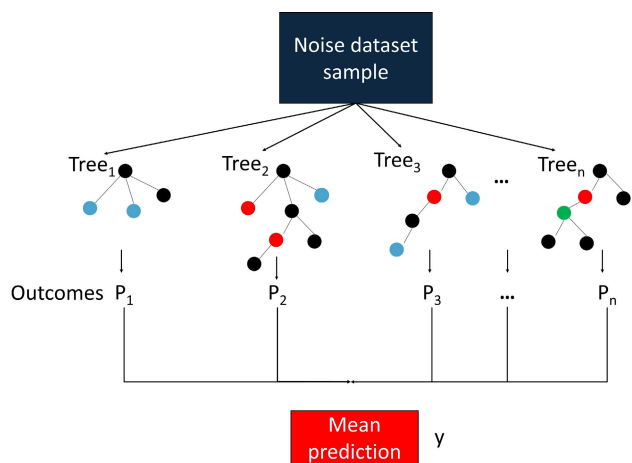


FIGURE 5. RF.

an expandable end-to-end tree-boosting system that has been widely embraced by data scientists for its ability to deliver state-of-the-art results in many ML challenges. They developed an innovative algorithm designed for efficiently handling sparse data and introduced a weighted quantile sketch technique for approximate tree learning. Additionally, they explored cache access patterns, data compression strategies, and sharding, demonstrating that XGBoost is scalable enough to handle datasets with billions of examples while using significantly fewer computing resources than existing systems.

XGBoost shares a common feature with RF as both are ensemble algorithms that combine the predictions of multiple individual models, typically decision trees, to produce a more accurate and robust final prediction. However, the methods by which XGBoost and RF build and combine these trees differ. While RF averages the predictions from each tree, XGBoost builds decision trees sequentially seen in **FIGURE 6**, with each new tree aiming to correct the errors of the previous trees. This sequential construction passes the residuals, or differences between the actual values and the predictions of prior trees, to the newly created tree, refining the model with each iteration.

Robustness in AI refers to the system’s ability to perform reliably under a wide range of conditions such as unforeseen scenarios and adversarial attacks, and the implementation of security measures throughout the AI lifecycle. Certification and standardization of AI systems are pivotal. These processes ensure that AI systems adhere to defined safety and performance standards, thereby mitigating the risks associated with AI deployment [30].

By incorporating diverse environmental conditions, such as temperature fluctuations, wind speed, humidity, and precipitation, as well as operational factors like tram speed, crowd levels, and running directions, the models are trained to handle a wide range of scenarios. This diversity in training data ensures that the models are less likely to be biased or overfitted to specific conditions and can generalize well

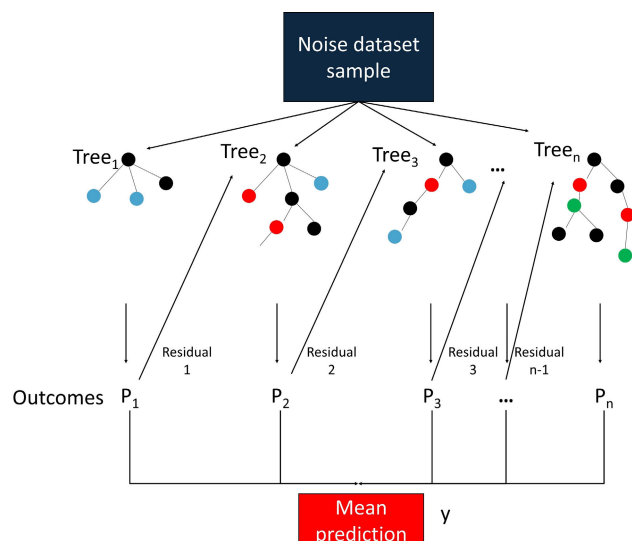


FIGURE 6. XGBoost.

to new, unseen data. Furthermore, robust and standardized steps are implemented such as the application of k-fold cross-validation and the isolation of the test set.

The application of k-fold cross-validation and rigorous testing procedures also ensures the robustness of the models. By splitting the dataset into multiple folds and ensuring that each fold is used for both training and validation, the models’ performance can be evaluated more thoroughly. This method helps in identifying any potential overfitting and ensures that the models maintain high predictive accuracy across different subsets of the data. The test set is completely isolated during the training process. This means no information leaks into the test data while tuning the model.

III. RESULT

The dataset (300 journeys) encompasses comprehensive weather data, including key parameters such as temperature (ranging from a minimum of 46.0°F to a maximum of 81.0°F), dew point (34.0°F to 57.0°F), humidity (39.0% to 93.0%), wind speed (6.0 mph to 15.0 mph), and pressure (29.42 inHg to 29.89 inHg). These variables provide a broad spectrum of environmental conditions under which the noise measurements were recorded. Additionally, the dataset captures categorical weather attributes such as wind direction, and weather conditions. The wind direction includes various orientations (e.g., SSW, WSW, WNW), while weather conditions span descriptions like Fair, Mostly Cloudy, Partly Cloudy, Showers in the Vicinity, Cloudy, and Haze. The dataset also accounts for the direction of the operating tram (a, b). Three levels of crowdedness are provided and the speed is calculated based on (1). Temporal data is recorded, with dates ranging in March, June, July, and August in 2024 and August in 2023 to provide a diverse set of observations across months and weather conditions. This collection allows for an in-depth analysis of how various weather factors influence the RMS values of the different types of noise.

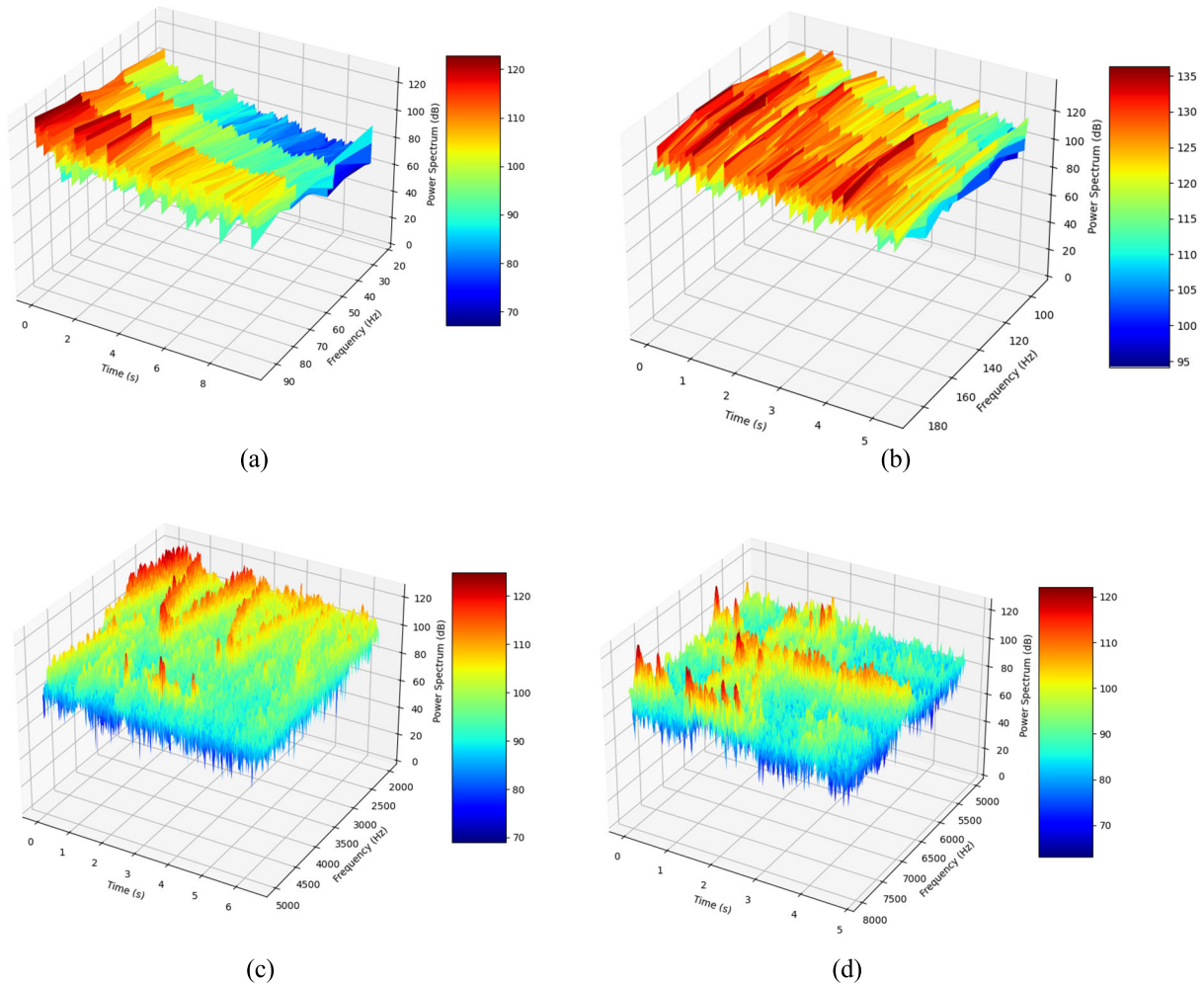


FIGURE 7. Noise extraction (a) Impact noise (b) Rolling noise (c) Flanging noise (d) Squeal noise.

A. NOISE EXTRACTION

To train a model that can predict the R.M.S value for each type of noise, the first step is to extract the noise of interest. **FIGURE 7** provides an example of how this study acquires different noise from the raw data. As can be seen in **FIGURE 7(a)**, distinct sharp peaks at various time intervals show the transient nature of railway impact noise, coming from abrupt interactions such as wheel-rail contact and track irregularities. High-power regions predominantly at higher frequencies highlight the significant high-frequency components which are the results of the rapid force during impacts. Variations in peak height and intensity over time might show the amplitude variation due to changes in train speed, wheel condition, or track conditions. Isolated high-power peaks indicate localized high-impact events, often caused by track irregularities. **FIGURE 7(b)** captures the characteristics of tram rolling noise across time consistent nature of rolling noise.

The temporal distribution shows intermittent high-power events which indicate the tram is navigating curves where the

flanging contact is the largest. The variability in amplitude with sharp peaks in power suggests changes in the intensity of the flange-rail interaction.

FIGURE 7(d) shows squeal noise which has prominent high-frequency components between 5000 Hz and 8000 Hz. The variability in amplitude, with distinct spikes in power, shows the nature of squeal noise of fluctuations in the intensity of the flange-rail interaction.

B. MODEL RESULTS

1) RESULTS

The two selected models are tuned by Optuna [31] to get optimal hyperparameters. K-fold cross-validation [32] is also applied to improve the robustness of the result. The dataset is split into 80% of the training set and 20% of the test set. The 80% training set is further divided into 80% of the training set for and 20% of the validation set for the 5-fold cross-validation used in this study. The reason for such segmentation is to avoid unintentional data leakage which might inflate the model's performance significantly [33].

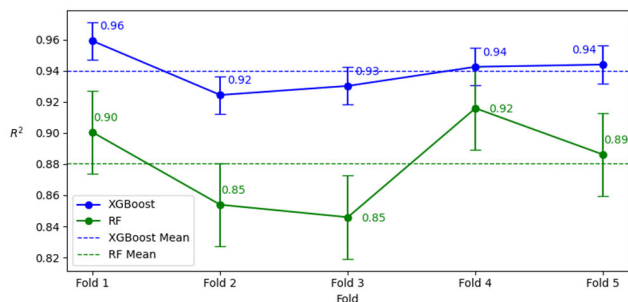


FIGURE 8. k-fold cross-validation results for XGBoost and RF.

In FIGURE 8, the XGBoost model consistently outperforms the RF model across all folds, with R^2 scores ranging from approximately 0.92 to 0.96. For the RF model, R^2 scores range from 0.85 to 0.90. The error bars, representing the standard deviation, are relatively narrow for both models which means that consistent performance across folds. The superior performance of the XGBoost model suggests it is more effective in capturing the underlying patterns in the data compared to the RF model. This consistent performance advantage highlights the robustness of XGBoost in modelling complex datasets, making it a preferable choice for tasks requiring high predictive accuracy. RF shows an R^2 of 0.91 using the test set while XGBoost presents a slightly larger R^2 of 0.94. The detailed performance on each sample for XGBoost is shown in FIGURE 9. For impact noise, the model demonstrates a reasonably good fit as the predicted values closely follow the actual values. The error bars indicate that the majority of predictions fall within a narrow error margin but a few outliers are present. In the case of rolling noise, the model maintains high predictive accuracy, as evidenced by the close alignment between the actual and predicted values.

The error bars are relatively small and consistent, indicating the model’s robustness in handling the variability inherent in rolling noise. The errors remain within a manageable range, although there is a slight increase in variability compared to the impact noise. The performance for flanging noise is promising, as the predicted values closely match the actual sound pressures. This indicates that the model performs well overall. Finally, the squeal noise predictions show the model’s capacity to track the actual values with a high degree of precision. The error bars are minimal which means that the model’s effectiveness in capturing the patterns associated with squeal noise. Another possible reason can be the inherent fluctuating nature of the squeal noise.

2) ROBUSTNESS EVALUATION

Evaluating the robustness of machine learning models is critical due to the potential vulnerabilities that can arise throughout their lifecycle. These vulnerabilities can stem from various sources, such as the contamination of training data, which can distort the model before training begins, or from perturbed data that can affect the model even after

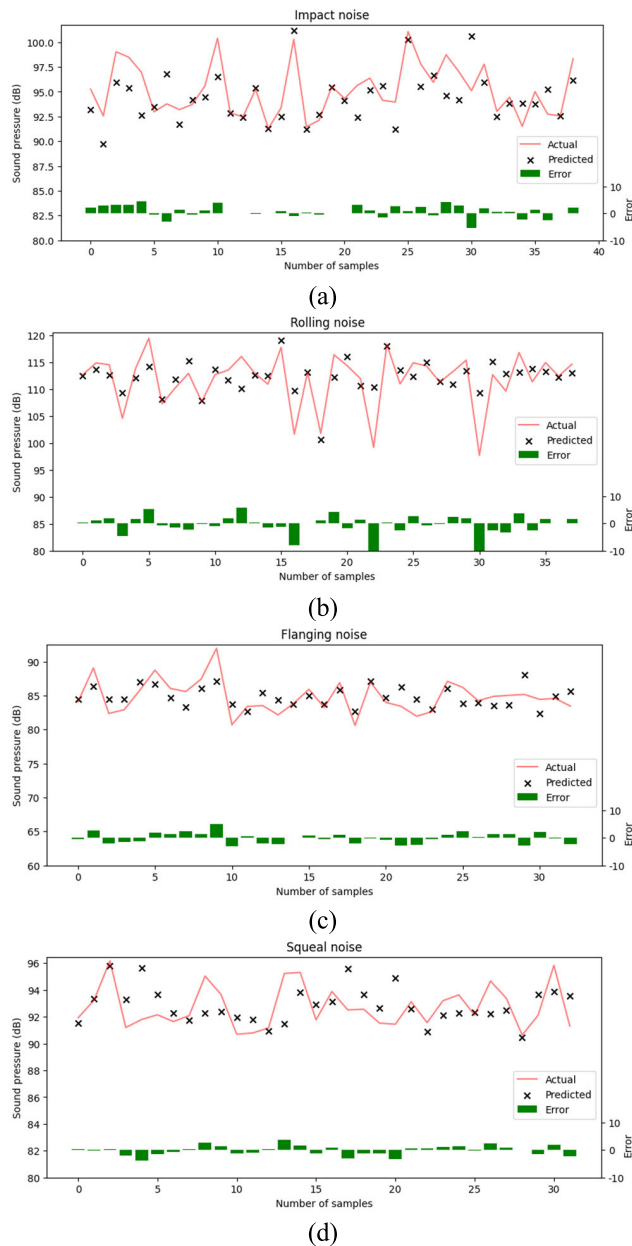


FIGURE 9. XGBoost performance using the test set.

deployment. To address these concerns, this study undertakes a comprehensive evaluation of the model’s robustness by introducing noise data both before and following the training process. This dual-phase evaluation aims to ascertain whether the model can accurately differentiate noise data before training and to test the model’s resilience against perturbed data post-training. Such rigorous testing ensures that the model maintains its integrity and performance in real-world applications, thereby enhancing its reliability and security against adversarial attacks [30].

To generate a noisy dataset, the entire original dataset is duplicated, and 20% noise is added to each feature in the duplicated dataset. Concurrently, an additional output label indicating ‘Noise’ or ‘Clean’ is appended to both datasets

before the noisy dataset is integrated into the clean dataset. The purpose of this is to assess the model's performance in maintaining the accuracy of prediction. By training the model with the noisy dataset, as illustrated in **FIGURE 10(a)**, the results demonstrate that the model effectively differentiates noise-affected data while maintaining high-performance levels. **FIGURE 10(b)** illustrates the impact of perturbing data with varying scale factors on the performance of the XGBoost, as measured by the R^2 . The scale factor is applied to each feature ranging from 0 to 0.5. Initially, the model's performance remains stable with minor perturbations (scale factors between 0.0 and 0.1), maintaining R^2 values around 0.92 to 0.93. However, as the scale factor increases beyond 0.1, the R^2 value begins to decline gradually.

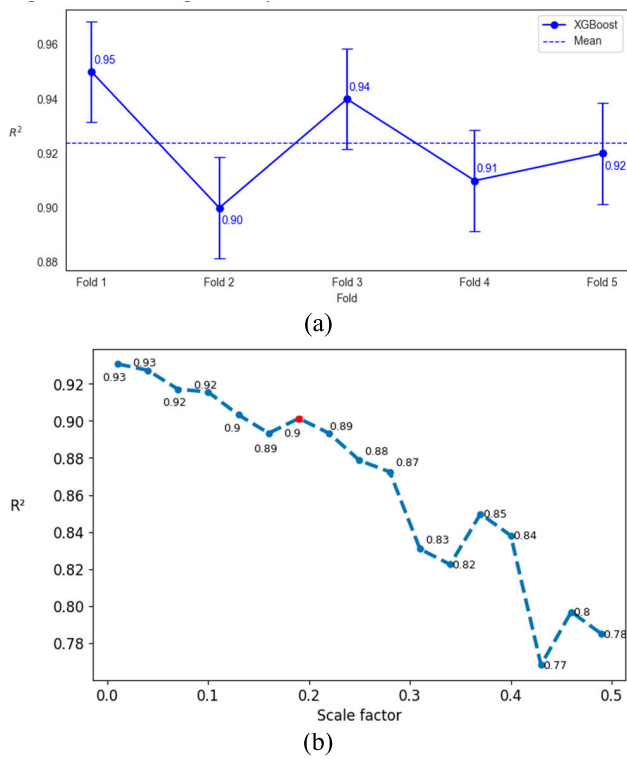


FIGURE 10. Performances for XGBoost (a) noise data added to train the model (b) perturbed data added to test the trained model.

This decline becomes faster with scale factors greater than 0.3, where the R^2 value drops sharply which suggests adverse effects on the model's accuracy. A notable point is highlighted at a scale factor of approximately 0.2 to show that the performance is lower than 0.9 if the scale factor is larger than 0.2.

Sensitivity analysis is also conducted to evaluate how the model responds to the variation of features. **FIGURE 11** reveals that speed is the most influential factor, significantly impacting the model's performance, followed closely by temperature and direction. These three features are important in determining the model's accuracy. Humidity, crowd level, and pressure also play important roles. Features such as wind, wind speed, condition, dew point, and precipitation have

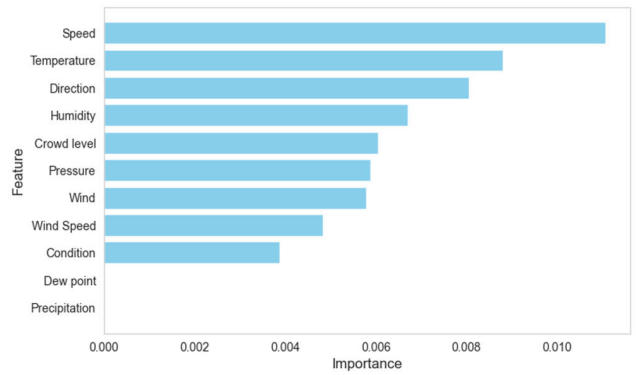


FIGURE 11. Sensitivity analysis for XGBoost.

lower importance scores which means that they contribute less impact to the model's predictions.

IV. CONCLUSION

This study presents an approach to predicting railway noise by integrating machine learning models with various environmental and operational factors. By incorporating these variables, the proposed model offers a significant advancement over existing models like TWINS. The proposed model can provide more insights into railway noise quantification and extend its applicability to a broader range of wheel and track designs. The ability to predict R.M.S levels for various types of railway noise based on features such as weather conditions, train speed, and crowd levels introduces a new dimension to noise management in the railway industry.

The implications of this research are profound for the engineering world. It offers a robust tool for railway operators to optimize noise control measures with features of more efficient and sustainable. By reducing noise pollution, the model contributes to better compliance with environmental regulations and improved quality of life for communities near rail tracks. Moreover, the interpretability of the tree-based models used in this study provides valuable insights into the decision-making process, fostering greater trust and adoption in industrial applications.

However, the study has certain limitations. The data collection is conducted in specific environmental conditions and geographical locations, which may not fully represent all possible scenarios. Future research should focus on expanding the dataset to include diverse conditions and locations to enhance the model's generalizability.

In conclusion, this research lays a solid foundation for future work in railway noise prediction and control. Future directions include integrating real-time data from sensors, exploring the impact of additional variables such as track maintenance activities, and refining the models for even higher accuracy. By addressing these areas, the proposed model can evolve into a more powerful tool, benefiting the railway industry and contributing to the development of quieter, more efficient rail systems.

ACKNOWLEDGMENT

The authors are grateful to the LORAM, Brazilian Railway Authority, China Academy of Railway Sciences (CARS), Network Rail, and Rail Safety and Standard Board (RSSB), U.K., for their assistance. The APC has been kindly sponsored by the University of Birmingham Library's Open Access Fund.

REFERENCES

- [1] C. G. Stanworth, "Consideration of some noise sources due to railway operation," *J. Sound Vib.*, vol. 87, no. 2, pp. 233–239, Mar. 1983.
- [2] D. J. Thompson and C. J. C. Jones, "A review of the modelling of wheel/rail noise generation," *J. Sound Vib.*, vol. 231, no. 3, pp. 519–536, Mar. 2000.
- [3] J. M. Fields and J. G. Walker, "The response to railway noise in residential areas in great Britain," *J. Sound Vib.*, vol. 85, no. 2, pp. 177–255, Nov. 1982.
- [4] U. Moehler, "Community response to railway noise: A review of social surveys," *J. Sound Vib.*, vol. 120, no. 2, pp. 321–332, Jan. 1988.
- [5] M. Sørensen, M. Hvidberg, B. Hoffmann, Z. J. Andersen, R. B. Nordsborg, K. G. Lillelund, J. Jakobsen, A. Tjønneland, K. Overvad, and O. Raaschou-Nielsen, "Exposure to road traffic and railway noise and associations with blood pressure and self-reported hypertension: A cohort study," *Environ. Health*, vol. 10, no. 1, pp. 1–11, Dec. 2011.
- [6] R. Grubliauskas, B. Strukčinskienė, J. Raistenskis, V. Strukčinskienė, R. Buckus, T. Januševičius, and P. A. da Silva Pereira, "Effects of urban rail noise level in a residential area," *J. Vibroeng.*, vol. 16, pp. 987–996, 2014.
- [7] P. S. Lucas, R. G. de Carvalho, and C. Grilo, "Railway disturbances on wildlife: Types, effects, and mitigation measures," in *Railway Ecology*, L. Borda-de-Água, R. Barrientos, P. Beja and H.M. Pereira. Cham, Switzerland: Springer, Sep. 2017, pp. 81–99, doi: [10.1007/978-3-319-57496-7_6](https://doi.org/10.1007/978-3-319-57496-7_6).
- [8] P. J. Remington, "Wheel/rail noise—Part I: Characterization of the wheel/rail dynamic system," *J. Sound Vib.*, vol. 46, no. 3, pp. 359–379, Jun. 1976.
- [9] P. J. Remington, "Wheel/rail noise—Part IV: Rolling noise," *J. Sound Vib.*, vol. 46, no. 3, pp. 419–436, Jun. 1976.
- [10] P. J. Remington, "Wheel/rail rolling noise. I: Theoretical analysis," *J. Acoust. Soc. Amer.*, vol. 81, no. 6, pp. 1805–1823, Jun. 1987.
- [11] P. J. Remington, "Wheel/rail rolling noise. II: Validation of the theory," *J. Acoust. Soc. Amer.*, vol. 81, no. 6, pp. 1824–1832, Jun. 1987.
- [12] D. J. Thompson, "Wheel-rail noise generation. Part I: Introduction and interaction model," *J. Sound Vib.*, vol. 161, no. 3, pp. 387–400, Mar. 1993.
- [13] D. J. Thompson, B. Hemsforth, and N. Vincent, "Experimental validation of the twins prediction program for rolling noise. Part 1: Description of the model and method," *J. Sound Vib.*, vol. 193, no. 1, pp. 123–135, May 1996.
- [14] D. J. Thompson, P. Fodiman, and H. Mahé, "Experimental validation of the twins prediction program for rolling noise. Part 2: Results," *J. Sound Vib.*, vol. 193, no. 1, pp. 137–147, May 1996.
- [15] X. Garcia-Andrés, J. Gutiérrez-Gil, J. Martínez-Casas, and F. D. Denia, "Wheel shape optimization approaches to reduce railway rolling noise," *Structural Multidisciplinary Optim.*, vol. 62, no. 5, pp. 2555–2570, Nov. 2020.
- [16] Y. He, Q. Zhou, F. Xu, X. Sheng, Y. He, and J. Han, "An investigation into the effect of rubber design parameters of a resilient wheel on wheel-rail noise," *Appl. Acoust.*, vol. 205, Mar. 2023, Art. no. 109259.
- [17] V. T. Andrés, J. Martínez-Casas, F. D. Denia, and D. J. Thompson, "Influence study of rail geometry and track properties on railway rolling noise," *J. Sound Vib.*, vol. 525, May 2022, Art. no. 116701.
- [18] X. Sheng, G. Cheng, and D. Thompson, "Modelling wheel/rail rolling noise for a high-speed train running along an infinitely long periodic slab track," *J. Acoust. Soc. Amer.*, vol. 148, no. 1, pp. 174–190, Jul. 2020.
- [19] X. Zhang, "The directivity of railway noise at different speeds," *J. Sound Vib.*, vol. 329, no. 25, pp. 5273–5288, Dec. 2010.
- [20] M. Cavacece, "The noise of rail transport," *J. Phys., Conf. Ser.*, vol. 2590, no. 1, Sep. 2023, Art. no. 012008.
- [21] E. J. Palin, I. Stipanovic Oslakovic, K. Gavin, and A. Quinn, "Implications of climate change for railway infrastructure," *WIREs Climate Change*, vol. 12, no. 5, Sep. 2021, Art. no. e728.
- [22] P. Potti and M. Marinov, "Evaluation of actual timetables and utilization levels of west midlands metro using event-based simulations," *Urban Rail Transit*, vol. 6, no. 1, pp. 28–41, Mar. 2020.
- [23] H. Yan, S. Xie, K. Jing, and Z. Feng, "A review of recent research into the causes and control of noise during high-speed train movement," *Appl. Sci.*, vol. 12, no. 15, p. 7508, Jul. 2022.
- [24] J. Han, Y. He, X. Xiao, X. Sheng, G. Zhao, and X. Jin, "Effect of control measures on wheel/rail noise when the vehicle curves," *Appl. Sci.*, vol. 7, no. 11, p. 1144, Nov. 2017.
- [25] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
- [26] S. Hara and K. Hayashi, "Making tree ensembles interpretable: A Bayesian model selection approach," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2018, pp. 77–85.
- [27] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.
- [28] T. Hesterberg, "Bootstrap," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 3, no. 6, pp. 497–526, 2011.
- [29] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
- [30] R. Hamon, H. Junklewitz, and I. Sanchez, "Robustness and explainability of artificial intelligence," Publications Office Eur. Union, Brussels, Belgium, Tech. Rep. EUR: Vol. 30040, vol. 207, p. 2020.
- [31] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2623–2631.
- [32] T.-T. Wong and P.-Y. Yeh, "Reliable accuracy estimates from k-fold cross validation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1586–1594, Aug. 2020.
- [33] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, "Data leakage inflates prediction performance in connectome-based machine learning models," *Nature Commun.*, vol. 15, no. 1, p. 1829, Feb. 2024.



JUNHUI HUANG received the bachelor's degree in electrical and railway engineering and the master's degree in railway systems engineering and integration from the University of Birmingham, U.K., and the Ph.D. degree from the University of Birmingham. His research interest includes artificial intelligent applied in railway



HAO LIU received the Bachelor of Engineering degree in civil engineering management from Hubei University of Technology, in 2019, and the M.Sc. degree in advanced engineering management from the University of Birmingham, in 2023.



WENYAN XI received the Bachelor of Engineering degree in engineering management from Changzhou University, in 2023. She is currently pursuing the degree in advanced engineering management with the University of Birmingham. She aims to develop strategies to mitigate railway noise impact on daily life, with research interests in railway noise, its effects, and its classification.



SAKDIRAT KAEWUNRUEN received the Ph.D. degree in structural engineering from the University of Wollongong, Australia. He has expertise in transport infrastructure engineering and management and successfully dealing with all stages of infrastructure life cycle and assuring safety, reliability, resilience, and sustainability of rail infrastructure systems. He is a Chartered Engineer. He has over 600 technical publications.