Evaluation of Railway Passenger Comfort With Machine Learning

JUNHUI HUANG AND SAKDIRAT KAEWUNRUEN [ID]

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

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

ABSTRACT Railway passenger comfort has been considered a growingly important field to attract more passengers from other public transports such as air flights. To allow passengers and train companies to estimate the onboard passenger comfort level, we propose a phone-based hybrid machine learning (ML) model combining pre-train convolutional neural network as a feature extractor and support vector regressor as a predictor. To better demonstrate the capacity of the proposed model, two sub-models of the hybrid model and the same hybrid model but with non-pre-train feature extractor are adopted to be benchmarks. The raw field data is acquired from a corridor between the University of Birmingham station and Birmingham International station using phones, subsequently calculated to corresponding comfort level according to UIC 513. The four models are trained by the dataset in two domains – time domain and frequency domain, then optimized by random search and validated by 10-fold cross-validation. The proposed method yields the best performance with an R^2 of 0.988 ± 0.004, a root-mean-square error (RMSE) of 0.028 ± 0.015, and a mean-absolute-error (MAE) of 0.02 ± 0.005. The results of this study underpin the possibility that the railway passenger has the access to quantify the level of comfort and the real-time assistance for the train driver to calibrate the driving style from the proposed system.

INDEX TERMS Machine learning, smart phone, vibration, passenger comfort, crowdsensing.

I. INTRODUCTION

The train has gained popularity in many countries as its high speed and shorter time over air transportation before and after boarding, especially in short or medium distance travel [1]. With the more demanding requirement of passengers, they expect a more accurate timetable, higher speed, and more comfortable service. Passenger comfort is an increasingly notable area that attracts attention from train companies and academia to improve it, making railway transportation more competitive. Román *et al.* analyzed the potential competition of the high-speed train with air transport in the case of Madrid-Barcelona. They unveiled that passenger is willing to pay more if the level of comfort drops [2]. Broadly, passenger comfort is an overall evaluation of the train journey depending on train vibration, temperature, acoustic, lighting, interior design, windows, etc. There is no universal standard to reckon passenger comfort. Section II have summarized three traditional approaches quantifying railway passenger

The associate editor coordinating the review of this manuscript and approving it for publication was Jesus Felez [ID].

comfort. However, we consider that the implementation of the three methods is not easy for everyone.

Our research proposes a smartphone-based hybrid ML model, CNN, to extract informative features from the raw data and SVR to predict passenger comfort using the extracted features. The proposed model promises to enable the train passenger to quantify the comfort level easily and provide the potential to feedback to the train company. There is a potential that the train company can use this cost-efficiency method to monitor the abnormality of passenger comfort.

A. CONTRIBUTIONS

Our primary contribution is adopting the phone-based ML model, providing a handy way to measure passenger comfort. The length of the section is adjustable to fit the specific train lines. UIC 513 suggests that the passenger comfort is calculated every five minutes; however, the time gap between stations is less than five minutes. For instance, the train takes about 2 minutes from the University station to the Fiveways station in the UK. More importantly, this framework can be

implemented to be online assistance to monitor the driver's driving style. A shorter period than five minutes is beneficial to a more explicit monitor.

II. LITERATURE REVIEW

In this section, we have fleshed out the traditional methodologies evaluating the railway passenger comfort to date, the critical status of vibration in railway passenger comfort, and the highlight of the gaps that may occur in the use of machine learning in the context of railway passenger comfort evaluation.

Method 1 - Oboknb and Clarke used a questionnaire to procure how comfortable the passenger felt about the journey [3]. They admitted that many pitfalls could emerge in the interpretation of the data. However, valuable information can still be extracted if a delicate questionnaire and data interpretation are used. Shinkansen has conducted a questionnaire to define the impact of physical attributes such as vibration, acoustics, humidity, lighting, air freshness, air pressure, seat size, etc., on passenger comfort. It was concluded that passenger comfort has 70% correlation with vibration, acoustics, seat size, interior design, and air freshness.

Method 2 - the use of vibration to calculate passenger comfort has been well established in different regions, such as ISO 2631, Sperling's method, and UIC 513. ISO 2631 provides an R.M.S-based approach to evaluate people's whole body vibration [4] in various conditions such as in a building or a car, which has not been adjusted according to train passengers. Sperling's method is more suitable for comparative analysis between vehicles [5], [6]. UIC 513 is dedicated to guiding the train passenger comfort evaluation using train vibration.

Method 3 - some of the studies considered using more features such as ventilation and lighting than vibration [7] to assess the train passenger comfort. Rather than just mechanical attributes, [8] introduced biological parameters such as the variation in heart rate to determine the passenger's comfort. More physical features (noise, air pressure, temperature, humidity, and illumination) and physiological factors (body pressure distribution, Electroencephalography, and Electromyography) were taken into account in [9] to estimate the passenger comfort on a high-speed train. Many pieces of evidence support the use of physiological factors to reflect the state of humans. However, it is not practical and expensive to use these features to estimate passengers' comfort as additional sensors are needed, and it is suspicious to violate the passengers' privacy to obtain the physical feature from passengers. Moreover, there is no universally applicable evaluation standard for the physiological factors.

It is noted that the three existing methods have no potential to evaluate passenger comfort continuously. More importantly, the measurement policy is not user-friendly, which implies that the execution of the measurement has to be done by a specialist. For instance, method 1 needs an appropriate survey. The passenger cannot answer the tedious questions about every train journey; method 2 demands proper set-up

of accelerometers to allow vibration collection; method 3 requires experts to commission vibration measurement.

Most existing research recognizes vibrations' critical role in this comfort level domain. [10] has claimed the vibration relevant ride comfort for a passenger train and the vibration related to a safety concern for a freight train should be anticipated at the origin of the design stage. Other studies like [11] strived to enhance a bogie suspension, hence increase the passenger's comfort; [12] has exerted a numerical model validated with the field test, concluding the rail irregularities with short-wavelength redounds to passenger's comfort; subsequently, [13] has used a magneto-theological (MR) damper to reduce the disutility yielded by the large-amplitude vibration caused by rail irregularities. A full-scale examination with the MR damper was carried out at speeds from 80- 350 km/h, concluding that excellent comfort was performed according to UIC 513. [14] considered there is a strong relationship between the number and the distance of train/bus stops and the comfort level. In [11]–[14] the common motivation for their studies is the importance of vibrations as a better bogie suspension can isolate more vibration to car bodies; irregularities trigger wilder train dynamic that affects the train vibration; closer and more train stops lead to more acceleration/deceleration, subsequently influence the passenger comfort. It is rationale to adopt vibration to calculate the passenger comfort since the literature has well established the application and feasibility of using vibration.

ML techniques have been thrived to process vibration data in various fields such as structural damage detection [15]–[17], human activity recognition [18]–[20], and the prediction of road surface roughness [21]–[23]. These researches agree with the success of the ML techniques in vibration analysis. However, there is little study evaluation of ride comfort using ML in railway vehicles. Azzoug and Kaewunruen proposed an artificial neural network (ANN) to predict ride comfort [24]. There was a limitation that the running vehicle was only allowed to run at 5 - 17 mph. However, the main interest was testing the feasibility of a smartphone phone as a vibration collected tool by a comparative analysis of the data from a sophisticated accelerometer and two smartphones. It was concluded that the phone-based accelerometer is trustworthy. The fully connected ANN has a large number of parameters, making the model time-consumed to train, especially for the vibration dataset that each sample is significant in general. We think that ANN is computational cost to process such data we use. This issue can be more dominant if the trained ML model is implemented to a smartphone. Therefore, a more efficient way is required.

III. PASSENGER COMFORT MEASUREMENT
A. UIC 513 GUIDELINE
In 1988, European Rail Research Institute (ERRI) B153 rolled out a UIC standard to assess the passenger level based on ISO 2631 after decade research. In 1994, the
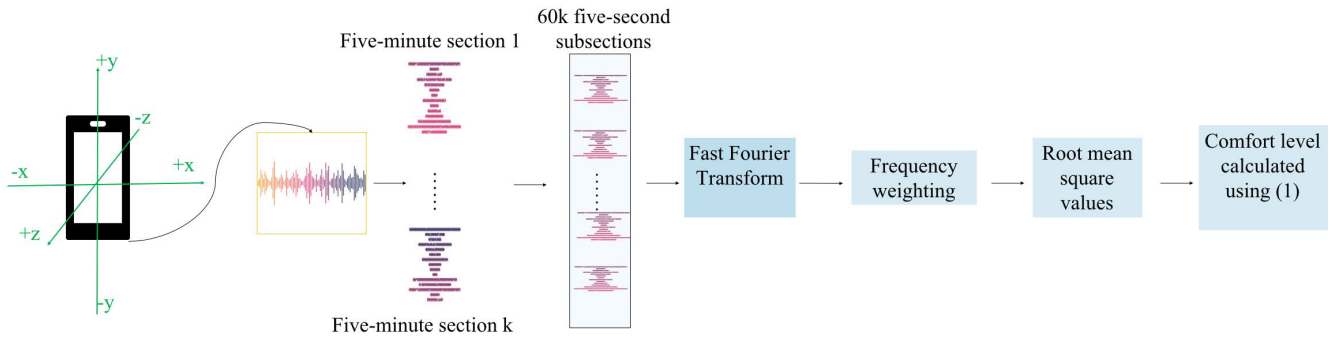
FIGURE 1. Flowchart to calculate the passenger comfort level.

TABLE 1. UIC comfort level.

Comfort level	Interval	Description
1	N<1	Very comfortable
2	1<N<2	Comfortable
3	2<N<4	Moderate comfortable
4	4<N<5	Poor comfortable
5	N>5	Very poor comfortable

standard was rolled out officially, namely UIC 513. This standard is a statistical method that integrates the relationship between vibration and human comfort and the impact of the vibration on ergonomics. UIC 513 is reliable as it focuses on the vibration characteristics, which predominantly affect passenger comfort. A comparative analysis based on ISO 2631 and UIC 513 has been conducted in [25] in a field testing from two corridors, Beijing to Chengdu and Shenzhen to Guangzhou, showing that the two standards have a large overlapped with UIC 513 produces a slightly larger area of uncomfortable than that in ISO 2631. This implies that UIC 513 is more sensitive to the uncomfortable zone.

Fig. 1 provides a flowchart to overview how the data flow from a smartphone to comfort level using UIC 513. UIC 513 specifies that passenger comfort is evaluated in every five-minute section. Each five-minute section is further split into 60 five-second sub-sections, and each subsection is transformed to frequency domain before being weighted. The RMS value of each subsection is calculated. Finally, the confidence probability of 95% is applied to calculate the passenger comfort using (1).

$$N_{mv} = 6 \cdot \sqrt{\left(a_{XP}^w d_{95}\right)^2 + \left(a_{YP}^w d_{95}\right)^2 + \left(a_{ZP}^w b_{95}\right)^2} \quad (1)$$

where:

N_{mv} - the passenger comfort of the simplified measurement.

a - acceleration.

$X, Y,$ and Z - the directions of the acceleration (See Fig. 1).

P - vehicle floor where the acceleration is collected.

W_d, W_b - frequency weighting where b is for Z direction and d is for X and Y.

Subscript 95 - confidence probability.

The ratings of comfort level are tabulated in Table 1. There are five levels from level one to five. UIC 513 recommends
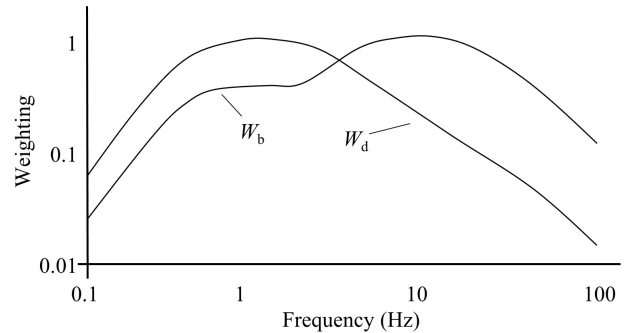
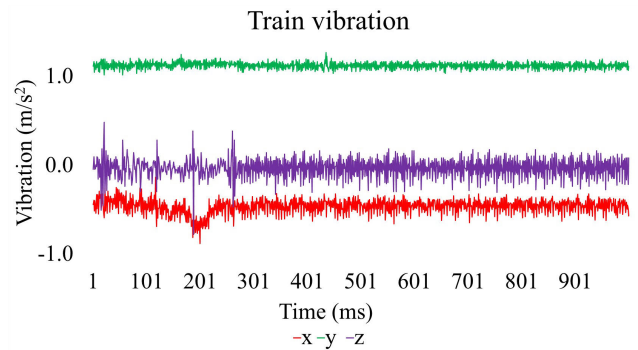

FIGURE 2. Frequency weighting curves.

FIGURE 3. An example of raw data.

the upper limits of the comfort level for rural trains, traditional trains, and posh trains are 4, 3, 2, respectively.

The correlation between the frequency of the vibration and the sensitivity of human beings is complex. To quantify the physical values of vibration to passengers' feelings, the frequency weighting curves in the vertical direction, longitudinal, and horizontal direction are given in Fig. 2. Several weighing factors are up to around one corresponding to some frequencies implying that people are more sensitive to that range of frequencies such as 0.5 - 5 Hz in the longitudinal and horizontal directions, 6 - 10 Hz in the vertical direction.

In our work, we have carried out a data collection with a smartphone on a return trip from University station to Birmingham International station. To make sure the acceleration sensed by the phone is only from the train, the

phone is asked to remain stable on the floor with no relative movement with the running train. A mobile app "Phyphox" is installed on the phone to allow access to the accelerometer embedded in the phone, visualizing and saving the vibrations. Fig. 3 provides an example of the raw data, showing the vibrations in horizontal and longitudinal directions are larger than those in the vertical direction.

As aforementioned, UIC 513 evaluates passenger comfort every five minutes. However, we use a slightly different scenario from UIC 513 since we promise to enable passengers to quantify the comfort level with their phones, which means that the passengers will need to keep their phones with no movement in five minutes. It is not practical for passengers to place their phones on the floor doing nothing for five minutes. More importantly, to estimate the passenger comfort in the section shorter than five minutes, we decide to evaluate the comfort level on a more detailed scale of five seconds, the so-called instance comfort level.

IV. HYBRID MODEL DEVELOPMENT

In this section, we discuss the details of how we configure the hybrid model, hyperparameter optimization, and 10-fold cross-validation

A. DATA PRE-PROCESSING

There are two datasets used: time-domain and frequency domain. The two datasets are processed in the same way. To predict passenger comfort in five seconds, the raw dataset is sliced into samples, and the first sample X_0 is given by

$$X_0 = \begin{bmatrix} a_1^x & a_1^y & a_1^z \\ \vdots & \vdots & \vdots \\ a_{500}^x & a_{500}^y & a_{500}^z \end{bmatrix} \quad (2)$$

where:

X_o: the first sample

a: acceleration

x, y, and z: the three directions shown in Fig. 1.

The subscripts 1 - 500: time steps (100 Hz sample rate is set, producing 500 timesteps in five seconds.

After the raw data is sliced into five-section windows, the dataset is split into ten elements with ten equal sizes. As 10-fold cross-validation is used, every nine elements are used to train the model, and left one-fold is used to validate the model. More details on 10-fold validation are discussed in Section IV (D). It is worth knowing that the validation set should hold-out and with no touch before the model is selected. Therefore, when applying the Min-Max normalization subsequently, only the information from the training set is employed to normalize the validation set.

B. HYBRID MODEL

CNN is initially designed for classifying images. However, we still use CNN in our work for two reasons: (1) the principle of CNN provides a more efficient way than ANN. This allows us to save computational costs when the proposed model is applied to phones. (2) CNN can be an auto feature extractor to avoid the usage of handcrafted features, which might introduce human bias if the dataset cannot be understood thoroughly.

Extensive research has shown that CNN has presented promising results in a variety of fields as sentences modelling [26], heartbeats classification [27], and a CNN for image classification dedicated to phone devices with significant computational cost reduction [28]. A substantial advantage of CNN is the reduction of parameters compared to ANN, which allows a more extensive and deeper model for a more complex task. To receive a five-second sample, one node of ANN produces 1,500 weights. One node might not be capable enough to achieve the required mission so that we might need to increase the number of nodes that considerably scale up the size of weights. This situation can become more dominant if the passenger comfort for a more extended period is estimated using ANN as the size of each sample is growing. Another advantage of using CNN is that it automatically allows the model to obtain abstract features. For instance, CNN can detect the edge and the simple shapes of the image in the first layers. The more detailed features can be extracted in the fowling layers. Our project can also introduce handcrafted features if the handcrafted features can be pre-defined by human experts. However, it is hard to cover all the features essential to the required task and challenging to avoid bias. UIC 513 infers passenger level based on the intensity of the vibration, which reflects that the peak of the vibration plays an important role. However, we should also consider the intensity at different frequencies as passengers can percept differently even though the intensity of the vibration is the same but at different frequencies.

In summary, it is sensible to use CNN as a feature extractor when the handcrafted features are hard to define. Fig. 4 presents how the data flow in the hybrid model with two sub-models: CNN and SVR. As shown in Fig. 4(a), CNN outputs the intermediate features before the prediction layer to the SVR predictor. Unlike some researches which replaced the prediction layer of CNN with SVR as shown in Fig. 4(b), such as [29] to predict guide RNA activity, [30] to predict wastewater index, and [31] to predict short-term traffic flow. These works concluded that the hybrid model performs better than either only CNN or SVR. However, the scenario in Fig. 4(b) does not present a satisfactory result in our work. The outcomes that compare the impact of the with and without CNN prediction layer are shown in SECTION V.

C. HYPER-PARAMETERS TUNING

Zhang and Wallace have carried out an extensive experiment on the effect of each CNN's hyperparameter for sentence classification [32]. Although the goal was different from ours, we can still use it as a guideline as [32] has included almost all the hyper-parameters of CNN. The tuned parameters are listed in Table 2, consisting of the boundary for each hyperparameter used in the proposed model. Random search, a strategy for hyper-parameter optimization, is dedicated to
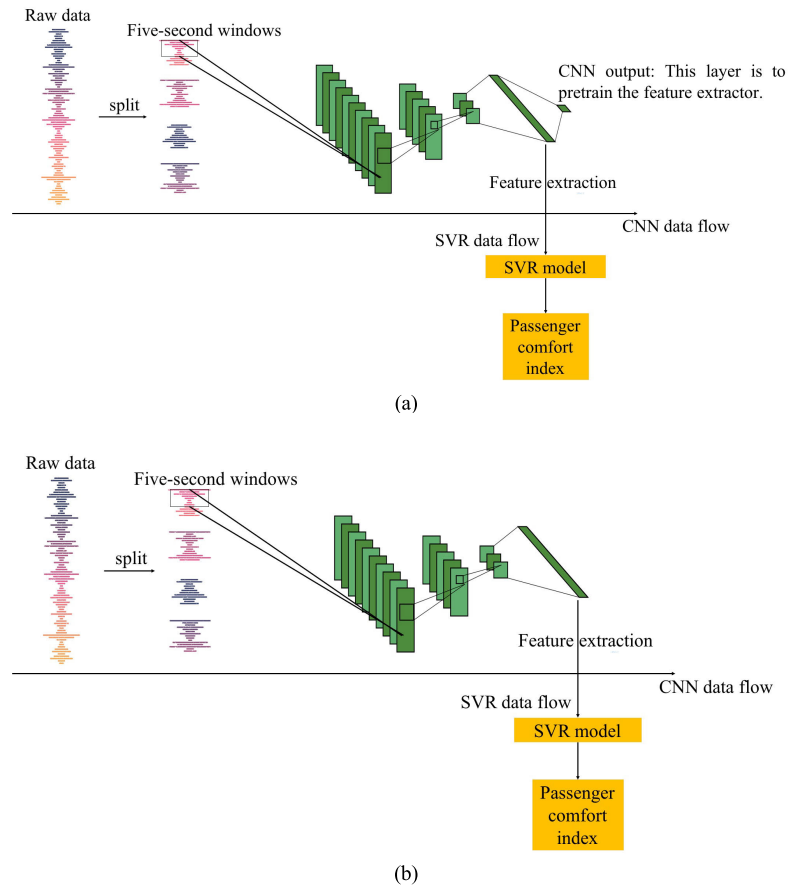
(a)

(b)

FIGURE 4. Two scenarios to implement feature extractor (a) pre-train feature extractor and (b) non-pre-train feature extractor.

TABLE 2. Hyper-parameter for the hybrid model.

Hyper-parameter	Boundary
The no. of convolutional layers	1 - 6
Filter size	1 - 15
The No. of filters	(32, 64, 128, 256, 512)
Activation function	Relu, tanh, sigmoid
Pooling strategy	Average pooling, max pooling
The No. of pooling layer	1 - 6
The No. of pooling size	1 -15
The No. of dense layer	1-3
Dense size	10-700
Regularization	$10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$
Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5, 0.6
Loss function	Mean square error, mean absolute error
Optimizer	Stochastic gradient descent, Adam
Momentum	0.7, 0.8, 0.9
Learning rate	$10^{-3}, 10^{-4}, 10^{-5}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}$
Weight initializer	Glorot uniform, he_uniform, he_normal, ramdom_normal
kernel	Linear, poly, rbf
Regularization parameter C	1-1000

searching the hyper-parameters with optimal scores [33]. It is noted that the dropout [34] and the regularization [35] are used to prevent overfitting. Learning rate decay by epochs is adopted to speed up the training process given by (3).

$$decay\ rate = \frac{learning\ rate}{epochs} \quad (3)$$

D. K-FOLD CROSS-VALIDATION AND EVALUATION MEASUREMENT

It is easy to yield an overoptimistic outcome if the same dataset was used to train and evaluate a model as early as 1930 [36]. This issue has been addressed by cross-validation estimating the model's performance with a new dataset [37].

To evaluate a model rigorously and minimize the bias related to randomly sampling, the way to split the data has been a major interest as the limited amount of data in practice. Our study introduces k-fold cross-validation to perform model selection and compare different learning algorithms. K-fold cross-validation splits the dataset into k segments with equal size or approximately equal size where *k-1* components are used to train the model. The kth element is employed to test the model performance [38]. The k is set to ten as [39] concluded *k = 10* is sensible if we aim to measure the model's error.

We adopt three indicators, the coefficient of determination R^2, root mean square error (RMSE), and mean absolute error (MAE). The definition of R^2 makes it useful to estimate the success of a model predicting the dependent variable from the independent variables [40]. Both RMSE and MAE are widely used to measure the average performance of an ML model, although there are arguments between two widely cited papers [41], [42]. The dispute is not our interest, so that we decide to deploy both of the two measures to include their advantages. The three metrics are given by:

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (p_i - y_i)^2}{\sum_{i=1}^{m} (y_i - \bar{y})^2} \qquad (4)$$

$$MAE = \frac{\sum_{i=1}^{m} |p_i - y_i|}{m} \qquad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m} (p_i - y_i)^2}{m}} \qquad (6)$$

where p_i and y_i are the predictions and the actual values; $\bar{y}$ is the mean of the label values in the validation set; m is the number of samples in the validation set.

V. RESULTS AND DISCUSSION

In this section, we perform the results produced by four models (CNN, SVR, pre-train CNN + SVR, non-pre-train CNN + SVR) based on two kinds of the dataset (time domain and frequency domain). This section commences with Table 3 presenting the optimal hyper-parameters searched by random search. The results follow the order: CNN, SVR, pre-train CNN + SVR, and non-pre-train CNN + SVR with time-domain data and frequency domain subsequently.

Fig. 5(a-c) plots the three measures R^2, MAE, and RMSE using time-domain data to compare the different performances between each fold. It is very apparent from Fig. 5(a-c) that the pre-train CNN + SVR model has gained the highest R^2, lowest MAE, and RMSE. The pre-train CNN + SVR model's R^2 appears to fluctuate mildly, unveiling a 0.04 standard deviation as shown in Table 4. The most interesting aspect in Fig. 5(b) and Fig. 5(c) is that the non-pre-train CNN + SVR procures a smaller MAE but larger RMSE than those of SVR in the fifth fold. It is due to the RMSE squares the error assigning a bigger weight to more considerable errors. This implies that the non-pre-train CNN + SVR produces some significant error points in the fifth fold. Although it has been widely argued that either

TABLE 3. Optimal hyper-parameters.

Hyper-parameter	Boundary
The no. of convolutional layers	4
Filter size	5,5,5,3 (layer1 to layer 4)
The no. of filters	64,64,64,512 (layer1 to layer 4)
Activation function	Relu
Pooling strategy	max pooling
The no. of pooling layer	4
The no. of pooling size	4,4,3,2 (layer1 to layer 4)
The no. of dense layer	1
Dense size	512
Regularization	10^{-3}
Dropout rate	0.4
Loss function	Mean square error
Optimizer	Adam
Momentum	Not applicable for Adam
Learning rate	1e-5
Weight initializer	he_normal
Kernel	linear
C	3

MAE or RMSE should be used [41], [42], it is sensible to use both RMSE and MAE as biases can emerge in certain conditions, such as the circumstance just mentioned.

Further analysis of the four models' performance using frequency domain data is provided in Fig. 5(d-f). It is noted that there is no performance for SVR due to the unavailable using a complex number. In the three subfigures Fig. 5(d-f), the pre-train CNN + SVR model gives the highest R^2 and lowest MAE and RMSE in every fold. It is concluded that the pre-trained CNN + SVR using the dataset in frequency domain outperforms all other scenarios, which is further proven by Table 4 providing the overall performance.

The results, calculated by the mean and the standard deviation in ten folds in the form of $\mu \pm \sigma$ as shown in Table 4 [43], indicate the overall performance comparison. The proposed model illustrates a significantly large average R^2 (0.988) than others and the slightest standard deviation (0.004), yielding a coefficient of variation of 0.4%. The proposed model is the best in terms of high performance and satisfactory stability. The average value for MAE and RMSE is 0.02 and 0.028, yielded by the proposed model, both indicating that the prediction error is minor. More details at the sample level can be seen from Fig. 6 in the appendix. Most of the actual values stay between the region from 0 to 1, which reveals the tested train line section is at an excellent comfort level, referring to Table 1. It is worth knowing that our work uses a regression model that predicts the comfort level's exact value, not the interval of the comfort level presented in Table 1. If we now turn to Fig. 6(i) - fold 9, the predictions at the two peaks offer an apparent gap from the two actual values; however, the true values and the predictions remain in the same comfort interval. It can be observed that the model's performance can be enhanced if the label of the dataset is transformed to the interval of the comfort level.

The three models, the non-pre-train CNN + SVR, SVR, and CNN, are also tuned by the same method random search. The difference between CNN and the proposed method is the prediction layer. The best R^2 CNN can perform is 0.706 ± 0.129. The reason can be that the prediction layer of the
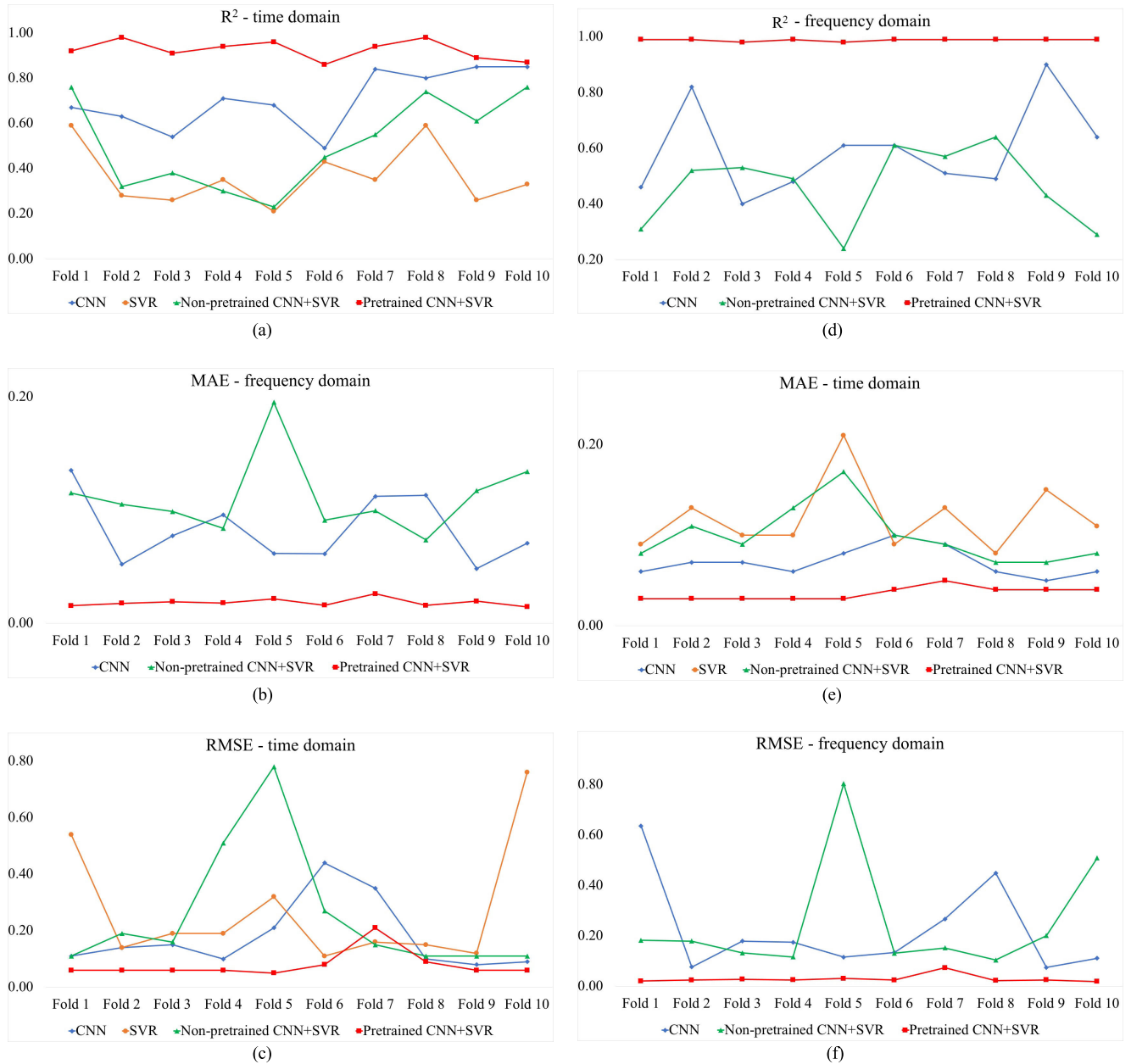
FIGURE 5. R^2, MAE, and RMSE for four models using time-domain data (a-b) and frequency-domain data (d-f).

TABLE 4. Overall results.

Data domain	R^2 Time	R^2 Frequency	MAE Time	MAE Frequency	RMSE Time	RMSE Frequency
CNN	0.706 ± 0.129	0.592 ± 0.161	0.07 ± 0.02	0.083 ± 0.031	0.177 ± 0.123	0.222 ± 0.185
SVR	0.365 ± 0.134	N/A	0.119 ± 0.04	N/A	0.268 ± 0.216	N/A
Non-pre-train CNN+SVR	0.51 ± 0.202	0.463 ± 00.14	0.099 ± 0.03	0.112 ± 0.036	0.250 ± 0.223	0.250 ± 0.226
Pre-train CNN+SVR	0.925 ± 0.043	0.988 ± 0.004	0.036 ± 0.007	0.02 ± 0.005	0.079 ± 0.047	0.028 ± 0.015

proposed model is SVR which is more potent than a dense layer used by CNN. The excellent generalization capability and accuracy that SVR can achieve have been proven [44].

The effectiveness of CNN to be a feature extractor can be seen from the comparison of non-pre-train CNN + SVR and SVR. Around a 10% increase of R^2 can be found from

SVR to non-pre-train CNN + SVR. A leap of R^2 has been observed from non-pre-train CNN + SVR to pre-train CNN + SVR. An explanation for this might be that the feature extractor shown in Fig. 4(a) has been trained to minimize the loss between the labels and the predictions during the CNN data flow. The features extracted by the pre-trained feature
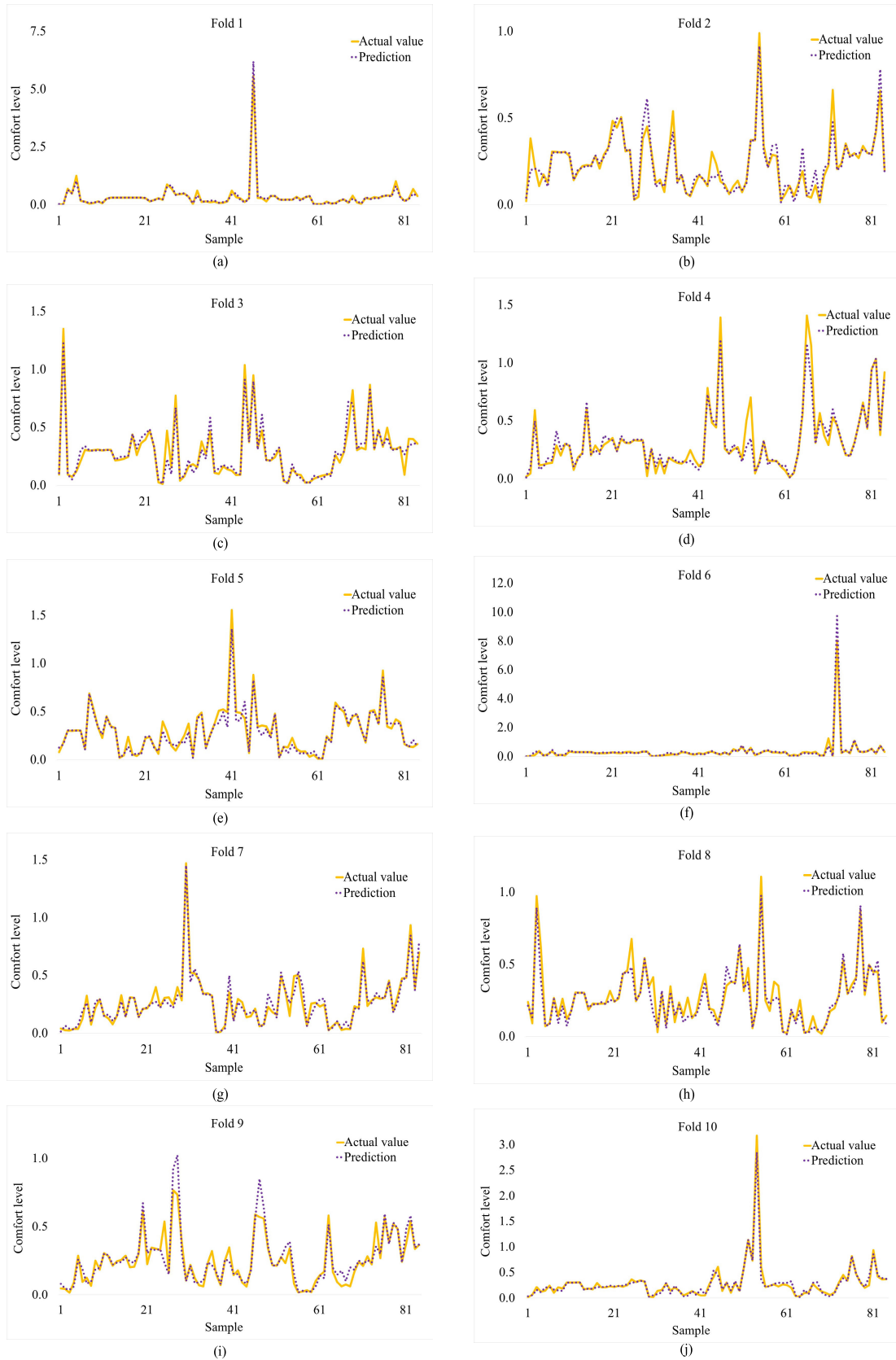
FIGURE 6. The actual values and predictions correspond to samples in ten folds. (a)-(j) are fold one to fold ten, respectively.

extractor are more informative than those extracted by a non-pre-train feature extractor shown in Fig. 4(b).

As mentioned in the literature review, the questionnaire needs massive effort from train crews and passengers, which may thwart the long-term implementation. Besides, it is challenging to get away from the false interpretation of the information collected. In contrast, our system is less likely to suffer from the effort and time consumed. Besides, the guideline we used, UIC 513, devotes itself to the vibration, avoiding subjective judgments in the questionnaire. Our implementation contrasts with that of method 2 discussed in the literature review requires an accelerometer to sense the train vibration and a computer to visualize and save vibrations. Smartphones provide us with a compact tool to obtain train vibrations. The results shown in this work also prove the feasibility of the proposed system at a higher speed than that of [24] conducted the field test at 5 – 17 mph.

VI. CONCLUSION

This study set out to measure train passenger comfort level using a phone-based machine learning model. We adopt a hybrid model combining CNN and SVR. Pre-train CNN extracts informative features, and SVR predicts the passenger comfort using the features. With a hyperparameter optimization method random search, optimal results are gained to four models. It has been inferred that the hybrid model shows superior performance than one of the sub-models. The most critical finding reveals that a pre-train feature extractor outperforms the non-pre-train one.

The finding will be of interest to train companies keen to improve the comfort level. The proposed solution can be integrated easily into an onboard driving-aided system that provides real-time feedback to the driver on how the current driving style impacts passenger comfort. The insights gained from this study may assist passengers with a simple way to quantify the level of comfort and engineers a straight ward way to calculate the comfort index as the traditional way is more complicated. A limitation of this study is that most of the actual values in the dataset fall into the interval from 0 - 1, as can be seen from Fig. 6. It is noticeable that this study is limited by the absence of more datasets with uncomfortable intervals. Despite its limitation, the proposed model presents a satisfactory result for the limited uncomfortable samples. The major effort we can put into making the result more convincing is using 10-fold cross-validation. Therefore, a larger number of field data with various comfort intervals could also be conducted to further determine the proposed model's effectiveness.

APPENDIX
See Fig. 6.

REFERENCES

[1] P. F. P. da Silva and J. Mendes, "Passengers comfort perception and demands on railway vehicles: A review," *KNE Eng.*, vol. 4, pp. 257–270, Jun. 2020.

[2] C. Román, R. Espino, and J. C. Martín, "Competition of high-speed train with air transport: The case of Madrid–Barcelona," *J. Air Transp. Manage.*, vol. 13, no. 5, pp. 277–284, Sep. 2007.

[3] D. J. Oboknb and M. J. Clarke, "The development of questionnaire surveys for the investigation of passenger comfort," *Ergonomics*, vol. 16, no. 6, pp. 855–869, Nov. 1973.

[4] *Mechanical Vibration and Shock-Evaluation of Human Exposure to Whole-Body Vibration-Part 1: General Requirements*, Standard ISO 2631-1, Geneva, Switzerland, 1997.

[5] K. Gangadharan, C. Sujatha, and V. Ramamurti, "Experimental and analytical ride comfort evaluation of a railway coach," in *Proc. SEM ORG Conf.*, 2004, pp. 1–15.

[6] R. C. Sharma, "Evaluation of passenger ride comfort of Indian rail and road vehicles with ISO 2631-1 standards: Part 2-simulation," *Int. J. Veh. Struct. Syst.*, vol. 8, no. 1, pp. 1–10, Apr. 2016.

[7] J. Winter, I. Windemut, N. Kevlishvili, D. Schmeling, and J. Maier, "Passenger comfort for the next generation train: Subject trials for ventilation and lighting," in *Proc. 3rd Int. Conf. Railway Technol., Res., Develop. Maintenance*, Stirlingshire, U.K., 2016, pp. 1–12.

[8] Y. Lee, K. Shin, S. Lee, Y. Song, S. Han, and M. Lee, "Ride comfort analysis with physiological parameters for an E-health train," *Telemed. E-Health*, vol. 15, no. 10, pp. 1010–1021, Dec. 2009.

[9] Z. Tang and J. Xu, "Research on the physiological and psychological factors' evaluation index framework of train ride comfort," in *Proc. IEEE Conf. Anthol.*, Jan. 2013, pp. 1–4.

[10] W.-S. Yoo, C.-H. Lee, W.-B. Jeong, and S.-H. Kim, "Development and application of new evaluation system for ride comfort and vibration on railway vehicles," *J. Mech. Sci. Technol.*, vol. 19, no. 7, pp. 1469–1477, Jul. 2005.

[11] C. M. Lee, V. N. Goverdovskiy, C. S. Sim, and J. H. Lee, "Ride comfort of a high-speed train through the structural upgrade of a bogie suspension," *J. Sound Vib.*, vol. 361, pp. 99–107, Jan. 2016.

[12] J. Sadeghi, S. Rabiee, and A. Khajehdezfuly, "Effect of rail irregularities on ride comfort of train moving over ballast-less tracks," *Int. J. Struct. Stability Dyn.*, vol. 19, no. 6, Jun. 2019, Art. no. 1950060.

[13] S. S. Alehashem, Y. Ni, and X. Liu, "A full-scale experimental investigation on ride comfort and rolling motion of high-speed train equipped with MR dampers," *IEEE Access*, vol. 9, pp. 118113–118123, 2021.

[14] L. Amador-Jimenez and A. Christopher, "A comfort index for public transportation: Case study of Montreal," in *Proc. IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Aug. 2016, pp. 3–7.

[15] P. Cawley and R. D. Adams, "The location of defects in structures from measurements of natural frequencies," *J. Strain Anal. Eng. Des.*, vol. 14, no. 2, pp. 49–57, Apr. 1979.

[16] C. G. Koh, L. Wu, and C. Liaw, "Distributed GA for large system identification problems," *Proc. SPIE Smart Nondestruct. Eval. Health Monitor. Struct. Biol. Syst.*, vol. 4702, pp. 438–445, May 2002.

[17] J. Ko, Y. Ni, X. Zhou, and J. Wang, "Structural damage alarming in Ting Kau Bridge using auto-associative neural networks," in *Proc. Int. Conf. Adv. Struct. Dyn.*, 2000, pp. 1021–1028.

[18] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.* Berlin, Germany: Springer, 2011, pp. 289–296.

[19] A. Bayat, M. Pomplun, and D. A. Tran, "A study on human activity recognition using accelerometer data from smartphones," *Proc. Comput. Sci.*, vol. 34, pp. 450–457, Jan. 2014.

[20] I. Andrey, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Appl. Soft Comput.*, vol. 62, pp. 915–922, Jan. 2017.

[21] M. Elangovan, N. Sakthivel, S. Saravanamurugan, B. B. Nair, and V. Sugumaran, "Machine learning approach to the prediction of surface roughness using statistical features of vibration signal acquired in turning," *Proc. Comput. Sci.*, vol. 50, pp. 282–288, Jan. 2015.

[22] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: Using a mobile sensor network for road surface monitoring," in *Proc. 6th Int. Conf. Mobile Syst., Appl., Services*, 2008, pp. 29–39.

[23] M. Perttunen, "Distributed road surface condition monitoring using mobile phones," in *Proc. Int. Conf. Ubiquitous Intell. Comput.* Berlin, Germany: Springer, 2011, pp. 64–78.

[24] A. Azzoug and S. Kaewunruen, "RideComfort: A development of crowdsourcing smartphones in measuring train ride quality," *Frontiers Built Environ.*, vol. 3, p. 3, Jan. 2017.

[25] T. Ding, J. Lin, and X. Chen, "Comfort evaluation and analysis of high-speed train," *J. Phys., Conf. Ser.*, vol. 1986, no. 1, Aug. 2021, Art. no. 012089.

[26] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," 2014, *arXiv:1404.2188*.

[27] U. Acharya, S. Oh, Y. Hagiwara, and J. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.

[28] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2018, pp. 6848–6856.

[29] G. Zhang, Z. Dai, and X. Dai, "A novel hybrid CNN-SVR for CRISPR/Cas9 guide RNA activity prediction," *Frontiers Genet.*, vol. 10, p. 1303, Jan. 2020.

[30] W. Fan and Z. Zhang, "A CNN-SVR hybrid prediction model for wastewater index measurement," in *Proc. 2nd Int. Conf. Adv. Comput. Technol., Inf. Sci. Commun. (CTISC)*, Mar. 2020, pp. 90–94.

[31] W. H. Luo, B. T. Dong, and Z. S. Wang, "Short-term traffic flow prediction based on CNN-SVR hybrid deep learning model," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 17, no. 5, pp. 68–74, Oct. 2017.

[32] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," 2015, *arXiv:1510.03820*.

[33] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 2, pp. 281–305, 2012.

[34] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[35] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, Feb. 2019, Art. no. 022022.

[36] S. C. Larson, "The shrinkage of the coefficient of multiple correlation," *J. Educ. Psychol.*, vol. 22, no. 1, p. 45, 1931.

[37] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc., B, Methodol.*, vol. 36, no. 2, pp. 111–133, Jan. 1974.

[38] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," *Encyclopedia Database Syst.*, vol. 5, pp. 532–538, Jan. 2009.

[39] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of K-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010.

[40] N. J. D. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, Sep. 1991.

[41] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)–arguments against avoiding RMSE in the literature," *Geosci. Model. Develop.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014.

[42] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Res.*, vol. 30, no. 1, pp. 79–82, Dec. 2005.

[43] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statist. Comput.*, vol. 21, no. 2, pp. 137–146, Apr. 2011.

[44] M. Awad and R. Khanna, "Support vector regression," in *Efficient Learning Machines*. Berlin, Germany: Springer, 2015, pp. 67–80.

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., where he is currently pursuing the Ph.D. degree. His research interest includes artificial intelligent applied in railway.

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 currently a Chartered Engineer. He has over 400 technical publications.

• • •