Applied Energy 333 (2023) 120608

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Development of surrogate models for evaluating energy transfer quality of high-speed railway pantograph-catenary system using physics-based model and machine learning

Guizao Huang, Guangning Wu *, Zefeng Yang, Xing Chen, Wenfu Wei

School of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China

HIGHLIGHTS

- Evaluation of energy transfer quality informs high-speed train running safety.
- Dataset is generated via the physics-based model followed by feature extraction.
- We used and compared 5 classification algorithms and 8 regression methods.
- GBDT is optimal algorithm for classification model and MLF-DNN for regression.
- The established surrogate models are expected to replace the traditional models.

ARTICLE INFO

Keywords:
Surrogate model
Machine learning
Physics-based model
Pantograph-catenary system
Energy transfer
Classification and regression

ABSTRACT

High-speed railway pantograph-catenary system is the only energy transfer pathway to drive a train operation. Energy transfer quality deteriorates with the increasing train speed and harsh service environment, thereby quickly and accurately evaluating the energy transfer quality is very important to guarantee the normal operation of a train. In this study, firstly, the physics-based model to simulate the dynamic interaction of pantograph-catenary system is established and validated. Eleven input parameters involve the essential line design and train operation parameters, and the output parameters that are crucially responsible for energy transfer quality are obtained by feature extraction. Secondly, a sampling strategy is employed to construct the input sampling points, based on which the outputs are computed via physics-based model, then combining them the dataset is obtained. Thirdly, five tree-based classification surrogate models are developed and compared to assess the level of energy transfer quality. Finally, eight regression surrogate models are developed in replacing physics-based model to evaluate the essential values of energy transfer quality. It is found that the gradient boosting decision tree (GBDT)-based surrogate model is the optimal classification model and the multi-layer feed-forward deep neural network (MLF-DNN)-based surrogate model for the optimal regression model. The two surrogate models are expected to quickly find the optimal design parameters and improve the operation control of trains of high-speed railway for the purpose of enhancing the energy transfer quality if coupled with optimization procedure.

1. Introduction

1.1. Background

Global warming is getting worse due to the increasing atmosphere environmental pollution. Facing this threat, low carbon economy has become the trend of international development. As one of the clean energy sources, electricity has been applied to various fields. Electrified high-speed railways have developed rapidly in recent decades because of not only their huge capacity to carry people and goods but also the environmental protection with no carbon emissions [1]. By 2021 in China, the length of high-speed railways (>200 km/h) in operation had exceeded 40,000 km, which is making a tremendous contribution to the carbon emission reduction of the world.

With the train speed continues to increase, the tractive energy power

* Corresponding author.
E-mail addresses: gzhuang@swjtu.edu.cn (G. Huang), gnwu@swjtu.edu.cn (G. Wu), yangzefeng@foxmail.com (Z. Yang), 742765770@qq.com (X. Chen), wfwei@home.swjtu.edu.cn (W. Wei).

https://doi.org/10.1016/j.apenergy.2022.120608
Received 21 August 2022; Received in revised form 29 November 2022; Accepted 28 December 2022
Available online 4 January 2023
0306-2619/

Nomenclature

L	Span length of catenary
N	Number of droppers in a span
R_{sag}	Rate between pre-sag of contact wire and span length
Sv	Stagger value
Ty_{cw}	Type of contact wire
T_{cw}	Tension in contact wire
Ty_{mw}	Type of messenger wire
T_{mw}	Tension in messenger wire
D	Vertical distance between contact wire and messenger wire at the supports
v	Running speed of a train
F_{uplift}	Vertical force exerted upward by the pantograph on the contact wire
F_m	Mean contact force
σ	Standard deviation of contact force
F_{max}	Maximum contact force
F_{min}	Minimum contact force
U_s	Maximum contact wire uplift at support
NQ	Percentage of arcing
R^2	Coefficient of determination
MAE	Mean absolute error
MSE	Mean squared error

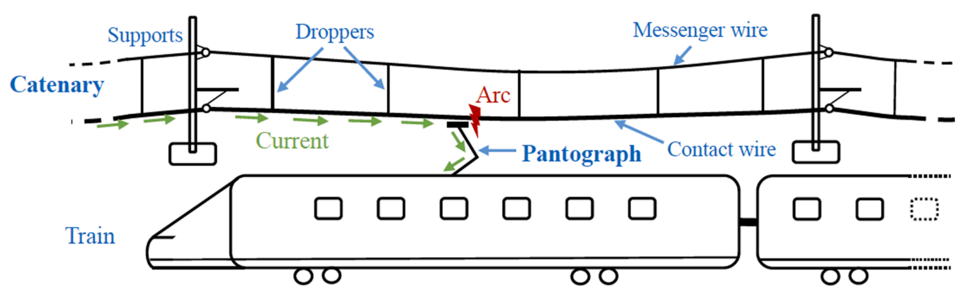

Fig. 1. Energy transfer via pantograph-catenary system.

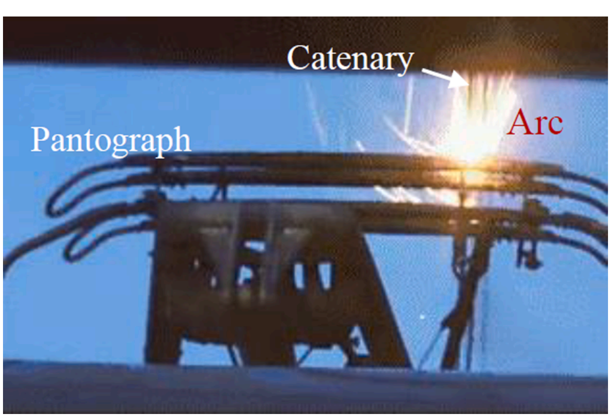

Fig. 2. Electric arc of pantograph-catenary system.

to drive a train has become increasing. For example, the tractive power of China CRH380A train in continuous operation at 350 km/h and maximum operation at 380 km/h is up to 9.8 MW. As is well known that the high-capacity energy is transmitted through the overhead catenary located upward the train. When a train is in operation, the pantograph installed on the train body roof is requested to uplift and contact with the catenary to receive energy to drive the train, as is shown in Fig. 1. Therefore, the contact quality between pantograph and catenary is crucially responsible for the high-capacity energy transfer.

Contact force is the most essential parameter to assess the contact quality, and has been revealed to be predominant in the evaluation of the energy transfer quality in the design of high speed railway overhead lines [2]. On the one hand, too large contact force will contribute to the drastic mechanical wear for pantograph strip and contact line, which reduces their service life and even damage themselves. On the other hand, too small or even vanished contact force will produce the electric arc between contact line and pantograph strip as shown in Fig. 2, which deteriorates the energy transfer quality and even breaks off the power transfer meanwhile brings noise and light pollution. Therefore, the maintenance of a proper and stable contact force plays an important role for the energy transfer quality.

1.2. Literature review

During the last several decades, great efforts have been made in studying the variation of contact force to evaluate the energy transfer quality of pantograph-catenary system by means of field measurements and numerical computations. Take a Norwegian contact line system including a 1500-m straight section and a 1301-m curved section as an example, the pantograph-catenary dynamic interaction was assessed to explore higher train velocities [3]. In [4], contact force time series were obtained via field measurement using wireless accelerometers and their influence on energy transfer was analyzed. Based on finite element method (FEM) and wave propagation analysis, a physics-based model for simulating pantograph-catenary dynamic interaction was established, and effective measures to improve current collection quality for double pantographs-catenary system were proposed [5]. Waves, modes and properties of catenary impacted by pantograph-catenary interaction were analyzed by means of spectral analysis for contact force variations [6]. To investigate the influence of crosswinds [7], ice coating [8], irregularity of track [9] and contact wire [10] on energy transfer quality, different physics-based models via FEM were set up to analyze the variations of contact force. Nevertheless, field measurement-based approaches are expensive and not easy for extensive analysis, and physics-based models computing (e.g. via FEM) is much costly.

To improve the computation efficiency, a moving mesh method integrated into a general FEM was developed to deal with cable structure subjected to moving loads [11]. The developed physics-based model was applied to simulate pantograph-catenary dynamic interaction, with more than 2.5 times faster for an equivalent accuracy than the traditional method. The method was also used in [12] to investigate the variation of contact force, and the simulation time cost is reduced by over 60 %. However, the improvement of the computation efficiency on the basis of physics-based model is still limited, hard to apply to the fast design of catenary structure and real-time control of operation status of trains.

In recent years, machine learning-based surrogate model is widely

Table 1
Input parameters in the physics-based model as well as surrogate model.

Parameter type	Symbol	Unit	Description
Design parameters of the system	L	m	Span length
	N	–	Number of droppers
	R_{sag}	‰	Ratio between contact wire sag at the middle span and L
	Sv	m	Stagger value
	D	m	Vertical distance between contact wire and messenger wire at the supports
	T_{cw}	kN	Tension in contact wire
	T_{mw}	kN	Tension in messenger wire
	Ty_{cw}	mm^2	Contact wire type (defined with cross-section area)
	Ty_{mw}	mm^2	Messenger wire type (defined with cross-section area)
Operation parameters of train	v	km/h	Train speed
	F_{uplift}	N	Vertical force exerted upward by the pantograph on contact wire at standstill

used to study complex, real world phenomena or solve challenging design problems. In many fields there is great interest in tools and techniques that facilitate the construction of surrogate models, while minimizing the computational cost and maximizing model accuracy [13]. For example, in the field of sustainable energy systems, five comprehensive artificial neural network surrogate models for the Integrated Regenerative Methanol Transcritical Cycle were developed, with the improvement of computing time by 95 % compared to physics-based model [14]. Aspect of environmental pollutants, different machine learning and data processing methods were used to establish surrogate models in evaluating the energy efficiency parameters of the electric arc furnace process, aiming to reduce the pollution of electric arc [15]. In addition, in the areas of energy conversion and conservation, twelve data-driven models to predict building thermal load were developed and compared with the data collected from a campus building, and their models were expected to apply to three implications for practice [16]. Combined dynamic programming and machine learning, the economic predictive control model for combined cooling, heating and power system management was established, and its convergence time was reduced by 93 % [17]. Therefore, surrogate models are compact and cheap to evaluate, and have proven very useful for tasks in terms of optimization, design space exploration and sensitivity analysis [13]. Similarity, it is possible using machine learning-based surrogate model to evaluate the energy transfer quality of high-speed railway pantograph-catenary system.

1.3. Contribution and organization of this paper

In the study of energy transfer quality of pantograph-catenary system, using physics-based model to evaluate the energy transfer quality is a mature and widely adopted method [4,18–20]. However, limited to the deficiency of computing efficiency of physics-based model, it is difficult to be applied on some high requirement of real-time scenarios. Different with the previous studies, this paper tries to develop an optimal classification surrogate model to assess the level of energy transfer quality and an optimal regression surrogate model to quantitatively evaluate the energy transfer quality. The established surrogate models are expected to replace the traditional physics-base models or even field test for evaluating the energy transfer quality. The results of this study can be extended to applications such as the preliminary designs of railway line and real-time evaluation and control on the operation of trains. The major contributions of this paper can be summarized as follows:

(1) A physics-based model is proposed and validated to collect data.
(2) Five tree-based classification surrogate models are developed and compared to assess the level of energy transfer quality of pantograph-catenary system.
(3) Eight machine learning algorithms are employed to develop the regression surrogate models aiming to quantitatively evaluate the essential values of energy transfer quality.

In Section 2, the physics-based model to simulate the pantograph-catenary interaction of high-speed railway was established, and its verification and accuracy was validated with the international benchmark. In Section 3, the dataset was collected via the physics-based model, and different machine learning algorithms to carry out surrogate models were briefly introduced. In Section 4, five tree-based classification surrogate models were established to assess the level of energy transfer quality, and in Section 5 eight regression surrogate models were developed to evaluate the values of energy transfer quality. Potential applications of the surrogate models are introduced briefly in Section 6 and the conclusions are drawn in Section 7.

2. Physics-based model of pantograph-catenary system

For evaluating the energy transfer quality of pantograph-catenary system of high-speed railway, the physics-based model is established by means of FEM and will be discussed in Section 2.1. It is worth mentioning that the purpose of the physics-based model is to collect the data. The input and output parameters of physics-based model as well as surrogate model will be discussed in Section 2.2 and Section 2.3, respectively. The validation of physics-based model will be introduced in Section 2.4.

2.1. Pantograph and catenary model

The physics-based model of pantograph-catenary system was introduced in detail in many previous studies, such as Refs. [4,19,20]. Here, the fundamental details of the model are introduced. For pantograph, it is a complex structure installed on the train body roof. To reduce the computing complexity, a pantograph is usually simplified as a discrete mass-spring-damper model with three lumped masses in terms of studying the pantograph-catenary interaction [18]. The physical parameters in [18] are adopted in the work. Moreover, the lumped masses are modelled with mass element, and the stiffness between lumped masses is modelled with spring element.

The catenary consists of three main parts, as shown in Fig. 1. First one is the contact wire, which is used to mainly transfer power by contacting with pantograph. The other two are messenger wire and droppers, which are used to support the contact wire and keep it maintaining target configuration. In the physics-based model, contact wire and messenger wire are modeled with beam elements, and the spatial two-node beam element is used to discretize them. The droppers are simplified as cable element to cater their slackness when the pantograph passes. In addition, the structural damping of the catenary is introduced with proportional damping model, whose constant values α and β are respectively set as 0.0125 and 0.0001, determined by a field line test [20]. Moreover, a 16 spans catenary model is employed to simulate the pantograph-catenary interaction.

Beam-to-Beam general contact is employed in the physics-based model to simulate the contact interaction between the contact wire and pantograph, with the contact and separation of contact pairs allowed during the dynamic analysis. The tangential contact behavior with frictionless coefficient suggested by [4] is assumed as the contact property.

2.2. Input parameters

The performance of energy transfer quality is influenced by many parameters including design parameters of pantograph-catenary system

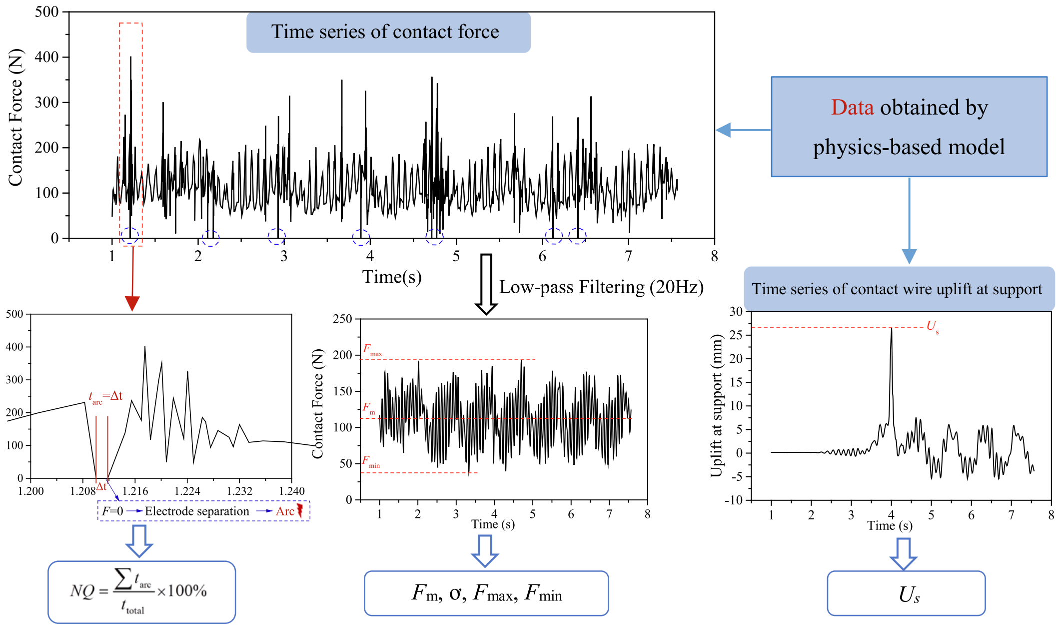

Fig. 3. Feature extraction for the data of physics-based model to obtain the key evaluation values of energy transfer quality.

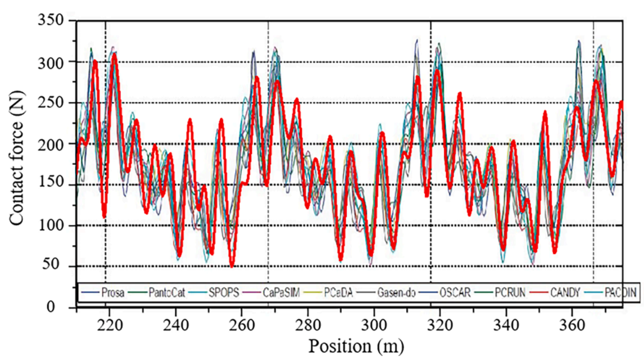

Fig. 4. Comparison of contact force time series obtained by physics-based model and benchmark models [20].

and operation parameters of trains. In this work, eleven parameters that may predominantly affect the pantograph-catenary dynamic interaction are employed as the input parameters in the physics-based model, as shown in Table 1. Nine design parameters and two operation parameters are included. It is worth to be mentioned that the input parameters of physics-based model are the same as those of surrogate model.

2.3. Output and feature extraction

Using the input parameters, the dynamic responses of pantograph-catenary interaction are calculated via physics-based model, and the output data including time series of contact force between pantograph and contact line and time series of contact wire uplift at support are obtained. The first output is responsible for the energy transfer quality and the second for operation security of train, so they are analyzed at the same time. However, the time series output can not directly represent the energy transfer quality.

Feature extraction for the time series results is carried out to obtain the key evaluation values of energy transfer quality and its procedure in detail is shown in Fig. 3. For the results of contact force time series, it can be seen that the contact force is dynamic variation with time. It is obviously that there are some time periods the contact force is equal to zero, marked with blue dash circle line, whose enlarged curve is also presented in the figure. In these specific time periods, the contact status between pantograph and catenary is broken and the electrode separation contributes to produce electric arc, which will considerably deteriorate the energy transfer quality and even break off the power transfer. The percentage of arcing (NQ) is employed to define the level of energy transfer quality [21], which is determined by

$$NQ = \frac{\sum t_{arc}}{t_{total}} \times 100\% \tag{1}$$

in which t_{arc} is the duration of arcs recorded exceeding 5 ms and t_{total} is the total time. Moreover, the time series of contact force is low-pass filtered with a cut-off frequency of 20 Hz according to the Standard EN 50317 [21], and the results of the filtered contact force time series are obtained. Based on the results, four essential values, including the mean value (F_m), maximum value (F_{max}), minimum value (F_{min}) and standard deviation (σ) of contact force, for evaluating energy transfer quality are calculated. In addition, for time series of contact wire uplift, too high contact wire uplift at support may induce the severe impact both on pantograph and steady arm, and even destroy the equipment. According to the Standard [22], the maximum contact wire uplift at support (U_s) is introduced to evaluate its affect.

2.4. Validation of the physics-based model

To validate the physics-based model, the results [20] of benchmark models to simulate pantograph-catenary dynamic interaction established by 10 research institutions around the world were introduced as the contrast. The parameters used in the physics-based model are the same as those in the benchmark models [20]. The comparison results are discussed as follows.

First of all, time series of contact force obtained by physics-based model and benchmark models were compared as shown in Fig. 4. The red curve is the results of physics-based model, and the other curves are the 10 results of the benchmark models. It is seen that the tendency of the red curve is consistent well with the other 10 curves.

Secondly, in the evaluation of energy transfer quality according to IEC Standard [2], five essential evaluation values including F_m,σ, F_{max}, F_{min} and U_s are crucially responsible. Consequently, feature extraction method discussed in Section 2.3 is used to deal with the time series data in Fig. 4, and the results of F_m,σ, F_{max} and F_{min} are obtained. The result of

Table 2
Comparison of key evaluation values obtained by physics-based model and benchmark models [20].

Parameters	Unit	Benchmark models [20]	Physics-based model
F_m	N	169	168.7
σ	N	50 to 62	59.0
F_{max}	N	281 to 325	309.3
F_{min}	N	47 to 72	50.4
U_s	mm	53 to 63	59.2

Table 3
Range of input parameters.

Input	Range	Refer to
L	[45.0, 70.0]	UIC CODE 799 [25]
N	[$int(L/9.5)$, $int(L/5.0)$]*	
R_{sag}	[0, 1.0]	
Sv	[0.1, 0.3]	
D	[1.2.2.0]	
T_{cw}	[15.0, 50.0]	
T_{mw}	[15.0, 50.0]	
Ty_{cw}	(85, 110, 120, 150)	–
Ty_{mw}	(50, 93, 117, 147)	–
V	[50, 400]	–
F_{uplift}	[$F_{m,min}$, $F_{m,max}$]#	EN 50367 [26]

Note: **int* represents round down and ignore the decimal.
$^{\#}F_{m,min} = 0.00047v^2 + 60$, $F_{m,max} = \begin{cases} 0.00047v^2 + 90 & v \leqslant 200\text{km/h} \\ 0.00097v^2 + 70 & v > 200\text{km/h} \end{cases}$, v is train speed.

U_s obtained by physics-based model is also obtained by means of the procedure presented in Fig. 3. The comparisons of the five evaluation values obtained by the physics-based model and benchmark models [20] are listed in Table 2. It is seen that the relative error of F_m between physics-based model and benchmark models is 0.18 % and the other four evaluation values are all in the range of the results of benchmark models. Therefore, it can be concluded that the physics-based model behaves efficient to evaluate the energy transfer quality of pantograph-catenary system.

3. Data and methods

In this Section, data and methods to establish surrogate models are discussed. Specifically, data collection of pantograph-catenary system will be discussed in Section 3.1. The machine learning algorithms will be briefly introduced in Section 3.2. The relationship between the classification surrogate model and regression surrogate model will be illustrated in Section 3.3.

3.1. Data collection

Data is the most important element to construct the surrogate models for evaluating the energy transfer quality of pantograph-catenary system. The collection of field test data of pantograph-catenary system is with highly economic expense and considerably difficult, so as the test data is very deficient. To collect enough and high-quality data, three steps were proceeded, including the sampling strategy, the range of input parameters and the batch computing via physics-based model.

At the first step, a decent sampling strategy was adopted. The selection of sampling points, namely the design of experiment (DOE), is responsible for the performance of surrogate model constructed by machine learning [23]. A favorable DOE is able to reflect full information of design space with limited sampling points as well as reduce the computing cost and meanwhile improve the generalization ability of surrogate model. Therefore, the strategy of sampling point selection plays an important role for the data collection of pantograph-catenary system.

Latin hypercube sampling (LHS) is a space-filling design method. It supposes that there are n variables and m sampling points. The range of each variable is divided into m parts, thus $m \times n$ zones can be obtained. Subsequently, the m sampling points are spread randomly inside one zone on the basis of the rule that only one sampling point is located at each part of the axis projection of a variable. LHS has advantage of avoiding the aggregation of sampling points [24], which is employed to design the sampling points in the work.

At the second step, reasonable ranges of input parameters were determined. The choice of the range of input parameters determines the prediction accuracy and generalization ability of surrogate models. In this work, according to relative design Standards and actual operation conditions, the ranges of the 11 input parameters listed in Table 1 were determined and illustrated in Table 3. Specifically, the ranges of parameter L, N, R_{sag}, Sv, D, T_{cw} and T_{mw} are set referring to the Standard UIC CODE 799 [25]. The Ty_{cw} and Ty_{mw} are the type of contact wire and messenger wire, respectively, which are defined using cross-section area and their ranges are chosen from the typical types widely used in China. Moreover, the v is the train speed, whose range involves from low-speed to hyper-speed trains in operation around the world. F_{uplift} is the vertical force exerted upward by the pantograph on contact wire at standstill, which is considerably important for pantograph-catenary dynamic interaction, and its range is set according to the suggestion of Standard EN 50367 [26].

At the third step, extensive and high-quality data was obtained via batch computing using physics-based model. A total of 1465 sampling points for the 11 input parameters were generated by means of LHS. Each sampling point was as an input into the physics-based model, and the time series outputs were calculated after about 8 h of computation at a high-performance platform with AMD Ryzen 5950X CPU. To improve the computation efficiency, the multi-core parallel batch computing technique was employed. The final outputs, including NQ, F_m,σ, F_{max}, F_{min} and U_s, were obtained by the feature extraction processing method discussed in Section 2.3. With the computation of two months, the output data of the 1465 sampling points are obtained. Afterwards, combined the input data and output data, the dataset to establish surrogate models is constructed.

3.2. Machine learning algorithms

Due to the differences in data structure and data complexity for the energy transfer quality of pantograph-catenary system, it is difficult to determine which machine learning behaves the most suitable for the target [27]. In this study, eleven machine learning algorithms were applied to evaluate the energy transfer quality. Specifically, five tree-based models, including decision tree (DT), random forest tree (RFT), extremely randomized tree (ERT), gradient boosting decision tree (GBDT) and extreme gradient boosting tree (XGBoost), were employed. In addition, support vector machine (SVM), multi-layer feed-forward artificial neural network (MLF) and multi-layer feed-forward deep residual neural network (MLF-DNN) were utilized.

Tree-based ensemble models are popularly used for the problem of classification [28] as well as regression [29–31], because the algorithm of tree-based structure is more intuitive and its recursive process is consistent with human decision-making process. Tree-based model is a supervised learning algorithm and able to process both nominal and numerical data, which is efficiently used both in classification and regression. They support categorical features and split the search space into separate regions by defining decision thresholds at every node of a decision tree and assigning branches to it [31]. Features importance are easily revealed due to their particular tree branch structure. The DT is highly effective in the sum of active leaf weights makes up the prediction, which has been considered to be a suitable tool for the dispatch problem [32]. The RFT developed by Ref. [33] consists of multiple unrelated decision trees and has advantages in dealing with high

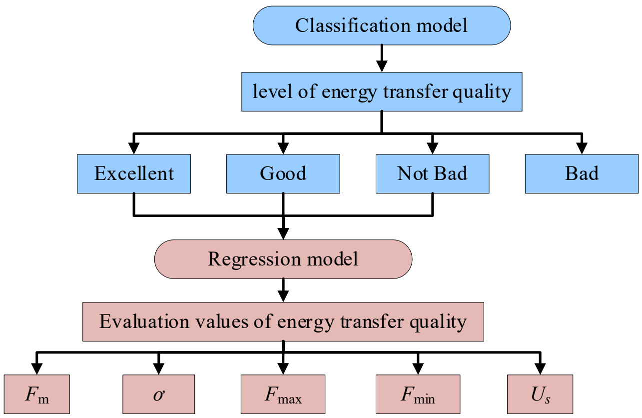

Fig. 5. Relationship between classification and regression surrogate model.

Table 4
Definition of the level of energy transfer quality according to the Standard IEC 62486 [2] and EN 50367 [26].

NQ	Definition	
	Level	label
≤0%	Excellent	0
0–0.2 %	Good	1
0.2 %-5.0 %	Not Bad	2
>5.0 %	Bad	3

dimension problems, meanwhile avoiding the overfitting easily occurred in the DT model. Different with RFT using a random subset to get the best bifurcation attribute, the ERT is completely random to get the bifurcation value in realizing the bifurcation of decision tree so that it behaves better performance. On the basis of RFT and ERT, GBDT is developed by continuously constructing trees in the direction of residual gradient decrease and thus each tree tries to correct the mistakes of the previous one. Due to the increase of model complexity, the GBDT behaves great prediction accuracy under proper hyper-parameters choice. XGBoost is an improved algorithm based on the GBDT algorithm and widely used by data scientists to achieve state-of-the-art results on many machine learning challenges [34]. The loss function of XGBoost is expanded to the second-order derivatives to further consider the trend of gradient variation so that the fitting efficiency is improved and the overfitting is avoided [35]. It also supports parallel learning to improve the computing efficiency.

SVM is a supervised machine learning method using statistical learning theory just like the other artificial neural networks and widely used for regression problems [36,37]. SVM regression is to find a regression hyperplane so that all the data of a set are closest to the plane. It has advantages of preventing the over-learning and is suitable for fitting of small sampling dataset.

Artificial neural network is the tendency of computing science with the rapid improvement of computer hardware and extensive algorithms. As one of the well-known neural networks, MLF is trained according to the error back propagation algorithm that including two processes: signal forward-propagation and error back-propagation [38]. MLF consists of many neurons that are involved into layers. The first layer is called the input layer, and the last layer is the output layer, and between them are hidden layers. The fitting of MLF is to continuously adjust the weights and biases among the neurons to reduce the error. Based on MLF, the MLF-DNN was developed. It allows training of deep networks up to more than 1000 layers [39]. MLF-DNN has very strong nonlinear fitting ability to adapt any complicated function and behaves considerably efficient to avoid overfitting because of its complex neural network with deep layers and wide neurons. However, MLF-DNN is not easy to train due to its complicated hyper-parameters, a lot of skill is needed to obtain the optimal surrogate model. The relevant training and testing techniques will be introduced in the later Sections.

3.3. Relationship between classification and regression surrogate models

In this study, we respectively proposed classification and regression surrogate models to evaluate the energy transfer quality of pantograph-catenary system. The relationship between classification model and regression model is illustrated as shown in Fig. 5. The classification surrogate model is to evaluate the level of energy transfer quality. The level is defined with four levels, i.e. 'Excellent', 'Good', 'Not Bad' and 'Bad', according to relevant Standards and will be illustrated in next Section in detail. For the first three levels, it is acceptable for normal operation of trains but the quantitative evaluation of energy transfer quality, including F_m,σ, F_{max}, F_{min} and U_s, is needed to be further studied, which will be dealt with by the establishment of regression model.

4. Classification surrogate models

The goal of the classification models is to classify the levels of energy transfer quality. Five tree-based algorithms, including DT, RFT, ERT, GBDT and XGBoost, are explored as the candidate classification surrogate models. Comparisons of the performance of these classification surrogate models are carried out to find the optimal one.

4.1. Definition of the level of energy transfer quality

The contact status of pantograph-catenary system is strongly associated with the energy transfer quality. As shown in Fig. 2, electric arc will be produced if the contact force is too small or to be zero, which crucially deteriorates the energy transfer quality or even interrupts the energy transfer. The *NQ* determined by formula (1) is generally employed to evaluate the level of energy transfer quality. It is noted that a slight electric arc, i.e. proper *NQ*, is allowed in the normal operation of trains for economic reasons.

The definition of the level of energy transfer quality is presented in Table 4. Four levels are classified based on the value of *NQ*. Specifically, the 'Excellent' level is defined if no arc occurred. It is called 'Good' level when *NQ* is in the range of 0 and 0.2 % according to the Standard IEC 62486 [2]. When *NQ* is in the range of 0.2 % and 5.0 %, we called 'Not Bad' level according to the suggestion of Standard EN 50367 [26] and 'Bad' level if the *NQ* is greater than 5.0 %.

4.2. Dataset pre-processing for classification

For input parameters, to eliminate the adverse effects caused by singular sample data in the training process of surrogate models, the 11 input parameters are all normalized between zero and one by min–max normalization method that is expressed by

$$x^{'} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2}$$

in which x is the origin parameter and x' is normalized parameter. Specifically, the 11 input parameters except for Ty_{cw} and Ty_{cw} are continuous values and directly normalized. The discrete parameters Ty_{cw} and Ty_{cw} are respectively labeled to 1, 2, 3 and 4 from smallest to largest in terms of cross-sectional area, and then normalized to between zero and one. Moreover, for output parameter, the levels of energy transfer quality are labeled by 0, 1, 2 and 3 as shown in Table 4.

4.3. Evaluation criteria

The R-squared value (R^2), mean squared error (*MSE*) and mean absolute error (*MAE*) were employed as indictors to evaluate the accuracy of surrogate models, and they are expressed by [40]

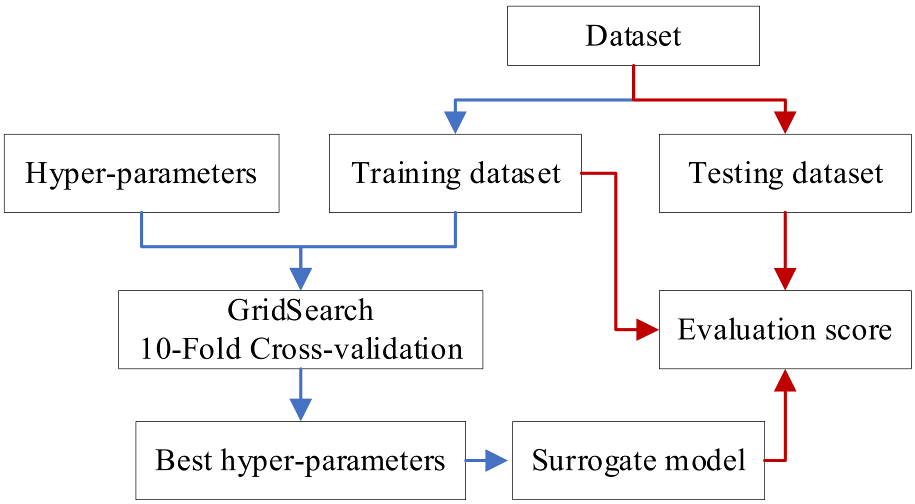

Fig. 6. Illustration of the process of surrogate model establishment.

Table 5
Evaluation scores of classification surrogate models.

Model name	Training dataset			Testing dataset		
	R^2	*MSE*	*MAE*	R^2	*MSE*	*MAE*
DT	0.834	0.071	0.055	0.693	0.122	0.088
RFT	0.976	0.010	0.010	0.718	0.112	0.078
ERT	0.990	0.004	0.004	0.710	0.116	0.082
GBDT	0.930	0.030	0.028	0.812	0.075	0.061
XGBoost	0.832	0.071	0.058	0.778	0.088	0.068

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2} \tag{3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

in which n is the number of data points, y is the real value, $\hat{y}$ is the predicted value and $\overline{y}$ is the average value. There is a simple but useful way to understand the three indictors that if a surrogated model performs perfectly, the R2 should be as close to 1.0 as possible while the *MSE* and *MAE* as close to 0.0 as possible.

4.4. Development of classification surrogate models

The process of the establishment of classification surrogate models is outlined in Fig. 6. Firstly, the prepared dataset was randomly divided into training dataset and testing dataset, with a ratio of 80 to 20. Then, five tree-based algorithms discussed in Section 3.3 were used to train the models at the training dataset. The grid search tool GridSearchCV provided by scikit-learn [41] was utilized for the hyper-parameters tuning, in which the 10-fold cross-validation was chosen to increase the accuracy and credibility of the surrogate model. The training dataset was used for hyper-parameters tuning to obtain the best hyper-parameters. Using the best hyper-parameters, the models were retrained using the whole training dataset and then validated on the testing dataset. The hyper-parameters tuning details, including the tuning parameters and their searching range as well as chosen parameters, for the five tree-based models are presented in the Appendix Table A.1 to A.5, respectively. Finally, the mature classification surrogate models were obtained.

The evaluation results of the five tree-based classification surrogate models both on training dataset and testing dataset are shown in Table 5. It is seen that the RFT and ERT model are of better scores at training dataset but fail for prediction at the testing dataset. This may be due to

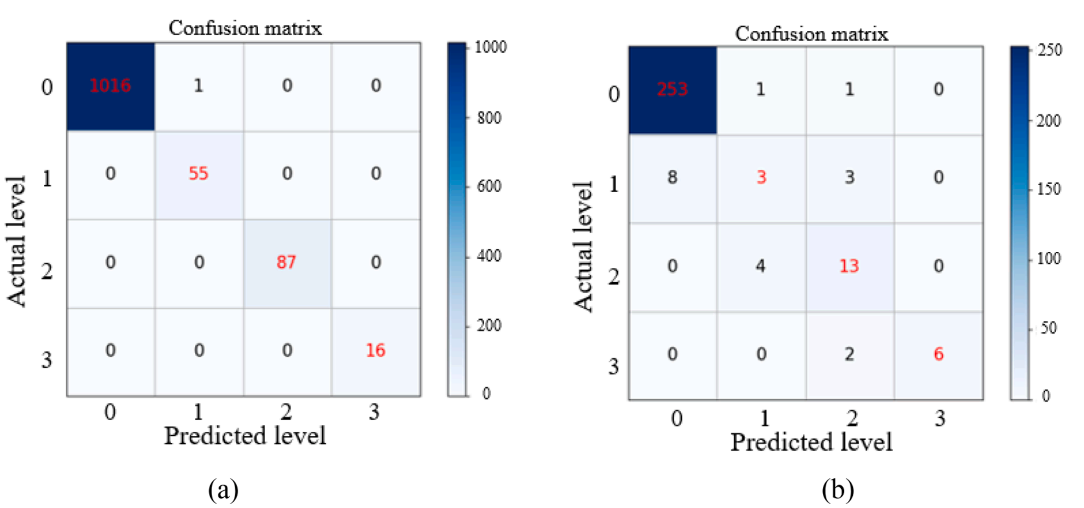

Fig. 7. Confusion matrix of classification model via GBDT, (a) Training dataset, (b) Testing dataset.

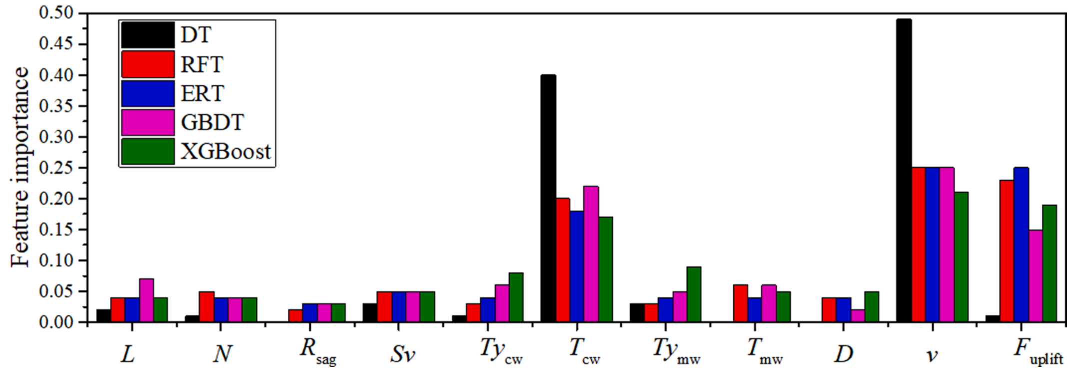

Fig. 8. Feature importance analysis for the input parameters of classification surrogate model.

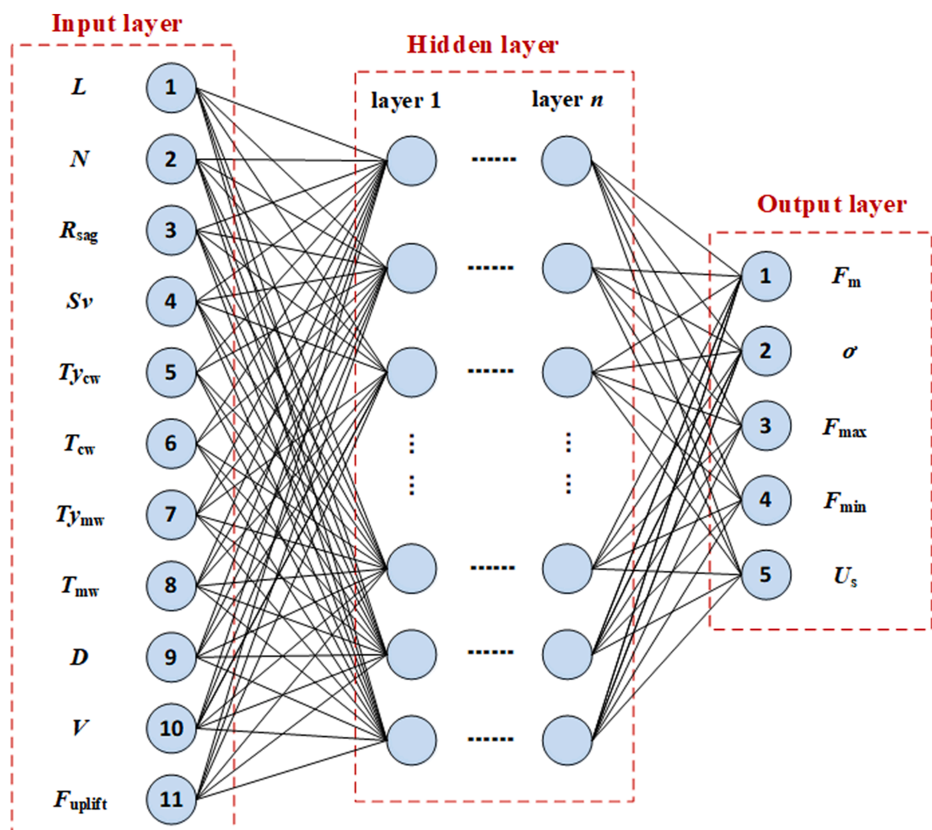

Fig. 9. Neural network framework of regression surrogate model.

the model overfitting during the training of ERT and RFT models. Moreover, although the scores of GBDT model at the training dataset is not the best, the scores at the testing dataset behaves optimal with the highest R^2 as well as the lowest *MSE* and *MAE*. Therefore, the GBDT-based model is the optimal surrogate model to classify the level of energy transfer quality.

Confusion matrix is an error matrix commonly used to visually evaluate the performance of supervised learning algorithms. Here, as the GBDT-based model for example, the confusion matrixes for training dataset and testing dataset are calculated and shown in Fig. 7. The levels of energy transfer quality are labeled to 0, 1, 2 and 3 according to the definition in Table 4. The horizontal axis is the predicted level and the vertical axis is the actual level. It is seen that the GBDT-based surrogate model behaves perfect in classifying the level of energy transfer quality in the training dataset, with only one sample is misclassified. In the testing dataset, the model is of great prediction accuracy for '0' label, i.e. 'Excellent' level, but not that good for the other three levels. This is because the sampling point number of the three levels is far less than that of 'Excellent' level and may be deficient. However, the selection of the dataset is consistent with the actual situation that 'Excellent' level is more common but the other three are sometimes happened in the normal operation of trains. Whatever, the GBDT-based classification surrogate model behaves optimal to classify the level of energy transfer quality of pantograph-catenary system.

4.5. Feature importance analysis

To evaluate the influence degree of the 11 input parameters on the level of energy transfer quality, the feature importance analysis was carried out. Feature importance is a basic property of tree-based model and it can be calculated by computing each feature contributes to each tree and then being averaged to compare the contribution between the features. The feature importance analysis for the 11 input parameters of the five tree-based classification surrogate models is presented in Fig. 8.

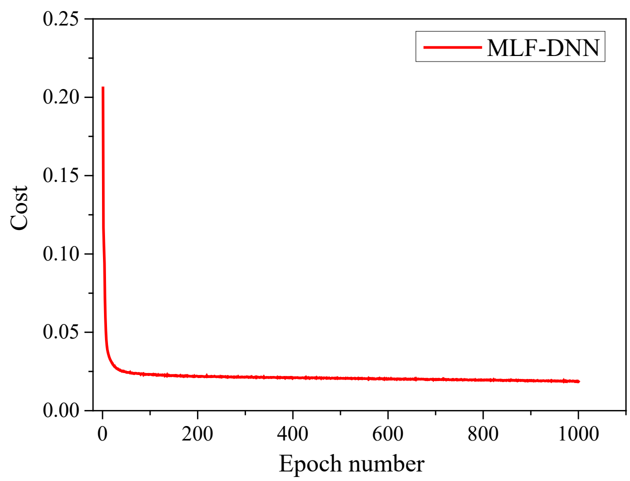

Fig. 10. Training process of the MLF-DNN-based surrogate model.

Table 6
Evaluation scores of regression surrogate models.

Model name	Training dataset			Testing dataset		
	R^2 [-]	*MSE* [%]	*MAE* [%]	R^2 [-]	*MSE* [%]	*MAE* [%]
DT	0.923	0.135	2.285	0.823	0.289	3.158
RFT	0.987	0.023	0.903	0.896	0.191	2.471
ERT	0.991	0.002	0.775	0.913	0.161	2.230
GBDT	0.988	0.022	0.798	0.929	0.137	2.212
XGBoost	0.972	0.055	1.617	0.922	0.149	2.326
SVM	0.880	0.231	3.834	0.856	0.252	3.787
MLF	0.985	0.273	1.040	0.877	0.263	2.796
MLF-DNN	0.963	0.075	1.800	0.950	0.099	2.077

In general, the T_{cw}, v and F_{uplift} are predominant to affect the output because of with higher feature importance scores, not considering the DT-based model due to it with the worst prediction accuracy so that its feature importance score may not be credible. For the GBDT-based model, the top three input parameters that have the greatest impact on output are v, T_{cw}, F_{uplift}, respectively.

5. Regression surrogate models

According to the suggestion of the Standard [2,22], five essential parameters, including F_m,σ, F_{max}, F_{min} and U_s, are crucially responsible for the energy transfer quality of pantograph-catenary system. So they are chosen as the evaluation values to assess the energy transfer quality.

The purpose of regression surrogate model is to quantitatively evaluate the energy transfer quality via the five evaluation values. To achieve this goal, in addition the five tree-based algorithms, the SVM, MLF and MLF-DNN algorithms discussed in Section 3.3 are employed to carry out the regression surrogate models, and then the optimal regression surrogate model is obtained by comparisons of the eight candidate regression models.

In the first step, dataset pre-processing for regression models was dealt with. For the eleven input parameters, the data pre-processing procedure is the same as the classification models. The five output parameters are respectively normalized between zero and one by min–max

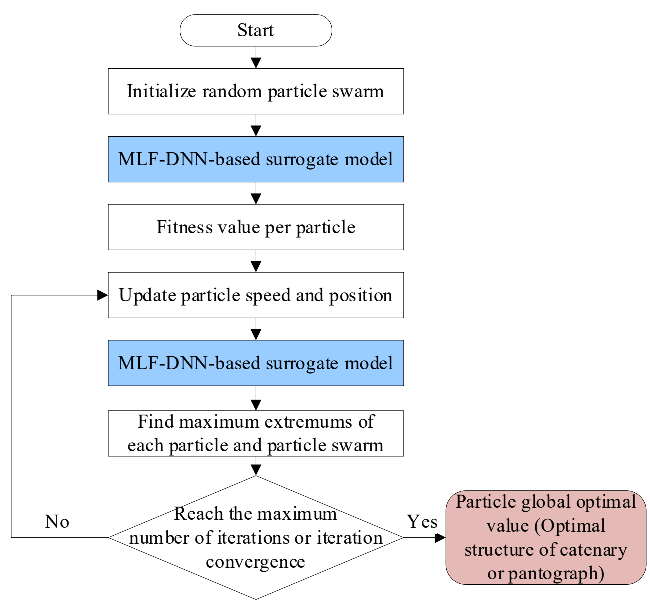

Fig. 12. Optimization procedure for the structure design parameters of catenary and pantograph system.

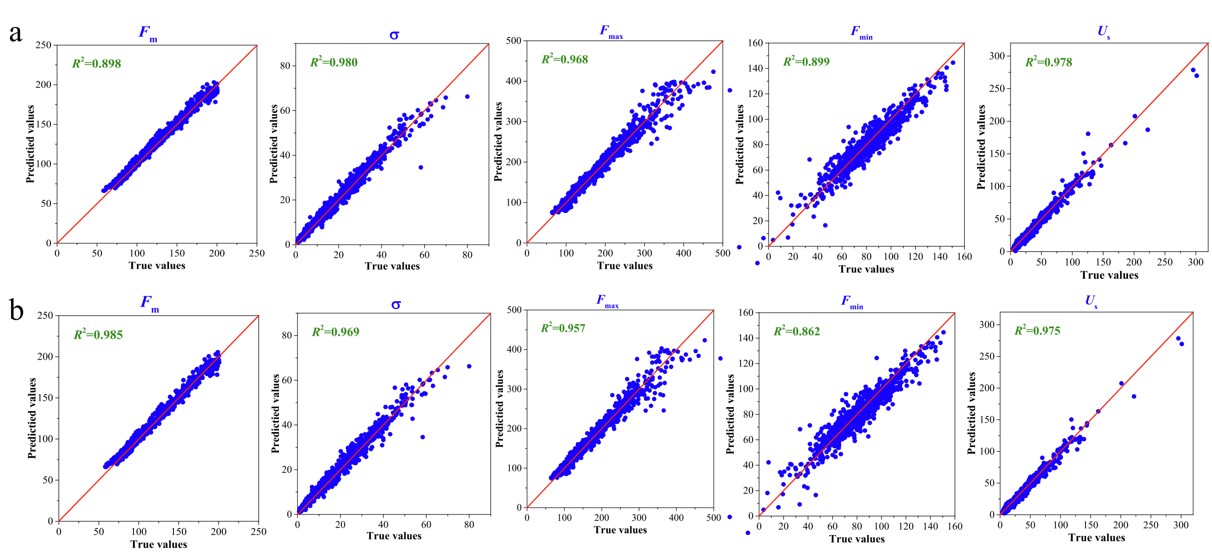

Fig. 11. Comparisons of the five evaluation values of energy transfer quality between true values and predicted values obtained by MLF-DNN-based surrogate model, a) for training dataset, b) for testing dataset.

Table A1
Grid search of hyper-parameters tuning for DT.

Hyper-parameters	Description	Searched	Selected	
			Classification	Regression
max_depth	Maximum depth of the tree	range(2, 50, 4)	14	46
min_samples_split	Minimum number of samples required to split inner nodes	range(2, 20, 2)	6	16
min_samples_leaf	Minimum number of samples required at leaf nodes	range(2, 10, 1)	5	3

Table A2
Grid search of hyper-parameters tuning for RFT.

Hyper-parameters	Description	Searched	Selected	
			Classification	Regression
n_estimators	Number of trees	[10, 20, 50, 100, 200]	50	200
min_samples_split	Minimum number of samples required to split inner nodes	range(2, 10, 2)	4	2
max_features	Number of features to consider when finding the best split	['auto', 'sqrt']	'auto'	'auto'
max_depth	Maximum depth of the tree	range(2, 50, 4)	30	30

Table A3
Grid search of hyper-parameters tuning for ERT.

Hyper-parameters	Description	Searched	Selected	
			Classification	Regression
n_estimators	Number of trees	[10, 20, 50, 100, 200]	10	200
min_samples_split	Minimum number of samples required to split inner nodes	range(2, 10, 2)	4	4
max_features	Number of features to consider when finding the best split	['auto', 'sqrt']	'sqrt'	'auto'
max_depth	Maximum depth of the tree	range(2, 50, 4)	34	14

normalization method by formula (2).

In the second step, the training and testing process for regression surrogate models were carried out. The process of the establishment of the regression model is the same as that of classification model as presented in Fig. 6. The prepared dataset was randomly divided into training dataset and testing dataset, with a ratio of 80 to 20, and the grid search tool GridSearchCV was employed to find the optimal hyper-parameters of each surrogate models. The hyper-parameters tuning details of the eight regression models are presented in the Appendix Table A.1 to A.7.

For neural network, the basic construction of neural network is presented in Fig. 9. The eleven input parameters are set as the input layer, and the five evaluation values of energy transfer quality are the output layer, between them is the hidden layer. In this study, we selected two hidden layers for MLF algorithm and six hidden layers for MLF-DNN algorithm. In the hyper-parameters tuning process, the neuron number, batch size, learning rate, and activation function are selected to be tuned. Using the optimal hyper-parameters, the MLF and MLF-DNN are retrained using the whole training dataset. The training process of the MLF-DNN-based surrogate Model is shown in Fig. 10. The vertical axis is

Table A4
Grid search of hyper-parameters tuning for GBDT.

Hyper-parameters	Description	Searched	Selected	
			Classification	Regression
n_estimators	Number of trees	range(5, 50, 5)	25	45
learning_rate	Step size shrinkage used in update to prevent overfitting	range(0.1, 1.5, 0.1)	0.2	0.2
max_depth	Maximum depth of the tree	range(2, 20, 4)	10	6
min_samples_leaf	Minimum number of samples required at leaf nodes	range(1, 10, 1)	3	3

Table A5
Grid search of hyper-parameters tuning for XGBoost.

Hyper-parameters	Description	Searched	Selected	
			Classification	Regression
max_depth	Maximum depth of the tree	range(2, 20, 4)	18	6
learning_rate	Step size shrinkage used in update to prevent overfitting	range (0.01, 0.2, 0.04)	0.17	0.17
n_estimators	Number of trees	range(5, 50, 10)	45	45
min_child_weight	Minimum sum of instance weight needed in a child	[0, 1, 3, 5]	1	3
gamma	Minimum loss reduction required to make a further partition on a leaf node of the tree	[0.01, 0.05, 0.1, 0.2, 0.3]	0.01	0.01
subsample	Subsample ratio of the training instances	[0.6, 0.7, 0.8, 0.9]	0.8	0.6
lambda	L2 regularization term on weights	[0.1, 0.3, 0.6, 0.9, 1.1]	0.3	0.1
colsamle_bytree	Subsample ratio of columns when constructing each tree	[0.5, 0.7, 0.9, 1.0]	1.0	0.9

Table A6
Grid search of hyper-parameters tuning for SVM.

Hyper-parameters	Description	Searched	Selected
kernel	kernel type used in the algorithm	['linear', 'poly', 'rbf', 'sigmoid']	'linear'
C	Penalty coefficient	[0.1, 0.5, 1.0, 5.0, 10.0, 50.0, 100.0, 1000.0]	1000.0

Table A7
Grid search of hyper-parameters tuning for MLF and MLF-DNN.

Hyper-parameters			MLF		MLF-DNN	
			Searched	Selected	Searched	Selected
Input layer		size	11	11	11	11
Hidden layers	1st	neuron	[256, 512]	256	[128, 256]	256
		activation	['relu','sigmoid', 'linear']	'relu'	['relu','sigmoid', 'linear']	'linear'
	2nd	neuron	[64, 128]	128	[64, 128]	128
		activation	['relu','sigmoid', 'linear']	'sigmoid'	['relu','sigmoid', 'linear']	'linear'
	3rd	neuron			[16, 32]	16
		activation			['relu','sigmoid', 'linear']	'relu'
	4th	neuron			[64, 128]	128
		activation			['relu','sigmoid', 'linear']	'sigmoid'
	5th	neuron			[32, 64]	32
		activation			['relu','sigmoid', 'linear']	'linear'
	6th	neuron			[8, 16]	8
		activation			['relu','sigmoid', 'linear']	'sigmoid'
Output layer		size	7	7	7	7
Optimizer		learning rate	[0.0001, 0.001, 0.01]	0.001	[0.0001, 0.001, 0.01]	0.0001
Batch_size			[2, 5, 10]	5	[2, 5, 10]	5
Epochs			1000	1000	1000	1000

the cost that represents the outcome value of the cost function, and the horizontal axis is the training epoch. It is worth to mention that the cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between the inputs and outputs. It is seen that the model is converged after 1000 epochs.

In the end, the evaluation scores of the eight regression surrogate models both on training dataset as well as testing dataset are presented in Table 6. It can be seen that the ERT is of the best score on the training dataset but fails on the testing dataset. Model overfitting may be produced in the training of the ERT model. However, the MLF-DNN-based regression surrogate model behaves the best with the highest $R^2 = 0.950$ and lowest $MSE = 0.099$ % and $MAE = 2.077$ % on the testing dataset. This is because the MLF-DNN-based model has deep layers and complex neural network, which is of strong nonlinear fitting ability to cater the extremely complicated physics-based model. In summary, the MLF-DNN-based regression surrogate model is the optimal model to evaluate the values of energy transfer quality of pantograph-catenary system.

Moreover, taking the MLF-DNN-based regression surrogate model as an example, the comparisons of the five evaluation values between true values and predicted values are carried out and shown in Fig. 11. It is seen that the predicted values obtained by MLF-DNN-based surrogated model are consistent with the true values obtained by physics-based model not only on the training dataset but also on the testing dataset. As for the testing dataset, the MLF-DNN-based surrogate model has the highest score to predict F_m with $R^2 = 0.985$, followed by U_s,σ, F_{max} and F_{min} in sequence, whose R^2 are 0.975, 0.969, 0.957 and 0.862, respectively.

6. Potential applications of the surrogate models

This work is the first trial to establish surrogate models to evaluate the energy transfer quality of pantograph-catenary system using machine learning in this field. The surrogate models are expected to replace the physics-based model or even the field test. Relative to the physics-based model with time cost of about 8 h for a case, the surrogate models achieved highly time saving with only several seconds in the case of enough prediction accuracy guaranteed. Therefore, the innovation of this work does not lie in the method we used, but provides a new idea for energy transfer problems of pantograph-catenary system of high speed railway.

With the advantages of high accuracy and high efficiency, the surrogate models are of extensive potential applications. For example, in the aspect of structure design of the catenary and pantograph system, the MLF-DNN-based regression surrogate model can be employed to quickly establish the mapping relation between structure parameters and energy transfer quality, which will greatly accelerate the optimization iteration. The optimal structure can be obtained by coupled with optimization procedure, such as particle swarm optimization (PSO) algorithm, aiming to improve the energy transfer quality. To illustrate this, the optimization procedure is briefly presented in Fig. 12. Moreover, in the aspect of operation control of trains, the real-time active control technique for the pantograph is expected to achieve by coupling the surrogate models for adjusting the uplift force to improve the energy transfer quality. These two potential applications are considerably helpful for the improvement of energy transfer quality of catenary and pantograph system, and they will be elaborately investigated in the further works.

7. Conclusion

This work aims to establish surrogate models to evaluate energy transfer quality of pantograph-catenary system of high-speed railway. To reach this goal, the physics-based model is proposed to collect the dataset. The inputs involve the essential parameters of pantograph-catenary system including nine design parameters and two operation parameters, and the outputs consist of several key evaluation values to judge the energy transfer quality. Five tree-based classification surrogate models including DT, RET, ERT, GBDT and XGBoost were developed to classify the level of energy transfer quality, and eight regression surrogate models including five tree-based models, SVM, MLF and MLF-DNN were carried out to quantitatively evaluate the energy transfer quality.

It was found that the GBDT-based model is the optimal classification surrogate model to evaluate the level of energy transfer quality. The three inputs v, T_{cw} and F_{uplift} are predominant parameters to affect the energy transfer quality. Therefore, to improve the energy transfer quality on the condition of specific train speed v, the most effective ways are changing the T_{cw} and adjusting the value of F_{uplift}. Moreover, the MLF-DNN-based regression surrogate model behaves optimal to evaluate the five essential values of energy transfer quality, with the highest comprehensive score of $R^2 = 0.950$, $MSE = 0.099$ % and $MAE = 2.077$ %. For the each output, the MLF-DNN-based model has the best accurate prediction accuracy for F_m with $R^2 = 0.985$, followed by U_s,σ, F_{max} and F_{min} in sequence respectively.

The established surrogate models are of high accuracy and high efficiency, which are expected to replace traditional physics-based model or even field test for evaluating the energy transfer quality of pantograph-catenary system. In addition, potential actual applications of the surrogate models are considerably extensive, but not deeply investigated in this paper, which points the way for the future works.

CRediT authorship contribution statement

Guizao Huang: Conceptualization, Methodology, Data curation, Investigation, Validation, Writing – original draft. **Guangning Wu:** Funding acquisition, Supervision. **Zefeng Yang:** Project administration, Resources. **Xing Chen:** Software, Visualization. **Wenfu Wei:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 52107169) and the Natural Science Foundation of Sichuan Province (No. 2023NSFSC0822).

Appendix

Input parameters and hyper-parameter tuning for surrogate models. See Appendix Tables A.1–A.7.

References

[1] Lin Y, Qin Y, Wu J, Xu M. Impact of high-speed rail on road traffic and greenhouse gas emissions. Nat Clim Change 2021;11:1–6. https://doi.org/10.1038/s41558-021-01190-8.
[2] IEC 62486. Railway applications - Current collection systems - Technical criteria for the interaction between pantograph and overhead contact line. International Electrical Commission; 2017.
[3] Rønnquist A, Nåvik P. Dynamic assessment of existing soft catenary systems using modal analysis to explore higher train velocities: a case study of a Norwegian contact line system. Veh Syst Dyn 2015;53:756–74. https://doi.org/10.1080/00423114.2015.1013040.
[4] Nåvik P, Rønnquist A, Stichel S. Variation in predicting pantograph–catenary interaction contact forces, numerical simulations and field measurements. Veh Syst Dyn 2017;55:1265–82. https://doi.org/10.1080/00423114.2017.1308523.
[5] Xu Z, Song Y, Liu Z. Effective measures to improve current collection quality for double pantographs and catenary based on wave propagation analysis. IEEE Trans Veh Technol 2020;69:6299–309. https://doi.org/10.1109/tvt.2020.2985382.
[6] Van OV, Massat J-P, Balmes E. Waves, modes and properties with a major impact on dynamic pantograph-catenary interaction. J Sound Vib 2017;402:51–69. https://doi.org/10.1016/j.jsv.2017.05.008.
[7] Song Y, Liu Z, Wang H, Lu X, Zhang J. Nonlinear analysis of wind-induced vibration of high-speed railway catenary and its influence on pantograph-catenary interaction. Veh Syst Dyn 2016;54:723–47. https://doi.org/10.1080/00423114.2016.1156134.
[8] Yao Y, Zhou N, Mei G, Zhang W. Dynamic analysis of pantograph-catenary system considering ice coating. Shock Vib 2020::8887609. https://doi.org/10.1155/2020/8887609.
[9] Wang Z, Song Y, Yin Z, Wang R, Zhang W. Random response analysis of axle-box bearing of a high-speed train excited by crosswinds and track irregularities. IEEE Trans Veh Technol 2019;68:10607–17. https://doi.org/10.1109/TVT.2019.2943376.
[10] Song Y, Liu Z, Ronnquist A, Navik P, Liu Z. Contact wire irregularity stochastics and effect on high-speed railway pantograph-catenary interactions. IEEE Trans Instrum Meas 2020;69:8196–206. https://doi.org/10.1109/tim.2020.2987457.
[11] Jimenez-Octavio JR, Carnicero A, Sanchez-Rebollo C, Such M. A moving mesh method to deal with cable structures subjected to moving loads and its application to the catenary–pantograph dynamic interaction. J Sound Vib 2015;349:216–29. https://doi.org/10.1016/j.jsv.2015.03.051.
[12] Song Y, Liu Z, Xu Z, Zhang J. Developed moving mesh method for high-speed railway pantograph-catenary interaction based on nonlinear finite element procedure. Int J Rail Transp 2019;7:173–90. https://doi.org/10.1080/23248378.2018.1532330.
[13] Gorissen D, Couckuyt I, Demeester P, Dhaene T, Crombecq K. A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design. J Mach Learn Res 2010;11:2051–5. https://doi.org/10.1007/s10846-010-9395-x.
[14] Zhang Y, Bryan J, Richards G, Wang H. Development and comparative selection of surrogate models using artificial neural network for an integrated regenerative transcritical cycle. Appl Energy 2022;317:119146. https://doi.org/10.1016/j.apenergy.2022.119146.
[15] Manojlović V, Kamberović Ž, Korać M, Dotlić M. Machine learning analysis of electric arc furnace process for the evaluation of energy efficiency parameters. Appl Energy 2022;307:118209. https://doi.org/10.1016/j.apenergy.2021.118209.
[16] Wang Z, Hong T, Piette MA. Building thermal load prediction through shallow machine learning and deep learning. Appl Energy 2020:263. https://doi.org/10.1016/j.apenergy.2020.114683.
[17] Zhou Y, Wang J, Liu Y, Yan R, Ma Y. Incorporating deep learning of load predictions to enhance the optimal active energy management of combined cooling, heating and power system. Energy 2021:233. https://doi.org/10.1016/j.energy.2021.121134.
[18] BS EN 50318. Railway applications – Current collection systems – Validation of simulation of the dynamic interaction between pantograph and overhead contact line. British Standards Institution; 2018.
[19] Song Y, Liu Z, Wang H, Lu X, Zhang J. Nonlinear modelling of high-speed catenary based on analytical expressions of cable and truss elements. Veh Syst Dyn 2015;53:1455–79. https://doi.org/10.1080/00423114.2015.1051548.
[20] Bruni S, Ambrosio J, Carnicero A, Cho YH, Finner L, Ikeda M, et al. The results of the pantograph–catenary interaction benchmark. Veh Syst Dyn 2015;53:412–35. https://doi.org/10.1080/00423114.2014.953183.
[21] BS EN 50317. Railway applications. Current collection systems. Requirements for and validation of measurements of the dynamic interaction between pantograph and overhead contact line. British Standards Institution; 2012.
[22] BS EN 50119. Railway applications-Fixed installations-Electric traction overhead contact lines. British Standards Institution; 2020.
[23] Fernandez G, Park C, Kim N, Haftka R. Review of multi-fidelity models. 2017. https://doi.org/arXiv:1609.07196.
[24] Georgiou SD. Orthogonal Latin hypercube designs from generalized orthogonal designs. J Statist Plann Inference 2009;139:1530–40. https://doi.org/10.1016/j.jspi.2008.08.016.
[25] UIC CODE 799. Characteristics of a.c. overhead contact systems for high-speed lines worked at speeds of over 200 km/h. International Union of Railways; 2002.
[26] BS EN 50367. Railway applications. Fixed installations and rolling stock. Criteria to achieve technical compatibility between pantographs and overhead contact line. British Standards Institution; 2020.
[27] Wei Z, Zhang T, Yue B, Ding Y, Xiao R, Wang R, et al. Prediction of residential district heating load based on machine learning: A case study. Energy 2021:231. https://doi.org/10.1016/j.energy.2021.120950.
[28] Calzavara S, Cazzaro L, Lucchese C, Marcuzzi F, Orlando S. Beyond Robustness: Resilience Verification of Tree-Based Classifiers. Comput Secur 2022:102843. https://doi.org/10.1016/j.cose.2022.102843.
[29] Yang C, Du J, Zhang J, Wu C, Chen M, Wu J. Tree-based Data Augmentation and Mutual Learning for Offline Handwritten Mathematical Expression Recognition. Pattern Recognit 2022::108910. https://doi.org/10.1016/j.patcog.2022.108910.
[30] Huo Y, Bouffard F, Joós G. Decision tree-based optimization for flexibility management for sustainable energy microgrids. Appl Energy 2021;290:116772. https://doi.org/10.1016/j.apenergy.2021.116772.
[31] Thebelt A, Tsay C, Lee RM, Sudermann-Merx N, Walz D, Tranter T, et al. Multi-objective constrained optimization for energy applications via tree ensembles. Appl Energy 2022;306:118061. https://doi.org/10.1016/j.apenergy.2021.118061.
[32] Moutis P, Skarvelis-Kazakos S, Brucoli M. Decision tree aided planning and energy balancing of planned community microgrids. Appl Energy 2016;161:197–205. https://doi.org/10.1016/j.apenergy.2015.10.002.
[33] Breiman L. Random forests, machine learning 45. J Clin Microbiol 2001;2:199–228.
[34] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: Association for Computing Machinery; 2016. p. 785–94.
[35] Wang R, Lu S, Feng W. A novel improved model for building energy consumption prediction based on model integration. Appl Energy 2020;262:114561. https://doi.org/10.1016/j.apenergy.2020.114561.
[36] Rahimi M, Abbaspour-Fard MH, Rohani A. Machine learning approaches to rediscovery and optimization of hydrogen storage on porous bio-derived carbon. J Cleaner Prod 2021;329:129714. https://doi.org/10.1016/j.jclepro.2021.129714.
[37] Cheng Y, Zhu Q, Peng Y, Huang X-F, He L-Y. Multiple strategies for a novel hybrid forecasting algorithm of ozone based on data-driven models. J Cleaner Prod 2021;326:129451. https://doi.org/10.1016/j.jclepro.2021.129451.
[38] Huang G, Yan B, Mou Z, Wu K, Lv X. Surrogate model for torsional behavior of bundle conductors and its application. IEEE Trans Power Delivery 2021;37:67–75. https://doi.org/10.1109/TPWRD.2021.3053341.
[39] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770–8.
[40] Chung WH, Gu YH, Yoo SJ. District heater load forecasting based on machine learning and parallel CNN-LSTM attention. Energy 2022:246. https://doi.org/10.1016/j.energy.2022.123350.
[41] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2012:12. https://doi.org/10.1524/auto.2011.0951.