

Received 26 November 2022, accepted 8 December 2022, date of publication 12 December 2022,
date of current version 15 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3228441

APPLIED RESEARCH

A Novel Multi-Model Stacking Ensemble Learning Method for Metro Traction Energy Prediction

SHAN LIN¹, XINGZHONG NONG¹, JIANQIANG LUO^{1,2} , (Member, IEEE), AND CHEN'EN WANG²

¹Guangzhou Metro Design and Research Institute Company Ltd., Guangzhou, Guangdong 510010, China

²School of Automation, Guangdong University of Technology, Guangzhou, Guangdong 510006, China

Corresponding authors: Jianqiang Luo (jqlo@gdut.edu.cn) and Chen'en Wang (1084676593@qq.com)

ABSTRACT Metro traction energy prediction is the basis of abnormal monitoring and plays an indispensable role in the planning and operation of the metro system. However, current studies rarely offer a satisfactory prediction performance. To improve the prediction accuracy, a novel prediction method for metro traction energy consumption is proposed based on gradient penalty Wasserstein generative adversarial network (WGAN-GP) and stacking ensemble learning with multi-model integration. Firstly, aiming to collect effective train data, WGAN-GP is used to generate characteristic data of traction energy consumption. Then, various algorithms like BP, SVM, ELM, and XGBoost are employed to preliminarily disclose the relationship between traction energy consumption and characteristic data of traction energy consumption via K-fold verification. Thereafter, the XGBoost algorithm is implemented as the meta model to construct a stacking ensemble learning prediction model. Finally, the proposed method is verified with data from Guangzhou Metro Line 13, and the results substantiate the effectiveness of the prediction model.


INDEX TERMS Metro traction energy, gradient penalty, Wasserstein generative adversarial network, ensemble learning, prediction model.

I. INTRODUCTION

In 2022, China's rail transit mileage exceeds 9000km, and Shanghai, Beijing, Guangzhou, Chengdu, and Wuhan rank in the top five cities in the world in terms of subway mileage. The rapid development of metro rail transit has brought challenges in high energy consumption. To achieve the carbon peak and neutrality targets, it is necessary to reduce the energy consumption of the metro system. In the metro system, 40%~60% of the total energy is consumed by traction [1]. Therefore, accurate traction energy consumption prediction is conducive to reducing the total energy consumption in the metro system. Usually, the abnormal monitoring of traction energy consumption is realized by analyzing the error between the predicted and the actual value [2]. Therefore, the accurate prediction of traction energy consumption plays a significant role and also provides a basis for abnormal monitoring [3], [4]. On the other hand, the energy optimization of subway train operation also needs the prediction result of traction energy consumption [5], [6], [7]. Current prediction

works on metro traction energy consumption are mainly based on historical data via either the physical model [8] or the statistical model [9]. The physical model is to establish the dynamics model of the metro train through force analysis. The calculation process is complex and inaccurate, especially considering multiple trains running on the line [10]. The statistical model is mainly based on a machine learning regression model that can quickly evaluate the traction energy consumption [11]. For example, Lü et al [12]. proposed a prediction model based on support vector regression and random forest regression. This model can accurately describe the relationship between traction energy consumption and related influencing factors with large-scale operation data. Tang et al [13]. established the prediction model of migration energy consumption and total energy consumption. By binary linear regression fitting and support vector regression with two years' operation data, the model can achieve high accuracy on multiple lines.

However, all the above studies are based on existing lines and require a large amount of historical operation data, making them difficult to directly apply to new lines with short operation times. Moreover, the existing studies are all based

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Pau .

on a single prediction model. The learning ability of a single prediction model is limited, and the complex relationship in the feature data of traction energy consumption cannot be fully extracted [14]. This makes accurate predictions hard to be achieved.

To get a better prediction model, ensemble learning can be used to train different prediction models and select appropriate combination methods [15]. Ensemble learning methods can integrate different prediction models or the same type of prediction models. The commonly used ensemble learning methods are Stacking [16], Boosting [17], and Bagging [18]. They have shown excellent performance in load prediction and other fields. For instance, Dong et al [19]. proposed a wind power prediction method based on Stacking ensemble learning and achieved higher accuracy and stability than those single prediction models. Al-Hajj et al [20] proposed A global solar short-term prediction model with integrated stacking learning and compared and analyzed multiple stacking integrated learning structures and cycle models to achieve one-year solar radiation assessment and analysis. Zhang et al [21] proposed four different photovoltaic prediction models of stacking ensemble learning based on random forest, extreme gradient boosting, and other base models. Experimental results showed that the prediction performance of the stacking integrated learning model was better than that of a single model. Therefore, given the excellent prediction performance, it is promising to apply the ensemble learning method in predicting traction energy consumption.

In addition, it is hard to obtain enough characteristic data on traction energy consumption, it is difficult to build a traction energy consumption prediction model with high accuracy. A few data modeling methods, such as transfer learning [22] generative adversarial network (GAN) [23], and other methods have been successfully applied in the field of wind power prediction.

The GAN can learn the distribution of the historical data, and the generated data with GAN have the same statistical properties as historical data [24], [25]. This method can provide a novel solution for the above problems. Based on the game theory, GAN trains both generators and discriminators. Therein, generators use noises to generate new data that match the mathematical distribution of historical data, while discriminators are employed to distinguish the generated data from the original data [26]. Once proposed, GAN has been widely used in many fields of power systems, building power demand prediction [27] daily power demand prediction [28], prediction of the Motor State [29], and battery state prediction [30]. For example, Chen et al. [31]. proposed a data-driven GAN to generate scenarios that capture the spatial and temporal correlations of renewable power plants. This method can generate real and diverse wind and photovoltaic power distribution maps. Whereas, traditional GAN also has the disadvantages of poor generation diversity and schema collapse.

To emphasize the novelty of this work, two concerns regarding the research gaps are further clarified.

1) The first concern is the data shortage issue in new subway lines. The abnormal monitoring of metro lines is based on the difference between predicted and measured values. Therefore, it requires a high prediction accuracy, which is conducive to developing a more energy-saving driving scheduling plan. However, in new subway lines, owing to data shortage, it is difficult to establish a prediction model with high accuracy. To overcome data insufficiency, the WGAN-GP model is first proposed to enhance the data samples so as to address the few-shot learning problem in new subway lines. Compared with the original GAN, the improved GAN can generate high-quality characteristic data of traction energy consumption to meet the needs.

2) The second concern is how to improve prediction accuracy. Since deep learning algorithms require a large demand of data, single shallow-layer network algorithms such as SVM, BP, and random forest may lead to a large deviation in prediction accuracy. To fully utilize their respective advantages, the stacking ensemble learning is first applied to construct a traction energy consumption prediction model. The results in the experimental tests confirm the obvious advantage of our proposed ensemble learning method over the single shallow-layer network algorithms.

Therefore, in this paper, a prediction model for metro traction energy consumption is proposed based on the gradient penalty Wasserstein generative adversarial network (WGAN-GP) and Extreme Gradient Boosting (XGBoost). The main contributions are summarized as follows:

1) To address the data shortage problem, a novel WGAN-GP method for traction energy consumption is proposed. The limited data on traction energy consumption are amplified by the trained WGAN-GP. The data generated by WGAN-GP have the same statistical properties as the real data and thus can be applied in the prediction of traction energy consumption of new Metro lines. The method can effectively generate characteristic data and circumvent the dilemma of data shortage in the prediction of traction energy consumption.

2) To improve prediction accuracy, a stacking ensemble learning method is used to excavate the complex nonlinear relationship of metro traction energy consumption. The prediction model of traction energy consumption is established on the XGBoost ensemble learning model. The stacking ensemble learning model is formed by the concatenation of two-layer models. The model performance can be greatly improved by integrating multiple models.

3) Eleven combinations of four base models are tested with the Stacking ensemble learning method. With single models extracting different feature information, integrated Stacking ensemble learning can improve the prediction performance. The best prediction model, in which the combined algorithm with BP, SVM, ELM, and XGBoost as one layer base model and XGBoost algorithm as the second layer meta model, is proved to be more suitable for traction energy prediction.

The remaining of this paper is organized as follows. Section II introduces the WGAN-GP model. Section III

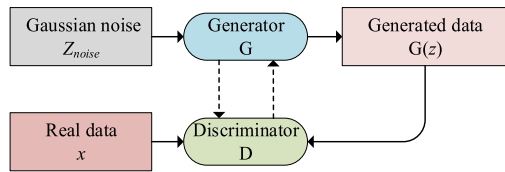


FIGURE 1. Basic structure of GAN.

presents the structure of stacking ensemble learning. In Section IV, the model parameter setting is conducted and the prediction process is demonstrated. Section V compares the effectiveness of WGAN-GP data generation and the performance of different basic models and meta models, and also summarizes the advantages of the proposed method with extensive case studies. Section VI concludes this paper.

II. DATA AMPLIFICATION MODEL

A. GAN MODEL

The basic architecture of GAN is shown in FIGURE 1. Generator G learns the relationship between random signal Z_{noise} and real data R through continuous training, while discriminator D is used to distinguish whether the input data is real data during training.

By training the GAN, the generated data by generator G have the same statistical properties as historical data, cheat discriminant D ; Discriminator D tries its best to identify whether its input data is real data or generated data, and the two constantly play games to finally reach Nash equilibrium.

The generator and discriminator interactively train and compete with each other, so that the network is constantly optimized. Finally, the trained generator can generate high-quality new sample data, but the discriminator cannot distinguish it from the real data. Therefore, the loss functions of the generator and discriminator are defined as:

$$L_G = E_{z \sim P_g(z)} [-\log D(G(z))] \quad (1)$$

$$L_D = E_{x \sim P_{data}} [\log D(x)] - E_{z \sim P_g(z)} [\log(1 - D(G(z)))] \quad (2)$$

where L_D is the generator loss function, L_G is the discriminator loss function, G is the generator, D is the discriminator, $P_g(x)$ is the true distribution of the original data, $z \sim P_g(z)$ is the noise data conforming to the normal distribution, E is the expectation, x is the real data, and z is the gaussian noise.

GAN is a kind of unsupervised learning neural network. It can train models with Gaussian noise as the input and output of the data with a similar distribution pattern to the real data. The training objectives of GAN are defined as follows:

$$\min_G \max_D (D, G) = E_{x \sim P_g(x)} |\log D(x)| + E_{x \sim P_z(z)} |\log[-D(G(z))]| \quad (3)$$

where P_Z is the generated data distribution.

B. WGAN-GP MODEL

Traditional GAN has problems like gradient vanish and cannot learn the distribution law of the characteristic data well.

To alleviate the gradient disappearance and enhance training stability in the original GAN, the Wasserstein generative adversarial network (WGAN) is proposed in [32]. WGAN can effectively utilize the real data and generate new data by minimizing the Wasserstein distance, which can be expressed as:

$$W(P_R, P_Z) = \inf_{\prod(P_R, P_Z)} E_{(x,y) \sim \prod(P_R, P_Z)} (\|x - y\|) \quad (4)$$

where $\prod(P_R, P_Z)$ represents the joint distribution of real data P_R and generated data P_Z , $(x, y) \sim \prod(P_R, P_Z)$ represents sampling from $\prod(P_R, P_Z)$, $W(P_R, P_Z)$ is the lowest value for $\|x - y\|$ of all the x and y that satisfy the distribution, namely the Wasserstein distance.

$$L_D = E_{x \sim P_Z} (f_w(x)) - E_{x \sim P_R} (f_w(x)) + \lambda E_{x \sim P_{\hat{x}}} [\ell (\|\nabla_{\hat{x}} f_w(\hat{x})\|_2 - 1)^2] \quad (5)$$

$$L_G = -E_{x \sim P_Z} (f_w(x)) \quad (6)$$

where L_D is the generator loss function, L_G is the discriminator loss function, $f_w(*)$ is the fitting function of the neural network, λ is the penalty coefficient, $\|*\|_2$ represents the binary norm, and $P_{\hat{x}}$ is the generated sample distribution.

III. ENSEMBLE LEARNING METHOD

A. MACHINE LEARNING ALGORITHM

Popular deep learning methods require massive data for training [33], [34], whereas the traction energy consumption data are limited in new lines. As a result, it would be very hard to build effective deep learning models. Therefore, some excellent machine learning methods are selected to establish the prediction model instead of a single algorithm.

1) XGBOOST ALGORITHM

XGBoost is a Boosting ensemble learning algorithm based on gradient lifting. XGBoost optimizes Boosting algorithm based on the gradient lifting decision tree, reducing the problem that the traditional gradient lifting decision tree is easy to overfit [35], [36], [37], and its model is shown in Equation (7).

$$\hat{y}(i) = \sum_{m=1}^M f_m(x_i), f_m \in F \quad (7)$$

where $\hat{y}(i)$ is the predicted value of the i th sample, M is the number of trees, $f_m(*)$ is the state of the m -th tree, x_i is the i th sample, and F is the set space of trees.

The objective function of XGBoost is shown in Equation (8):

$$\Gamma = \sum_{i=1}^n L[y(i), \hat{y}(i)] + \sum_{m=1}^M \Omega(f_m) \quad (8)$$

where $L[y(i), \hat{y}(i)]$ is the training error between the predicted value and the actual value, and $\Omega(f_m)$ is the tree complexity.

The complexity of the tree is calculated as shown in Equation (9):

$$\Omega(f_m) = \gamma T + \frac{\lambda}{2} \sum_{j=1}^T (\omega_j^2) \quad (9)$$

where γ is the control coefficient of the number of leaf nodes, T is the number of leaf nodes, λ is the control coefficient of leaf node fraction, and ω is the leaf node fraction.

XGBoost algorithm can realize parallel computing for nodes in each layer of the network, which is conducive to improving the training speed of the model. By inputting multiple characteristics of metro train traction energy consumption into the XGBoost model and aiming at measured traction energy consumption, XGBoost's metro train traction energy consumption prediction model can be established. Through several iterations, a good fitting effect can be obtained.

2) EXTREME LEARNING MACHINE

Extreme Learning Machine (ELM) is an improved model of the feedforward neural network, which has the advantages of fast training speed and simple parameter adjustment [38], [39], [40]. ELM, as a new type of feed-forward neural network, randomly generates input weights during training and keeps them unchanged. Only the weights of the training outputs are needed to make the network model constantly approximate the training samples.

3) SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a machine learning method based on statistical learning theory. It is mainly divided into two categories, which are commonly used in classification and nonlinear regression tasks [41], [42], [43]. When SVM is used in a regression task, low-dimensional samples can be mapped to high-dimensional vector space through a nonlinear mapping function $\varphi(x)$ to better solve the small-sample problem. The function relation of SVM is as follows:

$$f_{SVM}(x) = \omega_{SVM} \cdot \varphi(x) + b_{SVM} \quad (10)$$

where $f_{SVM}(x)$ is the predicted value for sample x , ω_{SVM} is the weight coefficient matrix of SVM, and b_{SVM} is the bias coefficient matrix of SVM.

4) ERROR BACKPROPAGATION ALGORITHM

Error back Propagation (BP) neural network is a neural network that adjusts the weights and bias parameters among network layers according to the training errors [44], [45], [46], [47]. At the same time, through different activation functions, the complex nonlinear relations in the features can be further extracted, so that the BP neural network can be infinitely close to the real distribution of features.

B. STACKING ENSEMBLE LEARNING

A single prediction model has limited prediction performance. Ensemble learning can integrate the advantages of

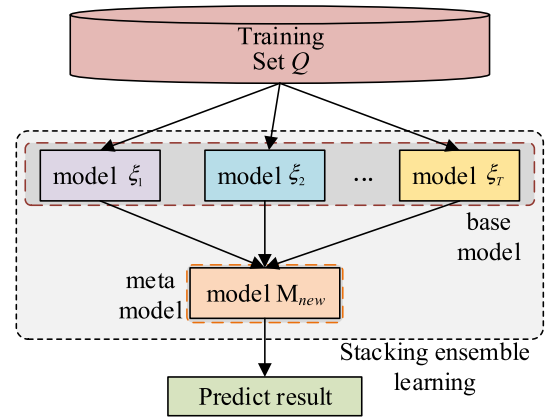


FIGURE 2. Structure of stacking ensemble learning.

single prediction models and build a better prediction model. A large number of studies show that Stacking ensemble learning has excellent prediction performance [48]. In stacking ensemble learning, the training set is divided into several subsets. The prediction results of the first layer model are obtained from the base learning model and implemented as the input features of the meta model. Then, the output of multiple base models is combined to form the input features of the meta model. The rules among these features are learned via the meta model, and the desired prediction results are finally obtained. Stacking ensemble learning improves the performance of the model by integrating multiple models. The stacking ensemble learning model for traction energy consumption prediction is shown in FIGURE 2.

For metro traction energy consumption characteristics of the sequence of $Q = \{(y_n, x_n)\}, n = 1, \dots, N, x_n$ for the first n samples of characteristics, y_n as the traction energy consumption; The sequence Q is randomly divided into K equal subsets S_1, S_2, \dots, S_k . Where $S_{-k} = S - S_k, S_k$ and S_{-k} are respectively defined as the k -th folding test set and training set in k -folding cross verification. For layer 1, the prediction algorithm contains K base learning [49], the training set S_{-k} training gets the base model using the first k algorithm $M_k, k = 1, \dots, K$.

For each sample x_n in the k -folding test set S_k in the k -folding cross validation, the prediction of it by the base learner M_k is expressed as z_{kn} . After completing the cross-validation process, the output data of K base learners is constituted into a new data sample, namely $S_{new} = \{(y_n, z_{1n}, \dots, z_{kn})\}, n = 1, \dots, N$. The second layer M_{new} is trained by using x_n as the input of the second layer Stacking ensemble learning learner.

The detailed implementation of the Stacking ensemble learning algorithm is illustrated in Algorithm 1.

IV. MODEL PARAMETER SETTING AND PREDICTION PROCESS

A. WGAN-GP PARAMETER SETTINGS

The WGAN-GP model of the metro traction energy consumption is shown in Fig.3.

Algorithm 1 Stacking Ensemble Learning Algorithm

Input:

Training data $Q = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, One layer learning algorithm ξ_1, \dots, ξ_T , Two-level learning algorithm ξ ;

Output:

Ensemble learning model M_{new} ;

1. **Step one:** The primary learner is generated by training a layer learning algorithm through the training set dataset Q ;
2. **for** $t = 1$ **to** K **do**
3. $h_t = \xi_t(Q)$;
4. **end for**
5. **Step two:** Create a new training data S_{new} ;
6. $S_{new} = \emptyset$;
7. **for** $i = 1$ **to** n **do**
8. **for** $t = 1$ **to** K **do**
9. $z_{it} = h_t(x_i)$;
10. **end for**
11. $S_{new} = S_{new} \cup \{y_n, z_{1n}, \dots, z_{kn}\}$;
12. **end for**
13. $M_{new} = \xi(S_{new})$;
14. **return** M_{new}

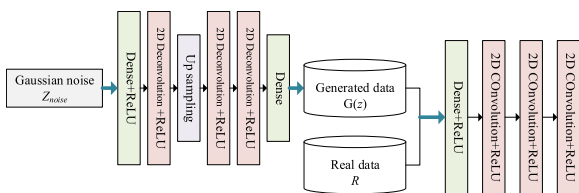


FIGURE 3. Metro traction energy consumption WGAN-GP model.

TABLE 1. The structure information of the generative adversarial network.

Model	Input dimension	Configuration	Output dimension
Generative network	1×100	$n_{Dconv}:3, kernel\ size: 3 \times 3, n_{up}:1, n_{fc}:5$	1×5
Adversarial network	1×5	$n_{conv}:3, kernel\ size: 3 \times 3, n_c:1$	1×1

In this paper, the deconvolution layer and the convolution layer are used to build the WGAN-GP model of metro traction energy consumption. The generator is composed of 5 hidden layers, while the discriminator D is composed of 4 hidden layers. The basic parameters are shown in **Table 1**.

where n_{Dconv} is the number of deconvolution layers, *kernel size* is the size of the convolution kernel, n_{up} is the number of upsampling layers, n_{fc} is the number of neurons in the fully connected layer, and n_{conv} is the number of convolution layers.

The input of generator G is gaussian noise, and the output is traction energy consumption data of the metro train. The input of discriminator D is the real data set of metro traction energy consumption and the generated data set to form a 1×5 tensor. In the training process, both generator G and discriminator D use the Adam optimization algorithm, and the initial learning rate is 0.001. After training, generator G can generate traction energy consumption data consistent

TABLE 2. Stacking ensemble learning parameter settings.

	model	parameter settings
first layer	XGBoost	<i>booster: gbtree, colsample_bytree: 0.8</i>
	SVM	<i>C: 0.01, kernel: rbf</i>
	BP	$n_c: 3, activation\ function: relu, optimizer: Adam$
	ELM	$n_{layers}: 3, kernel\ function: liner$
second layer	XGBsoost	<i>booster: gbtree, colsample_bytree: 0.6</i>

with the real train, so as to realize the data amplification of the traction energy consumption.

B. STACKING ENSEMBLE LEARNING PARAMETER SETTINGS

In this paper, the stacking ensemble learning model is divided into two layers. The first layer uses XGBoost, SVM, BP, and ELM models as the base learning model, and the second layer uses XGBoost as the meta-learning model.

Detailed parameter Settings of different models are shown in Table 2.

C. WGAN-GP AND STACKING PREDICTION PROCESS

The flow chart of traction energy consumption prediction is illustrated in **FIGURE 4** below.

According to **FIGURE 4**, the prediction process mainly consists of three steps: 1) data preprocessing, 2) WGAN-GP model training and data generation, and 3) base learning model training and meta-learning model training.

1) Data preprocessing: Eliminate the influence of outliers and extreme values through data processing, and provide an excellent data source for data generation and prediction model.

2) WGAN-GP model training and data generation: The WGAN-GP data generation model suitable for metro train traction energy consumption is trained and established to provide sufficient data for the prediction model.

3) Base learning model training and meta-learning model training: Fully mine the characteristic data of traction energy consumption, establish the stacking ensemble learning model of traction energy consumption prediction, and conduct experimental verification.

V. CASE STUDIES

To verify the effectiveness of the proposed prediction model in this paper, this section employs the measured data of Guangzhou Metro for verification. All simulation experiments are implemented with Keras deep learning framework under Python 3.8. The configuration of the simulation platform is Intel core i5-10210U processors running at 2.11 GHz with a memory capacity of 16 GB under Windows 11 Operating System.

A. DATA DESCRIPTION

In this paper, a total of 120 data pieces are collected including the daily train operation energy consumption, auxiliary equipment energy consumption, passenger flow, depot

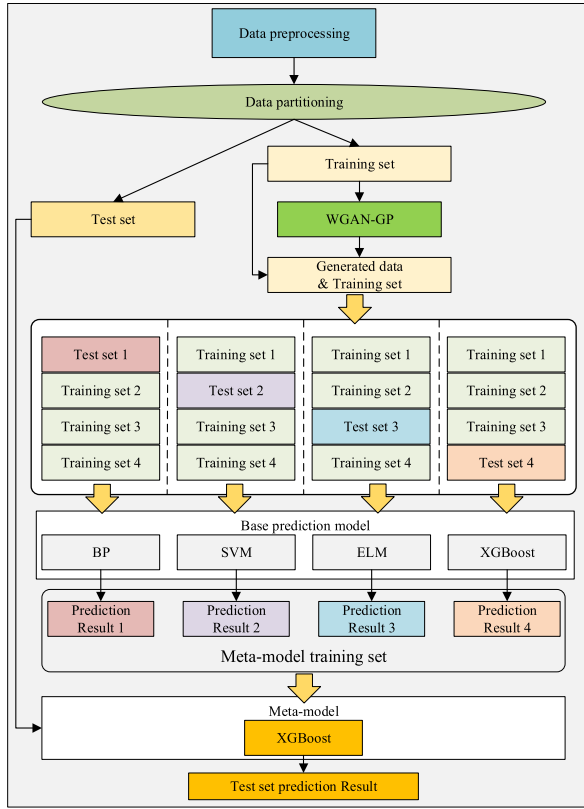


FIGURE 4. WGAN-GP-Stacking prediction process.

commissioning mileage, positive commissioning mileage, passenger car mileage, empty train number, and actual train number from January to February 2022 in Guangzhou Metro Line 13 (first phase, Yuzhu station – Xinsha Station).

The first 60 points are set as the training set and the last 60 points as the test set. All features are mapped to the interval [0,1] according to the following formula:

$$x_{std}^i = \frac{x^i - x_{min}^i}{x_{max}^i - x_{min}^i}(\mu_u - \mu_d) + \mu_d \quad (11)$$

where x_{std}^i is the value after mapping the i -th feature, x^i is the original value of the i -th feature, x_{max}^i and x_{min}^i are the maximum and minimum values of the i -th feature respectively, μ_u and μ_d are the lower and upper bounds of normalization, in this paper $\mu_u = 1$ and $\mu_d = 0$.

To reduce the difficulty of model training, the Pearson correlation coefficient is applied to select the above 8 features, as displayed in **Table 3**.

To select suitable variables as the input of the prediction model, the Pearson prod-moment correlation coefficient is calculated in **Table 2**. It can be figured that the correlation coefficient between some variables is high, such as actual train number, and passenger car mileage. But some variables have lower correlation coefficients, which indicates that these variables have less relation with the traction energy consumption and should be discarded. Thus, the top four variables with the largest correlation coefficients are selected

TABLE 3. Characteristic correlation coefficient.

feature	Pearson values
Traction energy consumption	1.0000
Energy consumption of auxiliary equipment	0.5047
Passenger flow	0.4886
Positive debugging mileage	0.0947
Depot commissioning mileage	0.1247
Passenger car mileage	0.6726
Empty train number	0.3642
Actual train number	0.6567

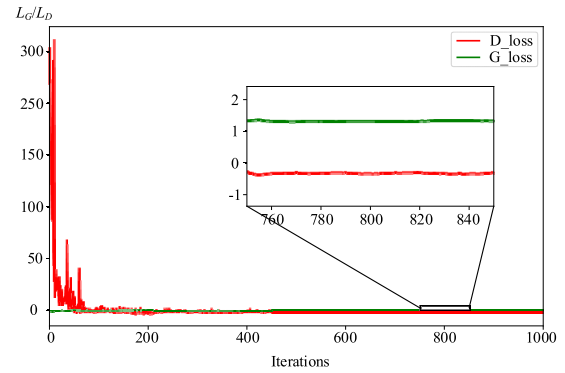


FIGURE 5. Training loss curve of WGAN-GP.

as input features, viz, the actual train number, passenger car mileage, passenger flow, and energy consumption of auxiliary equipment.

B. PREDICTION PERFORMANCE EVALUATION INDICES

To evaluate the prediction performance, mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) are chosen as the evaluation indices. The corresponding calculation formulas are expressed in Equations (12) - (14).

$$e_{MAE} = \frac{1}{N} \left(\sum_{i=1}^N |y(i) - \hat{y}(i)| \right) \quad (12)$$

$$e_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N |y(i) - \hat{y}(i)|^2} \quad (13)$$

$$e_{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y(i) - \hat{y}(i)}{y(i)} \right| \times 100\% \quad (14)$$

where N is the number of samples, $y(i)$ is the measured traction energy consumption value, and $\hat{y}(i)$ is the predicted traction energy consumption value.

As can be seen from formula (12)-(14), the smaller the values of e_{MAE} , e_{RMSE} and e_{MAPE} , the better the prediction effect.

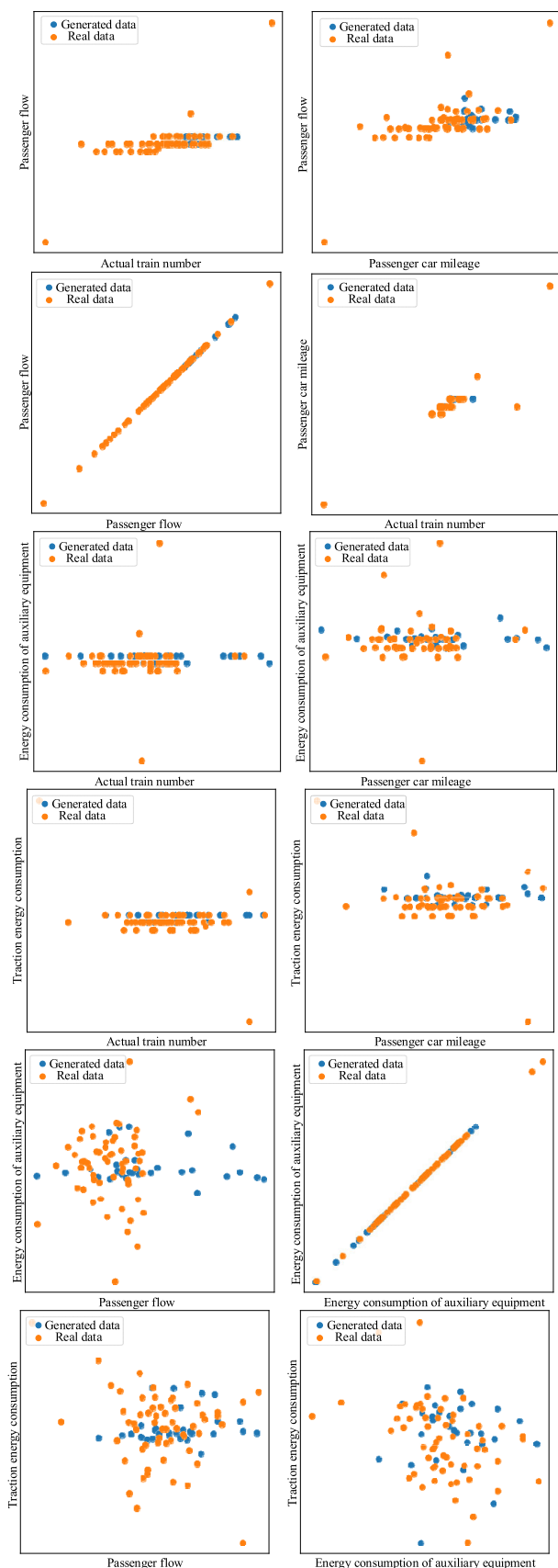


FIGURE 6. Characteristic distribution of traction energy consumption.

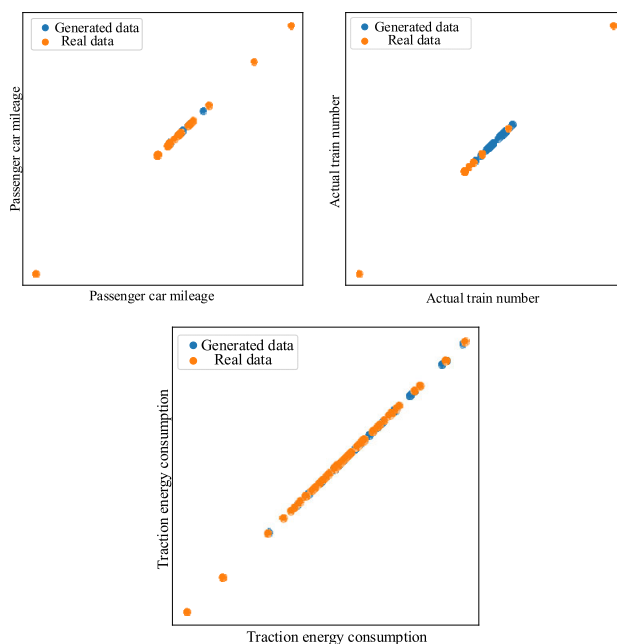


FIGURE 6. (Continued.) Characteristic distribution of traction energy consumption.

TABLE 4. Feature-similarity degree.

feature	Pearson values
Traction energy consumption	0.9071
Energy consumption of auxiliary equipment	0.7263
Passenger flow	0.8457
Passenger car mileage	0.7140
Actual train number	0.7494

TABLE 5. Prediction index results of the base model.

Data	Base model	ϵ_{RMSE} (kWh · 100 vehicles /km)	ϵ_{MAE} (kWh · 100 vehicles /km)	ϵ_{MAPE} (%)
Raw data	BP	15.55	10.57	5.19%
	SVM	14.15	11.10	5.59%
	ELM	19.75	14.15	6.88%
	XGBoost	13.45	10.02	5.03%
Data generated by GAN	BP	14.32	10.15	5.11%
	SVM	14.10	10.77	5.21%
	ELM	19.27	13.98	6.72%
	XGBoost	13.41	9.79	4.89%
Data generated by WGAN-GP	BP	13.85	10.04	4.96%
	SVM	13.2	10.1	5.07%
	ELM	19.22	13.74	6.68%
	XGBoost	13.36	9.54	4.70%

C. CASE 1: GENERATION AND ANALYSIS OF WGAN-GP TRACTION ENERGY CONSUMPTION DATA

To verify the effectiveness of WGAN-GP, 30 data are generated with WGAN-GP to expand the data by 50%. Fig.5 shows

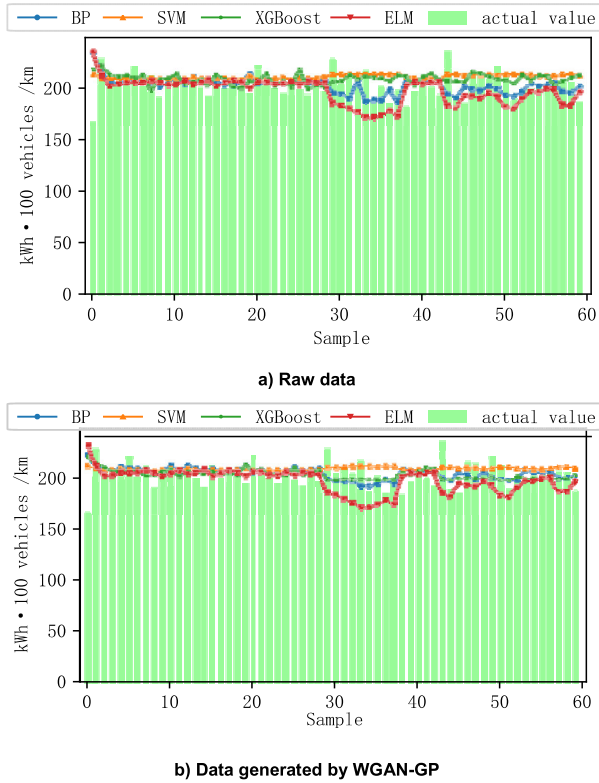


FIGURE 7. Prediction curves of four base models.

TABLE 6. Prediction index results of different meta model.

meta model	e_{RMSE} (kWh·100 vehicles /km)	e_{MAE} (kWh·100 vehicles /km)	e_{MAPE} (%)
BP	13.17	9.58	4.71%
SVM	13.13	9.98	5.01%
ELM	13.43	9.74	4.79%
XGBoost	12.05	8.72	4.31%

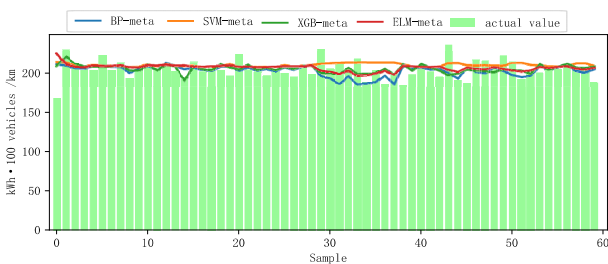


FIGURE 8. Different meta model predicted curve.

the training loss curve of the generator and adversarial, and Fig.6 presents the distribution of the generated data and real data.

As can be seen from Fig.5, in the multi-iteration training of the proposed WGAN-GP model, the loss of generator and adversarial tends to be stable after 800 iterations. The final generator loss function value is stable at 1.3 ~ 1.4, and the loss function value of the determination device is stable

TABLE 7. Stacking ensemble learning prediction index results of different base model combinations.

Model combination	e_{RMSE} (kWh·100 vehicles /km)	e_{MAE} (kWh·100 vehicles /km)	e_{MAPE} (%)
BP+SVM+ELM+XGBoost	12.05	8.72	4.31%
BP+SVM+XGBOOST	12.56	9.59	4.73%
BP+SVM+ELM	13.26	10.05	4.93%
BP+XGBOOST+ELM	13.09	9.83	4.83%
SVM+XGBOOST+ELM	12.05	8.76	4.32%
BP+SVM	13.84	10.11	5.00%
BP+XGBOOST	13.90	10.93	5.50%
BP+ELM	13.07	9.88	4.94%
SVM+XGBOOST	12.55	9.26	4.66%
SVM+ELM	12.49	9.22	4.57%
XGBOOST+ELM	12.61	9.36	4.62%

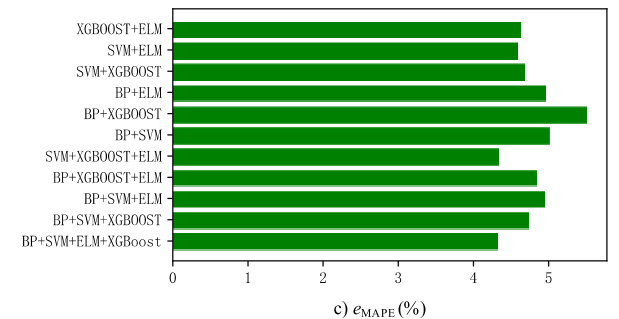
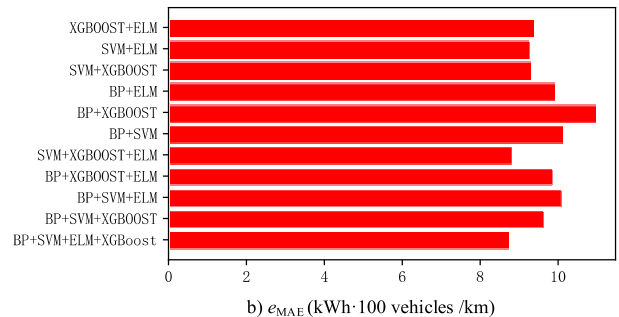
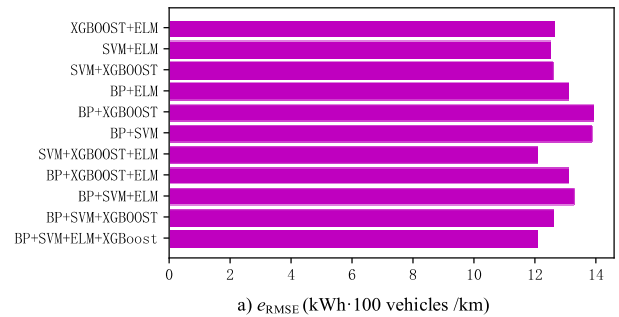


FIGURE 9. Prediction index results of different base model combinations.

between $-0.33 \sim -0.15$. Therefore, the iteration number of the WGAN-GP model is set to 800.

From FIGURE 6 characteristic distribution of traction energy consumption, the distribution of generated data basically consistent with that of real data, indicating that the data

TABLE 8. Time complexity.

Method	Base model	Time complexity	Time of training (s)	Time of prediction (s)
Raw data	BP	$O(9n)$	23.17	0.68
	SVM	$O(n^2)$	19.36	0.74
	ELM	$O(5n)$	17.82	0.69
	XGBoost	$O(45+5\log n)$	20.63	0.68
Data generated by GAN	BP	$O(9n)$	56.32	1.69
	SVM	$O(n^2)$	56.93	1.41
	ELM	$O(5n)$	56.26	1.76
	XGBoost	$O(45+5\log n)$	56.48	1.94
Data generated by WGAN-GP	BP	$O(9n)$	57.36	1.73
	SVM	$O(n^2)$	56.88	1.56
	ELM	$O(5n)$	57.36	1.81
	XGBoost	$O(45+5\log n)$	57.98	1.63
Data generated by WGAN-GP and different meta model	BP	$O(45+23n+n^2+5\log n)$	77.52	2.31
	SVM	$O(45+14n+2n^2+5\log n)$	76.85	2.33
	ELM	$O(45+19n+n^2+5\log n)$	76.42	2.29
	XGBoost	$O(90+14n+n^2+10\log n)$	76.98	2.35
Data generated by WGAN-GP and different base model combinations	BP+SVM+ELM+XGBoost	$O(90+14n+n^2+10\log n)$	76.98	2.35
	BP+SVM+XGBOOST	$O(90+9n+n^2+10\log n)$	75.49	2.33
	BP+SVM+ELM	$O(45+14n+n^2+5\log n)$	75.69	2.17
	BP+XGBOOST+ELM	$O(90+14n+10\log n)$	75.96	2.36
	SVM+XGBOOST+ELM	$O(90+5n+n^2+10\log n)$	75.51	2.22
	BP+SVM	$O(45+9n+n^2+5\log n)$	74.36	2.06
	BP+XGBOOST	$O(90+9n+10\log n)$	74.15	2.11
	BP+ELM	$O(45+14n+n^2+5\log n)$	74.10	1.98
	SVM+XGBOOST	$O(90+n^2+10\log n)$	74.73	2.05
	SVM+ELM	$O(45+14n+n^2+5\log n)$	74.36	2.13
	XGBOOST+ELM	$O(90+5n+10\log n)$	74.23	2.22

generation model is effective. From **Table 4**, the similarity between the generated features and the original features is higher than 0.7, which indicates that the generated data has a strong correlation with the original data.

D. CASE 2: EFFECTIVENESS ANALYSIS OF WGAN-GP GENERATED DATA

To further verify the data validity of WGAN-GP, four base models i.e., XGBoost, SVM, ELM, and BP are employed in the comparative analysis. They share the same input with the original metro train feature series and generate new feature series respectively. The prediction indicators of the four base models are shown in **Table 5**, and the predicted curves are drawn in **FIGURE 7**.

The following conclusions can be drawn from **Table 5** and **FIGURE 7**:

(1) Compared with BP, SVM and ELM, the XGBoost model achieves the best prediction performance with all three evaluation indices. Statistically speaking, in the prediction with original data, the e_{MAE} index is reduced by 5.19%, 9.72%, and 29.19%, respectively, the e_{MAPE} index is reduced by 3.14%, 10.05%, and 26.90%, respectively, and the e_{RMSE} index is decreased by 13.50%, 4.95%, and 31.90%, respectively. Therefore, the XGBoost model is selected to build the prediction model for traction energy consumption.

(2) Compared with the original data, the data generated by GAN play a more significant role in improving the prediction performance in all four base models. Take the MAE indicator as an example, the e_{MAE} indexes are reduced by 3.97%, 2.97%, 1.20%, and 2.30 with BP, the SVM, ELM and XGBoost respectively. In addition, the e_{RMSE} and e_{MAPE} indexes are also decreased with the data from GAN.

TABLE 9. Model stability verification with repetitive experiments.

Method	Base model	Index	Standard deviation e_{std}
Raw data	BP	e_{RMSE}	0.262
		e_{MAE}	0.253
		e_{MAPE}	0.256
	SVM	e_{RMSE}	0.211
		e_{MAE}	0.201
		e_{MAPE}	0.297
	ELM	e_{RMSE}	0.312
		e_{MAE}	0.290
		e_{MAPE}	0.293
	XGBOOST	e_{RMSE}	0.233
		e_{MAE}	0.194
		e_{MAPE}	0.216
Data generated by GAN	BP	e_{RMSE}	0.264
		e_{MAE}	0.267
		e_{MAPE}	0.258
	SVM	e_{RMSE}	0.231
		e_{MAE}	0.221
		e_{MAPE}	0.289
	ELM	e_{RMSE}	0.311
		e_{MAE}	0.287
		e_{MAPE}	0.265
	XGBOOST	e_{RMSE}	0.241
		e_{MAE}	0.187
		e_{MAPE}	0.232
Data generated by WGAN-GP	BP	e_{RMSE}	0.247
		e_{MAE}	0.248
		e_{MAPE}	0.267
	SVM	e_{RMSE}	0.261
		e_{MAE}	0.241
		e_{MAPE}	0.257
	ELM	e_{RMSE}	0.341
		e_{MAE}	0.247
		e_{MAPE}	0.235
	XGBOOST	e_{RMSE}	0.241
		e_{MAE}	0.227
		e_{MAPE}	0.236
Data generated by WGAN-GP and different meta model	BP	e_{RMSE}	0.255
		e_{MAE}	0.269
		e_{MAPE}	0.253
	SVM	e_{RMSE}	0.231
		e_{MAE}	0.256
		e_{MAPE}	0.259

TABLE 9. (Continued.) Model stability verification with repetitive experiments.

Data generated by WGAN-GP and different base model combinations	ELM	ϵ_{RMSE}	0.312
		ϵ_{MAE}	0.296
		ϵ_{MAPE}	0.287
	XGBOOST	ϵ_{RMSE}	0.233
		ϵ_{MAE}	0.236
		ϵ_{MAPE}	0.222
	BP+SVM+ELM+XGBoost	ϵ_{RMSE}	0.176
		ϵ_{MAE}	0.159
		ϵ_{MAPE}	0.168
	BP+SVM+XGBOOST	ϵ_{RMSE}	0.196
		ϵ_{MAE}	0.186
		ϵ_{MAPE}	0.187
	BP+SVM+ELM	ϵ_{RMSE}	0.195
		ϵ_{MAE}	0.189
		ϵ_{MAPE}	0.176
	BP+XGBOOST+ELM	ϵ_{RMSE}	0.199
		ϵ_{MAE}	0.201
		ϵ_{MAPE}	0.203
	SVM+XGBOOST+ELM	ϵ_{RMSE}	0.198
		ϵ_{MAE}	0.203
		ϵ_{MAPE}	0.210
	BP+SVM	ϵ_{RMSE}	0.232
		ϵ_{MAE}	0.233
		ϵ_{MAPE}	0.225
	BP+ELM	ϵ_{RMSE}	0.216
		ϵ_{MAE}	0.209
		ϵ_{MAPE}	0.217
SVM+XGBOOST	ϵ_{RMSE}	0.196	
	ϵ_{MAE}	0.200	
	ϵ_{MAPE}	0.197	
SVM+ELM	ϵ_{RMSE}	0.231	
	ϵ_{MAE}	0.253	
	ϵ_{MAPE}	0.234	
XGBOOST+ELM	ϵ_{RMSE}	0.207	
	ϵ_{MAE}	0.216	
	ϵ_{MAPE}	0.207	
BP+XGBOOST	ϵ_{RMSE}	0.219	
	ϵ_{MAE}	0.213	
	ϵ_{MAPE}	0.214	

(3) Compared with the GAN generated data, the data generated by WGAN-GP play a more significant role in improving the prediction performance in all four base models. Take the MAE indicator as an example, the e_{MAE} indexes are reduced by 1.08%, 6.22%, 1.72%, and 2.55% with BP, the SVM, ELM and XGBoost respectively. In addition, the e_{RMSE} and e_{MAPE} indexes are also decreased with the data from WGAN-GP. Therefore, the WGAN-GP can provide useful data for traction energy consumption and effectively improve the prediction performance.

E. CASE 3: PERFORMANCE VALIDATION OF DIFFERENT META MODELS

To verify the effect of different meta models, based on the data generated by the characteristics of metro train sequences, with four different models (XGBoost, SVM, ELM, and BP model) as a meta model to constitute a stacking ensemble learning model. **Table 5** shows the predicted index results, and **FIGURE 8** shows the predicted curve.

From **Table 6** and **FIGURE 8**, four meta models show different performances, and the best performance is the XGBoost meta model, and the worst effect is the ELM meta model. Compared with BP, SVM, and ELM, e_{RMSE} index of XGBoost decreased by 8.50%, 8.23%, and 10.28%, respectively. e_{RMSE} index decreased by 8.98%, 12.63% and 10.47%, respectively. e_{MAPE} index decreased by 8.49%, 13.97% and 10.02%, respectively. The experimental results show that XGBoost is more suitable as a stacking ensemble learning meta model for traction energy consumption prediction. This is because the XGBoost model selects nodes with the largest information gain to construct the Classification and Regression Tree, and reduces the overfitting. In addition, the model can train in parallel in feature granularity and improve the training speed.

F. CASE 4: STACKING ENSEMBLE LEARNING VALIDATION OF DIFFERENT BASE MODEL COMBINATIONS

To verify the combinations of different base models, using XGBoost as the meta model, different stacking ensemble models are built. The same data are used for analysis. The index results are shown in **Table 7**.

From **Table 7** and **FIGURE 9**, in all indexes, the first combination, viz, BP, SVM, ELM, and XGboost, achieves the best performance. Therefore, the stacking ensemble prediction model, in which the combined algorithm with BP, SVM, ELM, and XGBoost as one layer base model and XGBoost algorithm as the second layer meta model, is more suitable for traction energy consumption prediction. Different single models can extract different feature information, and then be integrated by Stacking ensemble learning, which is conducive to improving the prediction performance.

The model time complexity and running time of different models are shown in **Table 8**. It can be found that the time complexity of the generation model and ensemble learning prediction model proposed is $O(n^2)$, and thus guarantees a relatively fast runtime to obtain the prediction results.

G. CASE 5: STABILITY VERIFICATION

To further verify the stability of the prediction model, all models were repeated 30 times, and the standard deviation (e_{std}) of each evaluation index was calculated based on equation (15).

$$e_{std} = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n}} \quad (15)$$

Table 9 shows the experimental results of different methods involved in this paper. It can be seen that the standard deviation of all models is between 0.1 and 0.3, showing good stability, which confirms the effectiveness of the proposed method.

VI. CONCLUSION

In this paper, a novel prediction model for metro traction energy consumption is proposed. In addition to limited original data, extra data are generated by WGAN-GP using gradient penalty to alleviate the gradient disappearance and enhance training stability in the original GAN. The stacking ensemble learning with multiple models is employed to improve the prediction performance. Satisfactory predictions are obtained and verified by extensive experiments. Conclusions are drawn as follows:

- 1) The WGAN-GP model can learn the distribution patterns of traction energy consumption and the characteristics of metro trains. WGAN-GP can provide the data needed for prediction and effectively improve prediction accuracy. This is particularly suitable for predicting the traction energy consumption of new lines.
- 2) The stacking ensemble prediction model, in which the combined algorithm with BP, SVM, ELM, and XGBoost algorithms work as one layer base model and XGBoost algorithm as the second layer meta model, can offer the best prediction performance.
- 3) The ensemble learning model with other single prediction models can still achieve good results, and thus can also be applied in other prediction programs accordingly. For example, high-speed railway or load prediction of a new industrial park.

REFERENCES

- [1] A. González-Gil, R. Palacin, P. Batty, and J. P. Powell, "A systems approach to reduce urban rail energy consumption," *Energy Convers. Manag.*, vol. 80, pp. 509–524, Apr. 2014, doi: [10.1016/j.enconman.2014.01.060](https://doi.org/10.1016/j.enconman.2014.01.060).
- [2] H. Liu and B. Liu, "Research on the Beijing subway energy management platform," *Cyberspace Secur.*, vol. 7, no. 1, pp. 63–68&79, 2016.
- [3] X. Li and L. Zhang, "Outlier analysis of urban rail traction energy consumption based on times series," *J. Beijing Jiaotong Univ.*, vol. 45, no. 5, pp. 30–36, 2021.
- [4] L. Zhang, J. Xun, W. Zhang, X. Li, and Y. Zhang, "Pattern recognition of traction energy consumption for urban rail transit by using symbolic aggregate approximation," in *Proc. IEEE 10th Data Driven Control Learn. Syst. Conf. (DDCLS)*, May 2021 pp. 242–247.
- [5] C. A. De Sousa, S. L. Pereira, and E. M. Dias, "Improvement of the energy efficiency of subway traction systems through the use of genetic algorithm in traffic control," *J. Control Autom. Elect. Syst.*, vol. 30, no. 1, pp. 85–94, Feb. 2019, doi: [10.1007/s40313-018-00429-8](https://doi.org/10.1007/s40313-018-00429-8).

- [6] C. Wang, X. Wang, and Y. Wang, "Research on improvement of urban rail transit energy consumption model," in *Proc. Int. Conf. Smart Transp. City Eng.*, Nov. 2021, pp. 517–522, doi: [10.1117/12.2613940](https://doi.org/10.1117/12.2613940).
- [7] X. Wang, T. Tang, S. Su, J. Yin, Z. Gao, and N. Lv, "An integrated energy-efficient train operation approach based on the space-time-speed network methodology," *Transp. Res. E, Logistics Transp. Rev.*, vol. 150, Jun. 2021, Art. no. 102323, doi: [10.1016/j.tre.2021.102323](https://doi.org/10.1016/j.tre.2021.102323).
- [8] J. Lesel, D. Bourdon, G. Claisse, P. Debay, and B. Robyns, "Real time electrical power estimation for the energy management of automatic metro lines," in *Proc. 11th Int. Conf. Model. Simul. Electr. Mach. Convert. Syst.*, vol. 131, Jan. 2017, pp. 3–20, doi: [10.1016/j.matcom.2016.06.003](https://doi.org/10.1016/j.matcom.2016.06.003).
- [9] P. C. M. Leung and E. W. M. Lee, "Estimation of electrical power consumption in subway station design by intelligent approach," *Appl. Energy*, vol. 101, pp. 634–643, Jan. 2013, doi: [10.1016/j.apenergy.2012.07.017](https://doi.org/10.1016/j.apenergy.2012.07.017).
- [10] S. Zhou, Y. Bai, B. Yuan, Q. Wang, and J. Li, "Traction energy consumption prediction of new metro lines based on simulation combining support vector regression," *J. Railway Sci. Eng.*, vol. 18, no. 10, pp. 2733–2740, 2021.
- [11] P. M. Fernández, C. G. Román, and R. I. Franco, "Modelling electric trains energy consumption using neural networks," *Transp. Res. Proc.*, vol. 18, pp. 59–65, Jan. 2016.
- [12] H. Lyu and Y. Zhang, "Research on prediction of traction energy-consumption of subway train based on machine learning," *J. Railway Sci. Eng.*, vol. 16, no. 7, pp. 1833–1841, 2019.
- [13] Z. Tang, H. Yin, C. Yang, J. Yu, and H. Guo, "Predicting the electricity consumption of urban rail transit based on binary nonlinear fitting regression and support vector regression," *Sustain. Cities Soc.*, vol. 66, Mar. 2021, Art. no. 102690, doi: [10.1016/j.scs.2020.102690](https://doi.org/10.1016/j.scs.2020.102690).
- [14] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers Comput. Sci.*, vol. 14, no. 2, pp. 241–258, 2020, doi: [10.1007/s11704-019-8208-z](https://doi.org/10.1007/s11704-019-8208-z).
- [15] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, and V. Makarenkov, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021, doi: [10.1016/j.inffus.2021.05.008](https://doi.org/10.1016/j.inffus.2021.05.008).
- [16] J. Dou, A. P. Yunus, D. T. Bui, A. Merghadi, M. Sahana, Z. Zhu, C.-W. Chen, Z. Han, and B. T. Pham, "Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan," *Landslides*, vol. 17, no. 3, pp. 641–658, Mar. 2020, doi: [10.1007/s10346-019-01286-5](https://doi.org/10.1007/s10346-019-01286-5).
- [17] M. H. D. M. Ribeiro, R. G. Da Silva, S. R. Moreno, V. C. Mariani, and L. D. S. Coelho, "Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 136, Mar. 2022, Art. no. 107712, doi: [10.1016/j.ijepes.2021.107712](https://doi.org/10.1016/j.ijepes.2021.107712).
- [18] W. Wu and M. Peng, "A data mining approach combining K -means clustering with bagging neural network for short-term wind power forecasting," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 979–986, Aug. 2017, doi: [10.1109/JIOT.2017.2677578](https://doi.org/10.1109/JIOT.2017.2677578).
- [19] Y. Dong, H. Zhang, C. Wang, and X. Zhou, "Wind power forecasting based on stacking ensemble model, decomposition and intelligent optimization algorithm," *Neurocomputing*, vol. 462, pp. 169–184, Oct. 2021, doi: [10.1016/j.neucom.2021.07.084](https://doi.org/10.1016/j.neucom.2021.07.084).
- [20] R. Al-Hajj, A. Assi, and M. Fouad, "Short-term prediction of global solar radiation energy using weather data and machine learning ensembles: A comparative study," *J. Sol. Energy Eng.*, vol. 143, no. 5, Oct. 2021, Art. no. 051003, doi: [10.1115/1.4049624](https://doi.org/10.1115/1.4049624).
- [21] H. Zhang and T. Zhu, "Stacking model for photovoltaic-power-generation prediction," *Sustainability*, vol. 14, no. 9, p. 5669, May 2022, doi: [10.3390/su14095669](https://doi.org/10.3390/su14095669).
- [22] H. Yin, Z. Ou, J. Fu, Y. Cai, S. Chen, and A. Meng, "A novel transfer learning approach for wind power prediction based on a serio-parallel deep learning architecture," *Energy*, vol. 234, Nov. 2021, Art. no. 121271, doi: [10.1016/j.energy.2021.121271](https://doi.org/10.1016/j.energy.2021.121271).
- [23] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018, doi: [10.1109/TGRS.2018.2805286](https://doi.org/10.1109/TGRS.2018.2805286).
- [24] N. Othberdout, M. Daoudi, A. Kacem, L. Ballihi, and S. Berretti, "Dynamic facial expression generation on Hilbert hypersphere with conditional Wasserstein generative adversarial nets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 848–863, Feb. 2022, doi: [10.1109/TPAMI.2020.3002500](https://doi.org/10.1109/TPAMI.2020.3002500).
- [25] K. Yan, J. Su, J. Huang, and Y. Mo, "Chiller fault diagnosis based on VAE-enabled generative adversarial networks," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 1, pp. 387–395, Jan. 2022, doi: [10.1109/TASE.2020.3035620](https://doi.org/10.1109/TASE.2020.3035620).
- [26] H. Yin, Z. Ou, Z. Zhu, X. Xu, J. Fan, and A. Meng, "A novel asexual-reproduction evolutionary neural network for wind power prediction based on generative adversarial networks," *Energy Convers. Manag.*, vol. 247, Nov. 2021, Art. no. 114714, doi: [10.1016/j.enconman.2021.114714](https://doi.org/10.1016/j.enconman.2021.114714).
- [27] Y. Ye, M. Strong, Y. Lou, C. A. Faulkner, W. Zuo, and S. Upadhyaya, "Evaluating performance of different generative adversarial networks for large-scale building power demand prediction," *Energy Buildings*, vol. 269, Aug. 2022, Art. no. 112247, doi: [10.1016/j.enbuild.2022.112247](https://doi.org/10.1016/j.enbuild.2022.112247).
- [28] C. Tian, Y. Ye, Y. Lou, W. Zuo, G. Zhang, and C. Li, "Daily power demand prediction for buildings at a large scale using a hybrid of physics-based model and generative adversarial network," *Building Simul.*, vol. 15, no. 9, pp. 1685–1701, Sep. 2022, doi: [10.1007/s12273-022-0887-y](https://doi.org/10.1007/s12273-022-0887-y).
- [29] C. Shi, G. Tang, Y. Li, and X. Hu, "The prediction of the motor state of the shore bridge based on the generative resistance network GAN," in *Proc. 2nd Int. Conf. Artif. Intell. Appl. Technol.*, vol. 435, 2018, Art. no. 012010, doi: [10.1088/1757-899X/435/1/012010](https://doi.org/10.1088/1757-899X/435/1/012010).
- [30] H. Zhang, W. Tang, W. Na, P.-Y. Lee, and J. Kim, "Implementation of generative adversarial network-CLS combined with bidirectional long short-term memory for lithium-ion battery state prediction," *J. Energy Storage*, vol. 31, Oct. 2020, Art. no. 101489, doi: [10.1016/j.est.2020.101489](https://doi.org/10.1016/j.est.2020.101489).
- [31] Y. Chen, Y. Wang, D. S. Kirschen, and B. Zhang, "Model-free renewable scenario generation using generative adversarial networks," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3265–3275, May 2018, doi: [10.1109/TPWRS.2018.2794541](https://doi.org/10.1109/TPWRS.2018.2794541).
- [32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
- [33] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019, doi: [10.1109/ACCESS.2019.2912200](https://doi.org/10.1109/ACCESS.2019.2912200).
- [34] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Convers. Manag.*, vol. 198, Oct. 2019, Art. no. 111799, doi: [10.1016/j.enconman.2019.111799](https://doi.org/10.1016/j.enconman.2019.111799).
- [35] J. Fan, X. Wang, L. Wu, H. Zhou, F. Zhang, X. Yu, X. Lu, and Y. Xiang, "Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China," *Energy Convers. Manag.*, vol. 164, pp. 102–111, May 2018, doi: [10.1016/j.enconman.2018.02.087](https://doi.org/10.1016/j.enconman.2018.02.087).
- [36] X. Shi, Y. D. Wong, M. Z.-F. Li, C. Palanisamy, and C. Chai, "A feature learning approach based on XGBoost for driving assessment and risk prediction," *Accident Anal. Prevention*, vol. 129, pp. 170–179, Aug. 2019, doi: [10.1016/j.aap.2019.05.005](https://doi.org/10.1016/j.aap.2019.05.005).
- [37] Z. Wang, T. Hong, and M. A. Piette, "Building thermal load prediction through shallow machine learning and deep learning," *Appl. Energy*, vol. 263, Apr. 2020, Art. no. 114683, doi: [10.1016/j.apenergy.2020.114683](https://doi.org/10.1016/j.apenergy.2020.114683).
- [38] S. Ding, H. Zhao, Y. Zhang, X. Xu, and R. Nie, "Extreme learning machine: Algorithm, theory and applications," *Artif. Intell. Rev.*, vol. 44, no. 1, pp. 103–115, 2015, doi: [10.1007/s10462-013-9405-z](https://doi.org/10.1007/s10462-013-9405-z).
- [39] J. Fan, L. Wu, F. Zhang, H. Cai, W. Zeng, X. Wang, and H. Zou, "Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China," *Renew. Sustain. Energy Rev.*, vol. 100, pp. 186–212, Feb. 2019, doi: [10.1016/j.rser.2018.10.018](https://doi.org/10.1016/j.rser.2018.10.018).
- [40] C. Wu, J. Wang, X. Chen, P. Du, and W. Yang, "A novel hybrid system based on multi-objective optimization for wind speed forecasting," *Renew. Energy*, vol. 146, pp. 149–165, Feb. 2020, doi: [10.1016/j.renene.2019.04.157](https://doi.org/10.1016/j.renene.2019.04.157).
- [41] J. Nalepa and M. Kawulok, "Selecting training sets for support vector machines: A review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 857–900, Aug. 2019, doi: [10.1007/s10462-017-9611-1](https://doi.org/10.1007/s10462-017-9611-1).
- [42] J. Ma, D. Xia, H. Guo, Y. Wang, X. Niu, Z. Liu, and S. Jiang, "Metaheuristic-based support vector regression for landslide displacement prediction: A comparative study," *Landslides*, vol. 19, no. 10, pp. 2489–2511, Oct. 2022, doi: [10.1007/s10346-022-01923-6](https://doi.org/10.1007/s10346-022-01923-6).

- [43] A. Zendejboudi, M. A. Baseer, and R. Saidur, "Application of support vector machine models for forecasting solar and wind energy resources: A review," *J. Cleaner Prod.*, vol. 199, pp. 272–285, Oct. 2018, doi: [10.1016/j.jclepro.2018.07.164](https://doi.org/10.1016/j.jclepro.2018.07.164).
- [44] L. Xu, X. Yu, and T. A. Gulliver, "Intelligent outage probability prediction for mobile IoT networks based on an IGWO-Elman neural network," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1365–1375, Feb. 2021, doi: [10.1109/TVT.2021.3051966](https://doi.org/10.1109/TVT.2021.3051966).
- [45] D. Zhang and S. Lou, "The application research of neural network and BP algorithm in stock price pattern classification and prediction," *Future Gener. Comput. Syst.*, vol. 115, pp. 872–879, Feb. 2021, doi: [10.1016/j.future.2020.10.009](https://doi.org/10.1016/j.future.2020.10.009).
- [46] Y. Zhang, B. Chen, G. Pan, and Y. Zhao, "A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting," *Energy Convers. Manag.*, vol. 195, pp. 180–197, Sep. 2019, doi: [10.1016/j.enconman.2019.05.005](https://doi.org/10.1016/j.enconman.2019.05.005).
- [47] Y. Zhang, G. Pan, B. Chen, J. Han, Y. Zhao, and C. Zhang, "Short-term wind speed prediction model based on GA-ANN improved by VMD," *Renew. Energy*, vol. 156, pp. 1373–1388, Aug. 2020, doi: [10.1016/j.renene.2019.12.047](https://doi.org/10.1016/j.renene.2019.12.047).
- [48] Q. Zhang, J. Wu, Y. Ma, G. Li, J. Ma, and C. Wang, "Short-term load forecasting method with variational mode decomposition and stacking model fusion," *Sustain. Energy, Grids Netw.*, vol. 30, Jun. 2022, Art. no. 100622, doi: [10.1016/j.segan.2022.100622](https://doi.org/10.1016/j.segan.2022.100622).
- [49] J. Ma, D. Xia, Y. Wang, X. Niu, S. Jiang, Z. Liu, and H. Guo, "A comprehensive comparison among metaheuristics (MHs) for geohazard modeling using machine learning: Insights from a case study of landslide displacement prediction," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105150, doi: [10.1016/j.engappai.2022.105150](https://doi.org/10.1016/j.engappai.2022.105150).



SHAN LIN received the B.Eng. degree from East China Jiaotong University, in 1993. She is currently a Senior Engineer with Guangzhou Metro Design and Research Institute Company Ltd. She was awarded an Outstanding Engineering Survey Designer in Guangdong Province, China. Her research interests include the design and research of power supply, electrical, and intelligent engineering.



XINGZHONG NONG received the B.Eng. degree from Tsinghua University, in 1992. He is currently the Director of the Science and Technology Committee, Guangzhou Metro Design and Research Institute Company Ltd. He received the awards of Leading Talent of Technology Entrepreneurship in Guangdong Tezhi Program and Survey and Design Master of Guangdong Province, China.



JIANQIANG LUO (Member, IEEE) received the B.Eng. and M.Eng. degrees in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, and the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong. He was a Visiting Doctoral Student with Imperial College London, U.K. He is currently an Associate Professor with the Guangdong University of Technology, and also a National Registered Electrical Engineer and an Intermediate Economist in China. His research interests include power system stability analysis and the application of artificial intelligence in power systems.



CHEN'EN WANG received the B.Eng. degree in electrical engineering from Jishou University, Jishou, China. He is currently pursuing the M.Eng. degree in electrical engineering with the Guangdong University of Technology, Guangzhou, China. He received the Merit Student Award. His research interests include power system stability analysis and artificial intelligence.

...