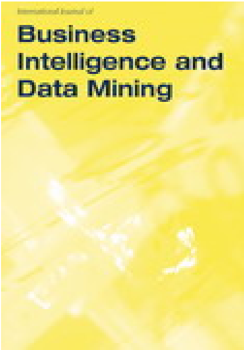
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Business Intelligence and Data Mining

ISSN online: 1743-8195 - ISSN print: 1743-8187
https://www.inderscience.com/ijbidm

Prediction of carbon emissions throughout the lifecycle of zero carbon substations based on Lasso-GRNN neural network model

Ting Zeng, Yueqing Chen, Liuhuo Wang, Mingpeng Yuan, Zhangqi Lv, Dianbin Wang

DOI: 10.1504/IJBIDM.2026.10076223

Article History:

Received:	25 April 2025
Last revised:	30 September 2025
Accepted:	30 September 2025
Published online:	14 May 2026

Prediction of carbon emissions throughout the lifecycle of zero carbon substations based on Lasso-GRNN neural network model

Ting Zeng and Yueqing Chen

Shantou Power Supply Bureau of Guangdong Power Grid Co., Ltd.,
Shantou, 515041, Guangdong, China
Email: DDDL0505@163.com
Email: gdjgcb8324@126.com

Liuhuo Wang

Guangdong Power Grid Co., Ltd.,
Guangzhou, 510699, Guangdong, China
Email: 379876352@qq.com

Mingpeng Yuan and Zhangqi Lv

Shantou Power Supply Bureau of Guangdong Power Grid Co., Ltd.,
Shantou, 515041, Guangdong, China
Email: 381584265@qq.com
Email: 18200620550@139.com

Dianbin Wang*

Northeast Electric Power University,
Jilin, 132012, Jilin, China
Email: 18346800877@163.com
*Corresponding author

Abstract: In order to solve the problems of low recall rate, low prediction accuracy, and long prediction completion time of carbon emission prediction factors in traditional methods, a prediction method of carbon emissions throughout the lifecycle of zero carbon substations based on Lasso-GRNN neural network model is proposed. Lasso-GRNN neural network model is constructed by using the key indicators for predicting the carbon emissions of a zero carbon substation throughout its entire lifecycle as input variables and the carbon emission values as output variables. The model uses Lasso to screen key indicators and inputs them into the GRNN neural network to obtain accurate prediction results. Experimental results show that the proposed method has a maximum recall rate of 98.12% for the influencing factors of carbon emissions throughout the entire life cycle of zero carbon substations, a maximum prediction accuracy of 98.51%, and a minimum prediction completion time of 0.68s.

Keywords: Lasso-GRNN neural network model; zero carbon substation; full lifecycle; carbon emissions forecast; indicators.

Reference to this paper should be made as follows: Zeng, T., Chen, Y., Wang, L., Yuan, M., Lv, Z. and Wang, D. (2026) 'Prediction of carbon emissions throughout the lifecycle of zero carbon substations based on Lasso-GRNN neural network model', *Int. J. Business Intelligence and Data Mining*, Vol. 28, No. 8, pp.1–19.

Biographical notes: Ting Zeng is a Senior Engineer, Supervision Engineer, and Technical Expert. With over 30 years of practical experience in infrastructure construction project management, he has led multiple engineering and technological innovation projects which have received provincial and ministerial-level awards. His primary research focuses on transmission and substation engineering and civil engineering.

Yueqing Chen graduated from North China Electric Power University with a Bachelor's in Automatic Control. He is currently the General Manager of Infrastructure Department of Shantou Power Supply Bureau, and has been engaged in infrastructure project management and has rich experience in project management.

Liuhuo Wang received his Master's in Electrical Engineering from Xi'an Jiaotong University. Currently, he serves as the Project Management Section Manager in the Infrastructure Department of Guangdong Power Grid Corporation. His research interests include power grid construction technologies, among others.

Mingpeng Yuan received his Master's in Architecture and Civil Engineering from Shenzhen University. He is currently a quality and safety management specialist in the Infrastructure Department of Shantou Power Supply Bureau. His research interests include robotic construction and big data safety management.

Zhangqi Lv has received his Bachelor's in Electrical Engineering and Automation from Hunan University of Engineering; Joined Shantou Power Supply Bureau of Guangdong Power Grid Co., Ltd. in 2018 and currently a Senior Operator and Technician in the Main Grid Dispatch Automation Team. His research interests include network topology of power systems, application of artificial intelligence in scheduling, and network security protection.

Dianbin Wang received his Bachelor's in Electrical Engineering from Northeast Electric Power University in 2023. He is a PhD student currently studying at Northeast Electric Power University, his research direction is low-carbon technology in power systems, and the regulation of low-carbon charging and discharging of vehicle-grid interaction.

1 Introduction

As an important infrastructure in the power industry (Wang et al., 2023), the carbon emissions during the construction and operation of substations cannot be ignored, therefore higher requirements have been put forward for the design and operation of substations (Chen, 2025). As a representative of new and efficient substations, zero carbon substations aim to reduce carbon emissions during the construction and operation stages of substations by fully utilising energy-saving measures, renewable energy resources, and carbon neutrality measures such as carbon sinks, ultimately achieving the goal of 'zero carbon' throughout the entire lifecycle. Zero carbon substation is a complete set of technical solutions aimed at achieving near zero carbon emissions in substations. It combines the relevant concepts of zero carbon buildings, fully utilises energy-saving measures, renewable energy resources, and carbon neutrality measures such as carbon sinks in substations, aiming to reduce carbon emissions during the construction and operation stages of substations, and ultimately achieve the goal of 'zero carbon' throughout the entire life cycle (Che et al., 2024; Liu et al., 2024). Therefore, predicting the carbon emissions of zero carbon substations throughout their entire lifecycle not only helps promote green design and construction of substations, but also provides theoretical basis for the power industry to formulate scientific carbon reduction strategies.

Yuan et al. (2024) proposed a carbon emission prediction method based on transfer deep reinforcement learning. First, build an automatic carbon emission data collection system and complete the data collection work. Using a fully automatic encoder with neural networks to decode and reconstruct sample data, combined with deep transfer learning to extract carbon emission data features. Design a reinforcement learning support vector machine regression model, import the feature extracted data into a linear regression function, and perform nonlinear regression calculations. Calculate the fitting value and mean square error value of the sample set, normalise the factors affecting carbon emissions through deep reinforcement learning, and input them into the prediction model to achieve real-time prediction of carbon emissions. However, the accuracy of the factors affecting the carbon emissions throughout the entire lifecycle of zero carbon substations using this method is not high, which affects the quality of subsequent predictions. Zhang et al. (2024) proposed a carbon emission prediction method based on recurrent neural networks. In the process of predicting carbon emissions related to user side electricity, the electricity consumption is estimated first, and then the conversion is completed based on the electricity carbon conversion relationship. Design a recurrent neural network model with multimodal embedding functionality. At the same time, an innovative historical attention mechanism is proposed, which deeply considers the periodic patterns presented by users' electricity consumption habits, accurately captures the periodic influencing factors in the electricity consumption sequence, and uses a prediction model to obtain relevant prediction results. However, the prediction accuracy of this method is not high, and the gap between it and the expected target is relatively large. Zhao et al. (2024) proposed a carbon emission prediction method based on EABC algorithm optimised RFR model. Integrating genetic learning strategies into traditional artificial bee colony algorithms for improvement, constructing a stochastic forest regression prediction model based on evolutionary artificial bee colony algorithm optimisation. Firstly, relying on the expandable stochastic environmental impact assessment model, the influencing factors of carbon emissions are identified and used as input independent variables for the prediction model. Subsequently, the evolutionary artificial bee colony algorithm was used to optimise the random forest regression model, in order to avoid the negative impact of subjective parameter settings on prediction accuracy. Finally, the parameter optimised model is used to predict the carbon emissions of the power industry. However, the implementation process of this method is relatively complex, which increases the prediction completion time.

A prediction method of carbon emissions throughout the lifecycle of zero carbon substations based on Lasso-GRNN neural network model is proposed with the expected goal of solving various problems existing in the above methods. The technical route of this article is as follows:

1 Determine the carbon emission prediction indicators for the entire lifecycle of the substation, collect relevant data, and perform clustering processing on the data to obtain high-quality data. Fully considering the complex characteristics of zero carbon substations at all stages of their lifecycle, ensuring the completeness and pertinence of the indicator system. After data collection, clustering processing techniques are used to optimise the data. Unlike conventional data cleaning methods, clustering processing can effectively remove noise and redundant information from the data structure and distribution characteristics, mine the inherent rules of the data, and obtain high-quality data, providing a solid and reliable data foundation for subsequent model training, greatly improving the data's ability to support the model.

2 A Lasso-GRNN neural network model is constructed by taking the key indicators for predicting the carbon emissions of zero carbon substations throughout their entire lifecycle as input variables and the carbon emissions values as output variables. The model uses Lasso to screen key indicators and inputs them into the GRNN neural network to obtain accurate predictions of the carbon emissions of zero carbon substations throughout their lifecycle. The introduction of Lasso algorithm is a major innovation point, which can accurately screen key indicators from numerous predictive indicators, effectively solve the problem of indicator redundancy, avoid the interference of irrelevant variables on the model, and greatly improve the efficiency and accuracy of the model. GRNN neural network, with its powerful nonlinear mapping ability and adaptive learning ability, can better fit the complex relationship between input variables and output variables. By inputting the key indicators selected by Lasso into the GRNN neural network, the advantages of both methods were fully utilised to achieve accurate prediction of carbon emissions throughout the entire lifecycle of zero carbon substations, providing a more scientific and forward-looking decision-making basis for carbon management in substations.

3 The effectiveness of this method was tested by selecting the factors affecting the carbon emissions of zero carbon substations throughout their entire lifecycle, including the recall rate, prediction accuracy, and prediction completion time, as experimental indicators.

2 Design of carbon emission prediction method for the whole life cycle of zero carbon substation

2.1 Collection of carbon emission prediction indicators for the entire lifecycle of substations based on K-modes algorithm

The entire life cycle of a substation covers multiple stages of building materials production, transportation, construction, operation, and demolition. The carbon emission sources vary significantly in each stage, and targeted key indicators need to be collected, as shown in Table 1.

Therefore, based on the above analysis, it can be determined that the key indicators in the production stage of building materials include material carbon emission factors, production process parameters, etc; The key indicators during the transportation phase of building materials include transportation distance, transportation mode, fuel type, and consumption, etc; The key indicators during the construction phase include energy consumption of construction machinery, labour input, and waste disposal methods (such as differences in emissions between on-site incineration and recycling); The key indicators during the operation phase include demolition energy consumption, waste disposal methods, and material recovery rate.

Table 1 Prediction indicators for carbon emissions throughout the whole life cycle of substations

Substation lifecycle stages	*Source of indicators*	*Indicator description*
Building materials production stage	Direct discharge	CO_2 emissions generated from energy consumption (such as coal and electricity) during the production process of building materials such as steel and concrete.
	Indirect emissions	Emissions from upstream processes such as raw material mining and transportation (such as blasting emissions from iron ore mining)
Construction material transportation stage	Emissions from transportation vehicles	Emissions from fuel consumption (such as diesel and gasoline) of transportation vehicles such as trucks and ships
	Transportation distance impact	The longer the transportation distance, the higher the total carbon emissions (such as the difference in emissions between inter provincial and intra provincial transportation)
Construction phase	Construction machinery emissions	Diesel consumption and emissions of construction machinery such as cranes and excavators
	Artificial activity emissions	Indirect emissions generated by commuting and accommodation of construction personnel
Operate phase	Equipment energy consumption and emissions	Power consumption emissions of transformers, switchgear and other equipment (affected by equipment energy efficiency levels)
	Maintain activity emissions	Material consumption and transportation emissions generated from equipment maintenance and replacement
Demolition stage	Dismantling energy consumption and emissions	Equipment energy consumption data, operating load curve, maintenance activity frequency, and material usage
	Waste disposal and discharge	Differences in emissions from landfilling, incineration, or recycling of waste (such as methane generated from landfilling, reduced emissions from recycling)

Adaptive synthetic sampling (ADASYN) is a sampling algorithm used to deal with class imbalance problems (Zhong et al., 2025; Qing et al., 2022), which avoids overfitting caused by simple oversampling. Therefore, using this method to determine key indicator data is of great significance.

Assuming that the sample size of the carbon emission prediction index data for the entire lifecycle of the substation to be generated is N_{new}, where N_1 represents the number of majority class samples, N_2 represents the number of minority class samples, and α represents a random number with a value between 0–1, the following relationship holds:

$$N_{new} = (N_1 - N_2) \times \alpha \tag{1}$$

Assuming that the proportion of majority classes in the K nearest neighbour sample points among the minority class sample points is Γ_i, N_i^+ is the number of majority class samples in the 44 nearest neighbours, and i = 1, 2, 3, …, N_2, the following relationship holds:

$$\Gamma_i = N_i^+ / K \tag{2}$$

The standardised proportion parameter $\hat{\Gamma}_i$ is obtained by processing the proportion of the majority class samples α for Γ_i, and its calculation formula is as follows:

$$\hat{\Gamma}_i = \Gamma_i / \sum_{i=1}^{N_2} \Gamma_i \tag{3}$$

Based on the standardised proportion coefficient $\hat{\Gamma}_i$, the number of new samples required to be generated for each minority class in the sample can be calculated as N_i^+, which satisfies the following relationship:

$$n_i^+ = \hat{\Gamma}_i \times N_{new} \tag{5}$$

K-modes algorithm, as one of the widely used classification attribute clustering algorithms (Kuo et al., 2024), is a partition clustering algorithm Set a sample point set X = $\{x_1, x_2, x_3, \ldots, x_n\}$ for classification attributes, where $x_i (i = 1, 2, \ldots, n)$ is the sample point, each sample point is represented as $x_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,m}]$, its attribute is $\{A_1, A_2, \ldots, A_m\}$, and the value range of each attribute is $Dom(Ai) = \{a_1^i, a_2^i, ?, a_r^i\} (r \geq 2)$. Before executing the algorithm, the number of clusters for predicting the carbon emissions of the entire lifecycle of the substation is determined to be k. During the execution process, the algorithm divides the n sample points of the carbon emissions prediction indicators of the entire lifecycle of the substation into k clusters, and ensures that the obtained partition results satisfy the condition of minimising the sum of distances within all groups This process is usually described as the following mathematical problem (Gillard et al., 2023; Jiang et al., 2023), and the formula for minimising the objective function $P(W, Z)$ is as follows:

$$P(W, Z) = \sum_{l=1}^{k} \sum_{i=1}^{n} \omega_{i,l} d(x_i, z_l) \tag{6}$$

In the above formula, W represents a binary membership matrix of $n \times k$, $Z = \{z_1, z_2, \ldots, z_k\}$ represents a matrix of $k \times m$ containing k centroids, $d(x_i, z_l)$ represents the distance from the sample point to a centroid, and $\omega_{i,l}$ represents the weight parameter.

The clustering process of carbon emission prediction indicators for the entire lifecycle of substations based on k-modes algorithm is described as follows:

Step 1 After determining the number of clusters k, randomly select k points from the dataset as initial centroids to form the initial $Z^{(1)}$. Determine $Z^{(1)}$ so that $P(W, Z^{(1)})$ takes the minimum value, and set $t = 1$

Step 2 Update the centroid, determine $Z^{(t+1)}$ to minimise $P(W^{(t)}, Z^{(t+1)})$, and stop if the condition $P(W^{(t)}, Z^{(t+1)}) = P(W^{(t+1)}, Z^{(t+1)})$ is met; Otherwise, continue with the third step

Step 3 Update $W^{(t+1)}$ to minimise $P(W^{(t)}, Z^{(t+1)})$. If $P(W^{(t)}, Z^{(t+1} = P(W^{(t+1)}, Z^{(t+1)})$, stop; Otherwise, let $t = t + 1$ and proceed to the second step until the clustering of carbon emission prediction indicators for the entire lifecycle of the substation is completed (Suryanarayana et al., 2022).

2.2 Carbon emission prediction based on lasso GRNN-neural network model

2.2.1 Selection of key indicators for carbon emission prediction based on LASSO regression

The principle of least absolute shrinkage and selection operator (LASSO) regression is to introduce L1 regularisation term, which minimises the sum of squared residuals while applying absolute value constraints to regression coefficients, causing some coefficients to shrink to zero, thus achieving feature selection. It can automatically screen key indicators and output variables corresponding to non-zero coefficients without manually setting thresholds; regularising the compression coefficient range through L1 effectively avoids overfitting and improves the model's generalisation ability; At the same time, LASSO regression has the ability to handle high-dimensional data, especially suitable for scenarios where the number of variables exceeds the sample size, that is, the screening of key indicators for carbon emission prediction.

The objective function of LASSO regression is represented by the following formula:

$$J(\beta) = \sum (y - X\beta)^2 + \lambda \|\beta\|_1 = \sum (y - X\beta)^2 + \sum \lambda |\beta| = ESS(\beta) + \lambda l_1(\beta) \tag{7}$$

In the above formula, y is a $n \times 1$ dimensional column vector used to represent the dependent variable in the model, $X = (X_1, X_2, \ldots, X_p)$ is a $n \times p$ dimensional matrix used to represent the independent variables in the model, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ is a $p \times 1$ dimensional column vector used to represent the regression coefficients to be estimated, $ESS(\beta)$ is the sum of squared errors, $\lambda l_1(\beta)$ is the penalty term, and λ is the non negative regularisation parameter. Derive the regression coefficients of the LASSO regression model, and the objective function of similar LASSO regression can be represented by the following formula:

$$J(\beta) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_i x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{8}$$

In the above formula, x_{ij} represents the sample data of carbon emission prediction indicators, y_i represents the value of carbon emission indicators, β_i and β_j represent different regression coefficients.

From equation (8), we can obtain:

$$ESS(\beta)=\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{p}\beta_i x_{ij}\right)^2=\sum_{i=1}^{n}\left(y_i^2+\left(\sum_{j=1}^{p}\beta_i x_{ij}\right)^2-2y_i\left(\sum_{j=1}^{p}\beta_i x_{ij}\right)\right) \quad (9)$$

Calculate the partial derivative of $ESS(\beta)$ in equation (9), and the result obtained is represented by the following formula:

$$\frac{\partial ESS(\beta)}{\partial\beta_j}=-2\sum_{i=1}^{n}x_{ij}\left(y_i-\sum_{k\neq j}\beta_j x_{ij}-\beta_j x_{ij}\right)=-\sum_{i=1}^{n}x_{ij}\left(y_i-\sum_{k\neq j}\beta_j x_{ij}\right)+2\beta_j\sum_{i=1}^{n}x_{ij}^2 \quad (10)$$

So the penalty term is expressed as the derivative of β_j using the following formula:

$$\frac{\partial\lambda l_1(\beta)}{\partial\beta_j}=\begin{cases}\lambda, & when\ \beta_j>0\\ [-\lambda,\lambda], & when\ \ \beta_j=0\\ -\lambda, & when\ \beta j<0\end{cases} \quad (11)$$

To obtain the estimation formula for the regression coefficients of the LASSO regression model, the following calculation is required: add the derivatives of $ESS(\beta)$ and $\lambda l_1(\beta)$, and make their derivative equal to 0. The calculation formula is as follows:

$$ESS(\beta)\frac{\partial ESS(\beta)}{\partial\beta_j}+\frac{\partial\lambda l_1(\beta)}{\partial\beta_j}=\begin{cases}-2m_j+2\beta_j n_j+\lambda=0\\ \left[-2m_j-\lambda,-2m_j+\lambda\right]=0\\ -2m_j+2\beta_j n_j-\lambda=0\end{cases} \quad 12)$$

The calculation formula for the regression coefficients of the LASSO regression model can be obtained from equation (12) as follows:

$$\beta_j=\begin{cases}\dfrac{m_j-\dfrac{\lambda}{2}}{n_j}, & m_j>\dfrac{\lambda}{2}\\ 0, & m_j\in\left[-\dfrac{\lambda}{2},\dfrac{\lambda}{2}\right]\\ \dfrac{m_j-\dfrac{\lambda}{2}}{n_j}, & m_j<\dfrac{\lambda}{2}\end{cases} \quad (13)$$

Let $\hat{\beta}=\left(\hat{\beta}_j\right)^T$, $j=1,2,\ldots,p,$ the LASSO parameter estimation result is expressed by the following formula:

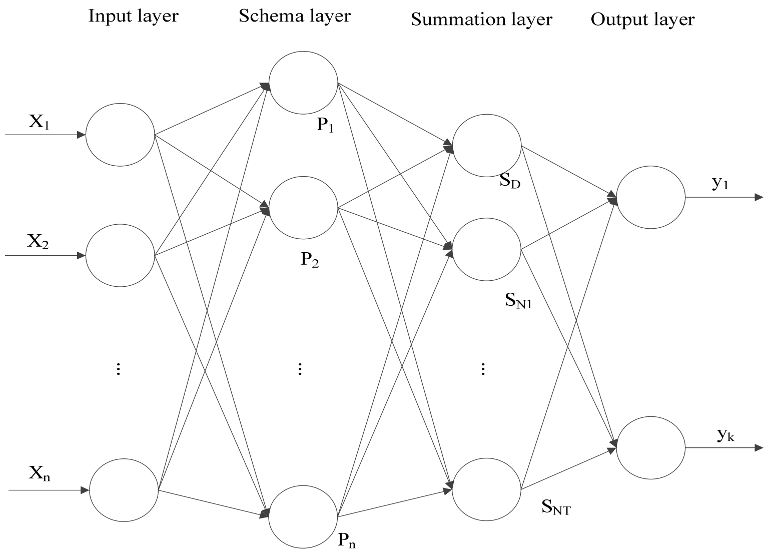
$$\left(\hat{\alpha},\hat{\beta}\right)=argmin_\beta\left\{\sum_{j=1}^{p}\left(y_j-a-\sum_{j=1}^{p}\beta_j x_j\right)^2\right\}+\lambda\sum_{j=1}^{P}\left|\beta_j\right| \quad (14)$$

The objective function of LASSO regression is equivalent to the following form using convex optimisation method, and the key indicators for carbon emission prediction can be selected using this formula. The formula is expressed as follows:

$$\begin{cases} Q = argmin\sum(y - X\beta)^2 \\ \sum|\beta| \leq t \end{cases} \tag{15}$$

In the above formula, t represents any non negative number.

Figure 1 GRNN neural network topology structure



2.2.2 *Prediction of carbon emissions throughout the lifecycle of zero carbon substations based on GRNN neural network*

Generalised regression neural network (GRNN) is a neural network based on radial basis function (RBF) (Chakraborty and Mukherjee, 2025; Sridevi and Samath, 2024), which belongs to non parametric estimation methods. The core idea is to calculate the similarity between input samples and training samples through kernel functions (such as Gaussian functions), estimate the probability density function of the target variable, and ultimately achieve regression prediction. GRNN does not require iterative training and can optimise model performance by adjusting the smoothing factor (spread). Applicability analysis of GRNN: with its nonlinear modelling capability, GRNN can effectively capture the complex nonlinear relationship between carbon emissions and equipment parameters, operation and maintenance strategies, and other factors. In addition, GRNN also has high-dimensional data processing capabilities, which can handle the joint impact of multiple factors such as equipment type, operation and maintenance frequency, and energy structure on carbon emissions, thereby ensuring the accuracy of prediction results. In the prediction of carbon emissions throughout the entire lifecycle of zero carbon substations, GRNN neural networks demonstrate significant innovation. It accurately captures the complex dynamic characteristics of carbon emissions at various stages of the entire lifecycle of substations through nonlinear mapping capabilities. Combined with key indicator screening techniques such as Lasso regression for carbon emission prediction, it can extract key influencing factors from massive operating data and construct high-precision prediction models. Compared with traditional statistical models, GRNN has stronger generalisation ability when dealing with multi-source heterogeneous data, especially suitable for complex systems such as zero carbon substations that integrate new equipment, comprehensive energy compensation, and multi-dimensional carbon reduction technologies such as building energy conservation. GRNN dynamically adjusts network parameters to achieve adaptive learning of real-time data such as equipment operating losses, photovoltaic power generation, and energy storage regulation, effectively solving the problem of insufficient accuracy in traditional methods for predicting nonlinear carbon emission trends and providing scientific decision-making basis for the full lifecycle carbon management of substations.

The topology structure of GRNN neural network is shown in Figure 1.

In Figure 1, the main structure of GRNN neural network includes input layer, schema layer, summation layer, and output layer. The input layer receives input variables and passes the data to the pattern layer. Each neuron in the pattern layer corresponds to a training sample, and the input data is nonlinearly mapped through the neuron transfer function; The summation layer performs weighted summation on the output of the pattern layer, which includes two different types of neurons used for arithmetic summation and weighted summation of the pattern layer output; Finally, the output layer normalises the results of the summation layer and outputs the final prediction result. Its structure enables GRNN to have strong ability and relatively efficient training when dealing with nonlinear problems. The GRNN neural network shown in Figure 1 significantly improves prediction performance through a four layer topology structure. The input layer receives key indicator vectors filtered by Lasso, achieving high-dimensional feature denoising; The pattern layer adopts an improved Gaussian RBF to dynamically adjust the neural receptive field through an adaptive smoothing factor σ, accurately capturing the nonlinear relationship between carbon emissions and equipment parameters and operating conditions; The innovative fusion of arithmetic summation and weighted summation in the summation layer preserves global statistical characteristics while highlighting key feature contributions; The output layer achieves conditional mean prediction through probability density estimation, and its non parametric characteristics effectively avoid the assumptions and limitations of traditional neural networks on data distribution. This structure is particularly adept at handling the complex coupling relationships between multi-source heterogeneous data such as building material production and equipment operation during the lifecycle of substations, ensuring the accuracy of prediction results. The input variable of GRNN neural network is $X = [x_1, x_2, \ldots, x_n]^T$, which represents the key indicator for predicting the carbon emissions of zero carbon substations throughout their entire life cycle. The output is $Y = [y_1, y_2, \ldots, y_k]^T$, which represents the carbon emissions value of zero carbon substations throughout their entire life cycle. The calculation formula for the transfer function of neurons in the pattern layer is as follows:

$$P_i = exp\left[\frac{-(X - X_i)^T (X - X_i)}{2\sigma^2}\right], i = 1, 2, \ldots, n \qquad 16)$$

In the above formula, n represents the number of key indicators for predicting the carbon emissions of zero carbon substations throughout their entire lifecycle, X_i represents the learning sample corresponding to the i^{th} neuron, and σ represents the smoothing factor.

There are two calculation methods for the neuron transfer function of the summation layer, and the calculation formulas are as follows:

$$S_D = \sum_{i=1}^{n} P_i \quad 17)$$

$$S_{Nj} = \sum_{i=1}^{n} y_{ij} P_i,\ j = 1, 2, ..., k \quad 18)$$

In the above formula, y_{ij} represents the connection weight between neurons in the pattern layer, and k represents the dimension of the output vector of the learning sample.

The output layer divides the sum layer outputs of each neuron, and the output of neuron j corresponds to the j^{th} result element. The calculation formula is as follows:

$$y_j = \frac{S_{Nj}}{S_D},\ j = 1, 2, \cdots, k \quad 19)$$

Assuming the joint probability density function of random variable x, y is $f(x, y)$, the observed values of x are X, and y are regressed relative to X, the conditional mean is expressed by the following formula:

$$\hat{Y} = E(y / X) = \frac{\int_{-\infty}^{\infty} yf(X, y)dy}{\int_{-\infty}^{\infty} yf(X, y)dy} \quad (20)$$

In the above formula, $\hat{Y}$ represents the predicted output of Y. The probability density function $f(X, y)$ can be obtained by non parametric estimation, and the calculation formula is as follows:

$$\hat{f}(X, y) = \frac{1}{n(2\pi)^{\frac{(p+1)}{2}p+1}} \sum_{i=1}^{n} exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right] exp\left[-\frac{(Y - Y_i)^2}{2\sigma^2}\right] \quad (21)$$

Substituting $\hat{f}(X, y)$ instead of $f(X, y)$ into equation (21) yields the following result:

$$\hat{Y}(X) - \frac{\sum_{i=1}^{n} exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right] \int_{-\infty}^{\infty} yexp - \left[\frac{(Y - Y_i)^2}{2\sigma^2}\right] dy}{\sum_{i=1}^{n} exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right] \int_{-\infty}^{\infty} exp - \left[\frac{(Y - Y_i)^2}{2\sigma^2}\right] dy} \quad (22)$$

Due to the integral $\int_{-\infty}^{\infty} xe^{-x^2} dx = 0$, the output of the GRNN neural network can be obtained, and the specific calculation formula is as follows:

$$\hat{Y}(X) - \frac{\sum_{i=1}^{n} yexp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right]}{\sum_{i=1}^{n} exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right]} \quad (23)$$

3 Experimental design

3.1 Experimental scheme

3.1.1 Experimental data

In the experiment of predicting carbon emissions throughout the life cycle of zero carbon substations, the experimental data types cover multiple stages and dimensions, as shown in Figure 2.

To improve the reliability of simulation experiments, this study adopted a systematic data acquisition and parameter optimisation method. In the data preparation stage, based on the characteristics of each link in the entire lifecycle of the substation (building materials production, transportation, construction, operation, demolition), a standardised data collection template containing 28 key indicators was constructed, and the accuracy of the original data was ensured through a multi-source data verification mechanism (such as cross verification of transportation data and logistics documents). To address data quality issues, a hierarchical processing strategy is adopted: for missing values, multiple imputation or mean imputation is used based on field importance; correct outliers by combining the 3 σ criterion with business rules; for imbalanced data, the ADASYN algorithm is used to dynamically generate minority class samples and maintain balanced class proportions. In the parameter design phase, a three-stage optimisation process is adopted: firstly, key parameters are initialised based on domain knowledge; secondly, the adjustment order of each parameter priority is determined through sensitivity analysis; finally, an improved Bayesian optimisation algorithm is used for automated parameter search within the preset range, and the stability of parameter combinations is evaluated through 5-fold cross validation in each iteration. The entire process adopts version control to record parameter adjustment paths, ensuring the reproducibility of experiments and improving the accuracy of simulation experiment results.

Table 2 Experimental data content

Data type	*Data composition*	*Descriptives*
Carbon emission factor data	Material production carbon emission factor	Used to calculate the carbon emissions during the material production process of substation construction, such as steel, concrete, insulation materials, etc
	Transportation carbon emission factor	Consider the transportation distance and mode of transportation (such as road, railway, sea) of materials and equipment, and calculate the carbon emissions during transportation
	Equipment operation carbon emission factor	The operating energy consumption and corresponding carbon emission factors of major equipment including transformers, GIS equipment, switchgear, etc
	Carbon emission factors during construction process	Energy consumption and carbon emissions related to construction machinery (such as excavators and cranes)

Table 2 Experimental data content (continued)

Data type	*Data composition*	*Descriptives*
Engineering quantity data	Material usage	The quantity of various materials required for substation construction, such as steel, concrete, cables, etc
	Equipment specifications and quantity	The model, power, quantity, etc. of the main equipment in the substation, such as transformer capacity, GIS equipment quantity, etc
Energy consumption data	power consumption	The power demand during the operation of the substation, including equipment operation, lighting, air conditioning, etc
	Renewable energy generation capacity	The power generation of renewable energy sources such as rooftop photovoltaics and wind power
	Fossil energy consumption	Fuel consumption of backup diesel generators
Carbon sink data	Green Area	The green area inside the substation is used to calculate carbon sequestration
	Carbon sink calculation parameters	Such as the carbon sequestration rate of vegetation, soil carbon sequestration capacity, etc
Equipment lifecycle data	Equipment life	The expected service life of major equipment, such as transformers, GIS equipment, etc
	Equipment replacement frequency	The frequency and timing of equipment replacement throughout its entire lifecycle
	Equipment energy efficiency	Energy efficiency rating and operational efficiency of the equipment
Building lifecycle data	Architectural design parameters	Building area, structural form, thermal performance of enclosure structure, etc
	Building operation data	Indoor temperature and humidity, lighting energy consumption, air conditioning energy consumption, etc
	Data on building demolition and recycling	Monitoring and measurement of energy consumption, waste disposal, data processing methods, and recycling rates during the demolition process
Monitoring and measured data	Real-time monitoring data	Real-time monitoring of energy consumption, carbon emissions, and other data in substations through sensors
	Historical data	Historical operating data of similar substations
Environmental parameters	geographic location	The geographical location of the substation affects climate conditions, energy supply, etc
	Climate data	The local climate conditions such as temperature, humidity, and wind speed affect building energy consumption and equipment operation

Selecting the Yuan et al. (2024) method, Zhang et al. (2024) method, and the proposed method as experimental comparison methods, the actual application effects of different methods are tested by comparing the recall rate of factors affecting the full life cycle carbon emissions of zero carbon substations, the accuracy of predicting the full life cycle carbon emissions of zero carbon substations, and the completion time of predicting the full life cycle carbon emissions of zero carbon substations.

1 The recall rate is an important indicator for evaluating the integrity of information retrieval or data analysis, referring to the proportion of correctly retrieved relevant items to the actual total number of relevant items. In the study of carbon emissions throughout the lifecycle of zero carbon substations, the recall rate is used to measure the comprehensiveness of identifying influencing factors, that is, the proportion of identified influencing factors to all potential influencing factors.

2 Prediction accuracy is the core indicator for measuring the effectiveness of carbon emission prediction methods, which refers to the degree of closeness between the predicted values of the method and the actual observed values. In the prediction of carbon emissions throughout the life cycle of zero carbon substations, the accuracy index directly reflects the model's ability to quantify carbon emissions in building materials production, equipment operation, carbon sinks, and other aspects, and is a key basis for evaluating the reliability of the evaluation method.

3 The predicted completion time indicator is a key parameter for measuring the efficiency of carbon emission assessment throughout the life cycle of zero carbon substations. It refers to the entire process time from data collection to result output. This indicator directly affects the timeliness of project decision-making, technology selection, and policy formulation, and is an important dimension for evaluating the feasibility of zero carbon substations.

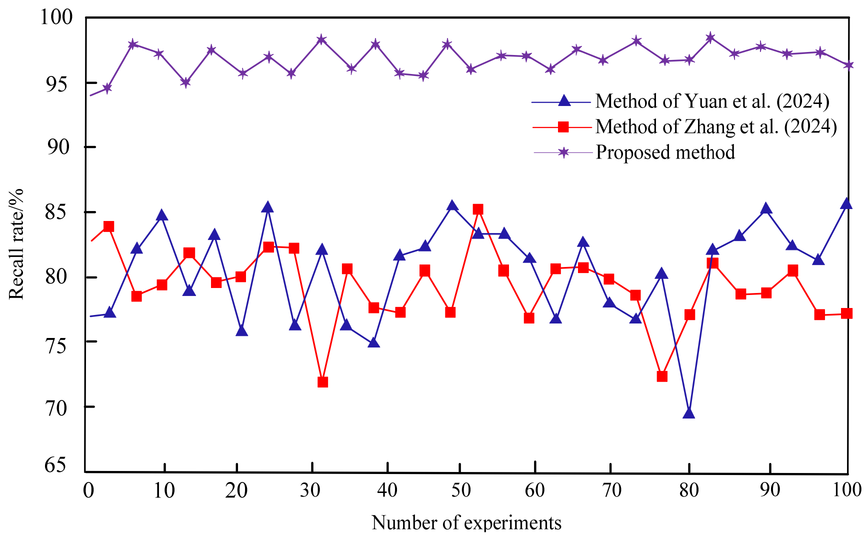
3.2 Experimental results

3.2.1 Verification rate of influencing factors

The comparative results of the factors influencing the carbon emissions throughout the life cycle of zero carbon substations are shown in Figure 2.

According to the analysis of the data in Figure 2, the maximum recall rate of factors affecting the carbon emissions of zero carbon substations throughout their entire life cycle using the Yuan et al. (2024) method is 86.69%, while the maximum recall rate of factors affecting the carbon emissions of zero carbon substations throughout their entire life cycle using the Zhang et al. (2024) method is 85.52%. The maximum recall rate of factors affecting the carbon emissions of zero carbon substations throughout their entire life cycle using the proposed method is 98.12%, which is 11.43% and 12.60% higher than the Yuan et al. (2024) method and Zhang et al. (2024) method, respectively. The minimum recall rate of factors affecting the carbon emissions of zero carbon substations throughout the entire lifecycle of the Yuan et al. (2024) method is 68.47%, and the minimum recall rate of factors affecting the carbon emissions of zero carbon substations throughout the lifecycle of the Zhang et al. (2024) method is 72.03%. The minimum recall rate of factors affecting the carbon emissions of zero carbon substations throughout the entire lifecycle of the proposed method is 94.32%, which is 25.85% and 22.29% higher than the Yuan et al. (2024) method and the Zhang et al. (2024) method, respectively. After testing, it was found that the proposed method has a high recall rate of factors affecting the carbon emissions of zero carbon substations throughout their entire lifecycle, with comprehensive coverage of influencing factors. This reduces the error in subsequent carbon emission predictions and makes the prediction results closer to the true values. The main reason for the high completeness of the influencing factors in the proposed method is that it first determines comprehensive carbon emission prediction indicators and collects relevant data. Then, high-quality data is obtained through clustering processing. Based on this, the Lasso-GRNN neural network model is used to screen key indicators using the Lasso algorithm. This screening process can accurately identify the key factors that have a significant impact on carbon emissions from numerous prediction indicators, avoiding the omission of important variables and ensuring the coverage of various important influencing factors for the full life cycle carbon emissions of zero carbon substations. This improves the completeness of influencing factors and lays a solid foundation for obtaining accurate prediction results through the GRNN neural network in the future.

Figure 2 Full recall rate of influencing factors (see online version for colours)

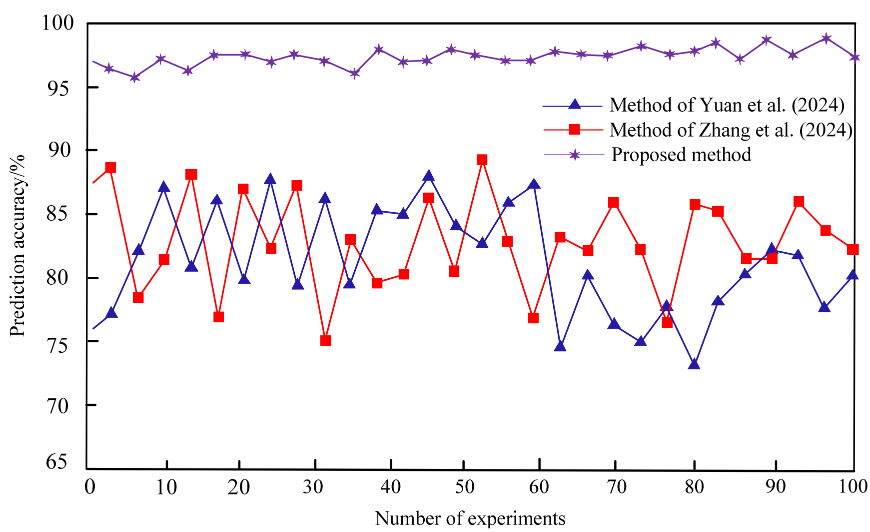


3.2.2 Prediction accuracy

The comparative results of the accuracy of predicting carbon emissions throughout the entire lifecycle of zero carbon substations are shown in Figure 3.

From the analysis of the results in Figure 3, it can be seen that the maximum accuracy of the zero carbon substation life cycle carbon emissions prediction using the Yuan et al. (2024) method is 87.71%, while the maximum accuracy of the zero carbon substation life cycle carbon emissions prediction using the Zhang et al. (2024) method is 88.86%. The maximum accuracy of the zero carbon substation life cycle carbon emissions prediction using the proposed method is 98.51%, which is 10.80% and 9.65% higher than the Yuan et al. (2024) method and Zhang et al. (2024) method, respectively. The minimum accuracy of predicting the full life cycle carbon emissions of zero carbon substations using the Yuan et al. (2024) method is 72.69%, and the minimum accuracy of predicting the full life cycle carbon emissions of zero carbon substations using the Zhang et al. (2024) method is 75.02%. The minimum accuracy of predicting the full life cycle carbon emissions of zero carbon substations using the proposed method is 95.86%, which is 23.17% and 23.17% higher than the Yuan et al. (2024) method and the Zhang et al. (2024) method, respectively. After comparison, it can be concluded that the proposed method has higher prediction accuracy and can provide important data foundation for the construction and operation of zero carbon substations. The main reason for the high accuracy of the proposed method for predicting the full life cycle carbon emissions of zero carbon substations is that, on the one hand, by determining comprehensive prediction indicators and collecting relevant data for clustering processing, noise and redundant information in the data are effectively removed, and high-quality data is obtained, providing a reliable basis for model training; On the other hand, the constructed Lasso-GRNN neural network model first uses the Lasso algorithm to screen key indicators and accurately extract key factors that have a significant impact on carbon emissions, avoiding the interference of irrelevant or redundant variables on the model. Then, the screened key indicators are input into the GRNN neural network. With its powerful nonlinear mapping ability and adaptive learning ability, the GRNN neural network can better fit the complex relationship between input variables and output variables, thereby obtaining accurate prediction results.

Figure 3 Prediction accuracy (see online version for colours)



3.2.3 Predict completion time

The comparison of the completion time for predicting the full lifecycle carbon emissions of zero carbon substations is shown in Table 3.

After testing, it was found that the maximum completion time for predicting the full life cycle carbon emissions of zero carbon substations using the Yuan et al. (2024) method is 5.78s, while the maximum completion time for predicting the full life cycle carbon emissions of zero carbon substations using the Zhang et al. (2024) method is 3.94 s. The maximum completion time for predicting the full life cycle carbon emissions of zero carbon substations using the proposed method is 0.97 s, which is 4.81 s and 2.97 s lower than the Yuan et al. (2024) method and the Zhang et al. (2024) method, respectively. The minimum completion time for predicting the full lifecycle carbon emissions of a zero carbon substation using the Yuan et al. (2024) method is 4.16 s, and the minimum completion time for predicting the full lifecycle carbon emissions of a zero carbon substation using the Zhang et al. (2024) method is 2.63 s. The minimum completion time for predicting the full lifecycle carbon emissions of a zero carbon substation using the proposed method is 0.68 s, which is 3.48 s and 1.95 s lower than the Yuan et al. (2024) method and the Zhang et al. (2024) method, respectively. After comparison, it can be seen that the proposed method has a shorter completion time and higher execution efficiency for predicting the carbon emissions of zero carbon substations throughout their entire lifecycle. The main reason for the short completion time of the proposed method for predicting the carbon emissions of zero carbon substations throughout their entire life cycle is that in the data pre-processing stage, clustering processing is used to optimise the collected relevant data, quickly removing noise and redundant information from the data, obtaining high-quality data, and reducing the amount of data processed by subsequent models; at the same time, in the constructed Lasso-GRNN neural network model, the Lasso algorithm can efficiently screen key indicators, accurately remove irrelevant or redundant variables, greatly simplify the input variables, and enable the GRNN neural network to only learn and calculate for the selected key indicators, avoiding ineffective operations on a large number of irrelevant variables and effectively shortening the completion time of the entire prediction process.

Table 3 Predicted completion time

Number of experiments	*Predict completion time/s*		
	Yuan et al. (2024) method	*Zhang et al. (2024) method*	*Proposed method*
10	5.26	2.88	0.69
20	4.47	3.65	0.97
30	4.86	3.29	0.78
40	5.53	3.94	0.68
50	5.36	3.17	0.74
60	4.16	3.89	0.72
70	5.17	2.74	0.71
80	4.66	2.63	0.92
90	4.18	3.78	0.86
100	5.78	3.12	0.72

4 Conclusions

With the rapid development of new energy technologies and the proposal of the ‘dual carbon’ goal, the construction and operation of zero carbon substations have become an important direction for the green transformation of the power industry. In addition, how to accurately predict the carbon emissions of zero carbon substations throughout their entire lifecycle is one of the urgent problems to be solved. Therefore, a zero carbon substation lifecycle carbon emissions prediction method based on Lasso-GRNN neural network model is proposed. In the data processing stage, comprehensively and accurately determine predictive indicators and use clustering processing to optimise data, effectively improving data quality; in terms of model construction, the innovative fusion of Lasso algorithm and GRNN neural network allows Lasso to accurately screen key indicators and GRNN to efficiently fit complex relationships. The main achievements are reflected in the successful acquisition of accurate predictions of carbon emissions throughout the entire lifecycle of zero carbon substations, providing scientific, reliable, and forward-looking decision support for carbon management in substations, and taking an important step in the field of carbon emission prediction for zero carbon substations. The experimental results show that the proposed method has a maximum recall rate of 98.12% for the influencing factors of carbon emissions throughout the entire life cycle of zero carbon substations, a maximum prediction accuracy of 98.51%, and a minimum prediction completion time of 0.68 seconds. This article has made significant contributions in theory, practice, technological innovation, and economics. It not only helps promote the green transformation and energy conservation and emission reduction of the power industry, but also promotes innovation and application of related technologies, improves economic efficiency, and drives industrial development. In the future, we should continue to strengthen relevant research and technological innovation, promote the widespread application and sustainable development of zero carbon substations.

Acknowledgements

This work is funded by the Science and Technology Project of China Southem Power Grid Company (Project number: GDKJXM20231217).

Declarations

All authors declare that they have no conflicts of interest.

References

Chakraborty, M. and Mukherjee, S. (2025) 'Does parental migration affect children's health in rural India? A fresh look at retrospective data using LASSO regression method', *SN Social Sciences*, Vol. 5, No. 1, pp.1–10.

Che, W., Wang, Y. and Zhu, W. (2024) 'A review of carbon emission reduction during the operation stage of substations', *Sustainability* (2071–1050), Vol. 16, No. 22, pp.10017–10028.

Chen, H. (2025) 'Life cycle evaluation of substations with carbon reduction based on analytic hierarchy process considering multiple performance measures', *Energies*, Vol. 18, No. 4, pp.800–811.

Gillard, J., Knight, V. and Wilde, H. (2023) 'A novel initialisation based on hospital-resident assignment for the k-modes algorithm', *Soft Computing: A Fusion of Foundations, methodologies and Applications*, Vol. 27, No. 14, pp.9441–9457.

Jiang, Z., Liu, X. and Zang, W. (2023) 'A kernel-based intuitionistic weight fuzzy k-modes algorithm using coupled chained P system combines DNA genetic rules for categorical data', *Neurocomputing*, Vol. 528, No.8, pp.84–96.

Kuo, R.J., Cendana, M. and Nguyen, T.P.Q. (2024) 'A multivariate fuzzy weighted K-modes algorithm with probabilistic distance for categorical data', *Journal of ICT Research and Applications*, Vol. 18, No. 2, pp.93–103.

Liu, T., Wu, Z. and Chen, C. (2024) 'Carbon emission accounting during the construction of typical 500 kv power transmissions and substations using the carbon emission factor approach', *Buildings*, Vol. 14, No. 1, pp.1–13.

Qing, Z., Zeng, Q. and Wang, H. (2022) 'ADASYN-LOF algorithm for imbalanced tornado samples', *Atmosphere*, Vol. 13, No. 4, pp.1–10.

Sridevi, V. and Samath, J.A. (2024) 'A combined deep CNN-lasso regression feature fusion and classification of MLO and CC view mammogram image', *International Journal of Systems Assurance Engineering and Management*, Vol. 15, No. 1, pp.553–563.

Suryanarayana, G., Lnc, P.K. and Mahesh, P.C.S. (2022) 'Novel dynamic k-modes clustering of categorical and non categorical dataset with optimized genetic algorithm based feature selection', *Multimedia Tools and Applications*, Vol. 81, No. 17, pp.24399–24418.

Wang, K., Yang, X. and Zhu, L. (2023) 'Study on life-cycle carbon footprint assessment of substation', *Proceedings of SPIE*, Vol. 12804, No. 1, pp.1–11.

Yuan, P., Tan, C. and Li, F. (2024) 'Real time carbon emission prediction method for thermal power units based on transfer deep reinforcement learning', *Industrial Heating*, Vol. 53, No. 7, pp.65–69+75.

Zhang, L., Yan, J.H. and Wang, L.Y. (2024) 'Prediction method for carbon emissions of power on user side', *Computer Simulation*, Vol. 41, No. 9, pp.494–499.

Zhao, Z.H., Zhang, X.H. and Wang, T. (2024) 'Forecast of electricity industry carbon emission based on EABC algorithm optimized RFR model', *Shandong Electric Power*, Vol. 51, No. 1, pp.77–84.

Zhong, K., Tan, X. and Liu, S. (2025) 'Prediction of slope failure probability based on machine learning with genetic-ADASYN algorithm', *Engineering Geology*, Vol. 346, No. 1, pp.1–15.