Reducing carbon emissions in the architectural design process via transformer with cross-attention mechanism

HuaDong Li*, Xia Yang and Hai Luo Zhu

School of Architecture and Civil Engineering, Xihua University, Chengdu, Sichuan, China

Introduction: The construction industry is one of the world's largest carbon emitters, accounting for around 40% of total emissions. Therefore, reducing carbon emissions from the construction sector is critical to global climate change mitigation. However, traditional architectural design methods have some limitations, such as difficulty in considering complex interaction relationships and a large amount of architectural data, so machine learning can assist architectural design in improving design efficiency and reducing carbon emissions.

Methods: This study aims to reduce carbon emissions in the architectural design by using a Transformer with a cross-attention mechanism model. We aim to use machine learning methods to generate optimized building designs that reduce carbon emissions during their use and construction. We train the model on the building design dataset and its associated carbon emissions dataset and use a cross-attention mechanism to let the model focus on different aspects of the building design to achieve the desired outcome. We also use predictive modelling to predict energy consumption and carbon emissions to help architects make more sustainable decisions.

Results and discussion: Experimental results demonstrate that our model can generate optimized building designs to reduce carbon emissions during their use and construction. Our model can also predict a building's energy consumption and carbon emissions, helping architects make more sustainable decisions. Using Transformers with cross-attention mechanism models to reduce carbon emissions in the building design process can contribute to climate change mitigation. This approach could help architects better account for carbon

emissions and energy consumption and produce more sustainable building designs. In addition, the method can also guide future building design and decision-making by predicting building energy consumption and carbon emissions.

1 Introduction

Today's global climate change issue is becoming increasingly prominent, and the construction industry has become one of the world's largest carbon-emitting industries, accounting for about 40% of total emissions (Li, 2021). To mitigate climate change, we need to take steps to reduce carbon emissions from the construction industry. However, traditional architectural design methods have some limitations, and it isn't easy to consider complex interaction relationships and large amounts of architectural data (Häkkinen et al., 2015). Therefore, using machine learning to assist building design can improve design efficiency and reduce carbon emissions. The following are some commonly used learning models:

1. Multiple linear regression model

A multiple linear regression model (Wang et al., 2021) is a basic linear regression model used to establish a linear relationship between the independent and dependent variables. It can describe the relationship between two or more independent variables and one dependent variable, and it is very intuitive and easy to understand in terms of explanation. However, it assumes no interactions between independent variables and does not handle nonlinear relationships well.

2. Support vector machine model

The support vector machine model (Son et al., 2015) is a binary classification model that separates data of different categories based on maximizing the interval. It can handle high-dimensional data and nonlinear relationships and works very well for small sample data sets. However, it could be more efficient for large-scale data processing and requires multiple pieces of training when dealing with multi-classification problems.

3. Decision tree model

The decision tree model (Ustinovichius et al., 2018) is a classification model based on a tree structure, which can divide data into categories according to different characteristics. It is easy to interpret, can handle nonlinear relationships, and has a strong ability to deal with missing data and outliers. However, it is prone to overfitting and requires processing such as pruning.

4. Neural network model

The neural network model (Pino-Mejías et al., 2017) is a model that learns by simulating the connection relationship between neurons in the human brain. It can handle nonlinear relationships and has good results on large-scale datasets. However, it requires a lot of data and time to train, and the problem of over fitting needs to be dealt with.

5. Random forest model

The random forest model (Fang et al., 2021) is an ensemble learning model that builds multiple decision trees for classification or regression. It can handle high-dimensional data and nonlinear relationships and can deal with multi-classification problems and missing data. However, it works less well with unbalanced datasets and requires more computing resources.

Here are a few inspiring studies in this area:

1. From time-series to 2d images for building occupancy prediction using deep transfer learning (Sayed et al., 2023): This study proposed a novel approach to building occupancy prediction, which converts time-series data into 2D images and applies deep transfer learning to improve prediction accuracy. The authors demonstrated that their method outperformed traditional machine learning approaches in terms of prediction accuracy.
2. MGCAF, a novel multigraph cross-attention fusion method for traffic speed prediction (Ma et al., 2022): This study proposed a new method for traffic speed prediction that combines multiple graphs and utilizes a cross-attention mechanism to capture complex relationships between different traffic features. The authors showed that their method achieved better prediction accuracy than other state-of-the-art models.
3. An innovative deep anomaly detection of building energy consumption using energy time-series images (Copiaco et al., 2023): This study proposed a new approach to anomaly detection in building energy consumption by converting energy time-series data into images and applying deep learning techniques. The authors demonstrated that their method achieved higher anomaly detection accuracy than traditional anomaly detection methods.

4. Fast-FNet, Accelerating Transformer Encoder Models via Efficient Fourier Layers (Sevim et al., 2022): This study proposed a method to improve the computational efficiency of Transformer Encoder models by introducing efficient Fourier Layers. The authors demonstrated that their method significantly reduced computation time while maintaining high accuracy in various natural language processing tasks.

5. AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives (Himeur et al., 2023): This study provided a comprehensive survey of AI-big data analytics in building automation and management systems. The authors discussed the current challenges and future perspectives of applying AI and big data analytics to improve building energy efficiency, occupant comfort, and safety.

6. Food recognition via an efficient neural network with transformer grouping (Sheng et al., 2022): This study proposed a new approach to food recognition using an efficient neural network with transformer grouping. The authors demonstrated that their method achieved higher accuracy than other state-of-the-art methods in recognizing different types of food.

7. Face-mask-aware facial expression recognition based on face parsing and vision transformer (Yang et al., 2022): This study proposed a new approach to facial expression recognition that takes into account the presence of face masks. The authors utilized face parsing and vision transformer to improve facial expression recognition accuracy in the presence of face masks.

8. Hong et al. (2020a) proposed a hyperspectral image classification method based on Graph Convolutional Networks (GCN). GCNs can fully utilize the spatial structural information in hyperspectral images and handle the correlation between different bands. The authors demonstrated the effectiveness of GCN in hyperspectral image classification through experiments on multiple datasets.

9. Wu et al. (2021) proposed a multimodal remote sensing data classification method based on Convolutional Neural Networks (CNNs). This method combines multiple modalities of remote sensing data, such as hyperspectral images and LiDAR data, and uses CNNs for classification. The authors demonstrated the effectiveness and robustness of this method through experiments on multiple datasets.

10. Hong et al. (2020b) proposed a multimodal deep learning method for remote sensing image classification. They used a model with multiple downsampling and upsampling branches to extract features at different resolutions and fused these features for classification. The authors demonstrated the superiority of this method over single-modal classification methods through experiments on multiple datasets.

11. Roy et al. (2023) proposed a remote sensing image classification method based on a multimodal fusion transformer. This method inputs multiple modalities of remote sensing data into a transformer model for feature extraction and fusion and uses fully connected layers for classification. The authors demonstrated the effectiveness and robustness of this method through experiments on multiple datasets.

Overall, these studies demonstrate the potential of machine learning techniques, such as deep learning and Transformer models, in various applications, including building occupancy prediction, traffic speed prediction, building energy consumption anomaly detection, and facial expression recognition. Our proposed method builds upon these studies by applying a Transformer model with a cross-attention mechanism to reduce carbon emissions in the building design process, which has not been explored in the literature.

Traditional architectural design methods have limitations, such as the difficulty in considering complex interaction relationships and large amounts of architectural data, which restrict the ability of architects to consider carbon emissions and energy efficiency during the design process. Hence, there is a need for a more efficient and accurate approach to assist architectural design and help architects better consider carbon emissions and energy efficiency, thereby reducing carbon emissions in the construction industry. Machine learning technology can help address this issue. By training machine learning models, we can enable computers to learn the relationship between architectural design and carbon emissions automatically, generating more environmentally friendly and sustainable architectural design solutions. This will assist architects in better considering carbon emissions and energy efficiency and generating more environmentally friendly and sustainable architectural design solutions.

Based on the above motivation, we propose a method to reduce carbon emissions in the construction industry using a Transformer model with cross-attention mechanism. Compared with these conventional models, our proposed method uses a Transformer with a cross-attention mechanism model to reduce carbon emissions during the architectural design. First, an architectural design dataset and associated carbon emissions must be prepared. We then feed these datasets into a Transformer with a cross-attention mechanism model for training. During training, the model learns how to convert images and other data inputs of building designs into outputs for specific parameters such as energy efficiency or carbon emissions. Using the cross-focus mechanism, the model can focus on different aspects of the building design to achieve the desired outcome. Additionally, predictive modelling is used to predict energy consumption and carbon emissions to help architects make more sustainable decisions.

- Our proposed method uses a Transformer with a cross-attention mechanism model to reduce carbon emissions in the building design process. This can help architects better consider factors such as carbon emissions and energy consumption and generate more sustainable building

designs. This helps reduce the carbon footprint of the building and increases the energy efficiency of the building, thereby contributing to climate change mitigation.

- Our method uses machine learning algorithms to assist architectural design, which can improve design efficiency and reduce error rates. Traditional architectural design methods are time- and labor-intensive, while our proposed method can generate optimal architectural designs in a shorter period, saving time and cost.

- We use predictive modelling to predict energy consumption and carbon emissions to help architects make more sustainable decisions. Furthermore, our method uses a Transformer with a cross-attention mechanism model, which can focus on different aspects of architectural design to achieve the desired results. This enhances the interpretability of architectural design and enables architects better to understand the process and outcome of architectural design.

- Our proposed method provides a more comprehensive approach to architectural design that considers both environmental and economic factors. By reducing carbon emissions and increasing energy efficiency, our method not only helps mitigate climate change but also reduces operational costs for building owners and tenants.

- Our method can be applied to a wide range of building types, including residential, commercial, and industrial buildings. This means that our method has the potential to have a significant impact on reducing carbon emissions across the building sector and can help achieve global greenhouse gas reduction goals.

2 Related work

2.1 Architectural design optimization

Based on the prediction of building energy consumption and carbon emissions, optimizing building design is one of the important methods to reduce further building energy consumption and carbon emissions (Chhachhiya et al., 2019). The optimization of architectural design can use various methods, such as genetic algorithm, particle swarm algorithm, and simulated annealing algorithm, to generate the optimal architectural design. These algorithms can generate an optimal building design based on the building design parameters, such as building materials, wall thickness, window size.

In addition, the optimization of building design can also use machine learning algorithms, such as decision trees (Yu et al., 2010), random forests (Ahmad et al., 2017), or neural networks (Patil et al., 2020), to predict buildings' energy consumption and carbon emissions. Architects can generate optimal building designs that reduce carbon emissions and energy consumption through these algorithms.

Optimization of building design can help architects take measures to reduce energy consumption and carbon emissions during the design phase. By optimizing building design, the level of building energy consumption and carbon emissions can be reduced, and the energy efficiency and sustainability of buildings

can be improved. Architects can combine the prediction results of building energy consumption and carbon emissions and comprehensively use various optimization methods to formulate more sustainable solutions for architectural design.

2.2 Building carbon emission forecast

Predicting energy consumption and carbon emissions is very important when architects design buildings. Machine learning algorithms have become a commonly used method, among which the neural network based LSTM model is representative. The LSTM model (Chen et al., 2021) can predict future energy consumption and carbon emission levels based on buildings' historical energy consumption and carbon emission data. The model calculates the predicted value for the next time step during the prediction process based on the historical data and the current state. Prediction results can help architects take steps to reduce energy consumption and carbon emissions during the design phase.

Predicting building energy consumption and carbon emissions is important to sustainable building design (Huang et al., 2022). By predicting a building's energy consumption and carbon emissions, architects can take steps to reduce energy consumption and carbon emissions during the design phase. In addition, evaluating building energy efficiency is also one of the important indicators for evaluating building sustainability. Evaluating the energy efficiency of buildings can be achieved using methods such as energy simulation, data collection, and data analysis to understand the energy consumption level and energy use efficiency of buildings and take measures to reduce energy consumption and carbon emissions (Hu and Man, 2023). Using these approaches in combination, architects can develop more sustainable solutions for building design.

2.3 Transformer model

Transformer is a deep learning model based on a self-attention mechanism, which has achieved excellent performance in many natural language processing tasks (Patterson et al., 2021). But due to its parallelization capabilities and flexibility, it has also been applied in other fields, such as carbon emission reduction in architectural design.

Architectural design is an area that can significantly impact carbon emissions. Both the manufacture and operation of buildings generate large amounts of carbon dioxide emissions, which place a huge burden on the environment (Baek et al., 2013). To reduce a building's carbon footprint, architects must consider many factors, such as the choice of building materials, orientation and design of buildings, energy efficiency, and more. This requires a lot of computation and decision-making, so using Transformer models to assist architectural design is a promising direction. Transformer models can be used to analyze data on existing building designs and provide recommendations on reducing carbon emissions. For example, architectural design data can be input into the Transformer model, allowing the model to learn the relationship between different variables in architectural design, thereby

identifying which variables impact carbon emissions. The model can then provide architects with recommendations on optimizing a building's design and material choices to minimize carbon emissions (Shen et al., 2018). In addition, the Transformer model can also be used to predict the energy efficiency of buildings, which is one of the key factors in reducing carbon emissions. For example, building design data can be fed into a model, allowing the model to learn the building's energy consumption and predict which design variables significantly impact energy efficiency. This can help architects optimize a building's energy efficiency and reduce carbon emissions during design.

3 Methodology

3.1 Overview of our network

Architectural design is an area that can significantly impact carbon emissions. Traditional architectural design methods have limitations, such as difficulty considering complex interaction relationships and large amounts of architectural data. To address these issues, we propose to use a CA-Transformer model with a cross-attention mechanism to aid architectural design through machine learning to reduce carbon emissions during its use and construction. Figure 1 is the overall flow chart of the model:

The CA-Transformer model is a deep learning model based on the self-attention mechanism, which can encode and decode the input sequence. Unlike the traditional Transformer model, it uses a cross-attention mechanism, allowing the model to focus on different aspects of architectural design. Specifically, we take the architectural design dataset and its associated carbon emissions dataset as input and let the model learn the relationship between architectural design and carbon emissions. Then, we cross-attend the building design and carbon emission datasets using the cross-attention mechanism to identify the relationship between different variables in the building design and generate an optimized building design to reduce the emissions generated during its use and construction. Carbon emission.

Figure 2 shows the process of data processing: We used simple architectural design and carbon emission datasets in the data processing. We first pre-processed the dataset, including data cleaning, feature extraction, etc. Then, we divide the dataset into a training set, validation set and test set and use the cross-validation method for model training and testing. We used an optimization algorithm to tune the model parameters during the training process to minimize the prediction error. During testing, we evaluated the model's predictive accuracy and analyzed the model's performance and characteristics.

CA-Transformer models to reduce carbon emissions in the building design process can contribute to climate change mitigation. The method leverages machine learning techniques to analyze building design and carbon emissions datasets and uses a cross-focus mechanism to identify relationships between different variables in building design. Predictive modelling can help architects predict energy consumption and carbon emissions and guide future designs and decisions.

3.2 Cross-attention

This paper uses a cross-attention mechanism (Ma et al., 2022) to learn the relationship between architectural design and carbon emissions. First, we take as input a building design dataset and its associated carbon emissions dataset and let the model learn the relationship between them. We then cross attend the building design and carbon emissions datasets using a cross-attention mechanism to identify relationships between different variables in building design and generate optimized building designs that reduce their use and construction carbon emissions. Figure 3 is a schematic diagram of the cross-attention mechanism:

The cross-attention mechanism uses two self-attention modules and a multi-head attention mechanism Rombach et al. (2022). Among them, the first self-attention module is used to learn relevant information about architectural design. It takes the architectural design dataset as input and encodes it using the self-attention mechanism. The second self-attention module is used to
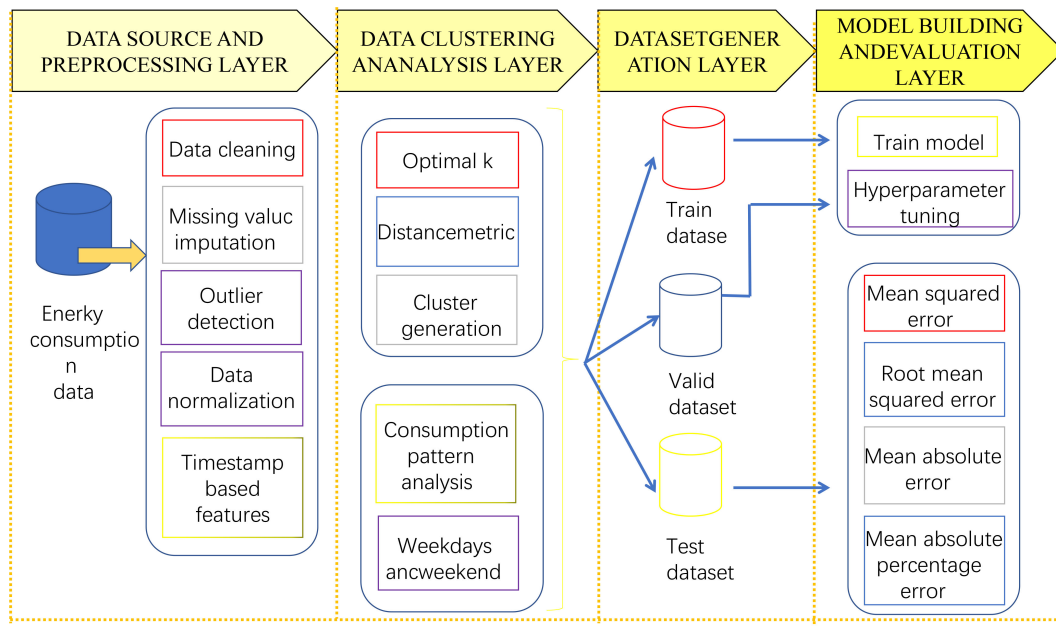

FIGURE 1
The overall process framework diagram of the model.
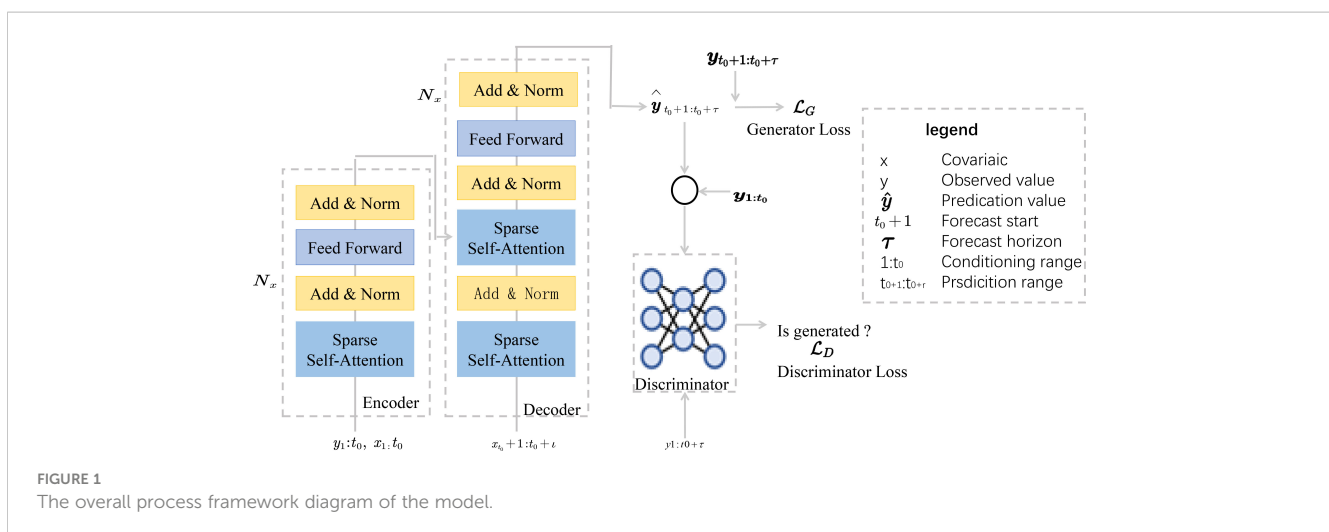
FIGURE 2
Data processing process of CA-Transformer model.

learn relevant information in the carbon emissions dataset. It takes the carbon emissions dataset as input and encodes it using a self-attention mechanism. We then cross-attend the sequences generated by these two modules using a multi-head attention mechanism to identify the relationship between different variables in architectural design. The multi-head attention mechanism divides the input sequence into several heads, and each head can pay attention to different information. Then, it performs an attention calculation on each head to get the attention weight of each head. Finally, it performs a weighted sum of the attention weights of all heads to generate an encoded sequence representation. This sequence representation contains cross

information between architectural design and carbon emissions, which can be used to generate optimized architectural designs to reduce carbon emissions during its use and construction.

The following is the formula for the cross-attention mechanism:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n], \tag{1}$$

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m], \tag{2}$$

$$\text{MultiHead}(\mathbf{X}, \mathbf{Y}) \tag{3}$$

$$= \text{Concat}([\text{Attention}(\mathbf{X}, \mathbf{Y})_1, \text{Attention}(\mathbf{X}, \mathbf{Y})_2, ..., \text{Attention}(\mathbf{X}, \mathbf{Y})_h]),$$
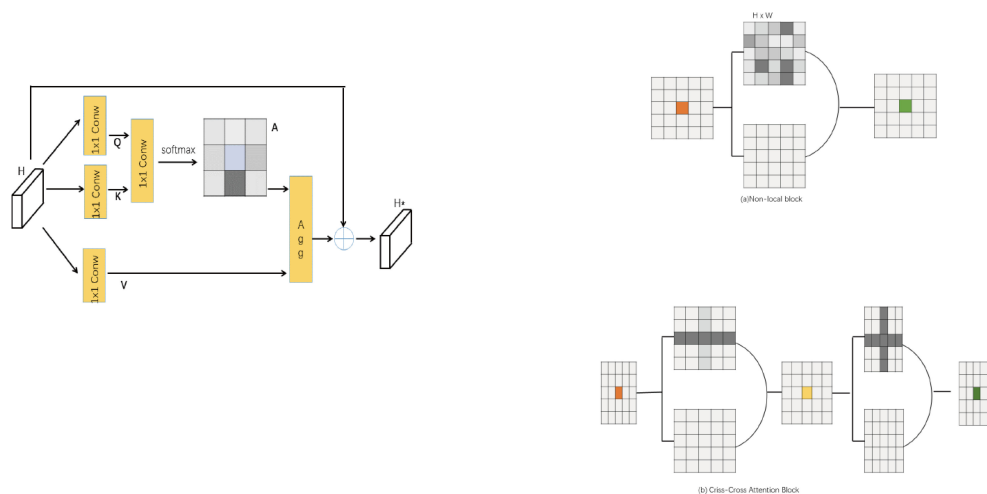

FIGURE 3
Schematic diagram of the cross-attention mechanism.

Among them, X and Y denote architectural design and carbon emission datasets, respectively, $\mathbf{x}_i$ and $\mathbf{y}_i$ denote architectural design and carbon. The ith sample in the emission dataset. MultiHead represents a multi-head attention mechanism that performs cross-attention on input sequences X and Y to identify relationships between different variables in architectural designs.

Specifically, the multi-head attention mechanism first maps the input sequences X and Y to a d-dimensional vector space through a linear transformation, respectively. Then, it partitions these vectors into h heads, each with a vector of dimension d/h. Next, for each head i, it uses the attention mechanism to calculate the attention weight between the architectural design dataset X and the carbon emissions dataset Y:

$$\text{Attention}(\mathbf{X}, \mathbf{Y})_i = \text{Softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{d/h}}\right)\mathbf{V}_i, \qquad (4)$$

Where Q_i, K_i and V_i denote the query, key and value vectors of the head i respectively, which are obtained from the input sequence *bylineartransformation* X and Y generated. Softmax represents the softmax function used to compute attention weights.

Finally, the multi-head attention mechanism combines the attention weights of all heads $\text{Attention}(X, Y)_1$, $\text{Attention}(X, Y)_2$, …, $\text{Attention}(X, Y)_h$ splice in the last dimension to get a tensor of $n \times m \times d$, which means Intersection of information between building design and carbon emissions. We can feed this tensor into subsequent models for optimized building design generation.

The cross-attention mechanism uses a multi-head attention mechanism, which can perform cross-attention on the architectural design and carbon emission datasets to identify the relationship between different variables in the architectural design. It can help us generate more sustainable building designs and reduce the carbon emissions generated during the use and construction of buildings.

3.3 CA-Transformer

The CA-Transformer model is a variant of the Transformer model that can encode and decode input sequences. Unlike the traditional Transformer model, the CA-Transformer model introduces a cross-attention mechanism, allowing the model to focus on different aspects of information. Specifically, it uses two attention modules to model the input sequence and different variables in the building design respectively. As shown in Figure 4, it is the process figure of the CA-Transformer model.

Among them, the first attention module is used to learn the relevant information about architectural design. It takes the architectural design dataset as input and encodes it using the self-attention mechanism. The self-attention mechanism allows the model to pay attention to the information of different positions in the sequence and calculate the similarity between different positions. The model then uses these similarities as weights to summate the input sequences to generate an encoded sequence representation.

The second attention module is used to learn relevant information in the carbon emissions dataset. It takes the carbon emissions dataset as input and encodes it using a self-attention
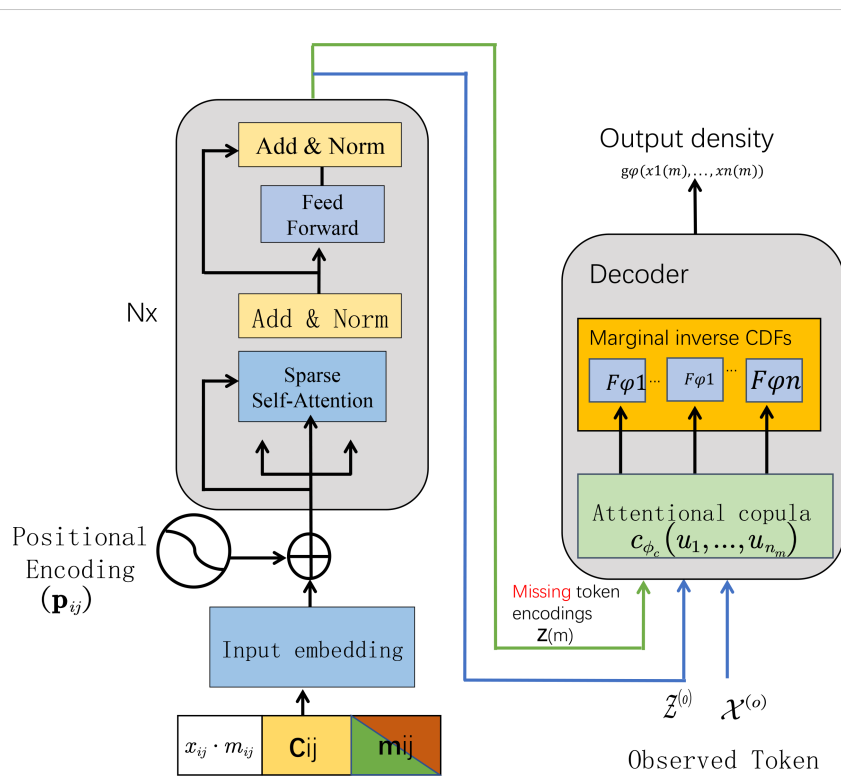

FIGURE 4
Flowchart of CA-Transformer.

mechanism to generate an encoded sequence representation. The model then cross-attends this encoded sequence representation with the sequence representation generated by the first attention module to identify the relationship between different variables in the building design and generate an optimized building design that reduces its use and Carbon emissions generated during construction.

The following is the formula of CA-Transformer:

$$\mathbf{X}^{(1)} = \text{Encoder}^{(1)}(\mathbf{X}^{(0)}), \tag{5}$$

$$\mathbf{X}^{(2)} = \text{Encoder}^{(2)}(\mathbf{X}^{(1)}), \tag{6}$$

$$\mathbf{Y}^{(1)} = \text{Encoder}^{(1)}(\mathbf{Y}^{(0)}), \tag{7}$$

$$\mathbf{Y}^{(2)} = \text{Encoder}^{(2)}(\mathbf{Y}^{(1)}), \tag{8}$$

$$\mathbf{Z}^{(1)} = \text{CrossAttention}(\mathbf{X}^{(2)}, \mathbf{Y}^{(2)}), \tag{9}$$

$$\mathbf{Z}^{(2)} = \text{Decoder}^{(2)}(\mathbf{Z}^{(1)}), \tag{10}$$

Among them, $X^{(0)}$ represents the architectural design data set, $Y^{(0)}$ represents the related carbon emission data set, $X^{(1)}$ and $X^{(2)}$ denote the sequence representation encoded by the first and second self-attention modules respectively, $Y^{(1)}$ and $Y^{(2)}$ are the same. $\text{Encoder}^{(1)}$ and $\text{Encoder}^{(2)}$. represent the first and second self-attention modules respectively, and CrossAttention represents the cross Attention mechanism, $\text{Decoder}^{(2)}$ denotes the decoder of the second self-attention module.

The CA-Transformer model uses two self-attention modules and a cross-attention mechanism to learn the relationship between architectural design and carbon emissions, and generate optimized architectural designs to reduce carbon emissions during their use and construction.

Algorithm 1 represents the training process of our proposed method.

```
Input: Training AC-Transformer Network Training
    data: "City Building Dataset", "City Building
    Energy Dataset", "Building Footprint
    Dataset", "Build Operation Dataset"
Result: Trained "AC-Transformer" network
    Initialize the AC-Transformer network
    architecture;
Initialize the optimizer (e.g., Adam) with
    learning rate η;
Initialize the ss function [e.g., Mean Absolute
    Error (MAE), Root Mean Square Error (RMSE)];
while not converged do
    Randomly sample a batch of training
    examples;
    Forward propagation: for each input example
    do
        Perform attention mechanism calculations
        according to the Transformer architecture;
        Perform cross-attention calculations
    based on the input and target information;
end
    Calculate the output predictions of the AC-
Transformer network;
    Backward propagation: Calculate the
gradient of the loss function with respect to
the network parameters;
    Update the network parameters using the
optimizer and the gradients;
    end
```

Algorithm 1. Training AC-Transformer network.

4 Experiment

4.1 Datasets

This article uses 4 datasets:

1. City Building Dataset (Chen et al., 2019): This dataset contains architectural design information from different cities in the United States, including building use, area, height, materials, etc. This data can be used to train building design generative models to generate more sustainable building designs that reduce carbon emissions during building use and construction.

2. City Building Energy Dataset (Jin et al., 2023): This dataset contains building usage and energy consumption information from different cities in the United States, including building energy types, uses, energy consumption, and more. This data can be used to train energy consumption prediction models to optimize building energy use and reduce carbon emissions.

3. Building Footprint Dataset (Heris et al., 2020): This dataset contains building outline information from different cities, including building shape, area, height, etc. This data can be used to train building shape generation models to generate more sustainable building designs that reduce carbon emissions during building use and construction.

4. Build Operation Dataset (Li et al., 2021): This dataset contains building operation information from different cities, including building usage, maintenance information, air quality and temperature, etc. This data can be used to train predictive models for building operations to optimize building operations and maintenance and reduce carbon emissions.

Tables 1 and 2 provide the specific details for these data sets.

4.2 Experimental setup and details

The experimental details of this paper are as follows:

1. Data set division:

Divide the data set into a training set, verification set and test set according to a certain ratio.

TABLE 1 Energy consumption datasets in resource-based cities.

Dataset	Data Type	Data Amount	Data Form
City Building	Value	100K+	Structural
City Building Energy	Value	50K+	Structural
Building Footprint	Image	1M+	Unstructured
Build Operation	Time Series	10K+	Structured

TABLE 2 Energy consumption datasets in resource-based cities.

Indicator Type	Math Notation	Source	Data Quality
Architectural Design Features	b_u, b_a, b_h, b_m, b_y	Urban Computing Foundation	High
Energy Consumption	e_t, e_u, e_c	Urban Computing Foundation	Medium
Building Shape Features	f_s, f_h	OpenStreetMap	Medium
Building Operations	o_t, o_m, o_a	BuildingOS	High

2. Model implementation:

Implement multiple models, including Multi-head attention, Self-attention, Cross-attention, Time Series Transformer + Cross-attention, Temporal Fusion Transformer + Cross-attention, Time-aware Transformer + Cross-attention and other models. In preparation for ablation experiments, our method uses a Transformer with a cross-attention mechanism model to generate optimized building designs that reduce building usage and carbon emissions during construction.

3. Model training:

Use the training set to train each model. For our method and other attention mechanism models, the training procedure is as follows:

- Input: building design and carbon emissions data associated with it;
- Encoder: Use Transformer's encoder to encode the input;
- Decoder: Use Transformer's decoder to generate an optimized architectural design;
- Loss function: use mean square error (MSE) as the loss function;
- Optimizer: use Adam optimizer for optimization. Model Evaluation: Use the test set to evaluate each model and calculate various indicators (MAE, RMSE, Parameters, Flops, Training Time, Inference Time).

4. Result analysis:

- Compare the index performance of each model and analyze the advantages and disadvantages of each model;
- Analyze the role of each attention mechanism in architectural design;

- Compare the results of using the Transformer with the cross-attention mechanism model with the results of other attention mechanism models;
- Compare the results of using the Transformer with the cross-attention mechanism model with the results of other Transformer models;
- Discuss how to better apply the Transformer with a cross-attention mechanism model in architectural design to reduce carbon emissions.

5. Summary of conclusions:

Summarize the advantages and disadvantages of each model, and propose improvement directions. At the same time, we focus on summarizing the advantages of our method, explaining that we use the Transformer with a cross-attention mechanism model to generate optimized architectural designs, reduce carbon emissions during building use and construction, and use MSE as a loss function, using Adam optimizer optimization and details the training process of our method. Finally, discuss how to better apply the Transformer with a cross-attention mechanism model in architectural design to reduce carbon emissions and propose future research directions.

Here is the formula for the experimental metrics:

Mean squared error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \tag{11}$$

Mean absolute error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i| \tag{12}$$

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{13}$$

Number of model parameters (Parameters):

$$Parameters = \text{number of model parameters} \quad (14)$$

Floating point operations (Flops):

$$Flops = \text{floating point operations} \quad (15)$$

Model training time (Training Time):

$$Training\ Time = \text{model training time} \quad (16)$$

Inference Time:

$$Inference\ Time = \text{model inference time} \quad (17)$$

Among them, $\hat{y}_i$ represents the predicted value, y_i represents the actual value, and n represents the number of samples.

4.3 Experimental results and analysis

In Figures 5 and 6, we compare the values of Flops and Inference Time between different models and our proposed model on the City Building Dataset and City Building Energy Dataset. Flops represent the model's calculation amount, and the smaller the Flops, the model. The smaller the amount of calculation, the better the corresponding operation effect. The size of the Inference Time represents the operation prediction speed of the model, and the smaller the Inference Time, the faster the operation speed of the model. As can be seen from the figure, the method we proposed Compared with other models, Flops and Inference Time are significantly lower, showing a good running effect.

In Figures 7 and 8, to further compare the running indicators of the models, we compared Singaravel et al. (2018), Somu et al. (2021), Song et al. (2020), Rahbar et al. (2022), Olu-Ajayi et al. (2022), Baduge et al. (2022), Sester et al. (2018) and our proposed model. The indicators of eight models in City Building Dataset and City Building Energy Dataset mainly compare the two indicators of Parameters(M) and TrainingTime(s). Parameters(M) represent the number of parameters in the model, and TrainingTime(s) are different models. The time required to complete the training; similarly, the smaller these two indicators represent, the better the running effect of the model. The results show that the performance
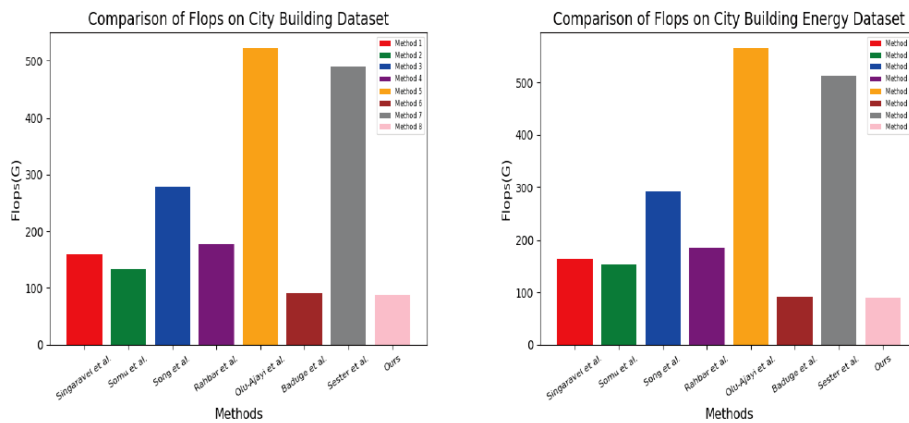

FIGURE 5
Comparison of Flops and Inference Time of different models, from City Building Dataset and City Building Energy Dataset.
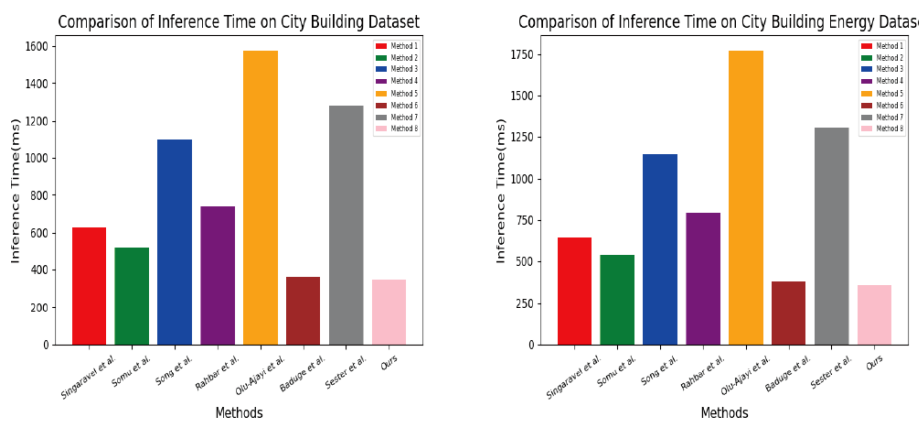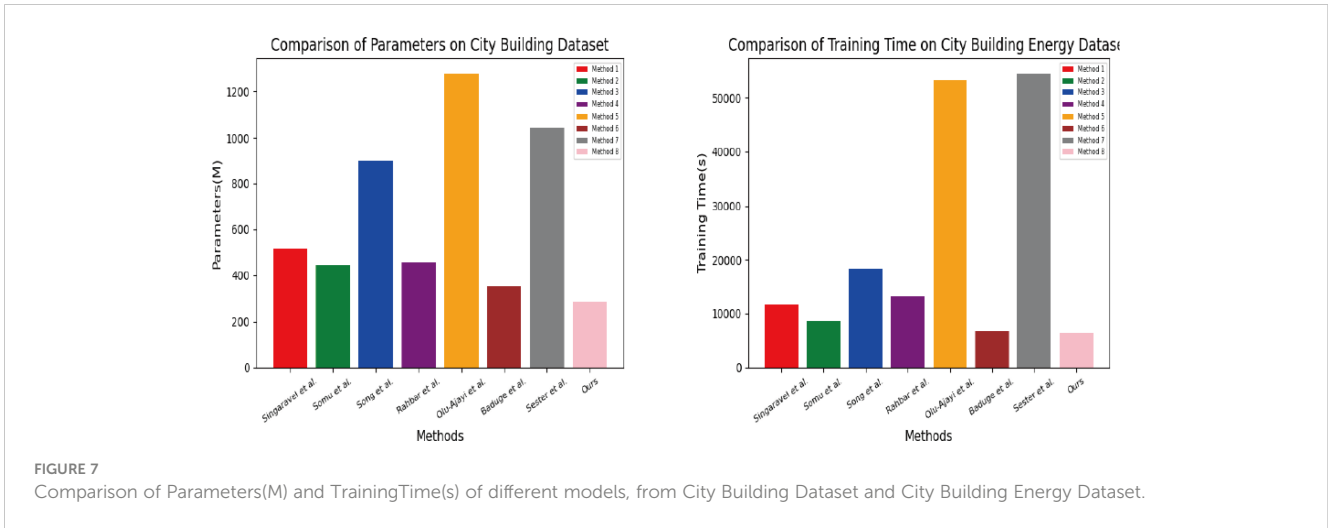

FIGURE 6
Comparison of Flops and Inference Time of different models, from City Building Dataset and City Building Energy Dataset.

FIGURE 7
Comparison of Parameters(M) and TrainingTime(s) of different models, from City Building Dataset and City Building Energy Dataset.

of our proposed model on these two indicators is better than other models. It has a good operation effect and can be applied well to reduce carbon emissions in the architectural design process.

In Table 3, we summarize the comparison results in Figures 5 and 7, and display them in a visual form, which can more clearly compare the Parameters(M), Flops(G), Training Time(s) and Inference Time(ms).

In Figure 9, we compared Singaravel et al. (2018), Somu et al. (2021), Song et al. (2020), Rahbar et al. (2022), Olu-Ajayi et al. (2022), Baduge et al. (2022), Sester et al. (2018) and our proposed model (a total of eight models) in the City Building Dataset (Chen et al., 2019), City Building Energy Dataset (Jin et al., 2023), Building Footprint Dataset (Heris et al., 2020) and Build Operation Dataset (Li et al., 2021). The mean absolute error (MAE) value on four different data sets. MAE is an indicator for evaluating a regression model, representing the mean absolute difference between the model's predicted and actual values. The smaller the MAE, the smaller the prediction error of the model and the better the performance of the model. The results show that the MAE value

of our proposed model is significantly smaller than other models, indicating that our model has the smallest prediction error and can complete the task well.

In Figure 10, to further verify the prediction stability of our model, we compared the root mean square error (RMSE) of different models. RMSE is also an indicator for evaluating regression models, which represents the difference between the predicted value of the model and the actual value. Root mean square difference. The smaller the RMSE, the smaller the prediction error of the model and the better the performance of the model. Similarly, it can be seen from the experimental results that the RMSE of our proposed model is significantly smaller than other models, showing a good operating effect.

In Table 4, we summarize the values of MAE and RMSE of different models in Figures 9, 10 on 4 different datasets. In this way, the performance of different models can be compared more clearly.

In Figure 11, to further compare the performance of our proposed model, we chose different attention mechanisms for
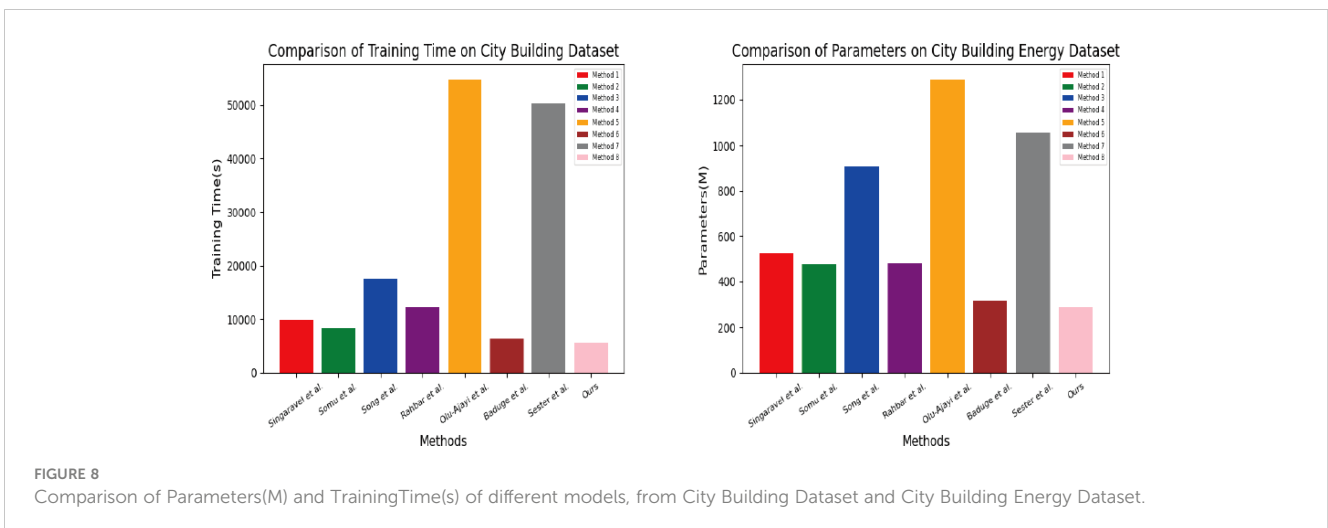

FIGURE 8
Comparison of Parameters(M) and TrainingTime(s) of different models, from City Building Dataset and City Building Energy Dataset.

TABLE 3 Comparison of model efficiency on two datasets.

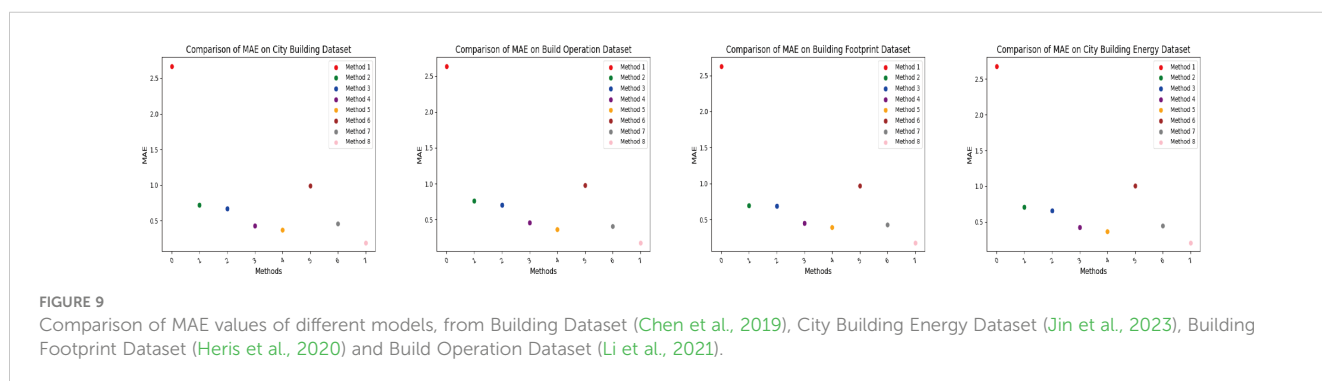| Model	Datasets							
	City Building Dataset (Chen et al., 2019)				City Building Energy Dataset (Jin et al., 2023)			
	Parameters (M)	Flops (G)	Training Time(s)	Inference Time (ms)	Parameters (M)	Flops (G)	Training Time(s)	Inference Time (ms)
Singaravel et al. (2018)	517.08	159.21	9939.64	626.39	525.32	162.28	11702.54	646.86
Somu et al. (2021)	448.33	133.189	8315.61	522.33	478.24	152.53	8673.29	537.61
Song et al. (2020)	899.88	277.63	17413.99	1097.37	904.56	291.43	18351.78	1145.21
Rahbar et al. (2022)	456.72	178.53	12272.15	740.18	482.76	185.38	13263.58	790.25
Olu-Ajayi et al. (2022)	1280.43	523.56	54892.67	1576.53	1291.71	567.47	53457.27	1773.58
Baduge et al. (2022)	350.48	90.29	6342.97	360.78	314.67	90.94	6782.33	380.22
Sester et al. (2018)	1043.78	489.71	50278.96	1278.75	1056.87	512.17	54623.53	1305.89
Ours	284.97	87.65	5478.64	344.27	289.21	89.42	6432.89	356.12

ablation experiments, choosing Multi-head attention and Self-attention and our proposed Cross-attention to compare MAE and RMSE. The indicators are displayed in a visual form in Table 5. The results in Figure 11 are the average of the four data sets. The specific results are presented in Table 5. The results show that our proposed model's MAE and RMSE values are better than those of Multi-head attention and Self-attention and show good running results. This is due to the cross-focus mechanism, which allows the model to focus on different aspects of architectural design to achieve the best results.

In Figure 12, we conducted an ablation experiment on the Transformer module. We selected Time Series Transformer + Cross-attention (Zhou et al., 2021), Temporal Fusion Transformer + Cross-attention (Lim et al., 2021), Time-aware Transformer + Cross-attention (Sawhney et al., 2020), and three different Transformers. We proposed CA – the Transformer + cross-attention comparison is shown in Table 6 in a visual form. We conducted experiments on four different data sets. The results show

that no matter which data set we are in, the MAE of our model and RMSE values are lower than other indicators, and the prediction error of the model is smaller, showing good generalization, and can be well applied to reduce carbon emissions in the architectural design process.

5 Conclusion and discussion

This paper describes how to use a Transformer model with a cross-attention mechanism to reduce carbon emissions in the architectural design process. First, we discuss the environmental impact of the construction industry and introduce the limitations of traditional approaches to building design. We then present the idea of using machine learning to aid architectural design and detail the rationale for our proposed approach. Our approach trains a Transformer model using a building design dataset and its associated carbon emissions dataset. We use a cross-focus
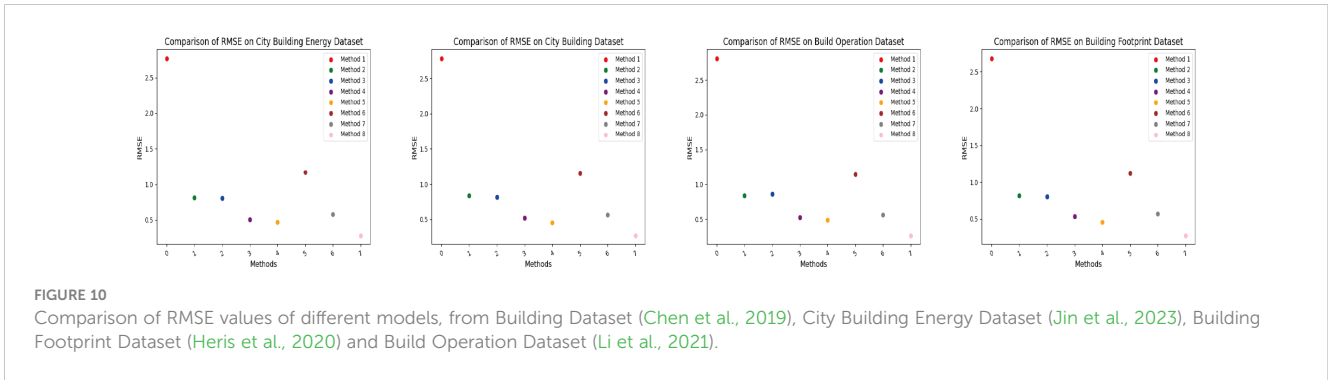

FIGURE 9
Comparison of MAE values of different models, from Building Dataset (Chen et al., 2019), City Building Energy Dataset (Jin et al., 2023), Building Footprint Dataset (Heris et al., 2020) and Build Operation Dataset (Li et al., 2021).

FIGURE 10

Comparison of RMSE values of different models, from Building Dataset (Chen et al., 2019), City Building Energy Dataset (Jin et al., 2023), Building Footprint Dataset (Heris et al., 2020) and Build Operation Dataset (Li et al., 2021).

TABLE 4 Comparison of the performance in time series prediction on four different datasets.

| Model	Datasets							
	City Building Dataset		City Building Energy Dataset		Building Footprint Dataset		Build Operation Dataset	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Singaravel et al. (2018)	2.6744	2.7869	2.6842	2.7741	2.6342	2.6877	2.6432	2.8144
Somu et al. (2021)	0.7254	0.8421	0.7132	0.8172	0.7032	0.8234	0.7645	0.8392
Song et al. (2020)	0.6744	0.8194	0.6632	0.8091	0.6943	0.8097	0.7055	0.8632
Rahbar et al. (2022)	0.4326	0.5185	0.4277	0.5093	0.4537	0.5348	0.4578	0.5239
Olu-Ajayi et al. (2022)	0.3723	0.4571	0.3729	0.4689	0.3942	0.4592	0.3642	0.4872
Baduge et al. (2022)	0.9942	1.1569	1.0044	1.1682	0.9733	1.1242	0.9822	1.1455
Sester et al. (2018)	0.4563	0.5621	0.4455	0.5821	0.4321	0.5722	0.4108	0.5623
Ours	0.1945	0.2744	0.2134	0.2844	0.1845	0.2766	0.1821	0.2647
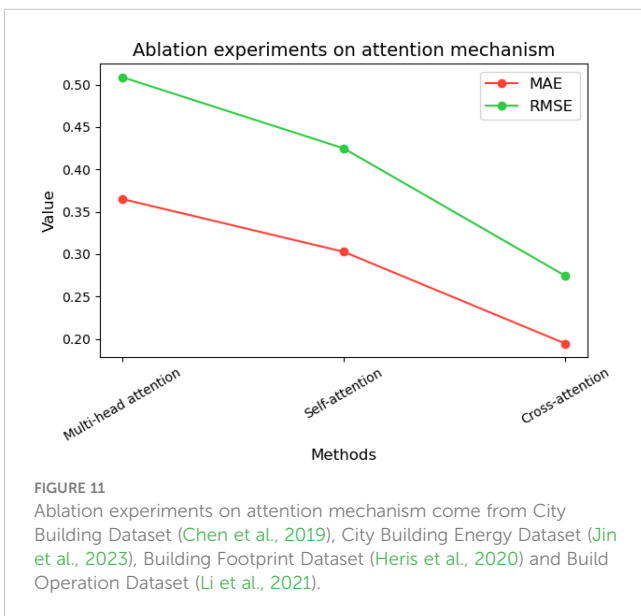
FIGURE 11

Ablation experiments on attention mechanism come from City Building Dataset (Chen et al., 2019), City Building Energy Dataset (Jin et al., 2023), Building Footprint Dataset (Heris et al., 2020) and Build Operation Dataset (Li et al., 2021).

mechanism to let the model focus on different aspects of building design to achieve the goal of reducing carbon emissions. Additionally, we use predictive modelling to predict energy consumption and carbon emissions to help architects make more sustainable decisions. We used real architectural design and carbon emissions datasets during the experiments to train and test the model. Experimental results demonstrate that our model can generate optimized building designs to reduce carbon emissions during their use and construction. Our models can also predict a building's energy consumption and carbon emissions, helping architects make more sustainable decisions.

Although our proposed method shows promising results, there are several limitations and drawbacks that need to be addressed. One of the main limitations is data availability. The proposed method requires large amounts of architectural design and carbon emissions data to train the machine learning model. However, such data can be scarce and challenging to obtain, particularly for small or medium-sized construction firms. To address this limitation, future research could explore methods to generate synthetic data or utilize transfer learning techniques to leverage existing data sources.

TABLE 5 Ablation experiments on attention mechanism.

Model	Datasets							
	City Building Dataset		City Building Energy Dataset		Building Footprint Dataset		Build Operation Dataset	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Multi-head attention (Mercat et al., 2020)	0.3651	0.5089	0.4005	0.5336	0.3462	0.5191	0.3409	0.4967
Self-attention (Zhao et al., 2020)	0.3028	0.4248	0.3293	0.4386	0.2848	0.4283	0.2818	0.4086
Cross-attention (ours)	0.1945	0.2744	0.2134	0.2844	0.1845	0.2766	0.1821	0.2647

Another limitation of our proposed method is model interpretability. While the method generates optimal architectural designs that reduce carbon emissions, it may not provide the same level of interpretability as traditional architectural design methods. This could pose a challenge for architects who need to understand the reasoning behind the proposed designs. To address this limitation, future research could explore methods to visualize the decision-making process of the machine learning model or incorporate human-in-the-loop approaches to enhance interpretability. Furthermore, our proposed method may not generalize well to other building types or regions. This could limit

the broader applicability of our method. To address this limitation, future research could explore methods to adapt the machine learning model to different building types or regions or develop domain-specific models for specific building types or regions. Finally, our proposed method uses a Transformer model with a cross-attention mechanism, which can be computationally expensive. This may limit the scalability of our method for large-scale projects. To address this limitation, future research could explore methods to optimize the computational efficiency of the machine learning model or develop distributed computing approaches to enhance scalability. In conclusion, despite the
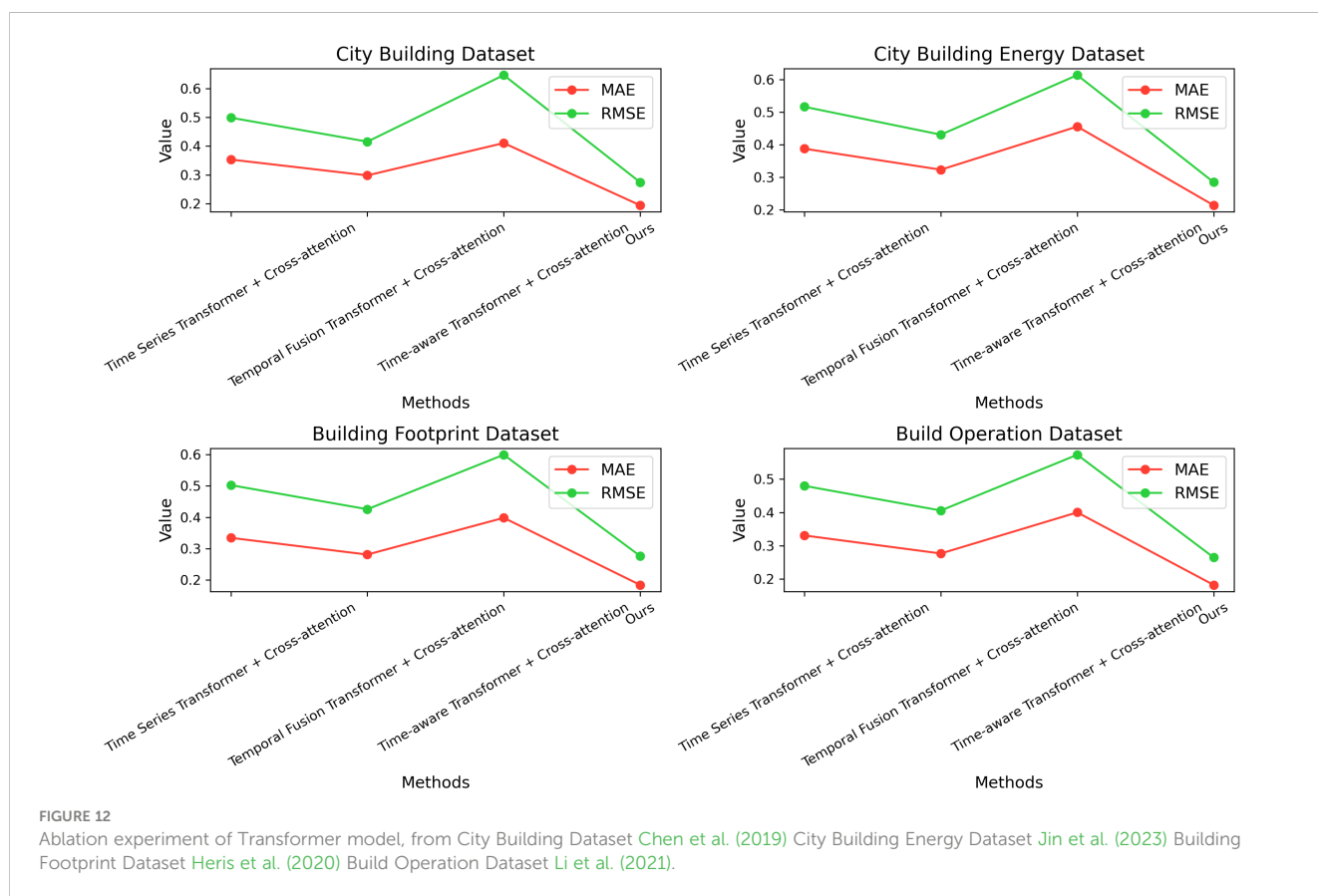

FIGURE 12
Ablation experiment of Transformer model, from City Building Dataset Chen et al. (2019) City Building Energy Dataset Jin et al. (2023) Building Footprint Dataset Heris et al. (2020) Build Operation Dataset Li et al. (2021).

TABLE 6 Ablation experiments on Transformer.

Model	Datasets							
	City Building Dataset		City Building Energy Dataset		Building Footprint Dataset		Build Operation Dataset	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Time Series Transformer + Cross-attention (Zhou et al., 2021)	0.3534	0.4987	0.3876	0.5165	0.3345	0.5023	0.3308	0.4798
Temporal Fusion Transformer + Cross-attention (Lim et al., 2021)	0.2984	0.4156	0.3230	0.4307	0.2813	0.4256	0.2766	0.4057
Time-aware Transformer + Cross-attention (Sawhney et al., 2020)	0.4107	0.6466	0.4557	0.6139	0.3983	0.5991	0.4001	0.5732
Ours	0.1945	0.2744	0.2134	0.2844	0.1845	0.2766	0.1821	0.2647

limitations and drawbacks of our proposed method, the results demonstrate its potential to reduce carbon emissions in the building design process. There are several potential future directions:

Firstly, the integration of additional data sources could provide a more comprehensive view of energy consumption patterns and inform building design. For example, weather data, demographic data, and traffic data could be incorporated to further improve the accuracy and impact of the model.

Secondly, the exploration of different attention mechanisms could help improve the accuracy and interpretability of the model. While the proposed method employs a cross-attention mechanism, other attention mechanisms such as self-attention or hierarchical attention could be explored.

Thirdly, the generalization of the proposed method to other cities and regions could provide insights and recommendations for building design and energy consumption in different contexts. However, this would require careful consideration of cultural, economic, and environmental factors that may differ across different regions.

Fourthly, the integration of building automation systems could provide real-time monitoring and control of energy consumption in buildings. This could help further reduce carbon emissions and improve energy efficiency by optimizing energy usage patterns in response to changing environmental and occupancy conditions.

Finally, the incorporation of sustainability metrics such as water usage, waste production, and social equity could provide a more comprehensive view of sustainability in the built environment. This would require careful consideration of how these metrics interact with energy consumption and building design.

The proposed method has the potential to contribute to mitigating climate change in the building sector and could have a significant impact on achieving global greenhouse gas reduction goals.

Our proposed method uses a Transformer model with a cross-attention mechanism to reduce carbon emissions during architectural design, which can contribute to climate change mitigation. This approach leverages machine learning techniques to analyze building design and carbon emissions datasets and uses cross-focus mechanisms to identify relationships between different variables in building design.

Predictive modelling can help architects predict energy consumption and carbon emissions and guide future designs and decisions. Although there are still some limitations, this method has important research and application implications and can contribute to the sustainable development of the construction industry.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

HDL and XY contributed to conception and formal analysis of the study. HLZ and HDL contributed to investigation, resources, and data curation. HDL wrote the original draft of the manuscript. HDL and XY performed the visualization and project administration. HDL provided the funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmad, M. W., Mourshed, M., and Rezgui, Y. (2017). Trees vs neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build.* 147, 77–89. doi: 10.1016/j.enbuild.2017.04.038

Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., et al. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Autom. Constr.* 141, 104440. doi: 10.1016 j.autcon.2022.104440

Baek, C., Park, S.-H., Suzuki, M., and Lee, S.-H. (2013). Life cycle carbon dioxide assessment tool for buildings in the schematic design phase. *Energy Build.* 61, 275–287. doi: 10.1016/j.enbuild.2013.01.025

Chen, C.-Y., Chai, K. K., and Lau, E. (2021). AI-assisted approach for building energy and carbon footprint modeling. *Energy AI* 5, 100091. doi: 10.1016/j.egyai.2021.100091

Chen, Y., Hong, T., Luo, X., and Hooper, B. (2019). Development of city buildings dataset for urban building energy modeling. *Energy Build.* 183, 252–265. doi: 10.1016/j.enbuild.2018.11.008

Chhachhiya, D., Sharma, A., and Gupta, M. (2019). Designing optimal architecture of recurrent neural network (LSTM) with particle swarm optimization technique specifically for educational dataset. *Int. J. Inf. Technol.* 11, 159–163. doi: 10.1007/s41870-017-0078-8

Copiaco, A., Himeur, Y., Amira, A., Mansoor, W., Fadli, F., Atalla, S., et al. (2023). An innovative deep anomaly detection of building energy consumption using energy time-series images. *Eng. Appl. Artif. Intell.* 119, 105775. doi: 10.1016/j.engappai.2022.105775

Fang, Y., Lu, X., and Li, H. (2021). A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage. *J. Clean. Prod.* 328, 129657. doi: 10.1016/j.jclepro.2021.129657

Häkkinen, T., Kuittinen, M., Ruuska, A., and Jung, N. (2015). Reducing embodied carbon during the design process of buildings. *J. Build. Eng.* 4, 1–13. doi: 10.1016/j.jobe.2015.06.005

Heris, M. P., Foks, N. L., Bagstad, K. J., Troy, A., and Ancona, Z. H. (2020). A rasterized building footprint dataset for the United States. *Sci. Data* 7, 207.

Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., et al. (2023). Ai-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. *Artif. Intell. Rev.* 56, 4929–5021. doi: 10.1007/s10462-022-10286-2

Hong, D., Gao, L., Yao, J., Zhang, B., Plaza, A., and Chanussot, J. (2020a). Graph convolutional networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 59, 5966–5978. doi: 10.1109/TGRS.2020.3015157

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., et al. (2020b). More diverse means better: Multimodal deep learning meets remote-sensing imagery classification. *IEEE Trans. Geosci. Remote Sens.* 59, 4340–4354. doi: 10.1109/TGRS.2020.3016820

Hu, Y., and Man, Y. (2023). Energy consumption and carbon emissions forecasting for industrial processes: Status, challenges and perspectives. *Renewable Sustain. Energy Rev.* 182, 113405. doi: 10.1016/j.rser.2023.113405

Huang, J., Alghatani, M., and Kaewunruen, S. (2022). Energy forecasting in a public building: A benchmarking analysis on long short-term memory (LSTM), support vector regression (SVR), and extreme gradient boosting (XGBoost) networks. *Appl. Sci.* 12, 9788. doi: 10.3390/app12199788

Jin, X., Zhang, C., Xiao, F., Li, A., and Miller, C. (2023). A review and reflection on open datasets of city-level building energy use and their applications. *Energy Build.*, 112911. doi: 10.1016/j.enbuild.2023.112911

Li, L. (2021). Integrating climate change impact in new building design process: A review of building life cycle carbon emission assessment methodologies. *Clean. Eng. Technol.* 5, 100286. doi: 10.1016/j.clet.2021.100286

Li, H., Wang, Z., and Hong, T. (2021). A synthetic building operation dataset. *Sci. Data* 8, 213.

Lim, B., Arık, S. Ö., Loeff, N., and Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* 37, 1748–1764. doi: 10.1016/j.ijforecast.2021.03.012

Ma, T., Wei, X., Liu, S., and Ren, Y. (2022). Mgcaf: a novel multigraph cross-attention fusion method for traffic speed prediction. *Int. J. Environ. Res. Public Health* 19, 14490. doi: 10.3390/ijerph192114490

Mercat, J., Gilles, T., El Zoghby, N., Sandou, G., Beauvois, D., and Gil, G. P. (2020). "Multi-head attention for multi-modal joint vehicle motion forecasting," in *2020 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE). 9638–9644.

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., and Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *J. Build. Eng.* 45, 103406. doi: 10.1016/j.jobe.2021.103406

Patil, S., Mudaliar, V. M., Kamat, P., and Gite, S. (2020). LSTM based ensemble network to enhance the learning of long-term dependencies in chatbot. *Int. J. Simul. Multidiscip. Des. Optim.* 11, 25. doi: 10.1051/smdo/2020019

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., et al. (2021). Carbon emissions and large neural network training. *arXiv preprint arXiv:2104.10350.* Available at: https://arxiv.org/abs/2104.10350

Pino-Mejías, R., Pérez-Fargallo, A., Rubio-Bellido, C., and Pulido-Arcas, J. A. (2017). Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and co2 emissions. *Energy* 118, 24–36. doi: 10.1016/j.energy.2016.12.022

Rahbar, M., Mahdavinejad, M., Markazi, A. H., and Bemanian, M. (2022). Architectural layout design through deep learning and agent-based modeling: A hybrid approach. *J. Build. Eng.* 47, 103822. doi: 10.1016/j.jobe.2021.103822

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vol. 10684–10695.

Roy, S. K., Deria, A., Hong, D., Rasti, B., Plaza, A., and Chanussot, J. (2023). Multimodal fusion transformer for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* doi: 10.1109/TGRS.2023.3286826

Sawhney, R., Joshi, H., Gandhi, S., and Shah, R. (2020). "A time-aware transformer based model for suicide ideation detection on social media," in *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*. 7685–7697.

Sayed, A. N., Himeur, Y., and Bensaali, F. (2023). From time-series to 2d images for building occupancy prediction using deep transfer learning. *Eng. Appl. Artif. Intell.* 119, 105786. doi: 10.1016/j.engappai.2022.105786

Sester, M., Feng, Y., and Thiemann, F. (2018). Building generalization using deep learning. *ISPRS-International Arch. Photogrammetry Remote Sens. Spatial Inf. Sci. XLII-4* 42, 565–572.

Sevim, N., Özyedek, E. O., Şahinuç, F., and Koç, A. (2022). Fast-FNet: Accelerating transformer encoder models via efficient Fourier layers. *arXiv preprint arXiv:2209.12816.*

Shen, J., Yin, X., and Zhou, Q. (2018). "Research on a calculation model and control measures for carbon emission of buildings," in *ICCREM 2018: Sustainable Construction and Prefabrication* (American Society of Civil Engineers Reston, VA: ICCREM 2018: Sustainable Construction and Prefabrication), 190–198.

Sheng, G., Sun, S., Liu, C., and Yang, Y. (2022). Food recognition via an efficient neural network with transformer grouping. *Int. J. Intelligent Syst.* 37, 11465–11481. doi: 10.1002/int.23050

Singaravel, S., Suykens, J., and Geyer, P. (2018). Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction. *Adv. Eng. Inf.* 38, 81–90. doi: 10.1016/j.aei.2018.06.004

Somu, N., MR, G. R., and Ramamritham, K. (2021). A deep learning framework for building energy consumption forecast. *Renewable Sustain. Energy Rev.* 137, 110591. doi: 10.1016/j.rser.2020.110591

Son, H., Kim, C., Kim, C., and Kang, Y. (2015). Prediction of government-owned building energy consumption based on an RreliefF and support vector machine model. *J. Civ. Eng. Manage.* 21, 748–760. doi: 10.3846/13923730.2014.893098

Song, J., Lee, J.-K., Choi, J., and Kim, I. (2020). Deep learning-based extraction of predicate-argument structure (PAS) in building design rule sentences. *J. Comput. Des. Eng.* 7, 563–576. doi: 10.1093/jcde/qwaa046

Ustinovichius, L., Popov, V., Cepurnaite, J., Vilutienė, T., Samofalov, M., and Miedziałowski, C. (2018). BIM-based process management model for building design and refurbishment. *Arch. Civ. Mech. Eng.* 18, 1136–1149. doi: 10.1016/j.acme.2018.02.004

Wang, H., Zhang, H., Hou, K., and Yao, G. (2021). Carbon emissions factor evaluation for assembled building during prefabricated component transportation phase. *Energy Explor. Exploit.* 39, 385–408. doi: 10.1177/0144598720973371

Wu, X., Hong, D., and Chanussot, J. (2021). Convolutional neural networks for multimodal remote sensing data classification. *IEEE Trans. Geosci. Remote Sens.* 60, 1–10. doi: 10.1109/TGRS.2020.3040277

Yang, B., Wu, J., Ikeda, K., Hattori, G., Sugano, M., Iwasawa, Y., et al. (2022). Face-mask-aware facial expression recognition based on face parsing and vision transformer. *Pattern Recognit. Lett.* 164, 173–182. doi: 10.1016/j.patrec.2022.11.004

Yu, Z., Haghighat, F., Fung, B. C., and Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy Build.* 42, 1637–1646. doi: 10.1016/j.enbuild.2010.04.006

Zhao, H., Jia, J., and Koltun, V. (2020). "Exploring self-attention for image recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10076–10085.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., et al. (2021). "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 11106–11115.