Research on Engineering Structures & Materials
www.jresm.org
MIM Research Group

Research Article

Integrating BIM and multi-source data with knowledge graph for predictive digital twin in building carbon management

Xiaochen Zhang [*,a], Shida Hou [b]

School of Civil Engineering and Architecture, Henan University of Science and Technology, Luoyang, 471000, China

Article Info

Article History:

Received 05 Aug 2026

Accepted 23 Sep 2026

Keywords:

Multi-source heterogeneous data;
Building carbon emission knowledge graph;
Building carbon emission prediction;
Deep learning prediction model;
Spatiotemporal laws of carbon emission

Abstract

Accurate carbon emission accounting and predictive analysis across the building lifecycle are critical for sustainable infrastructure management and smart city policymaking. However, conventional prediction models often suffer from low interpretability and insufficient integration with dynamic construction processes. This study proposes a digital twin-oriented framework that synergizes a domain-specific Knowledge Graph (KG) with a hybrid deep learning architecture, leveraging multi-source heterogeneous data including BIM, IoT sensor streams, and operational logs. A spatiotemporal KG encodes semantic relationships among building components, energy flows, and onsite activities. This KG is integrated with a Graph Convolutional Network–Temporal Convolutional Network (TCN-GCN) to capture spatial dependencies and temporal evolution patterns of carbon emissions, enabling real-time predictive analytics within a digital twin environment. The multi-source data input comprises BIM models, IoT sensor streams, and operational logs from a 12-month deployment in a Shanghai green building demonstration project, supplemented by the Global Infrastructure Emissions Database. The project is a public office building of approximately 20,000 m^2, equipped with a central HVAC system, a building automation system, and over 80 IoT submeters. The data span from January 2025 to December 2025, with sensor data sampled at 15-minute intervals and aligned to hourly timestamps. The framework enhances prediction accuracy and provides interpretable semantic support for emission traceability and scenario-based retrofitting decisions. Experimental results show that for entity and relation extraction in KG construction, the proposed method achieves 99.8% accuracy and 98.5% recall, with an F1-score of 0.95. The dataset was partitioned strictly by project to prevent data leakage.

© 2026 MIM Research Group. All rights reserved.

1. Introduction

Global climate change has made energy conservation and carbon reduction in the building sector a key link in achieving the “double carbon” goal. Energy consumption and carbon emissions throughout the building life cycle account for a large share [1]. However, current building carbon emission research and management face diverse sources and heterogeneous structures of data, covering design drawings, energy monitoring, building material lists, and equipment operation records, with a lack of semantic interconnection and holistic analysis capability [2]. Traditional prediction models mostly rely on single-dimensional historical energy or carbon data, making it difficult to reveal complex relationships among buildings, equipment, personnel behavior, and climate environment, and unable to achieve in-depth carbon emission traceability [3]. KG based on Multi-source Heterogeneous Data (MHD) can integrate static attributes, dynamic data, external environment, behavioral data, and implicit knowledge into a unified association network [4-5].

*Corresponding author: zhang.xiaoc@outlook.com
[a]orcid.org/0009-0005-6286-7763; [b]orcid.org/0009-0007-1812-5185
DOI: http://dx.doi.org/10.17515/resm2026-2117in0805rs

Therefore, this research proposes a building carbon emission KG construction and prediction method based on MHD, aiming to build a building carbon emission ontology model and KG for traceability and visual analysis of key carbon emission paths. The research integrates top-down and bottom-up methods to construct the KG, realizes knowledge fusion and storage through multi-source data processing, ontology modeling, and large language model knowledge extraction, and fuses TCN-GCN to integrate spatiotemporal features and graph semantic information for accurate carbon emission prediction and interpretable analysis. The main contributions are as follows:

- Methodological contribution: A systematic framework for constructing a building carbon emission KG from MHD, integrating data fusion, ontology modeling, and automated knowledge extraction using Large Language Models (LLMs), effectively breaking down data silos. This study focuses on the operation and maintenance stage, as the multi-source data input—BIM models, IoT sensor streams, and operational logs—are primarily generated during this phase. The material production, construction, demolition, and recycling stages are excluded due to the mismatch between their static, inventory-based data nature and the dynamic, real-time sensor data required by the predictive digital twin framework. For operational carbon accounting, emission factors for electricity consumption are derived from the most recent China regional grid baseline emission factors released by the Ministry of Ecology and Environment, with the applicable year set to 2024.
- Modeling contribution: A novel hybrid deep learning model, TCN-GCN, is developed for accurate carbon emission prediction, combining TCN to capture long-term temporal dependencies and GCN to mine complex correlations among influencing factors within the KG. KG embeddings further enrich the model with external semantic knowledge.
- System and application contribution: An integrated explainable prediction and intelligent question-answering (QA) system is built upon the constructed KG and TCN-GCN model, providing accurate forecasts and semantic support and interpretability through SHAP and graph-based reasoning. SHAP values and KG subgraphs jointly form a dual-layer explanation mechanism of feature contribution and knowledge association, providing traceable evidence rather than strict causal inference. The framework was validated through a 12-month deployment in a Shanghai green building demonstration project—a public office building with a gross floor area of approximately 20,000 m^2, served by a central HVAC system and a building automation system, with over 80 IoT submeters monitoring HVAC, lighting, and plug loads. This building is representative of modern public buildings in China's hot summer and cold winter climate zone. While the core methodological framework is generic and transferable, the specific optimization strategies and quantitative findings are context-dependent and require further validation when applied to residential, industrial, or buildings in different climate zones. The representativeness and boundaries of this case are further elaborated in the experimental setup.

2. Related Works

Global warming has become a major challenge, and carbon emissions are the main factor leading to climate warming. Peng et al. built a carbon emission prediction model based on genetic algorithm optimization for the provincial construction industry [6]. Fei et al. built a graph network method combining prior knowledge and automatic feature extraction for material quantity estimation of reinforced concrete buildings [7]. Guo and Luo designed a hybrid method of particle swarm optimization–back propagation network and fuzzy cuckoo search–support vector machine for performance optimization of traditional machine learning models [8]. Pan and Wu built an integrated Bayesian optimization XGBoost for carbon emission prediction using 121 residential data samples [9]. Wei et al. explored the accessibility of the carbon peak target in Henan Province by combining the Tapio decoupling model and ridge regression [10].

Many scholars have conducted in-depth research on MHD. Xiong et al. built a federated generation model framework for distributed heterogeneous data generation [11]. Li et al. designed a prediction method combining meta-path attention and graph neural networks for stock market fluctuation prediction [12]. Guo et al. proposed a cloud-pipe-edge-end integrated management framework for the power Internet of Things [13]. Liu et al. designed a tool status identification

method based on multi-source data fusion, using compressed sensing and Dropout network to suppress overfitting [14]. Wang et al. proposed an intelligent prediction method based on MHD for surface residual stress, combining improved convolutional networks and Gaussian process regression [15].

Existing research has good effects on building carbon emission KG construction and prediction, but still has limitations such as reliance on literature data, disconnection from actual projects, low spatial resolution, and lag in data update. Building a building carbon emission KG based on MHD from "isolated features" to "associated context" enables deeper mechanism modeling and can integrate multi-source information to break data silos. Therefore, this study proposes a building carbon emission KG construction and prediction method based on MHD.

3. Building Carbon Emission KG Construction and Prediction Method

3.1 Construction of Building Carbon Emission KG Based on MHD

The construction industry is one of the main sources of global carbon emissions. Traditional construction carbon emission knowledge management relies on relatively independent databases or indicators, making it difficult to integrate MHD for global analysis and easy to form "data islands" [16]. Building a building carbon emission KG based on MHD can integrate originally separated data into a semantically related, machine-readable unified knowledge network, shifting from isolated data points to context-aware carbon flow simulation and dynamic assessment [17]. The study combines top-down and bottom-up methods to construct the KG, as shown in Fig. 1.

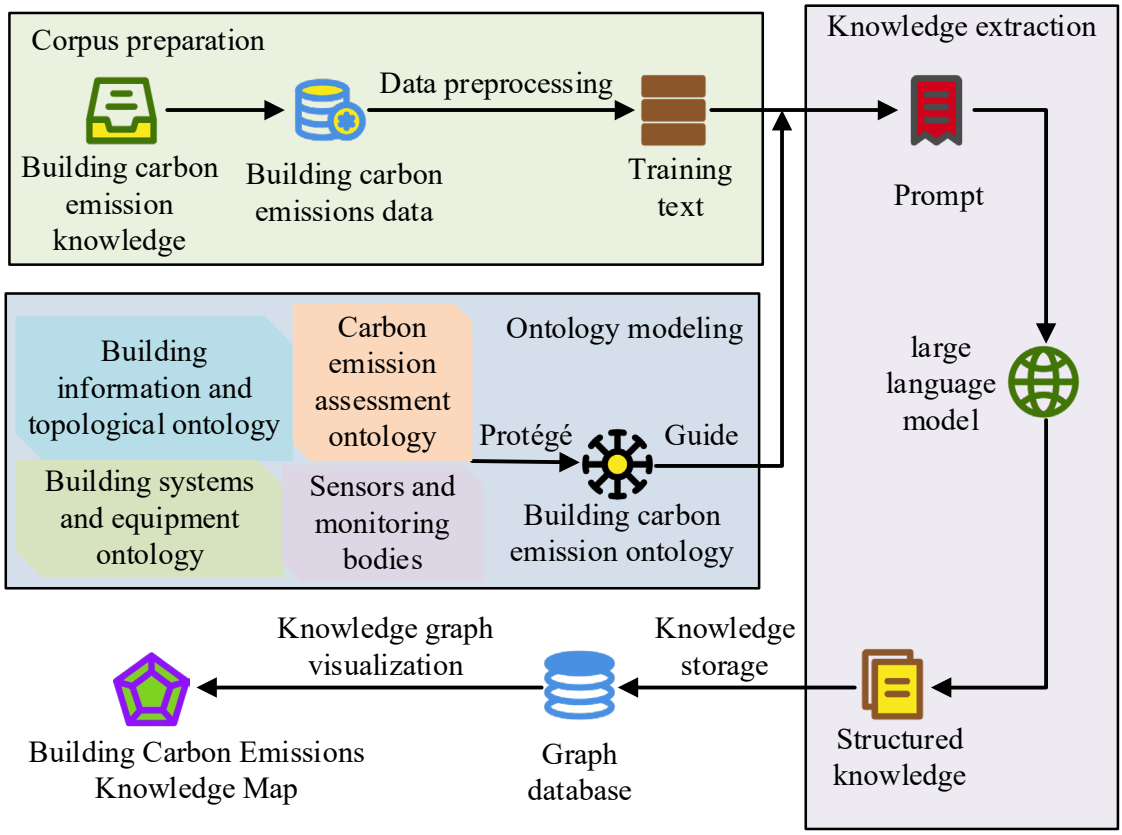


Fig. 1. Basic construction process of building carbon emission KG

From Fig. 1, for automated knowledge extraction, this study employs GPT-3.5-turbo with a few-shot prompting strategy to automatically extract entity and relation triples from unstructured text. Few-shot prompting is adopted instead of full-parameter fine-tuning due to its data efficiency and rapid adaptability to emerging carbon management terminologies, avoiding catastrophic forgetting. The prompt template is: "Extract all (head entity, relation, tail entity) triples relevant to building carbon emissions from the following text. Use relations from the predefined set: [contains, consumes, emits, adopts, locatedIn, affects, controls, hasParameter]. Text: {input_text}. Triples:".

This template is populated with three manually annotated example texts and their corresponding triple sets. The manual annotations, comprising 2,000 high-confidence triples verified by domain experts, serve as the gold-standard seed set. For error correction, a two-stage workflow is implemented: a rule-based validator filters extractions against predefined entity types and relation constraints; then, triples with a confidence score below 0.85 are flagged for manual review by domain experts weekly. To mitigate LLM hallucinations, a consistency-based filtering rule discards any extracted triple not supported by at least one other independent source or failing to align with ontology logical constraints. A cross-verification mechanism prompts the LLM to re-extract triples from the same text with slightly rephrased instructions; only triples consistently extracted across both iterations are retained. In data cleaning and preprocessing, the 3σ rule removes outliers in normally distributed data, while the Isolation Forest algorithm detects anomalies in high-dimensional heterogeneous data. During knowledge fusion, a weighted combination of embedding vector similarity, distribution similarity, and attribute similarity quantifies entity alignment. Multi-source heterogeneous data is aggregated from four main channels: building information models, IoT sensor networks, facility operation logs, and external databases. All stream data is resampled into uniform hourly intervals using linear interpolation, and the proportion of missing data in each channel is kept below 5%. The comprehensive similarity score is expressed in Equation (1).

$$\mathrm{Sim}(e_i, e_j) = \alpha \cdot \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{\| \mathbf{v}_i \| \| \mathbf{v}_j \|} + \beta \cdot \mathrm{JS}(D_i \parallel D_j) + \gamma \cdot \frac{| P_i \cap P_j |}{| P_i \cup P_j |} \tag{1}$$

In equation (1), $\mathrm{Sim}(e_i, e_j)$ is the comprehensive similarity score between entities e_i and e_j, which are generated using the ComplEx algorithm. ComplEx embeds entities into a complex vector space, enabling better modeling of symmetric and antisymmetric relationships commonly present in carbon emission knowledge graphs. $\mathbf{v}_i$ and $\mathbf{v}_j$ signify vector embeddings of e_i and e_j. $\mathrm{JS}(D_i \parallel D_j)$ is the Jensen-Shannon divergence between entity description texts D_i and D_j. P_i and P_j are sets of attributes for entities e_i and e_j,, respectively. α, β and γ are weight coefficients that satisfy $\alpha + \beta + \gamma = 1$. In the knowledge extraction stage, the study uses a few-shot prompting strategy to quickly adapt to domain-specific entity recognition tasks, as expressed in equation (2).

$$P(e \mid T, P_\theta) = \frac{\exp\left(f_{\mathrm{LLM}}(T, P_\theta)\right)}{\sum_{e' \in E_{\mathrm{candidate}}} \exp\left(f_{\mathrm{LLM}}(T, P_\theta)\right)} \tag{2}$$

In equation (2), $P(e \mid T, P_\theta)$ signifies the conditional probability of extracting entity e given unstructured text T and Few-shot prompt strategy P_θ. $f_{\mathrm{LLM}}(\cdot)$ is the feature mapping function of the large language model. $E_{\mathrm{candidate}}$ is the set of candidate entities. The core of building a carbon emission KG is ontology modeling, and the carbon emission assessment ontology is the core ontology connecting the assessment system, guidance system, and topology ontology. The building carbon emission field ontology and carbon emission assessment ontology are shown in Fig. 2.

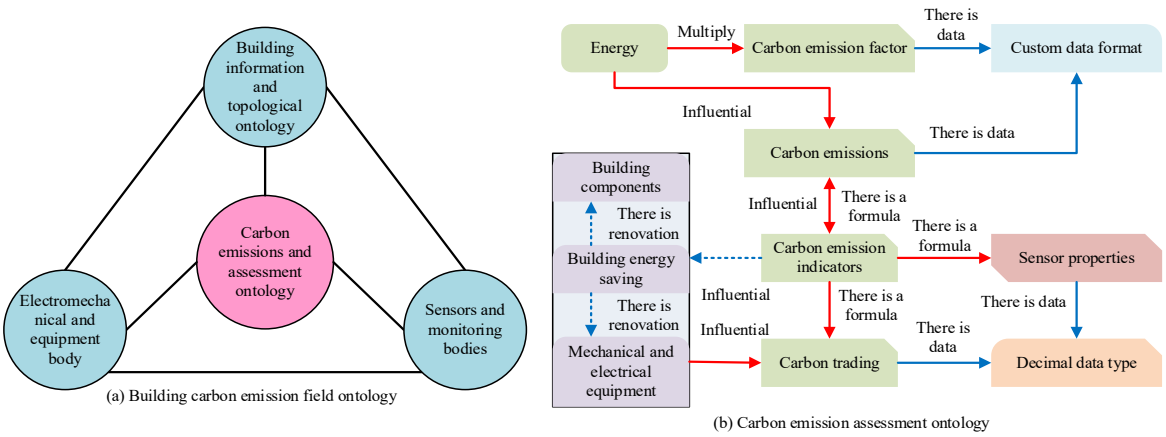


Fig. 2. Building carbon emission domain ontology and carbon emission assessment ontology

From Fig. 2(a), the core Carbon Emission and Assessment Ontology is linked to three specialized domain ontologies: the Building Space and Structure Ontology via the 'hasSpace' relation, the Mechanical/Electrical Facilities and Pipeline Ontology via the 'hasEquipment' relation, and the Sensor Monitoring and Data Ontology via the 'acquiresData' relation. From Fig. 2(b), the structured evaluation model uses 'there is data' to indicate that an activity requires data input; 'there is a formula' to denote that carbon emissions are determined through a formula; and 'influential' to mean that one factor exerts a promoting or inhibiting effect on another.

The ontology contains five main entity categories; BuildingSpace, EquipmentSystem, EnergyCarrier, ConstructionActivity, and CarbonMetric. Five core relationship types are defined as; locatedIn, consumes, emits, drivenBy, and hasMetric. All entities and relationships use CamelCase naming, with prefixes identifying the source domain. BuildingSpace should include area and functionType; EquipmentSystem should include ratedPower and efficiencyRating; EnergyCarrier must include emissionFactor and unit; ConstructionActivity should include duration and intensity; CarbonMetric needs to include value, unit, and timestamp. Representative examples of triples include <bim: HVAC_Unit_01, consumes, energy: Electricity>, <energy: Electricity, emits, carbon: CO2>, and <bim: Office_Zone_A, locatedIn, bim: Building_Structure>. For node-level constraint relationships, the system enforces: (1) Each BuildingSpace node must be linked to at least one EquipmentSystem node through locatedIn or has Equipment relationships; (2) Each EquipmentSystem node must be connected to exactly one EnergyCarrier node through a consumes relationship; (3) Each ConstructionActivity node must be associated with at least one EnergyCarrier node through a driven by or consumes relationship; (4) Each CarbonMetric node must be attached to exactly one source entity through a has Metric relationship. The system also imposes temporal validity constraints, requiring valid_Start and valid_TimeEnd properties. To capture complex relationships, this research combines entity triples to achieve end-to-end relationship extraction, as expressed in Equation (3).

$$P(h,r,t) = \sigma(W_r[h;t] + b_r) \cdot \prod_{i=1}^{n} P_{\text{mention}}(m_i \mid h,t) \tag{3}$$

In equation (3), $P(h,r,t)$ signifies the triplet probability composed of head entity h, relationship r, and tail entity t. h and t signify the vectors of the head entity and tail entity. W_r and b_r are the weight matrix and bias vector specific to the relationship r. $[h;t]$ is the vector splicing operation. $\sigma(\cdot)$ signifies the Sigmoid activation function. $\prod_{i=1}^{n} P_{\text{mention}}(m_i \mid h,t)$ signifies the probability of mentioning entity pair (h,t) at text position i. n is the total possible mention positions in the text. In the knowledge fusion stage, the research selects the most reliable knowledge statement, as expressed in Equation (4).

$$C_{\text{resolve}} = \underset{c}{\text{argmax}} \sum_{k=1}^{K} w_k \cdot \text{Sim}_{\text{claim}}(c, c_k) \cdot \text{Auth}(s_k) \cdot \text{TimeRecency}(t_k) \tag{4}$$

In equation (4), C_{resolve} signifies the knowledge statement selected after conflict resolution. c and c_k are candidate statements and the k-th candidate statement, respectively. w_k is the weight of the k-th information source. $\text{Sim}_{\text{claim}}(c, c_k)$ is the semantic similarity between statements c and c_k. $\text{Auth}(s_k)$ is the authoritative score of information source s_k. $\text{TimeRecency}(t_k)$ is the time freshness function. t_k is the time of knowledge generation. K is the total number of conflicting statements. To evaluate the rationality of links in the KG, the study introduces the ComplEx link prediction scoring equation, as expressed in Equation (5).

$$\phi(h,r,t) = \text{Re}(\langle h,r,\bar{t}\rangle) = \text{Re}\left(\sum_{i=1}^{d} h_i \cdot r_i \cdot \bar{t}_i\right) \tag{5}$$

In equation (5), $\phi(\boldsymbol{h},\boldsymbol{r},\boldsymbol{t})$ signifies the triplet score composed of head entity, relationship and tail entity. $\bar{\boldsymbol{t}}$ signifies the complex conjugate of tail entity embedding $\boldsymbol{t}$. $\langle\cdot,\cdot,\cdot\rangle$ signifies the trilinear dot

product operation. $\mathrm{Re}(\cdot)$ is the operation of taking the real part of complex numbers. d signifies the dimension of the embedding space. To realize relationship completion and entity disambiguation of the KG, the message passing mechanism of the graph convolution network is defined in Equation (6).

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{6}$$

In equation (6), $H^{(l)}$ signifies the characteristic matrix of all nodes in the l-th layer network. $\tilde{A}$ signifies an adjacency matrix with self-connection. $\widetilde{D}$ signifies the degree matrix. $W^{(l)}$ signifies the trainable weight matrix of the l-th layer. $\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}$ signifies a normalized symmetric adjacency matrix. Finally, the research evaluates the probabilities of all reasoning paths from the question to the answer to obtain the final answer, as expressed in Equation (7).

$$P(a \mid q, \mathrm{KG}) = \sum_{p \in \mathrm{Paths}(q,a)} \prod_{(e_i, r_i, e_{i+1}) \in p} P_{\mathrm{rel}}(r_i \mid e_i, e_{i+1}) \cdot P_{\mathrm{ent}}(e_{i+1} \mid e_i, r_i) \tag{7}$$

In equation (7), $P(a \mid q, \mathrm{KG})$ is the probability of answer a given question q and KG. $\mathrm{Paths}(q, a)$ is the set of all reasoning paths connecting question entities to answer entities. p is a single reasoning path. $P_{\mathrm{rel}}(r_i \mid e_i, e_{i+1})$ is the probability that there is a relationship r_i between entities e_i and e_{i+1}. $P_{\mathrm{ent}}(e_{i+1} \mid e_i, r_i)$ is the probability that the tail entity is e_{i+1} given the head entity e_i and relationship r_i. In summary, the building carbon emission KG construction system based on MHD is presented in Fig. 3.

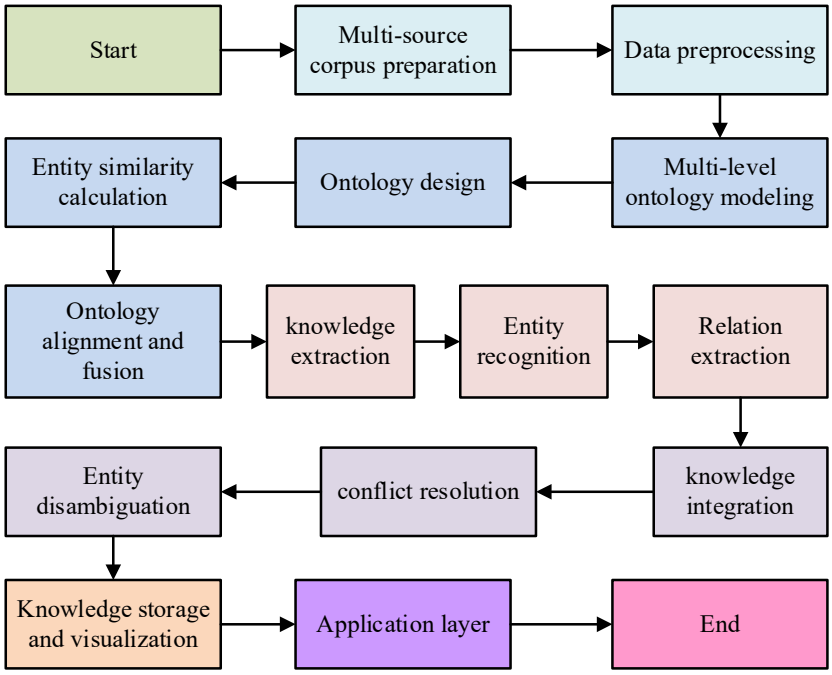


Fig. 3. Building carbon emission KG construction system based on MHD

From Fig. 3, the system employs GPT-3.5-turbo as the core extraction engine, automatically converting unstructured text from building operation manuals, equipment lists, and sensor logs into triples using few-shot prompting. Entity disambiguation and relation completion are achieved through ComplEx-based link prediction and GCN-based message passing. Incremental updates are driven by a dual mechanism of event triggers and timed scans. For instance, when SHAP identifies 'HVAC system energy consumption' as a key predictor, the KG can extract a subgraph containing this entity and its associated paths, revealing an interpretable semantic chain: 'temperature rise → increased HVAC load → elevated electricity consumption → carbon emission growth.' This association pathway provides a structural basis for understanding the prediction, but should be

interpreted as a correlational chain rather than a confirmed causal mechanism. The proposed framework establishes a closed-loop synchronization mechanism essential for a predictive digital twin. The physical building, its BIM model, and streaming IoT data are continuously aligned through an automated state-mapping pipeline. The outputs from the TCN-GCN predictor are visualized on a dashboard, where actionable insights are generated. Once a decision is implemented, the resulting operational changes are captured by the IoT network and fed back into the KG, thereby updating the digital twin and initiating the next prediction cycle.

To formally characterize the structural stability of the KG across iterative updates, the Consistency Score and Consistency Index are defined. Let $\mathcal{T}_v$ denote the set of core triples in version v, where the core triple set is strictly defined as all triples whose head entity belongs to the four primary entity categories and whose relation belongs to the five predefined relation types. The Consistency Score between consecutive versions (v-1) and v is given by: $\mathrm{CS(v)}= \frac{|T_v \cap T_{v-1}|}{|T_{v-1}|} \times 100\%$, where $|\cdot|$denotes set cardinality. The Consistency Index at version v is defined as the exponential moving average of historical Consistency Scores: $\mathrm{CI}(v) = \alpha \cdot \mathrm{S}(v) + (1-\alpha) \cdot \mathrm{CI}(v-1), \mathrm{CI}(0) = \mathrm{CS}(1)$, with smoothing factor $\alpha = 0.7$ assigning greater weight to recent versions to emphasize sustained stability. The minimum acceptable threshold for the Consistency Score is set at 80%, and for the Consistency Index at 0.80, below which the system triggers an alert and rolls back to the last stable version.

Version updates are governed by a dual-mechanism trigger strategy. An event-driven trigger initiates a new KG version when SHAP analysis identifies a feature whose importance score changes by more than 15% relative to the previous version, or when the prediction error RMSE exceeds 5.0 kgCO2e for three consecutive hours. A timed scan trigger produces a new version every 24 hours. From version V1.0 to V8.0, the KG evolved through four distinct phases. V1.0 represents the initial KG constructed from static BIM data and historical operational logs. V2.0 incorporates the first batch of streaming IoT sensor data after one week of deployment. V3.0 through V5.0 reflect incremental integration of equipment maintenance logs and occupant behavior patterns over the subsequent two months. V6.0 incorporates the first round of SHAP-driven event triggers, where the HVAC energy consumption feature was identified as a high-importance predictor. V7.0 and V8.0 represent continuous daily timed updates that further refine entity alignments and relation weights. This evolution process, with the Consistency Score rising from 84.3% at V1.0 to 97.4% at V8.0 and the Consistency Index increasing from 0.82 to 0.98, confirms the progressive stabilization of the KG structure.

3.2 Building Carbon Emission Explainable Prediction and Question-Answering System Based on KG

Developing an explainable prediction and question-answering system for building carbon emissions can achieve precise control and forward-looking decision-making, and improve the explainability and credibility of building carbon management [18]. The building carbon emission KG based on MHD can deeply integrate MHD, dynamic correlation, and feature engineering, while the KG-based explainable prediction and QA system can update the KG in real time [19-20]. The study combines GCN and TCN to build a building carbon emission prediction model, as presented in Fig. 4.

From Fig. 4, the TCN-GCN hybrid model accurately predicts building carbon emissions through a parallel dual-channel structure. The TCN branch uses gated dilation convolution to capture long-term laws of spatiotemporal evolution, and the GCN branch mines complex correlations between different influencing factors. The outputs of the two branches are fused via a fusion layer, where feature maps from each branch are concatenated along the feature dimension, followed by a learnable weighting mechanism. Entity embeddings extracted from the KG serve as external semantic priors and are integrated with spatiotemporal features via an attention mechanism, allowing the model to adjust predictions based on logical constraints, such as equipment efficiency labels or zone maintenance status, thus reducing data-driven biases. The fused features are subsequently passed through a fully connected layer to generate the final carbon emission prediction. This predicted value, along with SHAP-based feature importance analysis and the

interpretable KG subgraph, is formatted as a structured data stream for the visualization and decision support module of the digital twin platform. The platform renders these results on a user interface, enabling managers to identify high-emission zones and primary drivers. The system translates analytical results into recommended feedback actions, such as setpoints for HVAC systems or schedules for lighting and equipment. Once approved and executed on the physical asset, the resulting changes in operational data are captured by the next cycle of IoT data streams, closing the feedback loop. The graph structure data adjacency matrix is generated as shown in Equation (8).

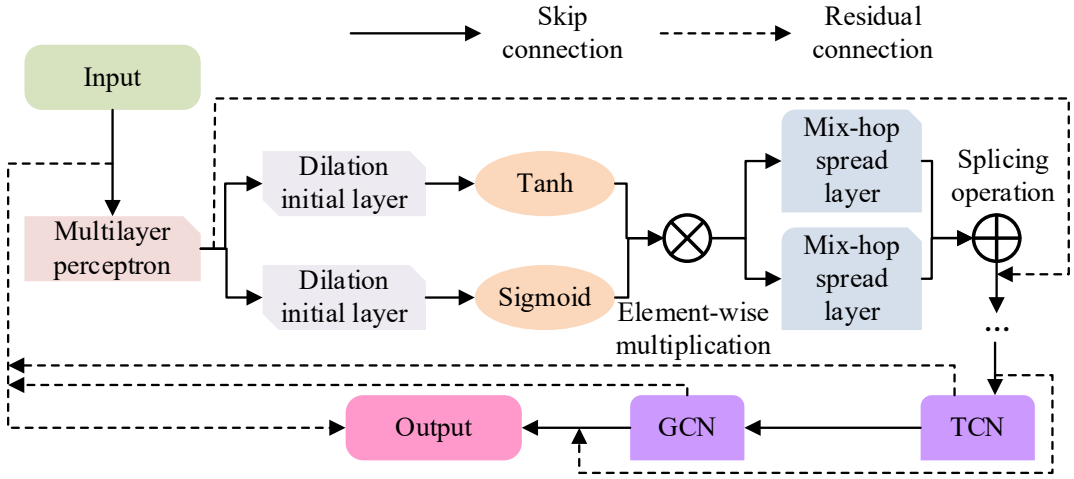


Fig. 4. Building carbon emission prediction model based on TCN-GCN algorithm

$$A = \sigma(ZZ^{\top}) + \lambda L_{\text{sparse}} \tag{8}$$

In equation (8), A signifies the adjacency matrix. Z signifies the node characteristic matrix. λ signifies the sparse regularization coefficient. L_{sparse} is the sparse loss term. To improve the accuracy of long-term prediction, TCN and GCN are fused, and KG embedding is introduced as external semantic information, as expressed in Equation (9).

$$\hat{Y}_{t+1:t+T} = \text{TCN}\big(\text{GCN}(X_{t-P:t}, A)\big) \oplus \text{KGEmb}(\text{KG}) \tag{9}$$

In equation (9), $\hat{Y}_{t+1:t+T}$ signifies the predicted value of carbon emissions in the future time step. $X_{t-P:t}$ is the input feature of the historical time step. $\text{GCN}(\cdot)$ signifies the graph convolution operation. $\text{TCN}(\cdot)$ signifies the temporal convolution operation. $\oplus$ is the feature splicing operation. $\text{KGEmb}(\text{KG})$ is the embedding of the KG. The research provides a transparent explanation of model decision-making based on SHAP theory, as expressed in Equation (10).

$$\phi_i(f,x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}[f(S \cup \{i\}) - f(S)] \tag{10}$$

In equation (10), ϕ_i signifies the SHAP of feature i. f is the trained prediction model. x signifies the input sample feature vector. N signifies the set of all features. S signifies a feature subset. $f(S)$ is the model output when only using subset S features. SHAP values quantify the contribution of each feature to the prediction, while the KG subgraph reflects the semantic relationships and influence paths among features. Combining the two enables a dual-layer explanation mechanism of "feature contribution – knowledge association." For instance, when predicting a sudden increase in a building's carbon emissions, SHAP analysis may identify "HVAC system energy consumption" and "outdoor temperature" as key contributing features. Simultaneously, extracting a subgraph from the KG containing these two entities and their relationships can further reveal the interpretive semantic pathway: temperature rise is statistically associated with increased HVAC load, which correlates with elevated electricity consumption and subsequent carbon emission growth. This example demonstrates an interpretive association derived from the model and knowledge graph,

not a validated causal chain. Next, the study identifies the key subgraph structure that affects a single prediction result, as expressed in Equation (11).

$$\max_{G_s} MI(Y, G_s) = \mu H(G_s) \tag{11}$$

In equation (11), G_s is the explanatory subgraph. $MI(Y, G_s)$ is the mutual information between the prediction result Y and the subgraph G_s. $H(G_s)$ is the entropy of the subgraph. μ is the balance coefficient. Subsequently, the study outputs a comprehensive carbon emission report from the question-and-answer system, as expressed in Equation (12).

$$C_{\text{fused}} = \alpha \cdot C_{\text{pred}} + \beta \cdot C_{\text{KG}} + \gamma \cdot C_{\text{shap}} \tag{12}$$

In equation (12), C_{fused} is the carbon emission assessment value after integration. C_{pred} is the predicted value of the spatiotemporal graph neural network. C_{KG} is the KG reasoning result. C_{shap} is the modified value after SHAP explanation weighting. Finally, the research provides a core matching mechanism for intelligent question answering, as expressed in Equation (13).

$$\text{Score}(q, a) = \text{BERT}(q)^{\top} R \cdot \text{KGEmb}(a) + \eta \cdot \text{Sim}_{\text{path}}(p_q, p_a) \tag{13}$$

In equation (13), $\text{Score}(q, a)$ is the comprehensive matching score between user question q and candidate answer a. $\text{BERT}(q)$ signifies the semantic vector of question q. R signifies the trainable alignment matrix. $\text{KGEmb}(a)$ is the embedding of answer a in the KG. Sim_{path} is the path similarity between the question and the answer in the KG. η is the path weight coefficient. To comprehensively evaluate the Q&A subsystem, a dedicated test protocol is designed. A set of user queries collected from the field deployment is manually annotated by domain experts, covering factual, causal, predictive, comparative, and action-oriented question types. Answer correctness is determined by expert majority voting, with exact matching for factual answers and a semantic similarity threshold for reasoning answers. The system supports ambiguous questions via entity disambiguation and context-aware embedding, and handles multi-hop follow-up queries through iterative subgraph traversal and path ranking. Performance metrics include response latency, correct answer rate per type, and failure case logging. In summary, the specific framework of the intelligent question and answer system designed in the research is shown in Fig. 5.

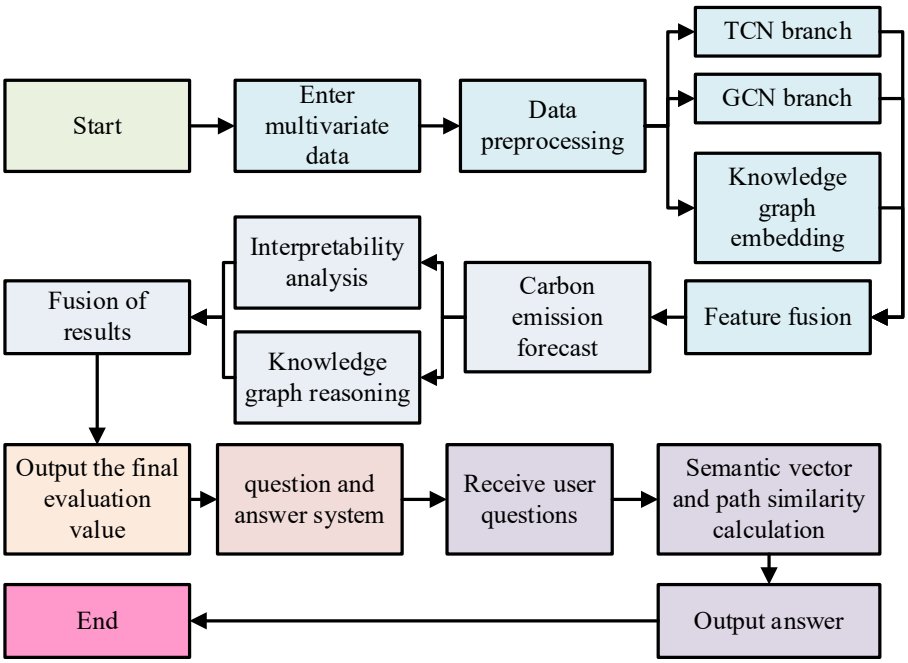


Fig. 5. Interpretable prediction and prediction question and answer system of building carbon emissions based on KG

As shown in Fig. 5, the system integrates prediction and question answering. The main prediction line extracts spatiotemporal features through the TCN-GCN dual-branch and integrates with the KG. Domain knowledge is used to optimize the data-driven model and then provide interpretability through SHAP. After the main prediction pathway outputs future carbon emission values, SHAP quantifies the marginal contribution of each feature, while the Q&A branch retrieves subgraphs from the KG related to high-contribution features, providing a dual-layer explanation of "numerical evidence + structured knowledge path." For example, when answering "Why will carbon emissions rise next month?", the system simultaneously presents the time-series forecast trend, SHAP values of key features, and the influence chain derived from the KG. Subsequently, the consistency between the visual output and the real distribution through the KL divergence constraint is obtained, as expressed in Equation (14).

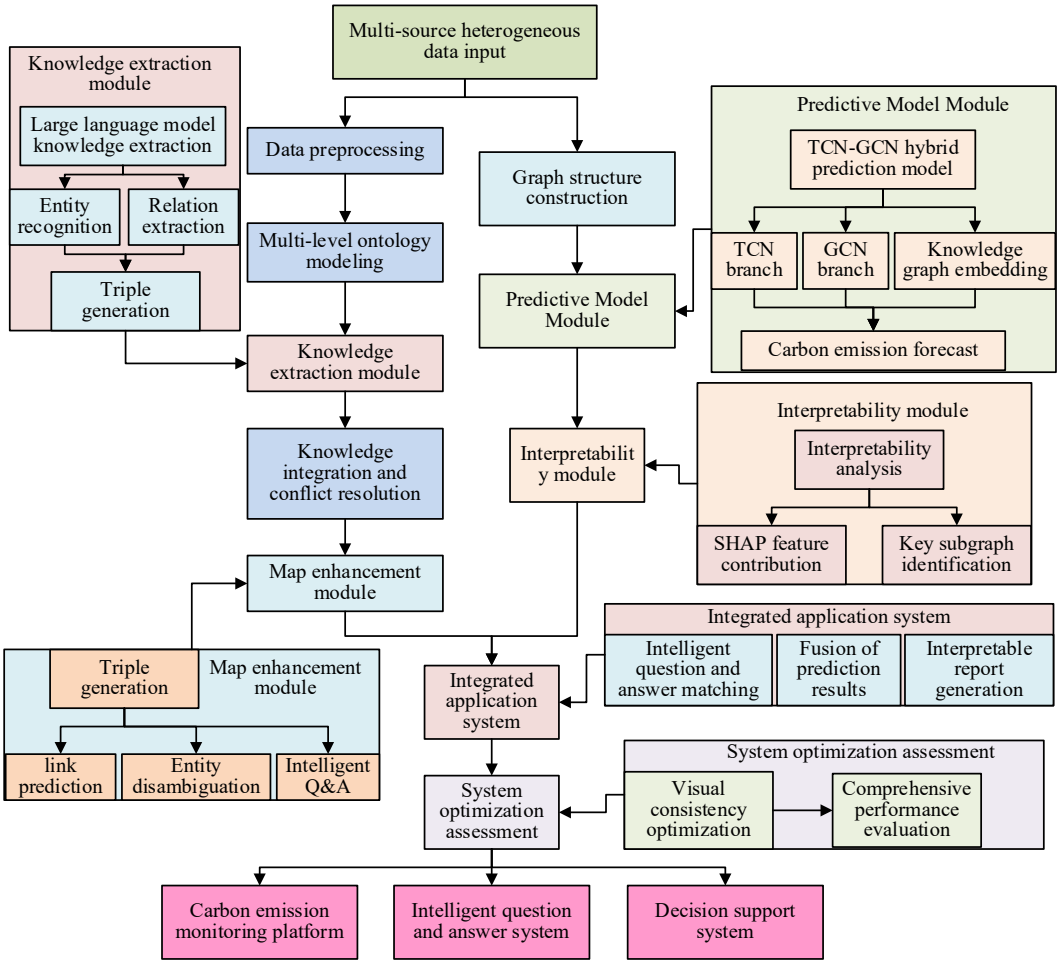


Fig. 6. Building carbon emission KG construction and prediction model based on MHD

$$L_{\text{sys}} = \mathrm{E}_{(D,M,O)}[\| O - Y \|_2^2 + \varphi \cdot \mathrm{KL}(p_{\text{UI}} \| p_{\text{true}})] \tag{14}$$

In Equation (14), L_{sys} is the total optimization objective function at the system level. E is the mathematical expectation. D, M and O are data layer input, model layer output and UI layer rendering result, respectively. Y is the real carbon emission value. $\mathrm{KL}(\cdot)$ is KL divergence. p_{UI} is the probability distribution of the UI layer visual data. p_{true} is the real data distribution. φ is the trade-off parameter. Finally, the research provides a multi-dimensional optimization goal for the question answering system, as expressed in Equation (15).

$$\text{Performance} = \frac{1}{Z}\sum_{i=1}^{Z}\left[\omega_1 \cdot \text{RMSE}_i + \omega_2 \cdot \text{ExplainScore}_i + \omega_3 \cdot \text{ResponseTime}_i^{-1}\right] \tag{15}$$

In equation (15), "Performence" is the comprehensive performance score of the system. RMSE_i

signifies the Root Mean Square Error (RMSE) of the i-th test sample. $\mathrm{ExplainScore}_i$ is the interpretability score. $\mathrm{ResponseTime}_i$ is the response time of the question-and-answer system. ω_1, ω_2 and ω_3 are weight coefficients, satisfying $\omega_1 + \omega_2 + \omega_3 = 1$. Z signifies the total test samples. Finally, the building carbon emission KG construction and prediction model based on MHD is presented in Fig. 6.

From Fig. 6, the framework first cleans and standardizes raw multi-source data, then feeds them into GPT-3.5-turbo for knowledge extraction, constructing an ontology-aligned KG. The graph-structured data are input to the TCN-GCN predictor, with pre-trained KG embeddings loaded as contextual priors. The prediction results, combined with SHAP explanations and KG reasoning paths, produce a traceable carbon emission report that supports equipment retrofitting, scheduling optimization, and other carbon management decisions. For instance, if the TCN branch predicts a spike in energy consumption for an HVAC unit based on historical data, but the KG indicates that the unit is currently linked to a zone undergoing scheduled maintenance, the model can attenuate the predicted consumption, correcting potential temporal biases introduced by missing contextual information.

3.3 Implementation Details of the TCN-GCN Model

To ensure reproducibility, this subsection specifies key implementation configurations. The adjacency matrix for graph convolution is constructed by computing cosine similarity between node feature vectors and retaining the top-K nearest neighbors per node (K=10), followed by summarization and self-loop addition. The TCN branch comprises three dilated convolutional layers with dilation factors [1, 2, 4], each having 64 filters and a kernel size of 3, using gated linear units for activation. The GCN branch consists of two graph convolutional layers, each with 64 hidden units and ReLU activation, applied on the same adjacency matrix. The outputs from both branches are fused via a learnable weighted summation, where the weights are automatically optimized during training along with the model parameters, instead of being fixed hyperparameters. The knowledge graph embeddings are generated with a dimension of 64 using the ComplEx algorithm, and they are integrated into the fusion layer through an attention mechanism that computes the relevance between spatiotemporal features and each entity embedding. The entire model is trained using the Adam optimizer with an initial learning rate of 0.001, a batch size of 64, and a random seed of 42 for all stochastic operations. Early stopping is applied when the validation loss does not decrease for 10 consecutive epochs, and the model checkpoint with the lowest validation error is retained for testing. All hyperparameters were selected via grid search on the validation set, and the reported results correspond to the best-performing configuration.

3.4 Comparative Model Configurations and Fair Comparison Principles

To ensure a rigorous and fair evaluation, all baseline models were implemented and tested under strictly controlled conditions mirroring the proposed TCN-GCN framework. Specifically, all models, including the Machine Learning-based Ensemble Prediction Model (MLEPM), the Top-down Input-Output Analysis based Macroeconomic Linkage Prediction (IO-MLP), and four state-of-the-art spatiotemporal models (STGCN, MTGNN, DyREP, and KnowEvolve), were configured with identical settings: an 80% to 20% training-testing data split, a maximum of 100 training epochs with an early stopping patience of 10 epochs, the Adam optimizer with a learning rate of 0.001, and Mean Squared Error as the loss function. Hyperparameter tuning for each model was performed using a unified grid search strategy on the validation set.

A critical aspect of fair comparison lies in the handling of KG semantic information. The proposed TCN-GCN model uniquely benefits from the integration of pre-trained KG entity embeddings as auxiliary features. To ensure a fair baseline, for models that cannot natively ingest graph-structured data, we adopted an input expansion strategy. The input feature vector for these models was augmented by concatenating the original statistical features with the 64-dimensional KG entity embeddings. This ensures that all models are provided with an equivalent volume of information, thereby isolating the performance gain attributable to the model's architectural superiority rather

than to a mere disparity in input data volume. The core configurations for each baseline model are detailed in Table 1.

Table 1. Core configurations of the baseline models

Model	Input Features	Key Parameters	KG Information Utilization
MLEPM	Original features + KG embeddings	Number of base learners: 5; Estimator: XGBoos	As additional input features
IO-MLP	Original features + KG embeddings	No. of hidden layers: 2; No. of units: [64, 32]	As additional input features
STGCN	Original features + KG embeddings	No. of Chebyshev filters: 3; Temporal kernel: 3; Activation: GLU	As additional input features
MTGNN	Original features + KG embeddings	No. of graph layers: 2; No. of temporal conv: 2	As additional input features
DyREP	Original features + KG embeddings	Embedding dimension: 64; Window size: 10	As additional input features
KnowEvolve	Original features + KG embeddings	Embedding dimension: 64; No. of time slices: 5	As additional input features
TCN-GCN (Proposed)	Original features + KG embeddings	TCN: 3 layers, dilation factors [1, 2, 4]; GCN: 2 layers, 64 units	Integrating via an attention mechanism into the fusion layer

4. Effectiveness Analysis of Building Carbon Emission KG Construction and Prediction Method

4.1 Performance Test of Building Carbon Emission KG Construction and Prediction Method

To verify the performance of the proposed method, a simulation platform was constructed. Table 2 presents the experimental configuration.

Table 2. Test environment and specific configuration

Testing environment	Specific configuration
Computing resources	Intel Xeon
Operating system	Ubuntu 22.04 LTS
Memory	More than 128GB
Storage	High-speed NVMe SSD
Deep learning framework	TensorFlow
Development environment	Python 3.8+ environment

As shown in Table 2, the study used the experimental environment and configuration for performance testing and used the global infrastructure emissions database as the experimental dataset. The ground-truth labels for entity and relation extraction tasks were derived from the manually annotated seed set of 2,000 high-confidence triples verified by domain experts. For the carbon emission prediction task, the labels were the measured on-site carbon emission values, calculated from energy consumption data using standard emission factors provided by the Global Infrastructure Emissions Database. The training dataset comprises operational data from the Shanghai demonstration project combined with the Global Infrastructure Emissions Database, spanning January 2025 to December 2025. Sensor data are recorded at 15-minute intervals, while BIM and operational log data are updated daily and, on an event-driven basis, respectively. All data streams are temporally aligned to hourly resolution using resampling and interpolation. The dataset covers key field types including energy consumption, environmental parameters, equipment status, and building physical attributes, with missing value rates maintained below 5% through a rule-based quality control pipeline. To rigorously prevent data leakage and evaluate generalization, the dataset was partitioned chronologically and by project. Specifically, the training set consisted of data from the first eight months (January to August 2025) of the Shanghai demonstration project. The test set, strictly held out from training, comprised data from the

subsequent four months (September to December 2025) for the same project, supplemented by an independent set of 500 samples from external projects entirely unseen during training. This strict temporal and project-level isolation ensures that no future information leaks into the model and that reported metrics genuinely reflect its ability to generalize to new, unseen data. Regarding the representativeness of the Shanghai demonstration project, it is a public office building with a gross floor area of approximately 20,000 m^2, located in Shanghai's hot summer and cold winter climate zone. The building is equipped with a central HVAC system consisting of two water-cooled chillers and three air-handling units, a building automation system for monitoring and control, and a dedicated IoT sensor network comprising over 80 submeters that track electricity consumption of HVAC, lighting, elevator, and plug-load systems at 15-minute intervals. The average annual operational energy use intensity of this building is approximately 85 kWh/m^2, which is typical for large-scale public office buildings in eastern China. The building operates under standard weekday occupancy schedules from 8:00 to 18:00, with weekend and holiday reduced-operation modes. The 12-month deployment covered a full seasonal cycle, capturing both peak cooling demand in summer and heating demand in winter, providing a comprehensive basis for model training and validation. While these characteristics make the project representative of a significant class of urban public buildings, the specific quantitative findings and optimization strategies—particularly those related to HVAC scheduling and equipment retrofit prioritization—are derived from this specific building system and climate context. The core methodological framework, including the KG construction pipeline, the TCN-GCN prediction model, and the interpretable QA system, is generic and transferable; however, their application to other building types such as residential complexes, industrial facilities, or buildings in different climate zones would require appropriate re-parameterization and validation using local data.

To ensure clear interpretation, evaluation metrics are categorized into four types: prediction error metrics (RMSE and MAE, in kgCO2e), knowledge graph construction metrics (precision, recall, F1-score), model performance metrics (recognition accuracy and extraction accuracy, in percentages), and system consistency metrics (Consistency Score and Consistency Index). The research method is compared with MLEPM and IO-MLP. MLEPM represents machine learning-based ensemble prediction, one of the mainstream data-driven methods in carbon prediction. IO-MLP represents Macro Correlation Forecasting based on Input-Output Analysis. These two methods cover the two mainstream prediction paradigms of "micro data-driven" and "macro system modeling." The F1 scores, recall rates, and precision rates are compared in Fig. 7.

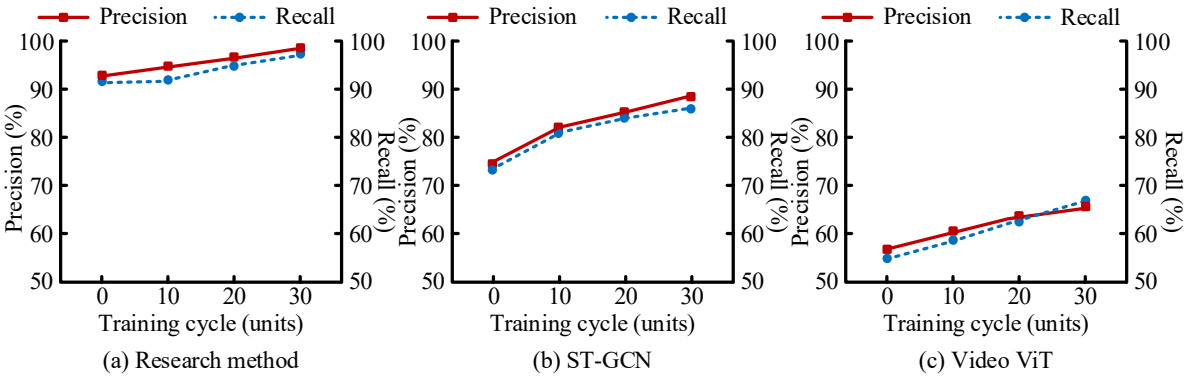


Fig. 7. Precision and recall of different methods for entity and relation extraction

From Fig. 7, precision and recall for entity and relation extraction tasks of the three methods showed an overall upward trend as training cycles increased. When the number of training cycles reached 30, precision and recall of the research method were 99.8% and 98.5%, respectively. The MLEPM method achieved 89.3% and 87.6%, respectively. The IO-MLP method was lower than 70%. The research method has higher precision, robustness, and generalization ability in the knowledge acquisition stage. Five-fold cross-validation was conducted to assess model stability. Furthermore, 500 independent samples from external projects were used to test generalization performance, with the proportion of samples achieving a matching score >85% serving as a generalization metric. The comparison results are shown in Table 3.

Table 3. Comparison of 5-fold cross-validation and generalization ability for KG construction

Metric	Research method	MLEPM	IO-MLP
5-fold CV average F1-score	0.948±0.012	0.872±0.025	0.761±0.041
5-fold CV average accuracy	99.6%±0.3%	88.7%±1.2%	75.2%±2.5%
Proportion of samples with match >85%	92.3%	78.5%	65.2%
Train:Test split ratio	80%:20%	80%:20%	80%:20%

From Table 3, the proposed method exhibits higher stability with lower variance in cross-validation and stronger generalization capability on unseen project data. The changing trends of RMSE and MAE for the carbon emission prediction task are compared in Fig. 8.

From Fig. 8, the RMSE and MAE values for carbon emission prediction of the three methods all showed a downward trend. When the iteration increased to 30, the RMSE value of the research method decreased to 3.1 kgCO2e. The RMSE values of the MLEPM method and the IO-MLP method were 14.8 kgCO2e and 25.2 kgCO2e respectively. When the iteration rounds increased to 60, the MAE value of the research method converged, with a stable value of 9.8. The MAE values of the other two methods had a smaller downward trend. The research method has better efficiency, convergence, stability, and prediction accuracy. To comprehensively evaluate the advancement of the proposed method, this study further selects four state-of-the-art models as comparative baselines: STGCN, MTGNN, DyREP, and KnowEvolve. These models were implemented and evaluated under the same fair comparison principle, ensuring that all models received equal information, including KG embeddings as additional input features. When the training period is 30, the accuracy of STGCN, MTGNN, DyREP, and KnowEvolve are 92.0%, 93.5%, 95.8%, and 96.5%, respectively, with recall rates of 90.0%, 91.2%, 94.3%, and 95.1%, all lower than the research method (99.8%/98.5%). The RMSE of STGCN, MTGNN, DyREP, and KnowEvolve were 8.2 kgCO2e, 7.5 kgCO2e, 5.8 kgCO2e, and 5.2 kgCO2e, respectively, and the MAE were 15.5 kgCO2e, 14.2 kgCO2e, 12.6 kgCO2e, and 11.9 kgCO2e, respectively, all higher than the method used in this study (3.1 kgCO2e/9.8 kgCO2e). This consistent performance gap underscores the superiority of the proposed architecture in effectively leveraging KG information for accurate prediction.

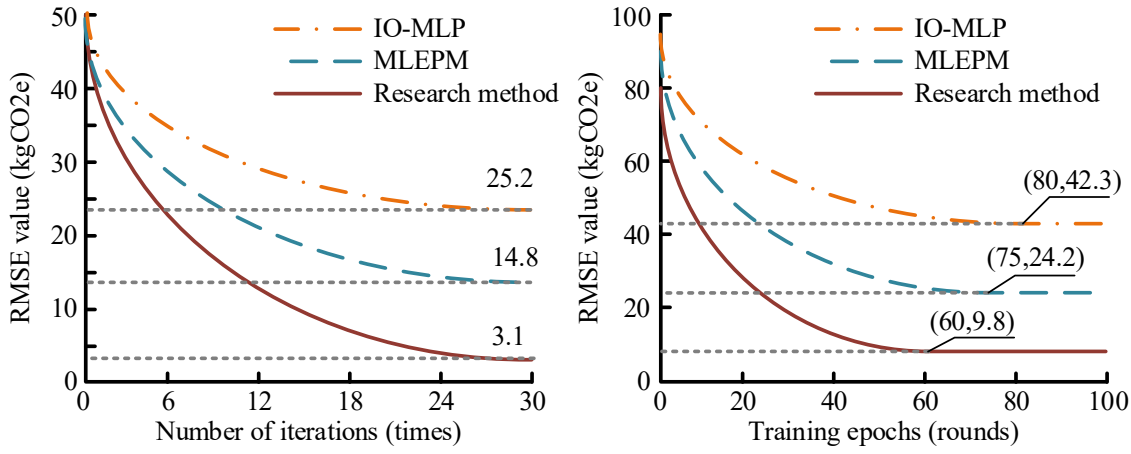


Fig. 8. RMSE and MAE trends for carbon emission prediction

To isolate the individual contributions of each core component, an ablation study was conducted. Five model variants were constructed by sequentially removing the KG embedding integration, the GCN branch, the TCN branch, the SHAP explanation module, and the QA module. For each variant, prediction error, F1-score, entity-relation extraction accuracy, and feature importance score were measured on the same test set. The results are summarized in Table 4.

Table 4. Ablation study results

Variant	RMSE (kgCO2e)	MAE	F1-score	Entity-relation extraction acc.	Feature importance score
Full model	3.1	9.8	0.948	92.3%	0.87
w/o KG embedding	5.8	12.4	0.912	83.6%	0.72
w/o GCN branch	7.6	15.2	0.888	88.1%	0.68
w/o TCN branch	8.4	16.3	0.875	86.7%	0.65
w/o SHAP module	3.3	10.2	0.943	91.8%	N/A
w/o QA module	3.1	9.8	0.946	79.4%	0.86

From Table 4, Removing the KG embedding caused the most pronounced degradation in entity-relation extraction accuracy, dropping from 92.3% to 83.6%, while RMSE increased by 87.1%, confirming that the semantic constraints provided by the KG are essential for maintaining structural coherence and improving prediction reliability. The GCN and TCN branches each contributed significantly to spatial and temporal feature extraction respectively; their removal elevated RMSE to 7.6 and 8.4 kgCO2e, demonstrating that the dual-channel architecture is critical for capturing the full spatiotemporal dynamics of carbon emissions. The SHAP module had negligible impact on prediction accuracy but was indispensable for generating feature-level interpretability, as the importance score became unavailable without it. The QA module did not affect prediction error, yet its removal reduced entity-relation extraction accuracy to 79.4%, indicating its role in refining knowledge retrieval and semantic alignment. These ablation results collectively verify that the predictive accuracy, interpretability, and knowledge reasoning capabilities of the proposed framework arise from the synergistic integration of all modules, rather than from any single component.

4.2 Practical Application Effect of Building Carbon Emission KG Construction and Prediction Method

After verifying the performance of the proposed method, the study further verified its practical application value. A prototype system named “Building Carbon Emission Knowledge Graph and Intelligent Prediction” was developed and deployed in the actual operating environment of a green building demonstration project in Shanghai. The system accesses multi-source data including building energy consumption monitoring, equipment operation logs, and indoor/outdoor environmental sensors through IoT gateways in real time, continuously constructs and updates the knowledge graph, and drives the TCN-GCN prediction model for online inference. The system adopts a cloud-edge collaborative architecture where data preprocessing, entity extraction, and local caching are performed at the edge side, while model inference, KG reasoning, and dashboard visualization are deployed on the cloud side. Data ingestion is implemented via MQTT protocol with JSON payloads, and a rule-based quality control pipeline performs schema validation, missing value interpolation, and outlier detection before data enters the KG construction flow. The single-round prediction latency from data arrival to visualized result is maintained below 900 milliseconds under standard workload. The knowledge graph is updated incrementally on an hourly basis, and a full graph version is persisted daily. To handle anomalous data points, the system employs a two-stage strategy comprising real-time statistical outlier detection followed by expert review of flagged anomalies. User feedback from the decision-support interface is periodically aggregated and used to fine-tune the LLM extraction prompts and update the model via an active learning loop. From January 2025 to December 2025, the system operated continuously and stably for 12 months, processing over 7.2 million measured data records cumulatively, and achieved a closed-loop verification covering the entire chain from data collection, knowledge extraction, and graph evolution to carbon emission prediction, visual traceability, and decision support.

In this deployment, the input data streams included energy consumption monitoring data at 15-minute intervals from 86 IoT submeters, equipment operational status logs updated every 5 minutes, indoor and outdoor environmental parameters sampled every 10 minutes, and BIM-derived static attributes synchronized weekly. These heterogeneous streams were aligned to an hourly temporal resolution for model inference, with missing value rates per data type ranging from 0.2% to 3.8% over the 12-month operation period. The Q&A subsystem was evaluated under the protocol described in Section 3.2, confirming its capability to correctly interpret ambiguous phrasing and to successfully traverse multi-hop reasoning chains within a response latency comparable to the prediction cycle. The feature importance scores of different features of the three methods are compared, as presented in Table 5.

Table 5. Feature importance scores of the three methods

Feature importance score (points)	Research method	MLEPM	IO-MLP	XGBoost (Baseline)
Building structural features	0.92	0.78	0.75	0.80
Energy system type	0.88	0.72	0.69	0.75
Use features	0.85	0.65	0.62	0.70
Material carbon footprint	0.90	0.70	0.68	0.78
Climate and environmental factors	0.82	0.60	0.58	0.55
Average feature importance	0.87	0.69	0.66	0.72

From Table 5, the feature importance scores are calculated based on the SHAP method. High-scoring features such as building structure characteristics and material carbon footprint are highly consistent with professional knowledge in the field of building carbon emissions. The relatively low scores of climates and environmental factors indicate that their impact is weak in the current model, which may be related to the granularity of data collection or model structure, but still conforms to their auxiliary role in long-term carbon emission trends. Table 5 also includes the built-in feature importance scores of XGBoost as a baseline reference. The proposed method achieves higher importance scores across all features compared to the XGBoost baseline. The average feature importance score of the research method was 0.87, with the highest score of 0.92 for building structure features and the lowest of 0.82 for climate and environmental factors. Among the other two methods, the MLEPM method had the highest feature importance score of 0.78 for building structure features, and the average feature importance score of the IO-MLP method was 0.66. The research method has better feature extraction capabilities and the ability to identify key influencing factors. The recognition accuracy of different test question types and the relationship extraction accuracy of different relationship types by the three methods are compared in Fig. 9.

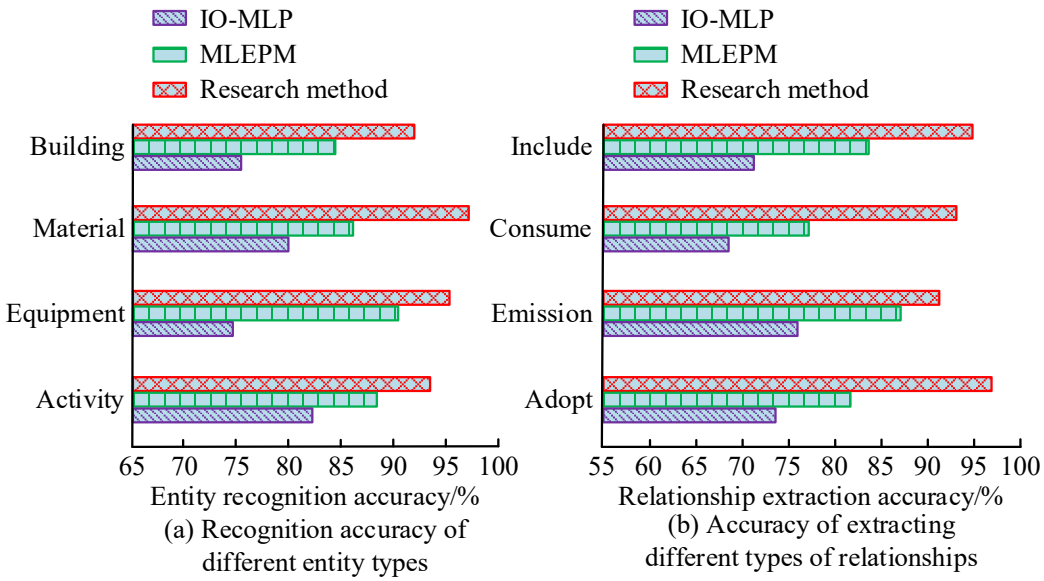


Fig. 9. Recognition accuracy for entity types and extraction accuracy for relation types

From Fig. 9, there were obvious differences in the recognition accuracy of different entity types and the extraction accuracy of different relationship types among the three methods. The overall recognition accuracy of the research method for different entity types was greater than 90%. The entity recognition accuracy was 91.6% for building, 97.5% for material, 95.8% for equipment, and 93.4% for activity. The other two methods had significantly lower entity recognition accuracy rates for each entity type than the research method. The research method had an overall relationship extraction accuracy of greater than 90% for different relationship types, with 94.9% for including, 93.3% for consuming, 91.4% for emission, and 96.5% for adopting. The research method has better generalization, accuracy, robustness, domain expertise, basic construction capabilities, and practicality. The consistency scores and consistency index of the three methods in different KG versions are compared in Fig. 10.

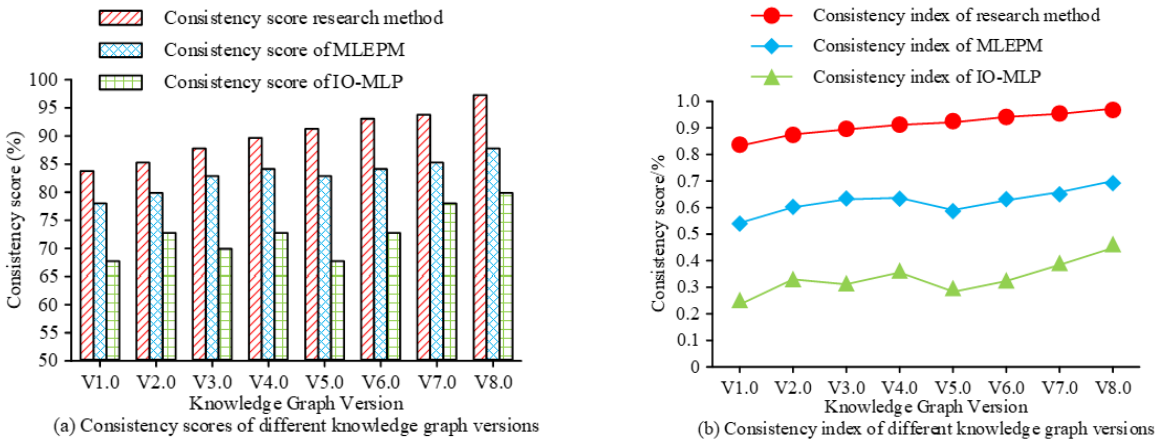


Fig. 10. Consistency score and consistency index test of different KG versions

From Fig. 10, the consistency scores and consistency indexes of the three methods showed an overall upward trend as the KG version increased. The overall consistency score of the research method was greater than 80%, and the overall consistency index was greater than 0.8. When the KG was V1.0, the consistency score of the research method was 84.3%. When the KG version was upgraded to V8.0, the consistency score of the research method increased to 97.4%, while the consistency scores of the other two comparison methods were significantly smaller. The consistency index of the research method increased from 0.82 to 0.98. The research method has better reasoning completeness, reliability, prediction accuracy, consistency, convergence, and stability. To further illustrate the practical utility of the KG in enhancing prediction accuracy, a case study was conducted on the HVAC system of a commercial complex. The standalone TCN-GCN model (without KG integration) predicted the system's weekly carbon emissions with an error of 15.2%. After integrating the KG and its embedded correlation paths (specifically, the HVAC Unit–consumes–Electricity–affectedBy–Outdoor Temperature and HVAC Unit–has EfficiencyRating–Coefficient of Performance (COP) relational chains), the prediction model incorporated real-time temperature data and equipment efficiency constraints, reducing the prediction error to 2.5%, which corresponds to a 12.7% reduction in prediction error attributable to the KG's contextual knowledge. This result was directly obtained from the online monitoring module of the prototype system.

To further evaluate the system's interpretability and diagnostic utility, a prediction error case was analyzed. In one instance, the model overestimated carbon emissions for a specific week by 18.3%. By examining the SHAP values, it was found that the feature 'outdoor temperature' had an unexpectedly high positive contribution. Concurrently, tracing the KG subgraph revealed that the temperature entity was linked to a cooling demand pathway that was temporarily inactive due to scheduled building maintenance. The semantic path in the KG offered a plausible explanation for the model's bias: it had learned a strong correlation from historical data that was temporarily invalid. This error analysis demonstrates that the combined SHAP-KG explanation mechanism is not only useful for understanding correct predictions, but also serves as a powerful tool for diagnosing model errors, identifying missing contextual information, and guiding targeted data

collection or model refinement. In summary, the proposed method has good feature extraction capabilities, generalization, accuracy, robustness, practicability, reliability, convergence and stability.

In summary, a partially visualized KG of building carbon emissions was constructed, and its specific structure is shown in Fig. 11.

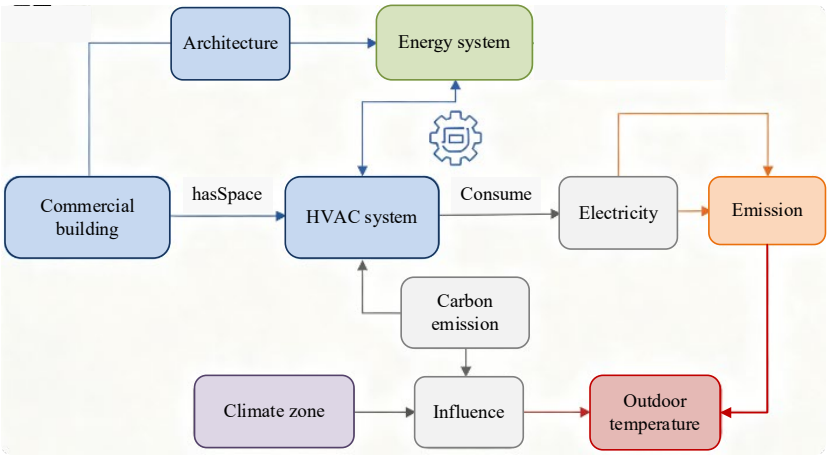


Fig. 11. Partial visualization of building carbon emissions KG

From Fig. 11, the graph presents the logic of carbon emissions in commercial buildings: buildings and climate zones respectively pass-through energy systems, HVAC systems, and other pathways, revealing the multi-factor interaction mechanism of building carbon emissions through electricity consumption, temperature effects, and carbon feedback. Then, to verify the support capacity of the research method for actual management decisions, three typical building carbon management scenarios were selected and quantitatively analyzed with traditional methods. The results are shown in Table 6.

Table 6. Comparison of quantified benefits in typical managerial decision scenarios

Managerial Decision Scenario	Key Metric	Research method	Baseline	Improvement
HVAC pre-regulation	Monthly emission reduction (tCO2e)	8.7	4.2	+107%
	Annual electricity cost savings (10k CNY)	9.4	4.5	+109%
Carbon pathway tracing & strategy decision	Root cause diagnostic time	15 min	3 weeks	-99.5%
	Decision convergence cycle	3 days	21 days	-85.7%
Equipment retrofit prioritization	Annual emission reduction (tCO2e)	35	13	+169%
	Unit abatement cost (CNY/tCO2e)	85	214	-60.3%

From Table 6, in the HVAC pre-regulation scenario, the proposed method improved monthly emission reduction by 107% and annual electricity cost savings by 109%. In the carbon pathway tracing scenario, the root cause diagnostic time was shortened from approximately three weeks to 15 minutes, and the decision convergence cycle was reduced by 85.7%. In the equipment retrofit prioritization scenario, the annual emission reduction increased by 169%, while the unit abatement cost decreased by 60.3%. These results demonstrate that the proposed method can effectively translate its technical superiority in prediction accuracy and knowledge reasoning into quantifiable managerial benefits. It should be emphasized that these quantified benefits are derived from the specific Shanghai demonstration project. The magnitude of improvements, particularly the emission reduction and cost savings, may vary when the framework is applied to buildings with different operational characteristics, climatic conditions, or equipment configurations. The decision support capabilities, including root cause tracing and retrofit

prioritization, are expected to remain effective across contexts, but the specific numerical outcomes should be interpreted as case-specific evidence rather than universal performance guarantees.

5. Conclusions

This research developed an integrated framework that constructs a dynamic KG from MHD and couples it with a spatiotemporal deep learning model for prediction and analysis. Through the constructed dynamic KG, which serves as a bridge between multi-source data and the prediction model, the research achieves enhanced interpretability of predictions and provides support for in-depth carbon emission traceability and intelligent decision-making. Experimental results show that the research method demonstrates robust KG construction ability, with an accuracy rate of over 90% for various types of entity recognition (such as 97.5% for material recognition), and an accuracy rate of over 90% for extracting relationships of different types. The consistency score of the constructed KG exceeds 80%, and the consistency index exceeds 0.8. The data preprocessing and temporal alignment strategies effectively handle multi-source heterogeneous inputs, maintaining missing value rates below 5% and ensuring reliable model performance. Furthermore, the credibility of these high-performance metrics is underpinned by a rigorous evaluation protocol that includes project-level data isolation and a strict temporal split, thereby ruling out data leakage. These results collectively confirm that the method has good accuracy, robustness, and stability in constructing carbon emission predictions and KG. The 12-month deployment in the Shanghai public office building, with its comprehensive sensor network and representative operational profile, provided a robust empirical basis for validation. However, as discussed in the experimental setup, the specific quantitative findings and optimization strategies are context-dependent and should be generalized with caution. The core methodological framework remains transferable to other building types and climatic regions, provided that appropriate local data and parameter adjustments are made.

Beyond empirical validation, this study provokes deeper considerations. First, the fusion of knowledge graphs with deep learning marks a paradigm shift from purely data-driven to knowledge-data dual-driven modeling, offering a blueprint for other complex systems where domain expertise and sparse data co-exist. Second, interpretability remains a cornerstone for carbon management decision-making; while SHAP and KG subgraphs provide traceable evidence, achieving true causal inference requires integrating counterfactual reasoning and temporal causal discovery. Third, the dynamic evolution of carbon emission KGs poses fundamental challenges: how to balance historical consistency with emerging knowledge, and how to enable lifelong learning without catastrophic forgetting. Addressing these questions will determine the long-term viability of intelligent carbon management systems.

This research does not remain at the level of theoretical design and simulated experiments; instead, the proposed knowledge graph construction and prediction method has been solidified into a deployable software system and undergone a 12-month online validation in a real building, with results fully demonstrating its effectiveness and reliability in practical carbon management scenarios. However, this study lacks relevant indicator data for long-term operation in real scenarios. In addition, the experiment was only tested on a single dataset and did not cover validation of generalization performance under different building types (such as residential, commercial, industrial), climate regions, and data source structures. Therefore, there are still limitations in cross-scenario adaptability. The next step will involve collaborating with the China Academy of Building Research to conduct an 18-month field verification in the Xiong'an New Area. In future research, long-term pilot deployments can be conducted in various real-world scenarios, collecting multi-category building data, multi-climate zone data, and multi-source heterogeneous data to further evaluate and optimize the adaptability and reliability of the model in different scenarios, thereby enhancing its practical applicability in dynamic and complex environments.

Data and Resources Availability

To enhance reproducibility and transparency, the following resources are provided as supplementary materials. The ontology schema is publicly available in OWL format at a permanent repository, containing the complete class hierarchy, property definitions, and cardinality constraints for the building carbon

emission knowledge graph. Representative sample triples are also listed in the supplementary file, illustrating the instantiation of the ontology with real building components, energy carriers, and carbon metrics. The complete training configuration for the TCN-GCN model, including the exact hyperparameter grid, the random seed list for all stochastic operations, and the data split indices, is documented in the supplementary materials. For raw data, desensitization is applied to all building-specific information, including precise geolocation, owner identities, and operational schedules, by replacing them with anonymized identifiers and generalized categories. The desensitized dataset, which preserves the statistical distributions and temporal patterns essential for model validation, is available upon reasonable request to the corresponding author. The Global Infrastructure Emissions Database, which serves as the primary external data source for emission factors, is publicly accessible through its official website, and the specific version and download date are recorded in the supplementary file.

Author Contributions

Xiaochen Zhang contributes to conceptualization, analysis, investigation, data collection, draft preparation, manuscript editing. Shida Hou contributes to conceptualization, methodology, manuscript editing, supervision, project administration. All authors have read and agreed with the manuscript before its submission and publication.

References

[1] Ziyuan C, Yibo Y, Simayi Z, Shengtian Y, Abulimiti M, Yuqing W. Carbon emissions index decomposition and carbon emissions prediction in Xinjiang from the perspective of population-related factors, based on the combination of STIRPAT model and neural network. Environmental Science and Pollution Research. 2022;29(21):31781-31796. https://doi.org/10.1007/s11356-021-17976-4

[2] Li X, Liu X, Huang Y, Li J, He J, Dai J. Evolutionary mechanism of green innovation behavior in construction enterprises: Evidence from the construction industry. Engineering, Construction and Architectural Management. 2024;31(1):159-178. https://doi.org/10.1108/ECAM-02-2022-0186

[3] Li T, Li Z, Dou Y. Diffusion prediction of prefabricated construction technology under multi-factor coupling. Building Research & Information. 2023;51(3):333-353. https://doi.org/10.1080/09613218.2022.2126343

[4] Zhao W, Fu Z, Fan T, Wang J. Ontology construction and mapping of multi-source heterogeneous data based on hybrid neural network and autoencoder. Neural Computing and Applications. 2023;35(36):25131-25141. https://doi.org/10.1007/s00521-023-08373-8

[5] Hu W, Hu R. Creating historical building models by deep fusion of multi-source heterogeneous data using residual 3D convolutional neural network. International Journal of Architectural Heritage. 2024;18(9):1377-1393. https://doi.org/10.1080/15583058.2023.2229253

[6] Peng S, Tan J, Ma H. Carbon emission prediction of construction industry in Sichuan Province based on the GA-BP model. Environmental Science and Pollution Research. 2024;31(16):24567-24583. https://doi.org/10.1007/s11356-024-32585-7

[7] Fei Y, Liao W, Lu X, Guan H. Knowledge-enhanced graph neural networks for construction material quantity estimation of reinforced concrete buildings. Computer-Aided Civil and Infrastructure Engineering. 2024;39(4):518-538. https://doi.org/10.1111/mice.13094

[8] Guo Z, Luo Y. A building carbon emission prediction model based on optimized machine learning model. Innovative Applications of AI. 2024;1(4):70-83. https://doi.org/10.70695/41ts7860

[9] Pan H, Wu C. Bayesian optimization + XGBoost based life cycle carbon emission prediction for residential buildings-An example from Chengdu, China. Building Simulation. 2023;16(8):1451-1466. https://doi.org/10.1007/s12273-023-1024-2

[10] Wei Z, Wei K, Liu J. Decoupling relationship between carbon emissions and economic development and prediction of carbon emissions in Henan Province: Based on Tapio method and STIRPAT model. Environmental Science and Pollution Research. 2023;30(18):52679-52691. https://doi.org/10.1007/s11356-023-26051-z

[11] Xiong Z, Li W, Cai Z. Federated generative model on multi-source heterogeneous data in IoT. Proceedings of the AAAI Conference on Artificial Intelligence. 2023;37(9):10537-10545. https://doi.org/10.1609/aaai.v37i9.26252

[12] Li X, Wang J, Tan J, Ji S, Jia H. A graph neural network-based stock forecasting method utilizing multi-source heterogeneous data fusion. Multimedia Tools and Applications. 2022;81(30):43753-43775. https://doi.org/10.1007/s11042-022-13231-1

[13] Guo P, Xiao K, Wang X, Li D. Multi-source heterogeneous data access management framework and key technologies for electric power Internet of Things. Global Energy Interconnection. 2024;7(1):94-105. https://doi.org/10.1016/j.gloei.2024.01.009

[14] Liu W, Tian Z, Jiang X, Liu S, Zhao B, Han Q, et al. A milling cutter state recognition method based on multi-source heterogeneous data fusion. The International Journal of Advanced Manufacturing Technology. 2022;122(7):3365-3378. https://doi.org/10.1007/s00170-022-10017-5
[15] Wang Z, Wang S, Wang S, Zhao Z, Tian Z. An intelligent prediction method of surface residual stresses based on multi-source heterogeneous data. Journal of Intelligent Manufacturing. 2025;36(1):441-457. https://doi.org/10.1007/s10845-023-02238-6
[16] Moshari A, Aslani A, Zolfaghari Z, Malekli M, Zahedi R. Forecasting and gap analysis of renewable energy integration in zero energy-carbon buildings: A comprehensive bibliometric and machine learning approach. Environmental Science and Pollution Research. 2023;30(40):91729-91745. https://doi.org/10.1007/s11356-023-28669-5
[17] Nsugbe E. Toward a self-supervised architecture for semen quality prediction using environmental and lifestyle factors. Artificial Intelligence and Applications. 2023;1(1):35-42. https://doi.org/10.47852/bonviewAIA2202303
[18] Yevu SK, Owusu EK, Chan APC, Oti-Sarpong K, Wuni IY, Tetteh MO. Systematic review on the integration of building information modelling and prefabrication construction for low-carbon building delivery. Building Research & Information. 2023;51(3):279-300. https://doi.org/10.1080/09613218.2022.2131504
[19] Ullah F, Olatunji O, Qayyum S. A scoping review of green Internet of Things in construction and smart cities: current applications, adoption strategies and future directions. Smart and Sustainable Built Environment. 2026;15(1):294-318. https://doi.org/10.1108/SASBE-11-2023-0349
[20] Sun B, Ju Z. Research on the promotion of new energy vehicles based on multi-source heterogeneous data: Consumer and manufacturer perspectives. Environmental Science and Pollution Research. 2023;30(11):28863-28873. https://doi.org/10.1007/s11356-022-24304-x