Applied Artificial Intelligence Research
INNOVIAIR

Can Public Health Policies Drive Carbon Reduction? —— Forecasting China's Carbon Emissions and Attribution of Policy Contributions Based on SHAP-BiLSTM

Haoyu Cheng[a], Yuanfang Du[a,*]

[a]College of Science, Xizang University, Lhasa, Xizang 850000, China

ARTICLEINFO

Keywords:
Public Health Policys
Carbon Emission Prediction
Policy Contribution
Machine Learning

ABSTRACT

Growing public health imperatives are accelerating carbon mitigation efforts, and carbon emission forecasting constitutes a critical foundation for assessing the feasibility of such measures. However, carbon emission data exhibit pronounced nonlinear and dynamically time-varying characteristics, rendering accurate prediction difficult for traditional econometric models, and existing studies lack quantitative assessment of public health policy contributions. This study proposes the SHAP-BiLSTM analytical approach to ensure both high prediction accuracy and policy interpretability. Based on a daily dataset comprising 2,466 observations spanning January 2019 to September 2025 to conduct quantitative analysis of policy contributions; two public health policy dummy variables and two carbon reduction policy dummy variables are established in chronological order in accordance with IPCC standards. The inclusion of carbon reduction policies helps control for other policy shocks during the implementation period of public health policies. The results indicated that the BiLSTM model significantly outperformed the comparative models in prediction accuracy, demonstrating that excessive algorithmic stacking does not necessarily enhance prediction accuracy. SHAP analysis revealed that public health policies contributed to carbon emission reduction to a certain extent,reflecting that the relevant policies played an effective role in slowing growth during high-emission periods.

1.Introduction

Global climate change has made carbon reduction a priority in global economic development[1,2], and accurate carbon emission forecasting is critical for effective government policymaking. China, a major emitter accounting for approximately one-third of global emissions, has committed to peaking carbon emissions by 2030 and achieving carbon neutrality by 2060—the "dual carbon" goals. However, emission data exhibit pronounced volatility and nonlinearity, challenging traditional econometric models[3]. Mainstream economic research is oriented toward post-hoc explanation rather than forward-looking prediction[4,5]: existing studies employ causal inference tools—difference-in-differences, synthetic control methods, and structural equation models—to evaluate low-carbon city pilot policies as quasi-natural experiments[6], but their reliance on linear relationships and strict exogeneity assumptions[7] conflicts with the nonlinear, time-varying nature of carbon emission systems, leaving the more policy-relevant task of predicting future trajectories understudied.

Although machine learning has been introduced for emission forecasting, two tendencies warrant caution. First, research scales are homogeneous: analyses of urban agglomerations[8,9] overlook regional disparities, carbon transfer, and spillover effects, impeding translation into Public Health policy. Second, ever more complex deep learning architectures are stacked to optimize performance metrics[10-12], yet their black-box outputs conflict with policy interpretation, creating an imbalance between predictive accuracy and policy interpretability.

This dilemma stems from the absence of an analytical framework integrating carbon emission physical principles with predictive accuracy sufficient for policy interpretation. BiLSTM (Bidirectional Long Short-Term Memory), which captures bidirectional sequence dependencies by combining forward-processing and backward-processing LSTM networks, offers theoretical feasibility for accommodating emission

* Corresponding author.
E-mail address: 13549014790@163.com.
https://doi.org/10.65455/wnr6ha05
Received 21 August 2026; Received in revised form 28 August 2026; Accepted 2 September 2026; Available 7 September 2026

dynamics under seasonal regularities and policy shifts[13]. SHAP (Shapley Additive Explanations), a cooperative game theory-based feature attribution method, calculates each feature's marginal contribution to predictions[14], furnishing locally accurate and globally consistent explanations[15] and thereby enabling the quantification of policy contributions.

This study therefore proposes a BiLSTM model under the SHAP framework. Using daily carbon emission data for China from the CarbonMonitor platform—76,416 province-day records from January 1, 2019, to September 30, 2025, aggregated into 2,466 national daily observations—we compare BiLSTM against LSTM and Attention-BiLSTM and quantify the contributions of four policy dummy variables under the SHAP framework. BiLSTM achieves mean reductions of 14.01% in MSE, 7.90% in RMSE, and 10.31% in MAE, averaged across the two comparative models and the three partition ratios. SHAP analysis reveals that the earliest public health policy (P1) is associated with emission declines, whereas more recent policies decelerate emission growth, with the dual carbon control policy (P4) registering the highest contribution ratio. The framework thus ensures both high prediction accuracy and policy interpretability, furnishing empirical foundations for optimizing China's carbon market policies and achieving the dual carbon goals.

2.Materials and methods

2.1.Policy design and date processing

The data employed in this study comprise daily carbon emission observations for China collected from the CarbonMonitor platform, which reports daily emissions for 31 provincial-level regions; the resulting 76,416 province-day records spanning the period from January 1, 2019, to September 30, 2025, are aggregated into a national daily series totaling 2,466 observations. Although the aggregated series covers only about 6.7 years, it consolidates information from more than seventy thousand underlying province-day records, and the chronological stability design described below verifies that this sample size supports stable identification of the BiLSTM. The aggregation from the provincial panel to a national series is a deliberate consequence of the research design rather than a loss of information: all four policies examined in this study were promulgated by the central government and implemented uniformly nationwide, so each policy dummy varies only over time and is constant across provinces on any given date; a provincial-level panel would therefore provide no additional identifying variation for the attribution of national policies, while introducing cross-sectional dependence and heteroskedasticity among provincial series. Because the objective of this study is to forecast national aggregate emissions and to quantify the contributions of nationally uniform policies, the national daily series aligns the data structure with the policy variables; the identification advantages of the provincial panel are relevant chiefly to regional heterogeneity, which is beyond the present scope and is explicitly left to future research in the limitations.

Regarding policy design, to avoid confounding shocks concurrent with policy implementation, carbon reduction-related policies are introduced on the basis of classifying public health policies according to IPCC standards. This classification follows the policy-instrument taxonomy of the IPCC Sixth Assessment Report, which distinguishes regulatory instruments from informational ones. Specifically, policies are classified into two categories—regulatory and informational—and policy dummy variables are established based on public health-related and carbon reduction-related policies promulgated by the Chinese government:

On July 18, 2021, the Chinese government issued the National Fitness Plan (2021–2025), whereby large-scale improvements in fitness infrastructure provided a long-term low-carbon lock-in effect. Accordingly, the first policy dummy variable, designated P1, is defined, with values set to 1 for observations on and after July 18, 2021, and 0 otherwise.

On April 27, 2022, the Chinese government issued the 14th Five-Year Plan for National Health, which explicitly incorporated the promotion of green development into health planning for the first time. Accordingly, the second policy dummy variable, designated P2, is defined, with values set to 1 for observations on and after April 27, 2022, and 0 otherwise. The dummy is set to the exact promulgation date of the policy document so that the variable accurately identifies the policy’s effective starting point; the same promulgation-date rule is applied uniformly to all four dummies, each of which is a sustained step function that switches from 0 to 1 on the issue date and remains 1 thereafter, thereby capturing the continuing institutional effect of the corresponding policy.

On October 24, 2021, the Chinese government proposed the top-level design concept for the dual carbon goals, incorporating carbon emission indicators into national planning. Accordingly, the third policy dummy variable, designated P3, is defined, with values set to 1 for observations on and after October 24, 2021, and 0 otherwise.

On July 11, 2023, the Chinese government introduced the dual control system for carbon emissions, shifting the core of government assessment toward direct control of total carbon emissions. Accordingly, the fourth policy dummy variable, designated P4, is defined, with values set to 1 for observations on and after July 11, 2023, and 0 otherwise.

The motivation underlying policies P1 and P2 originates from pure health concerns, ensuring their exogeneity and enabling more effective interpretation through SHAP. Policies P3 and P4 are top-level national policy documents formulated through established governmental decision-making procedures; their release timing is determined by the national policy agenda rather than by short-term emission fluctuations, and they are included in the model primarily to control for carbon-policy shocks concurrent with the implementation windows of the public health policies. It should be clarified that the SHAP analysis in this study measures correlative feature contributions to the model’s predictions rather than causal inference; accordingly, the policy attributions reported below are interpreted as model-based associations and are not overstated as causal policy effects.

Simultaneously, to preserve the principal characteristics of the carbon emission time series while reducing data noise, we preprocess the raw carbon emission data prior to predictive analysis: we standardize date fields through uniform formatting; to mitigate the interference of random noise on model learning, we apply a 7-day simple moving average to smooth the daily carbon emission series; since the BiLSTM constitutes a supervised learning model, we transform the unsupervised time series into a supervised sample set through sliding windows, whereby the carbon emission values at t consecutive preceding time points serve as input and the value at time point t+1 serves as output ; to determine the optimal window size, we extract candidate features including lag terms, difference terms, and rolling statistics from the original series. Following Lasso regression screening, the coefficients for la1 and diff1 are 2.84 and 0.79, respectively, substantially stronger than those of other features, whereas lag2 through lag18 are compressed to zero, thereby establishing a window size of t=7. We further employ the Min–Max method to normalize the data to [0,1], determining parameters exclusively from the training set to prevent data leakage . Finally, to enable a more profound assessment of model prediction accuracy, we chronologically partition the dataset into three training sets comprising 70%, 80%, and 90% of the data, with corresponding test sets of 30%, 20%, and 10%, respectively. We perform separate fitting for each partition to examine the sensitivity of model performance to the training data scale, with the primary experiment employing a 90% training set and a 10% test set. The model performs one-day-ahead (t+1) forecasts. After 7-day smoothing and sliding-window construction (t=7), approximately 2,450 supervised samples remain for model fitting; dropout regularization and Lasso-based feature screening further restrict the effective number of parameters. The three chronological partitions additionally serve as a stability check: as reported in Table 1, the ranking of model errors is largely consistent across partitions, indicating that 2,466 daily observations suffice for stable identification of the BiLSTM on this dataset, although the number of independent policy periods remains limited.

2.2.Processing model evaluation

In this study, the primary BiLSTM model and the comparative LSTM and Attention-BiLSTM models are constructed using the PyTorch framework, and three deep learning models are trained. The training set derived from the preprocessed data is employed to train the baseline models, and the test set is utilized to evaluate the fitted models. To enhance model robustness and prevent overfitting, dropout layers are incorporated into the model architecture. MSE, MAE, and RMSE are adopted to evaluate model performance on the test dataset; the specific values of the three error metrics under the three training configurations for the BiLSTM, LSTM, and Attention-BiLSTM models are reported in Table 1. The three models form an architecture chain of increasing complexity—unidirectional LSTM, bidirectional BiLSTM, and attention-stacked Attention-BiLSTM—so the comparison effectively constitutes an ablation over architectural stacking. Traditional benchmark models such as ARIMA and SARIMA are not included in the present experiments owing to the journal's space constraints; instead, we rely on established benchmark evidence: prior systematic comparisons have demonstrated that LSTM- and BiLSTM-type networks substantially outperform ARIMA-family models on nonlinear time-series forecasting tasks[3,13], so the claim that conventional econometric models attain insufficient accuracy on such data is supported by the published literature rather than by newly repeated experiments, and the present design focuses on the ablation among deep-learning architectures. According to the MSE metric, relative to Attention-BiLSTM, BiLSTM reduces MSE by 40.37%, 10.39%, and 3.10% under the 70:30, 80:20, and 90:10 partitions, respectively (mean 17.95%); relative to LSTM, BiLSTM reduces MSE by 39.01% and 3.05% under the 70:30 and 90:10 partitions, whereas under the 80:20 partition its MSE is 11.86% higher than that of LSTM (mean reduction across the three partitions: 10.07%). According to the RMSE metric, relative to Attention-BiLSTM, BiLSTM reduces RMSE by 22.78%, 5.34%, and 1.56% under the three partitions, respectively (mean 9.89%); relative to LSTM, the corresponding reductions are 21.91% and 1.54% under the 70:30 and 90:10 partitions, with a 5.76% increase under the 80:20 partition (mean reduction: 5.90%).

According to the MAE metric, relative to Attention-BiLSTM, BiLSTM reduces MAE by 28.13%, 6.26%, and 1.58% under the 70:30, 80:20, and 90:10 partitions, respectively (mean 11.99%); relative to LSTM, the corresponding reductions are 28.23% and 1.20% under the 70:30 and 90:10 partitions, with a 3.53% increase under the 80:20 partition (mean reduction: 8.63%). Notably, under the 80:20 partition, BiLSTM's MSE, RMSE, and MAE are 11.86%, 5.76%, and 3.53% higher than those of LSTM, respectively. This may be attributable to the influence of economic cycles and policy fluctuations on carbon emissions: under the 80:20 partition, most policy interventions were concentrated in the test set, whereas BiLSTM is more adept at capturing long-term dependencies, thereby yielding suboptimal fitting results under this partition. Nevertheless, across the other partition ratios, BiLSTM demonstrated superior performance relative to the other two models for all three parameters. Figures present the comparison between predicted and actual values for the BiLSTM, LSTM, and Attention-BiLSTM models under different partition ratios. These figures clearly demonstrate that the BiLSTM model achieves superior fitting effects and accuracy, with carbon emission trends closely matching actual changes. This fully demonstrates that BiLSTM exhibits significant performance advantages in carbon emission prediction tasks, with relatively smaller prediction errors and strong reliability and predictive capability. In particular, Attention-BiLSTM, which stacks an attention layer on top of BiLSTM, underperforms the plain BiLSTM in MSE under all three partitions (4.4934 vs. 2.6794 at 70:30, 2.4086 vs. 2.1583 at 80:20, and 2.1156 vs. 2.0501 at 90:10), providing direct numerical evidence that excessive algorithmic stacking does not necessarily enhance—and may even degrade—prediction accuracy on this dataset.

Table 1. Calculation results of model error indicators

Split	Model	MSE	RMSE	MAE
	Attention-BiLSTM	4.4934	2.1198	1.5650
70:30:00	BiLSTM	2.6794	1.6369	1.1247
	LSTM	4.3935	2.0961	1.5670
	Attention-BiLSTM	2.4086	1.5520	1.1851
80:20:00	BiLSTM	2.1583	1.4691	1.1109
	LSTM	1.9295	1.3891	1.0730
	Attention-BiLSTM	2.1156	1.4545	1.1590
90:10:00	BiLSTM	2.0501	1.4318	1.1407
	LSTM	2.1147	1.4542	1.1546

2.3.Policy contribution analysis

It should be emphasized that SHAP values quantify each feature's contribution to the model's predictions rather than counterfactual causal effects; the policy attributions reported below are therefore interpreted as model-based associations. Confounding is mitigated by design: the carbon-policy dummies P3 and P4 control for carbon-policy shocks concurrent with the implementation windows of the public-health policies, while the lagged and rolling features absorb temporal trends. Carbon emissions are predicted using the BiLSTM model, and the contributions of the four policies are quantified under the SHAP framework. As shown in Figure 4, the policies diverge markedly in both the direction and intensity of their effects on model predictions. P1 registers a mean SHAP value of −0.000679, the sole negative effect, indicating that—after controlling for temporal variables such as lag features and rolling means—its presence induces a downward adjustment in the model's CO_2 predictions, consistent with an inhibitory effect; its contribution ratio of 12.43% further validates the low-carbon lock-in effect at the data level. The remaining three policies exhibit positive SHAP values: P2 records 0.000435 (contribution 7.76%), P3 registers 0.003151 (37.74%), and P4 registers 0.001582 (42.06%). This likely stems from their implementation periods coinciding with China's carbon peaking phase, a high-emission period, rather than indicating that these policies drive emission growth. From the growth rate perspective, carbon emissions rose continuously from 11,148.4958 Mt in 2019 to 11,901.4367 Mt in 2024, an overall upward trend; yet after the implementation of P2, P3, and P4, the mean monthly increment decelerated from 40.5382 Mt in 2023 to 15.0676 Mt in 2024, a marked slowdown. Their positive SHAP values thus reflect short-term fluctuations during the institutional adaptation period. Meanwhile, implementation coincided with the post-pandemic economic recovery, when emissions were on an upward trajectory, and the model may partially misattribute temporal trend correlations as policy effects. Furthermore, short-term implementation entailed large-scale construction of medical institutions, health management centers, and green infrastructure, alongside industrial technological upgrading, whose construction-period energy consumption and emissions temporarily exceed the reduction benefits of the long-term operational period.

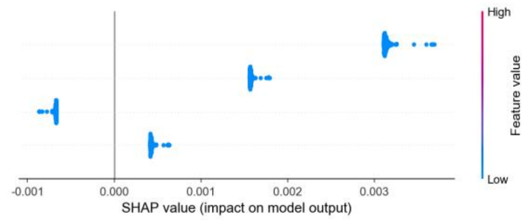

Fig 1. SHAP summary

3.Discussion

The present findings are consistent with prior evidence on the merits of deep learning for emission forecasting. The superiority of BiLSTM over the unidirectional LSTM and the attention-stacked variant observed here echoes the benchmark results of Siami-Namini et al.[13], who report that BiLSTM attains the highest forecasting accuracy among recurrent architectures, and accords with studies documenting the difficulty that conventional econometric and grey models face with nonlinear emission dynamics[3]. The accuracy degradation induced by attention stacking is likewise consistent with recent stacking-ensemble analyses[9], in which additional layers do not guarantee performance gains.

Our model-based attribution results diverge from quasi-experimental evaluations in informative ways. Difference-in-differences and PSM-DID studies of low-carbon city pilots, environmental regulations, and carbon trading schemes generally report significant emission-reduction effects[4-7], whereas the public health policies examined here exhibit only weak associations with predicted emissions. Three factors account for this discrepancy: the SHAP analysis quantifies correlative feature contributions rather than causal effects; public health policies target health outcomes, so carbon mitigation arises only as an indirect co-benefit transmitted through behavioral and infrastructural channels; and national-level aggregation masks the regional variation on which quasi-experimental identification relies.

Two counter-intuitive findings merit emphasis: the weak attributions of the public health policies P1/P2 and the positive SHAP values of the carbon policies P3/P4. The weak P1/P2 attributions are plausible because health policies influence emissions only indirectly and with long lags. The positive SHAP values of P3/P4 do not indicate that carbon policies raise emissions: their implementation windows coincide with China’s carbon peaking phase and the post-pandemic economic recovery, and potential confounding factors—the recovery of energy-intensive production, energyprice fluctuations, and concurrent construction activity—may be partially absorbed by the policy dummies despite the lagged and rolling controls. Data limitations, including the 6.7-year sample span and the absence of an untreated control group at the national scale, further constrain attribution. These caveats motivate the quasi-experimental validation with finer-grained regional or industry data identified as future research.

4.Conclusion

Addressing the dilemmas of existing carbon emission prediction research, this study proposes the SHAP-BiLSTM model, a unified analytical framework ensuring both high prediction accuracy and policy interpretability. Evaluated on daily carbon emission data with reduced model stacking, the BiLSTM model's predictions, combined with SHAP-based quantification of public health policies' contributions to carbon reduction, provide empirical foundations for optimizing China's public health policies and achieving carbon reduction targets.

Three policy recommendations follow, each tied directly to the empirical findings. First, because P1 is the only policy whose inclusion induces a downward adjustment in the model's emission predictions (mean SHAP value −0.000679; contribution ratio 12.43%), public health policies built around fitness infrastructure may serve as a complementary low-carbon channel; given the small magnitude of this association, however, such measures should be positioned as auxiliary rather than principal instruments, and their carbon co-benefits should be verified with finer-grained data before binding facility standards are considered. Second, because the dual carbon control policy P4 registers the highest contribution ratio (42.06%) among the four policies, regulatory instruments should remain the backbone of emission governance: the dual control system should be consolidated, and assessment mechanisms should distinguish short-term fluctuations during the institutional adaptation period from long-term reduction trends, so that long-term targets are not negated on the basis of short-term data. Third, the observed deceleration of the mean monthly emission increment—from 40.5382 Mt in 2023 to 15.0676 Mt in 2024 following the implementation of P2, P3, and P4—indicates that policy effects should be evaluated over sufficiently long horizons; differentiated dual carbon control pathways suited to regional economic development levels and industrial structures should be permitted, and health–carbon synergies should be strengthened in policy design.

Several limitations indicate future directions. The policy coverage was not exhaustive and could be expanded toward a more comprehensive evaluation system; the static assessment of policy contributions could be extended via time-varying parameter models or rolling window analysis to capture dynamic policy effects. Moreover, SHAP values measure features' "contribution" to predictions rather than strict causal effects; because all four policies were implemented nationwide, no untreated control group exists at the national scale, which precludes difference-in-differences or instrumental-variable designs on the present dataset, and quasi-experimental validation using finer-grained regional or industry data is left to future research. In addition, policy effect directions vary across samples, indicating significant heterogeneity that warrants differentiated interpretation by regional economic development and industrial structure.

References

[1] CHO H, ACKOM E. Artificial intelligence (AI)-driven approach to climate action and sustainable development. Nature Communications, 2025, 16(1).

[2] WU K, DING J, LIN J, et al. Big-data empowered traffic signal control could reduce urban carbon emission. Nature Communications, 2025, 16(1): 2013.

[3] WANG M. Predicting China's energy consumption and CO2 emissions by employing a novel grey model. Energies, 2024, 17(21): 5256.

[4] HUANG H, YI M. Impacts and mechanisms of heterogeneous environmental regulations on carbon emissions: An empirical research based on DID method. Environmental Impact Assessment Review, 2023, 99: 107039.

[5] ZHANG P, GUO Q. How do green municipal bonds affect carbon emissions in China? Evidence from the staggered difference-in-differences approach. Journal of Environmental Management, 2025, 373: 123692.

[6] XIAO Y, YANG H, XIANG X, et al. Lifting or lessening? Examining the effect of low-carbon city trial on synergistic reduction of pollution and carbon emissions. Energy, 2025, 314: 134195.

[7] JIANG H, LIU Z, CHEN Z. The effects of carbon emission rights trading pilot policy on corporate green innovation: Evidence from PSM-DID and policy insights. Sustainability, 2026, 18(3): 1207.

[8] JIN Y J, JIN X B, HONG X M, et al. Carbon emission prediction of Hainan Province based on Lasso-Transformer neural network model. Huanjing Kexue/Environmental Science, 2026, 47(2): 781-792.

[9] WANG M, YU J, ZHOU M, et al. Analysis and prediction of carbon emissions from public building operation based on stacking ensemble strategy: The case of Xi'an. Energy, 2026: 139890.

[10] WANG Q R, LIU X K, ZHU C F, et al. Carbon emission prediction in China based on LASSO-MEMD-SSA-LSTM. Journal of Safety and Environment, 2026, 26(06): 2475-2484.

[11] YANG Q, JIANG Y H, WU C Y, et al. Carbon peaking prediction for provincial transportation sectors in China based on CNN-LSTM-Attention. Journal of Safety and Environment, 2025, 25(10): 4064-4075.

[12] LI H W, WANG L J, LI S B, et al. Dynamic prediction method for electricity carbon emission intensity based on ISSA-CNN-BiLSTM. Electric Power, 2026,59(7): 66-77.

[13] SIAMI-NAMINI S, TAVAKOLI N, NAMIN A S. The performance of LSTM and BiLSTM in forecasting time series//2019 IEEE International Conference on Big Data. Piscataway, NJ: IEEE, 2019: 3285-3292.

[14] MOSCA E, SZIGETI F, TRAGIANNI S, et al. SHAP-based explanation methods: A review for NLP interpretability//Proceedings of the 29th International Conference on Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2022: 4593-4603.

[15] VAN DEN BROECK G, LYKOV A, SCHLEICH M, et al. On the tractability of SHAP explanations. Journal of Artificial Intelligence Research, 2022, 74: 851-886.