Asian Academy of Management Journal of Accounting and Finance

AAMJAF Vol. 22, No. 1, 91–120, 2026

HYBRID MACHINE LEARNING ANALYSIS OF EXOGENOUS FEATURES IN CHINA'S GUANGDONG CARBON MARKET PRICE PREDICTION

Chenyao Duan[1], Yuanfang Chen[2], Junlin He[3] and Kok-Haur Ng[1, 4]*

[1]Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[2]Faculty of Business and Economics, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[3]Institute for Advanced Studies, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[4]Centre of Research for Statistical Modelling and Methodology, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

*Corresponding author: kokhaur@um.edu.my

Submitted: 30 November 2025; Accepted: 27 February 2026; Published: 28 June 2026

To cite this article: Duan, C., Chen, Y., He, J., & Ng, K. H. (2026). Hybrid machine learning analysis of exogenous features in China's Guangdong carbon market price prediction. *Asian Academy of Management Journal of Accounting and Finance, 22*(1), 91–120. https://doi.org/10.21315/aamjaf2026.22.1.3

To link to this article: https://doi.org/10.21315/aamjaf2026.22.1.3

ABSTRACT

Carbon pricing is essential for the effective functioning of carbon markets, providing a basis for strategies that support sustainable green economic development. This study investigates the predictive power of intraday carbon price-related information, trading volume and crude oil prices within a hybrid model combining a convolutional neural network (CNN) and a long short-term memory (LSTM) network to forecast carbon emission allowance closing prices. The model is applied to Guangdong, China's largest and most active carbon-trading market, to improve forecasting accuracy. The CNN-LSTM models are benchmarked against a standalone LSTM model using multiple loss functions, with results showing the hybrid model's clear superiority in both in-sample and out-of-sample forecasts. Analysis of exogenous variable highlights that lagged intraday opening, highest, lowest and closing (OHLC) prices are the most critical predictors, with their exclusion significantly reducing accuracy. Brent crude oil prices and trading volume provide moderate contributions to the model, while quasi-likelihood analysis reaffirms lagged OHLC information and oil prices as key factors in carbon price prediction.

Keywords: Carbon trading prices, Price-related information, Long short-term memory, Convolutional neural network, Hybrid model

INTRODUCTION

Since the adoption of the Paris Agreement and its goal of limiting global temperature rise, emissions trading systems (ETS) have increasingly emerged as crucial policy instruments for countries seeking to fulfil their emission reduction commitments (Villoria-Sáez et al., 2016). Compared with traditional regulatory controls or tax-based mechanisms, ETS sets an overall emissions cap while allowing allowances to be traded in the market. It promotes compliance with emissions limits and gives firms flexibility to minimise abatement costs. China officially launched its national carbon market in 2021, covering 4.5 billion tonnes of carbon dioxide (CO_2) emissions from the power sector, thereby becoming the world's largest single carbon trading platform (Li et al., 2022). However, the effective operation of the national ETS has relied heavily on the pioneering experimentation and feedback provided by provincial pilot markets, particularly in areas such as institutional design, trading rules, and allowance allocation.

Nevertheless, even in mature pilots, achieving stable price discovery remains a challenge. Carbon prices are driven by multiple interacting factors, including energy price volatility (Tan et al., 2024), the pace of policy adjustments (Tsai, 2024), and broader macroeconomic cycles (Lu et al., 2025). Guangdong's carbon price exhibits pronounced seasonality and sudden price jumps, rendering its short-term dynamics difficult to capture through simple linear models. For government authorities, access to reliable short-term price forecasts enables early detection of potential market failures or excessive volatility, thereby facilitating timely intervention through price stabilisation mechanisms or allowance auctions. For emitting enterprises, accurate forecasts of carbon prices over the next few trading days are crucial for formulating dynamic inventory management strategies to optimise the timing of allowance purchases and minimise compliance costs. Meanwhile, for financial institutions and investors, the speed at which short-term forecast errors converge directly determines the effectiveness of arbitrage opportunities, hedging strategies, and the accuracy of derivative pricing and risk exposure assessments. Hence, the development of a high-precision short-term carbon price forecasting model tailored to the Guangdong ETS carries immediate and pressing practical significance for policymaking, corporate cost management, and financial risk control.

Early research on European Union (EU) ETS carbon price forecasting centred on linear econometric models, notably autoregressive integrated moving average (ARIMA) variants and vector autoregressive (VAR), supported by continuous spot and futures data (Zhang & Wu, 2022; Srinivasan et al., 2023; Jia et al., 2025). Autoregressive (AR)-type models capture short-term price inertia but rely on stationarity in mean and variance, which limits their ability to reflect structural breaks induced by policy shocks, energy price transmission or compliance deadlines (Shin et al., 2019; Srinivasan et al., 2023). VAR frameworks incorporate interdependencies among electricity prices, fuel spreads and climatic factors, yet their linear form leads to systematic underestimation of sharp spikes and heavy-tailed volatility (Bram et al., 2025). To address heteroskedasticity and higher-order nonlinearities, autoregressive conditional heteroskedastic (ARCH) and generalised ARCH (GARCH) families were adopted to model volatility clustering, which can partially account for intensified trading activity (Li & Lei, 2018; Otto & Schmid, 2023). Empirical studies nonetheless document parameter instability across regulatory phases and allocation periods, which forces frequent recalibration and these models provide limited explanatory power for exogenous variables such as oil and coal prices (Perifanis & Dagoumas, 2021; Wang et al., 2025).

With the rise of machine learning, nonparametric methods such as random forests and gradient boosting trees have improved fitness and offer some interpretability through feature importance, while the support vector regression (SVR) model shows robustness with small samples (Tange et al., 2017; Sahin, 2020; Adler & Painsky, 2022). However, the common sliding-window and feature-flattening pipeline weakens the sequential structure, making accuracy sensitive to window length and prone to dimensionality expansion. The SVR model is also highly sensitive to hyperparameter choices, which risks underfitting or overfitting without careful tuning (Rashid et al., 2022; Dhanka et al., 2025). To mitigate this issue, the simulated annealing (SA)-SVR framework by Goh et al. (2024) employs SA to optimise SVR hyperparameters, improving overall prediction accuracy and stability.

The rise of deep learning has opened new avenues for carbon price forecasting. Convolutional neural networks (CNNs), which leverage the local receptive fields of convolutional filters, have performed impressively in financial time series pattern recognition. For instance, Zhao et al. (2023) applied one-dimensional CNNs to price data from the EU ETS, demonstrating that convolution-based feature extraction outperforms traditional technical indicators. Nevertheless, researchers have observed that CNNs often suffer from memory discontinuity when attempting to capture long-term trends (Ghimire et al., 2019; Shih et al., 2019). Long short-term memory (LSTM) networks and gated recurrent units (GRUs) offer significant advantages over traditional RNNs in modelling long-term dependencies and lag effects. For example, Zhou et al. (2022) employed the LSTM model to forecast China's carbon futures prices and found that the model exhibits a slower accumulation of forecast errors in recursive multi-step predictions compared to ARIMA. However, a standalone LSTM primarily focuses on capturing global sequential structures and tends to underperform in recognising high-frequency local patterns such as intraday price spreads or spikes in trading volume, resulting in delayed responses to sudden events (Shah et al., 2022; Buratto et al., 2024).

Hence, scholars have begun experimenting with hybrid architectures that merge convolutional and recurrent layers to harness the strengths of both. Using power load and spot electricity prices as examples, Alsharekh et al. (2022) and Jin et al. (2022) showed in the United States (U.S.) Pennsylvania-New Jersey-Maryland Interconnection wholesale electricity market (PJM), and Korea Power Exchange (KPX) markets that CNN-LSTM models deliver markedly lower 24-hour forecasting errors, indicating that the local fluctuation features extracted by convolutional layers provide the LSTM with a more discriminating sequence representation. Ling and Cao (2024) introduced this hybrid architecture into China's carbon market and, using national average price samples, achieved marginal gains over a standalone LSTM. Nevertheless, these results were obtained in high-frequency or structurally stable data environments and focused primarily on national or cross-regional average prices. Institutional differences in allocation benchmarks, monthly compliance rhythms, and industrial composition give Guangdong's ETS price series sharper jumps and stronger seasonality, thus, the effectiveness of convolution-recurrent hybrids in this complex institutional context has yet to be systematically verified.

Although deep learning studies on carbon price forecasting are proliferating, most still treat exogenous information in a coarse manner, typically by adding broad macro-economic indicators (Zhou & Li, 2019; Song et al.,

2022) or headline energy prices (Krokida et al., 2020; Escribano et al., 2023; El Hokayem et al., 2024; Luo et al., 2024; Chen et al., 2025) to the input set, and they largely ignore the detailed and high-frequency signals contained within the carbon contract itself. Consequently, the literature provides no systematic evidence on whether intraday variables such as the opening, highest, lowest and closing (OHLC) prices of allowance trades possess predictive content that goes beyond the information already embedded in the end-of-day close. It is especially consequential for the Guangdong market, where the Guangdong ETS operates on a tight compliance schedule, requiring firms to complete allowance settlements within relatively short timeframes. At the same time, intraday trading activity within the market is highly vibrant (Liu & Jin, 2020). Under these conditions, the expectations of traders and enterprises are rapidly reflected in intraday price-related information, often much sooner than macro-level energy price signals such as international oil prices.

To address these gaps, this study shifts the analytical focus from external macro shocks to the market's own microstructure and employs a CNN-LSTM forecasting framework to quantify the incremental contribution of price-related information to short-horizon carbon price accuracy. This study offers a threefold contribution. First, it formalises intraday OHLC prices as well as their trading volume and Brent crude oil prices as core predictive inputs and designs a clear identification strategy using feature ablations and quasi-likelihood (QLIKE) evaluation to measure their marginal explanatory power. Second, it develops a compact CNN-LSTM model that integrates convolutional feature extraction with gated recurrent dynamics and pairs it with a transparent rolling expanding-window protocol, establishing a replicable and policy-relevant benchmark for carbon price prediction. Third, it builds a province-level empirical setting for Guangdong ETS, creating a policy-relevant testbed for regulators, compliance entities and investors. Overall, these contributions advance short-run carbon price forecasting and provide actionable insights for market oversight, such as early warning systems for volatility control and corporate procurement, enabling firms to transition from passive compliance to proactive, cost-minimising trading strategies.

This study selects the Guangdong ETS as the empirical setting not merely due to its status as one of the earliest pilot programmes, but because of its unique representativeness and data quality required for deep learning architectures. As the world's second-largest regional carbon market in terms of trading volume and value, trailing only China's unified national market (Wang et al., 2022), Guangdong offers a distinct advantage over other pilots that suffer from insufficient allowance transactions or discontinuous price series (Chang et al., 2018; Liu et al., 2020). Specifically, unlike the Shenzhen market, where frequent product differentiation complicates price unification, or inland markets characterised by thinner liquidity and heavy administrative guidance (Huang & Zhang, 2024; Dong & Zhang, 2024), Guangdong represents a highly market-oriented, coastal economy with a diversified industrial base encompassing advanced manufacturing, electronics, and petrochemicals. Its dense market participation and high trading frequency facilitate robust price discovery, ensuring that the price signals captured by our CNN-LSTM model reflect genuine market supply-and-demand dynamics rather than noise from liquidity vacuums. Furthermore, given the region's geographical and economic interconnections, the allowance price in Guangdong serves as a pivotal bellwether for carbon asset valuation across the broader Pan-Pearl River Delta region. It plays a crucial role in the "price discovery-policy feedback-corporate decision-making" loop, making it the ideal testbed for validating high-precision forecasting models intended for mature, information-efficient carbon markets, where technical signals are robust enough to support systematic trading and regulatory intervention (Wang et al., 2022; He et al., 2026).

METHODOLOGY

Data

The daily data on carbon emission allowance prices in Guangdong province is obtained from the Guangdong Emissions Allowance, ranging from 3 January 2017 to 29 December 2023. Five variables are collected, including opening price at time t (denoted as P_t^{open}), highest price at time t (denoted as P_t^{High}), lowest price at time t (denoted as P_t^{Low}), closing price at time t (denoted as P_t^{Close}), and trading volume at time t (denoted as V_t). In addition, referring to Lin and Jia (2019) and Zhou and Li (2019), the energy prices may have an impact on the carbon trading prices. Therefore, the Brent crude oil prices in U.S. dollars (USD) at time t (denoted as B_t) is considered one of the exogenous variables. The oil price, drawn from the U.S. Energy

Information Administration (EIA), is aligned to the carbon-market calendar. We remove dates with missing oil price data and align the remaining observations with the carbon price series. The resulting dataset contains a total of 1,549 observations over the full sample period. The training set spans from 3 January 2017 to 21 February 2023, comprising 1,349 observations, while the testing set covers 22 February 2023 to 29 December 2023, with 200 observations.

Figure 1 illustrates the time series plots of all carbon prices in Chinese Renminbi (RMB), Brent crude oil prices (USD), and scaled trading volume (volume × 10^{-5}). It is observed that all carbon prices increase rapidly after May 2021, stabilise from April 2022, and begin to decline after June 2023. In comparison, the crude oil price series generally fluctuates between USD40 and USD85, and exhibits a sharp drop around the end of April 2020. Subsequently, both the oil price series and the carbon trading price series exhibit a similar upward trend until around mid-2022, followed by a downward trend that continues until approximately September–October 2023, when a notable upward spike occurs within the study period. The volume shows several peaks around June 2017, between February 2018 and April 2019, and again in June 2021.

Notes: Open, High, Low and Close denote the opening, highest, lowest and closing carbon trading prices, respectively; Price denotes the Brent crude oil prices; and Volume (scaled) denotes the scaled carbon price trading volume.

FIGURE 1: Time series plots of OHLC carbon trading prices, Brent crude oil prices and scaled trading volume.

Feature Engineering

The forecasting target is the next-day closing price of Guangdong emission allowances, namely P_{t+1}^{Close}. Six raw variables, all directly observable in the market and closely related to short-term carbon-price movements, are used as inputs, including P_t^{Open}, P_t^{High}, P_t^{Low}, P_t^{Close}, V_t and B_t. The four price quotes and the volume capture the internal supply-and-demand conditions of the carbon market, whereas the Brent series reflects external energy-price dynamics.

All variables are rescaled to the [0,1] range by min-max normalisation. The scaling parameters are estimated on the training portion of each rolling cross-validation split and kept fixed on the corresponding validation and test portions, thereby eliminating any information leakage. Time dependence is incorporated through a sliding-window transformation. With a look-back length of $L = 10$ trading days, the observation at time $t(t \geq L)$ comprises the six-variable histories, such as P_{t-9}^{Open}, P_{t-8}^{Open}, ..., P_t^{Open}, stacked into a 10×6 matrix $\boldsymbol{X}_t$ that is used to predict P_{t+1}^{Close}. Accordingly, the tensors fed to the LSTM and CNN-LSTM models have the shape (batch, 10, 6).

To quantify the marginal contribution of each input variable, a series of ablation experiments is conducted on the CNN-LSTM model. Starting with the full six-variable model, the network is retrained systematically after removing the opening price, high price, low price, trading volume and oil price. Furthermore, two economically coherent blocks are excluded for comparison: (i) the entire variable set and (ii) the volume and oil pair.

Cross Validation

K-fold cross-validation (CV) is a commonly used model evaluation and selection method, aiming to make full use of the existing sample data and reduce bias caused by randomly dividing the training and validation sets (Jung, 2018). In this study, we divide the complete dataset into *k* mutually exclusive subsets (folds). In each iteration, one subset is selected as the verification set, and the other *k*–1 subsets are used as the training set. The validation results of each iteration are summarised and averaged to obtain a more robust estimate of the model's performance. In time series data, the observations are often interdependent, and the time sequence of data has an important impact on the prediction performance. If the sequence is randomly segmented like processing independent identically distributed data, the original time dependence will be destroyed, and then the evaluation of the real prediction ability of the model will be affected. Therefore, the cross-validation method suitable for time series data must be applied. By rolling the training set and validation set on the time axis, it is ensured that only the historical data earlier than the validation set is used in each model training. The data splitting method is shown in Figure 2. The first training uses the initial part of the sequence as the training set to verify the subsequent data segment. Then, the training set is rolled forward to contain more data at time points, and then the further backward time period is predicted. This process is repeated until the entire sequence is covered. This method can simulate the real prediction scenario, that is, it only relies on the known historical data to predict the future.

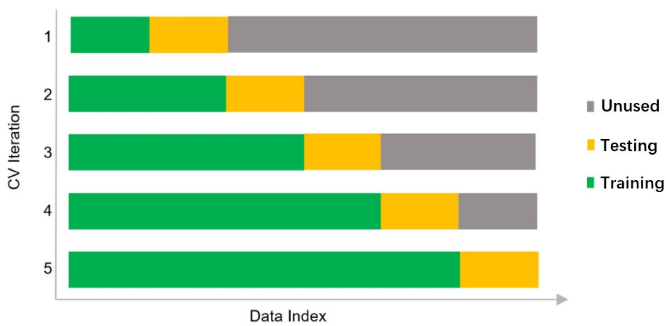


FIGURE 2: The splitting method for time series data

LSTM Model

The LSTM model is a specialised form of recurrent neural network (RNN) designed to effectively address the vanishing and exploding gradient issues encountered in training conventional RNNs on long sequences (Hochreiter & Schmidhuber, 1997). By incorporating forget, input and output gates in the hidden layers, LSTM can retain and update information over longer temporal spans, thus finding widespread applications in fields such as natural language processing, speech recognition and time series forecasting (Al-Selwi et al., 2024). LSTM's ability to maintain pertinent historical information grants it a competitive edge for modelling complex time series. For instance, in stock price or demand forecasting tasks, LSTM can effectively handle non-linearity and noisy data.

By incorporating memory cells and gating mechanisms, LSTM selectively retains or forgets information, significantly improving performance on sequence learning tasks. The fundamental unit of an LSTM is the "cell state", which is dynamically regulated by three gating mechanisms:

1. Forget Gate: Determines which information from the previous cell state should be retained or forgotten.
2. Input Gate: Determines how much of the new input at the current time step should be added to the cell state.
3. Output Gate: Determines how the information in the cell state is exposed to the hidden state for subsequent computations.

Mathematically, given the input vector $\boldsymbol{x}_t$ and weight matrix $\boldsymbol{W}_f$, $\boldsymbol{W}_i$, $\boldsymbol{W}_o$, and $\boldsymbol{W}_c$, the previous hidden state $\boldsymbol{h}_{t\text{-}1}$, and the previous cell state $\boldsymbol{c}_{t\text{-}1}$ at time t, the LSTM gating mechanisms can be expressed as:

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f[\boldsymbol{h}_{t-1},\boldsymbol{x}_t] + \boldsymbol{b}_f),$$

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i[\boldsymbol{h}_{t-1},\boldsymbol{x}_t] + \boldsymbol{b}_i),$$

$$\boldsymbol{O}_t = \sigma(\boldsymbol{W}_O[\boldsymbol{h}_{t-1},\boldsymbol{x}_t] + \boldsymbol{b}_O),$$

$$\tilde{\boldsymbol{c}}_t = \tanh(\boldsymbol{W}_c[\boldsymbol{h}_{t-1},\boldsymbol{x}_t] + \boldsymbol{b}_c),$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t,$$

$$\boldsymbol{h}_t = \boldsymbol{O}_t \odot \tanh(\boldsymbol{c}_t),$$

where σ denotes the sigmoid activation function, tanh denotes the hyperbolic tangent function, and $\odot$ represents element-wise multiplication. $\boldsymbol{W}_f$ projects the concatenated vector onto the forget gate subspace, after the sigmoid non-linearity, the resulting gate vector $\boldsymbol{f}_t$ scales the previous cell state $\boldsymbol{c}_{t\text{-}1}$ element-wise, thereby determining what proportion of past information is retained. $\boldsymbol{W}_i$ performs an analogous projection for the input gate, its output $\boldsymbol{i}_t$ modulates the degree to which newly computed content will be written into memory during the current time step. $\boldsymbol{W}_c$ linearly transforms the same concatenated input to produce the candidate memory vector $\tilde{\boldsymbol{c}}_t$ (subsequently passed through a hyperbolic-tangent non-linearity). This vector embodies fresh contextual information that, when weighted by the input gate, updates the cell state. $\boldsymbol{W}_o$ maps the concatenated vector to the output-gate subspace, the resulting $\boldsymbol{o}_t$ gate regulates how much of the updated cell content is exposed through the hidden state $\boldsymbol{h}_t$, completing the recurrent computation. The concatenation $[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t]$ is linearly transformed by four distinct weight matrices $\boldsymbol{W}_f$, $\boldsymbol{W}_i$, $\boldsymbol{W}_o$,$\boldsymbol{W}_c$, and their associated bias vectors $\boldsymbol{b}_f$, $\boldsymbol{b}_i$, $\boldsymbol{b}_o$, $\boldsymbol{b}_c$, which together parameterise the forget gate, input gate, output gate and candidate memory pathway. Figure 3 shows the structure of the LSTM networks.

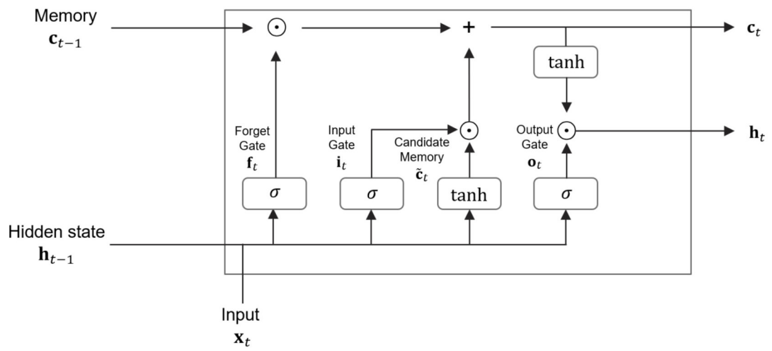


FIGURE 3: The structure of LSTM networks

Hybrid CNN-LSTM Model

Convolutional neural networks (CNNs) were originally developed for image recognition (Wang et al., 2019), yet their essential mechanism, namely the use of shared kernels to extract translation-invariant local features, is equally applicable to one-dimensional time-series data. In a one-dimensional convolution, the kernel slides along the temporal axis and performs weighted aggregation over short segments, enabling the network to learn automatically which combinations of variables and short-term patterns best explain forthcoming price movements (Montesinos López et al., 2022). Within a financial setting, the CNN can simultaneously capture the cross-sectional interactions among price, volume and the exogenous oil variable within a single trading day and detect local patterns across several days by stacking multiple convolutional layers.

However, CNN does not preserve longer-range sequential information and therefore cannot model trend persistence or lagged effects, which are hallmarks of recurrent neural networks. LSTM networks specialise in precisely these aspects. In contrast, the feature extraction function of CNN through the convolution operation can filter out noise information that is not beneficial for prediction in time series data, thereby solving the problem of introducing redundant features in LSTM. Hence, this research considers integrating one-dimensional CNN and LSTM to build a hybrid CNN-LSTM model, so as to make full use of the feature information extraction ability of CNN and the sensitivity of LSTM to time series data, and improve the effect of carbon price prediction.

The network of the hybrid CNN-LSTM model constructed in this study includes a one-dimensional convolutional layer (Conv1D), a max pooling layer (MaxPooling), an LSTM layer, and a fully connected layer (Dense), totalling 4 layers. The one-dimensional convolutional layer traverses the input carbon price data and performs convolution operations on local segments of the carbon price sequence using the weights within the convolutional kernel. This process generates an initial feature matrix, which exhibits stronger representational capability compared to the original time series matrix of carbon price data. The max pooling layer takes the feature matrix computed by the previous convolutional layer as input. It applies a pooling window that slides over the matrix sequence, extracting the maximum value within the window at each step. This operation produces a better feature matrix as output.

The entire CNN component can be regarded as a specialised data preprocessing or feature extraction structure. The convolutional layers extract local features from the input carbon price data using a sliding window mechanism, while the max pooling layers perform down-sampling on these local features to reduce the number of parameters and computational cost. This process further enhances the generalisation capability and training efficiency of the carbon price prediction model. The carbon price-related information processed by the CNN component is refined into inputs for the LSTM component, which is more sensitive to temporal sequence information. The LSTM layers model and integrate these local features along the temporal dimension, extracting higher-level feature representations. Finally, the fully connected layers map these feature representations to a single output, providing the carbon price prediction result. The overall structure of the hybrid CNN-LSTM model for carbon price prediction is shown in Figure 4.

Formally, each sample in the dataset is represented as a matrix $\boldsymbol{X}t$, $\in \mathbb{R}^{L \times d}$, where L denotes the look-back window length (number of timesteps) and d denotes the number of input features. Each timestep is represented as a feature vector $\boldsymbol{x}_i \in \mathbb{R}^d$, and the input matrix is constructed as:

$$X_t = \begin{bmatrix} \boldsymbol{x}_{t-L+1} \\ \boldsymbol{x}_{t-l+2} \\ \vdots \\ \boldsymbol{x}_t \end{bmatrix}$$

In a one-dimension CNN, when the kernel size is set to 1, each $\boldsymbol{x}_i$ is multiplied by a single weight vector $\boldsymbol{w}_j \in \mathbb{R}^d$ whose dimensionality is equal to the number of features. The output at timestep i for the j-th filter is given by:

$$c_{i,j} = \sigma(\boldsymbol{x}_i \cdot \mathbf{w}_j + b_j),$$

where σ is the activation function. By stacking the weights of the n filters into a matrix $\boldsymbol{w} \in \mathbb{R}^{n-d}$, this operation can be expressed compactly as:

$$\boldsymbol{C}_i = \sigma(\boldsymbol{W}\boldsymbol{x}_i + \boldsymbol{b}).$$

Although the convolution does not compress the temporal dimension, the subsequent pooling operation reduces the number of time steps. Let the matrix after the pooling layer be denoted as $\boldsymbol{C}'$. Following the procedure described in the next section for feeding data into the LSTM, $\boldsymbol{C}'$ replaces $\boldsymbol{X}_t$ as the input. In this way, the CNN and LSTM components can be connected and run jointly.

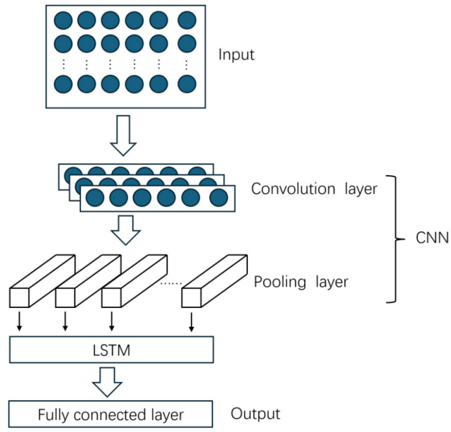


FIGURE 4: Structure of the hybrid CNN-LSTM model

RESULTS AND DISCUSSION

Performance of LSTM

Before model estimation, all variables are rescaled to the interval [0, 1] by the min-max normalisation technique in order to remove scale differences among features and to facilitate gradient convergence. The normalised input tensor comprises six time-series features: OHLC prices, trading volume, and the Brent crude oil prices, whereas the target variable is the next-day closing price.

Before the LSTM model is trained, the raw sequence must be reshaped so that each training instance contains a fixed history of past observations. Using a sliding-window scheme with a look-back length of L =10 trading days, the normalised feature vectors $\{\boldsymbol{X}_{t-L+1}, \boldsymbol{X}_{t-L} \ldots , \boldsymbol{X}_t\}$ are stacked into a matrix $\boldsymbol{X}_t$, and paired with the scalar label $y_{t+1}= P_{t+1}^{Close}$. This procedure converts the original unstructured time series into a supervised dataset while preserving local temporal dependencies, thereby providing the three-dimensional input tensor (batch, 10, 6) required by the LSTM architecture and enabling the network to learn mappings from multi-day patterns to next-day prices.

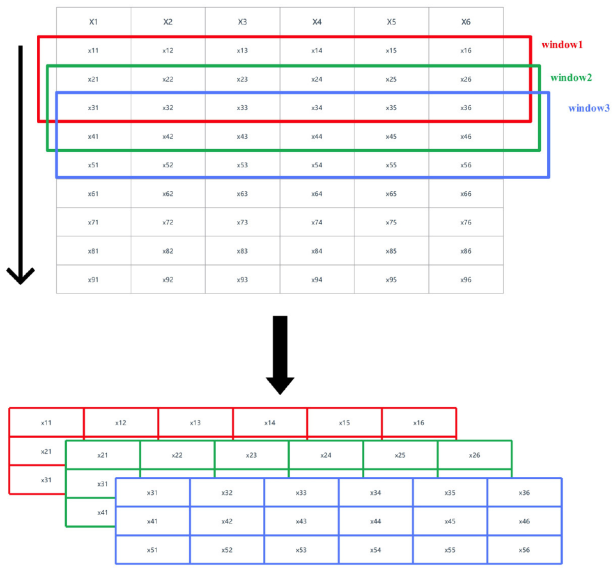


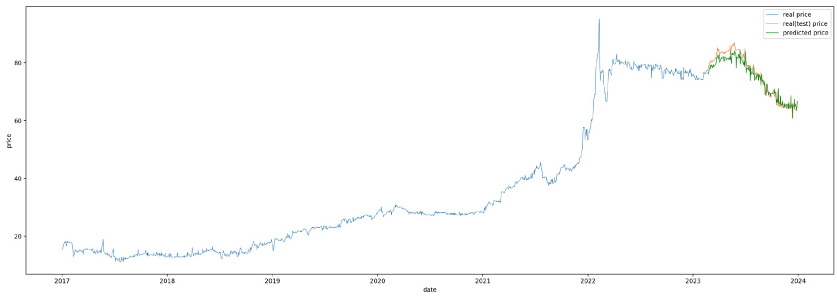
FIGURE 5: Rolling window operation

Figure 5 illustrates the sliding-window transformation applied to the six normalised features. For every trading day t ($t \geq 10$), the method concatenates the observations from the preceding 10 days into a matrix $\boldsymbol{X}_t$, each row corresponds to a single trading date, and each column to one of the six input features. Advancing the window by one day at a time yields an overlapping collection of supervisory pairs $\left(\boldsymbol{X}_{t,} P_{t+1}^{Close}\right)$, which jointly form the three-dimensional tensor (N, 10, 6), where N is the number of available samples after windowing. This representation preserves local temporal structure while recasting the original series as a standard supervised-learning dataset, enabling the LSTM to map multi-day patterns to the next-day closing price.

An LSTM model comprising a single recurrent layer with 64 hidden units and a linear output layer is trained for a maximum of 100 epochs. Hyperparameters, including the hidden-unit count, an initial learning rate of 0.005, and a batch size of 64, are selected from a preliminary random search that minimised validation mean squared error (MSE). Optimisation is performed with the Adam algorithm $\left(\beta_1 = 0.9, \beta_2 = 0.999\right)$ and the loss function is the MSE. Early stopping halts training if the validation loss fails to improve for 10 consecutive epochs, thereby mitigating over-fitting.

The data are partitioned by a rolling TimeSeriesSplit procedure with 5 folds and a fixed test horizon of 200 trading days per fold. Figure 6 contrasts the observed closing prices with the LSTM model forecasts. Although the network reproduces the broad trajectory of the carbon-price series, there are visible discrepancies around local turning points, motivating the development of the hybrid CNN-LSTM model.

In order to obtain a more accurate performance measure, this research performed 10 times on the fitting LSTM model and finally obtained the average of all testing results as the performance measure. This study uses MSE and mean absolute error (MAE) as performance metrics, and the results indicate the MAE of the LSTM as 1.57 and MSE as 3.67.



Note: The blue and orange curves depict the training and testing of the observed series, respectively, whereas the green curve represents the predicted series based on the LSTM model.

FIGURE 6: Observed and predicted closing prices using the LSTM model

Performance of Hybrid CNN-LSTM Model

In the hybrid CNN-LSTM model configuration, all six normalised input series are first arranged, via a sliding window of length $L = 10$ trading days, into tensors of shape (batch, 10, 6). The dataset is then partitioned with the five-fold rolling scheme defined in the cross-validation section, each fold holding a fixed test horizon of 200 observations. The convolutional front-end consists of a single one-dimensional layer with 32 kernels of size 1×6 and hyperbolic-tangent activation, followed by a max-pooling layer with window size two. The pooled feature maps are passed to an LSTM layer containing 64 hidden units, and the final hidden state is projected to a scalar output through a fully connected linear neuron. All network hyper-parameters, including kernel count, kernel size, hidden-unit count, batch size, learning rate (*lr*) and early-stopping patience, are summarised in Table 1. Optimisation is carried out with the Adam algorithm (initial *lr* = 1×10^{-3}) and the MSE loss. During training the convolutional kernels slide along the temporal dimension defined by the window, extracting local patterns that the LSTM subsequently integrates into a long-range representation. Weight updates are performed by back-propagation through time until the validation loss fails to improve for 10 consecutive epochs or the ceiling of 100 epochs is reached.

TABLE 1
Parameter settings of the hybrid CNN-LSTM model

Parameter	Value
Convolution layer filters	32
Convolution layer kernel_size	1 × 6
Convolution layer activation function	tanh
Convolution layer padding	same
Pooling layer pool_size	2
Pooling layer padding	same
Number of hidden units in LSTM layer	64
LSTM layer activation function	sigmoid
Time_step	10
Batch_size	64
Learning rate	0.001
Optimizer	Adam
Loss function	MSE
Epochs	100

Figure 7 plots the observed closing prices against the CNN-LSTM model forecasts on the test folds. The hybrid architecture reproduces both the medium-term trend and most short-term fluctuations of the carbon-price series, yielding lower MAE and MSE than the standalone LSTM. Residual deviations remain around abrupt inflection points. The predicted prices closely follow the actual test closing prices, indicating that the model generalises well without significant overfitting. Additionally, the model effectively captures trend changes, maintaining a consistent error margin between training and test data. It indicates that the model fits well with no evidence of overfitting. The model was trained and evaluated ten times, and the average of all testing results was used as the performance measure. The results show that the MAE of the hybrid CNN-LSTM model is 1.23, and the MSE is 2.33.

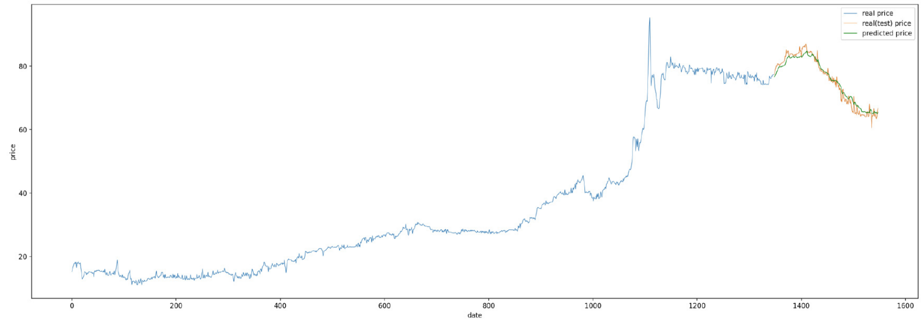


FIGURE 7: Observed and predicted closing prices using the hybrid CNN-LSTM model

Model Comparison

Both the LSTM and the hybrid CNN-LSTM models are fitted and used to predict the carbon emission allowance closing prices in Guangdong province. The results show that the hybrid CNN-LSTM model has better performance than the standalone LSTM model.

Figures 8 and 9 show the prediction performance of different models and the observed data on the training and testing sets, respectively. For the training set (see Figure 8), it is observed that the fitted series based on the hybrid CNN-LSTM model fluctuate around the observed series, while the fitted series based on the standalone LSTM model seem to deviate from the observed series and show overestimation of the prices. For the testing set (see Figure 9), the red line represents the observed series, the green line represents the predicted series based on the hybrid CNN-LSTM model, and the blue line represents the predicted series based on the standalone LSTM model. Again, it is observed that the hybrid CNN-LSTM model performs better in terms of prediction performance than its competitor.

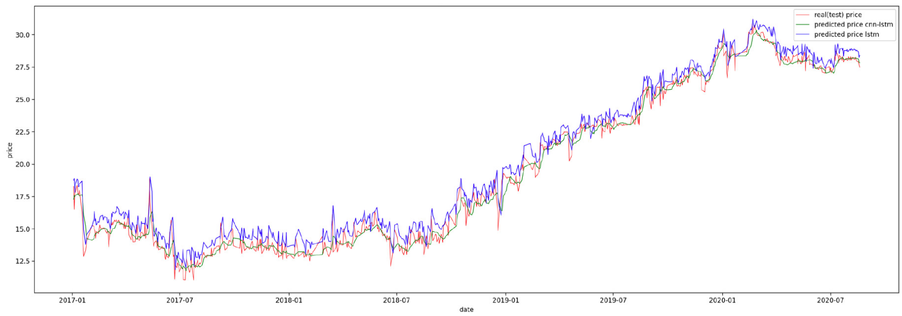


FIGURE 8: Observed and fitted closing prices from the standalone LSTM and hybrid CNN-LSTM models on the training set

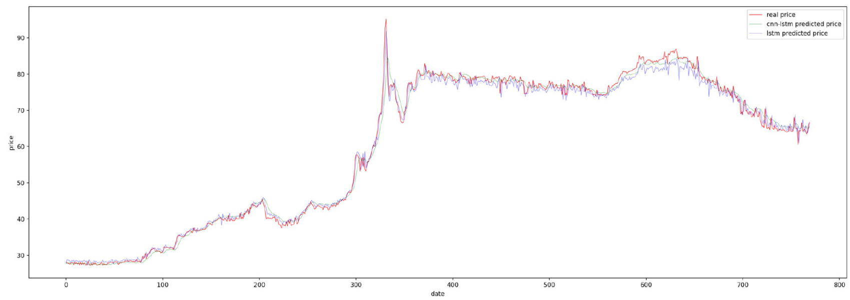


FIGURE 9: Observed and predicted closing prices from the standalone LSTM and hybrid CNN-LSTM models on the testing set

Because each model is fitted ten times and the number of splits (*n*-split parameter) is set to 5 for each fitting, 50 sets of performance metrics are obtained for each model, and a box plot can be drawn for comparison, the results are shown in Figure 10. The box plots show that the MAE and MSE of the obtained hybrid CNN-LSTM model, as well as the quantiles, are all smaller than those of the LSTM model.

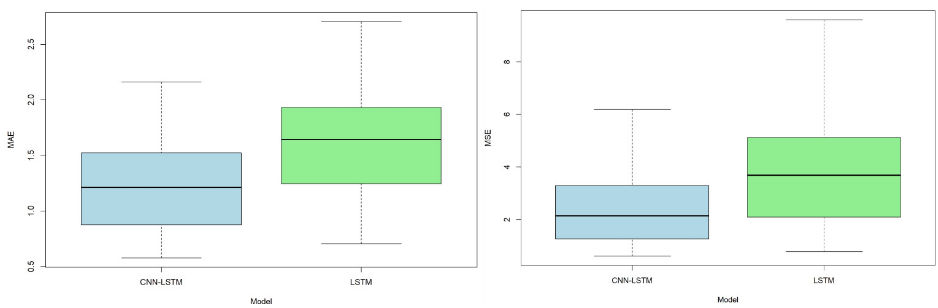


FIGURE 10: Comparison of MAE (a) and MSE (b) values between the CNN-LSTM and LSTM models

Exogenous Variables Identification

We next examine which feature is important for price prediction among the selected features. We adopted a "feature ablation" method. Specifically, we first train a baseline model using all features and then retrain the model with one feature removed under the same configuration, comparing the performance difference to infer the importance of that feature. The core of this method lies in the removing or masking of individual feature columns and comparing predictive performance. The procedure is illustrated as follows:

1. Full model training: Train a CNN-LSTM model using all six features, and record its performance metrics, namely MAE and MSE.
2. Removing (Masking) a feature: For any feature X_i or a combination of certain features of interest, construct a new dataset in which that feature or the combination of features is removed or masked in both the training and testing sets, while other features remain unchanged.
3. Retraining the model: Under the same network architecture, hyperparameter settings and random seed, retrain the CNN-LSTM model on the dataset with feature or features removed, and then perform predictions on the testing set.
4. Comparison of performance metrics: Compare the predictive performance of the model without a specific feature or combination of features with that of the full model. A larger performance drop indicates greater importance of that feature.

In addition to removing each feature separately, the OHLC prices are also removed and retained as a whole. The testing results of the CNN-LSTM model performance with different combinations of exogenous variables are presented in Table 2.

TABLE 2
CNN-LSTM model performance under different combinations of exogenous variables

Feature	MAE	MSE
Full model with all variables	1.23	2.33
All variables without opening prices (no open)	1.25	2.39
All variables without the highest price (no high)	1.42	2.95
All variables without the lowest price (no low)	1.35	2.83
All variables without volume (no volume)	1.30	2.58
All variables without oil prices (no oil price)	1.28	2.50
All variables without OHLC (no OHLC)	1.41	2.97
All variables without volume and oil prices (no volume and oil price)	1.39	2.74

From Table 2, it is observed that when the full model (including all features) is used, both MAE and MSE remain relatively low, indicating that the combined information from all features contributes to more accurate predictions. When certain individual features or feature combinations are removed, the predictive errors increase to varying degrees, suggesting that these features provide useful information for forecasting performance. Removing a single feature generally causes both MAE and MSE to rise compared with the baseline model, yet the extent of influence varies. Notably, excluding the "high" feature results in the most pronounced increase in MAE, implying that this feature may play an important role in capturing price dynamics and volatility patterns.

Furthermore, the "no volume" and "no oil price" models also lead to increased errors, although to a slightly lesser extent than the "no high" model. This suggests that trading volume and oil price contain additional predictive information, but some of their information may be partially compensated for by other features. When OHLC features are entirely excluded, the model produces the highest MSE, indicating that the model suffers the largest decline in performance due to the removal of essential price-based information. This finding also highlights that these price-related variables complement one another and jointly provide critical market signals for time series forecasting. Simultaneously removing both volume and oil price further increases the prediction error, although it remains slightly lower than the "no OHLC" model, suggesting that in certain contexts, trading volume and oil price may be interconnected, their joint absence deprives the model of important auxiliary information, thereby elevating the prediction error. Overall, the results indicate that removing key information leads to a deterioration in predictive performance with the "high" feature appearing to have the strongest individual contribution, while "volume" and "oil price" features also provide meaningful but partially substitutable information.

Feature Selection based on QLIKE Loss Function

Earlier experiments primarily adopted standard evaluation metrics such as MAE and MSE to assess predictive performance. However, in scenarios where sensitivity to price fluctuations or risk measurement is critical, these metrics may not adequately reflect the model's ability to capture extreme errors and nonlinear effects. Therefore, the QLIKE loss function is introduced as a robustness analysis to re-examine feature importance from a volatility-sensitive perspective.

In this study, a feature exclusion strategy is employed to evaluate the contribution of each feature to the predictive task, with QLIKE serving as the primary evaluation metric. Since QLIKE is widely used in financial applications such as volatility forecasting, and is particularly sensitive to extreme values and price fluctuations, its adoption helps emphasise the financial relevance of the model evaluation. For price-related features (OHLC), these features are closely interconnected from a financial theory perspective and typically used collectively to depict the full price range and volatility within a trading period. By treating these four features as an integrated set, either removing or retaining them together, the study captures the essence of price-related information more comprehensively.

TABLE 3
CNN-LSTM model performance under different combinations of exogenous variables using the QLIKE loss function

Feature	QLIKE
Full model	0.0217
No OHLC	0.0280
No volume	0.0220
No oil price	0.0240
No volume and oil price	0.0250

The QLIKE results for the Guangdong dataset are summarised in Table 3. When all features are included (the full model), the QLIKE is 0.0217, which is the lowest across all scenarios. This indicates that the combined information from price-related features (OHLC), trading volume and oil price yields the most accurate forecasts and the smallest penalties for extreme errors. After removing individual or combined features, the following observations are obtained.

1. Removing all price-related features causes a substantial increase in QLIKE from 0.0217 to 0.028. Since OHLC features provide the most direct information about market trends and price ranges, their absence forces the model to rely solely on other features (e.g., volume and oil price), thereby reducing its ability to capture price fluctuations accurately.
2. Removing trading volume results in a modest increase to 0.022, representing the smallest increment compared with the full model. In this dataset, the absence of volume does not substantially impact extreme-error penalty, possibly because price features (OHLC) and oil price already supply sufficient market information, or because volume information partially overlaps with these features.
3. Excluding oil price elevates QLIKE to 0.024, an increase of 0.0023 relative to the full model. Although the impact is smaller than that observed when removing OHLC, the increase is more pronounced than that resulting from removing volume alone, highlighting the importance of oil price in explaining fluctuations, particularly in contexts where macroeconomic or energy-related factors influence market dynamics.
4. When both trading volume and oil price are excluded, QLIKE rises to 0.025, which is 0.0033 higher than the full model. Compared with "no volume" (0.022) or "no oil price" (0.024) models, the simultaneous removal of both variables produces a greater negative impact, although it remains less severe than the removal of OHLC. This finding further reinforces the central role of price features in the forecasting task. Meanwhile, the combined absence of volume and oil price suggests some degree of redundancy or substitution effect between them.

Overall, the results indicate that OHLC features constitute the most critical information source for price prediction, and removing these core price signals leads to a marked increase in extreme-error penalties. Oil prices also play an important role, as evidenced by the noticeable increase in QLIKE when they are excluded. In contrast, trading volume appears relatively less influential under the present conditions, although the simultaneous removal of both volume and oil price still produces a larger deterioration in performance than removing either variable alone. Importantly, the conclusions obtained from the QLIKE analysis remain consistent with those derived from MAE and MSE, thereby confirming the robustness of the feature importance findings across different evaluation criteria.

CONCLUSION AND RECOMMENDATIONS

To more accurately forecast the complex and highly volatile short-term fluctuations in carbon trading markets, this study proposed a hybrid machine learning approach that integrates CNNs to capture local high-frequency price patterns and LSTM networks to model long-range temporal dependencies. Using data spanning 3 January 2017 to 29 December 2023, from the Guangdong ETS as the empirical sample, we compare the performance of the LSTM model and the hybrid CNN-LSTM model. The key findings are as follows:

1. The hybrid CNN-LSTM model consistently outperforms the LSTM model in rolling expanding-window tests, maintaining its advantage throughout the Guangdong ETS sample period.
2. Feature ablation and QLIKE diagnostics indicate that intraday OHLC quotes contain the most critical predictive signals; their removal leads to a sharp decline in forecasting accuracy. Among them, the daily high price emerges as particularly important.
3. Brent crude oil prices and trading volume contribute incremental value to the model's predictive performance. While these exogenous variables enhance forecasting accuracy, their importance remains secondary to the price-related information embedded in the carbon contracts themselves.
4. The model demonstrates the ability to track turning points and price jumps without overfitting, suggesting that combining convolutional feature extractors with gated recurrent units is well-suited for markets characterised by abrupt regulatory shocks and compliance-driven seasonality.

These findings carry important practical implications for market oversight and participant strategy. Based on the empirical results, we propose three targeted policy recommendations to enhance market stability and efficiency: First, regulators should upgrade current surveillance systems by integrating deep learning-based early warning mechanisms. Given the hybrid model's ability to capture sudden price jumps, authorities can utilise such tools to assess whether allowance prices remain aligned with marginal abatement costs and to monitor real-time volatility risks. When the model forecasts a price surge exceeding a safety threshold, ex-ante interventions, such as adjusting auction volumes or triggering price containment reserves, can be executed proactively rather than reactively. Second, exchanges should prioritise the transparency and granularity of intraday trading data. Our feature ablation results confirm that intraday OHLC are more critical predictive signals than external macro variables. Therefore, policymakers should enforce high-frequency data disclosure standards to reduce information asymmetry and improve overall price discovery efficiency. Third, compliance entities are encouraged to transition from passive compliance to active carbon asset management. Instead of concentrating purchases near deadlines, firms should leverage high-precision forecasts to execute "buy-low" algorithmic trading strategies. This approach not only soothes out market liquidity demand but also provides a reliable basis for timing allowance purchases, thereby minimising long-term hedging and compliance expenses. Finally, for financial institutions, such enhanced short-term forecasts serve as a foundation for more accurate valuation of carbon-linked derivatives and robust portfolio risk management.

Several limitations point to promising directions for future research. The current analysis focuses on next-day forecasts; extending the framework to intraday or week-ahead prediction windows could benefit both market makers and long-term hedgers. Incorporating additional high-frequency energy and macro-policy news feeds, potentially via attention mechanisms, may further enhance predictive performance. Finally, exploring explainable AI techniques, such as SHAP values, could provide deeper insights into how specific market signals are encoded within convolutional filters and LSTM gates, thereby improving model transparency for both practitioners and policymakers.

REFERENCES

Adler, A. I., & Painsky, A. (2022). Feature importance in gradient boosting trees with cross-validation feature selection. *Entropy, 24*(5), 687. https://doi.org/10.3390/e24050687

Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., Muneer, A., Sumiea, E. H., Alqushaibi, A., & Ragab, M. G. (2024). RNN-LSTM: From applications to modeling techniques and beyond-systematic review. *Journal of King Saud University-Computer and Information Sciences, 36*(5), 102068. https://doi.org/10.1016/j.jksuci.2024.102068

Alsharekh, M. F., Habib, S., Dewi, D. A., Albattah, W., Islam, M., & Albahli, S. (2022). Improving the efficiency of multistep short-term electricity load forecasting via R-CNN with ML-LSTM. *Sensors, 22*(18), 6913. https://doi.org/10.3390/s22186913

Bram, A. K., Ofori, C., Mangudhla, T., & Nuta, A. C. (2025). Exploring currency interdependence in West Africa: A time-varying parameter vector autoregression analysis. *The Journal of Risk Finance, 26*(2), 320–344. https://doi.org/10.1108/JRF-10-2024-0310

Buratto, W. G., Muniz, R. N., Nied, A., Barros, C. F. D. O., Cardoso, R., & Gonzalez, G. V. (2024). Wavelet CNN-LSTM time series forecasting of electricity power generation considering biomass thermal systems. *IET Generation, Transmission & Distribution, 18*(21), 3437–3451. https://doi.org/10.1049/gtd2.13292

Chang, K., Chen, R., & Chevallier, J. (2018). Market fragmentation, liquidity measures and improvement perspectives from China's emissions trading scheme pilots. *Energy Economics, 75*, 249–260. https://doi.org/10.1016/j.eneco.2018.07.010

Chen, T., Ye, J., Zhou, Y., Yu, Q., Wang, S., & Li, G. (2025). Prediction of carbon emission rights trading prices based on the CNN-LSTM Model in the context of carbon peak: Taking Guangdong Province as an example. *Concurrency and Computation: Practice and Experience, 37*(12–14), e70121. https://doi.org/10.1002/cpe.70121

Dhanka, S., Sharma, A., Kumar, A. Maini, S., & Vundavilli, H. (2025). Advancements in hybrid machine learning models for biomedical disease classification using integration of hyperparameter-tuning and feature selection methodologies: A comprehensive review. *Archives of Computational Methods in Engineering, 33*, 289–324. https://doi.org/10.1007/s11831-025-10309-5

Dong, X., & Zhang, J. F. (2024). Heterogeneity of regional carbon emission markets in China: Evidence from multidimensional determinants. *Energy Economics, 138*, 107835. https://doi.org/10.1016/j.eneco.2024.107835

El Hokayem, J., Jamali, I., & Hejase, A. (2024). A forecasting model for oil prices using a large set of economic indicators. *Journal of Forecasting, 43*(5), 1615–1624. https://doi.org/10.1002/for.3087

Escribano, A., Koczar, M. W., Jareño, F., & Esparcia, C. (2023). Shock transmission between crude oil prices and stock markets. *Resources Policy, 83*, 103754. https://doi.org/10.1016/j.resourpol.2023.103754

Ghimire, S., Deo, R. C., Raj, N., & Mi, J. (2019). Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy, 253*, 113541. https://doi.org/10.1016/j.apenergy.2019.113541

Goh, C. W., Chai, J., Rahman, A., & Ong, W. E. (2024). Crude palm oil price prediction using simulated annealing-based support vector regression (SA-SVR). *Asian Academy of Management Journal of Accounting and Finance, 20*(1), 305–333. https://doi.org/10.21315/aamjaf2024.20.1.10

He, J., Ng, K. H., Shelton, P., & David, A. (2026). Modelling volatility and return based on a two-stage Log-BiACARR framework and intraday information: Evidence from Guangdong and Hubei carbon emissions trading markets. *Physica A: Statistical Mechanics and its Applications, 681*, 131097. https://doi.org/10.1016/j.physa.2025.131097

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, Z., & Zhang, W. (2024). Forecasting carbon prices in China's pilot carbon market: A multi-source information approach with conditional generative adversarial networks. *Journal of Environmental Management, 359*, 120967. https://doi.org/10.1016/j.jenvman.2024.120967

Jia, S., Tan, Z., & Li, C. (2025). Carbon prices forecasting based on sliding time window and improved support vector regression. *Computing, 107*(2), 53. https://doi.org/10.1007/s00607-024-01404-9

Jin, Y., Acquah, M. A., Seo, M., & Han, S. (2022). Short-term electric load prediction using transfer learning with interval estimate adjustment. *Energy and Buildings, 258*, 111846. https://doi.org/10.1016/j.enbuild.2022.111846

Jung, Y. (2018). Multiple predicting K-fold cross-validation for model selection. *Journal of Nonparametric Statistics, 30*(1), 197–215. https://doi.org/10.1080/10485252.2017.1404598

Krokida, S. I., Lambertides, N., Savva, C. S., & Tsouknidis, D. A. (2020). The effects of oil price shocks on the prices of EU emission trading system and European stock returns. *The European Journal of Finance, 26*(1), 1–13. https://doi.org/10.1080/1351847X.2019.1637358

Li, H., & Lei, M. (2018). The influencing factors of China carbon price: A study based on carbon trading market in Hubei province. *IOP Conference Series: Earth and Environmental Science*, 121(5), 052073. https://doi.org/10.1088/1755-1315/121/5/052073

Li, J., Yao, Y., & Wang, X. (2022). The first compliance cycle of China's national emissions trading scheme: Insights and implications. *Carbon Neutrality, 1*(1), 34. https://doi.org/10.1007/s43979-022-00035-3

Lin, B., & Jia, Z. (2019). Impacts of carbon price level in carbon emission trading market. *Applied Energy, 239*, 157–170. https://doi.org/10.1016/j.apenergy.2019.01.194

Ling, M., & Cao, G. (2024). Carbon trading price forecasting based on parameter optimization VMD and deep network CNN-LSTM model. *International Journal of Financial Engineering, 11*(1), 2450002. https://doi.org/10.1142/S2424786324500026

Liu, X., Zhou, X., Zhu, B., & Wang, P. (2020). Measuring the efficiency of China's carbon market: A comparison between efficient and fractal market hypotheses. *Journal of Cleaner Production, 271*, 122885. https://doi.org/10.1016/j.jclepro.2020.122885

Liu, X., & Jin, Z. (2020). An analysis of the interactions between electricity, fossil fuel and carbon market prices in Guangdong, China. *Energy for Sustainable Development, 55*, 82–94. https://doi.org/10.1016/j.esd.2020.01.008

Lu, X., Yan, H., He, P., & Apergis, N. (2025). Multifractal relationship between decomposed oil price shocks and trading volume. *Humanities and Social Sciences Communications, 12*(1), 1–16. https://doi.org/10.1057/s41599-025-05227-7

Luo, E., Shi, L., Liu, J., Wu, Z., Duan, G., Zeng, L., & Huang, T. (2024). MNPM: Research on metabolic neural network prediction model for predicting carbon emission accuracy. *Environmental Research Communications, 6*(11), 115034. https://doi.org/10.1088/2515-7620/ad9086

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Convolutional neural networks. In *Multivariate statistical machine learning methods for genomic prediction* (pp. 533–577). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0_13

Otto, P., & Schmid, W. (2023). A general framework for spatial GARCH models. *Statistical Papers, 64*(5), 1721–1747. https://doi.org/10.1007/s00362-022-01357-1

Perifanis, T., & Dagoumas, A. (2021). Crude oil price determinants and multi-sectoral effects: A review. *Energy Sources, Part B: Economics, Planning, and Policy, 16*(9), 787–860. https://doi.org/10.1080/15567249.2021.1922956

Rashid, M., Kamruzzaman, J., Imam, T., Wibowo, S., & Gordon, S. (2022). A tree-based stacking ensemble technique with feature selection for network intrusion detection. *Applied Intelligence, 52*(9), 9768–9781. https://doi.org/10.1007/s10489-021-02968-1

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences, 2*(7), 1308. https://doi.org/10.1007/s42452-020-3060-1

Shah, J., Vaidya, D., & Shah, M. (2022). A comprehensive review on multiple hybrid deep learning approaches for stock prediction. *Intelligent Systems with Applications, 16*, 200111. https://doi.org/10.1016/j.iswa.2022.200111

Shih, S. Y., Sun, F. K., & Lee, H. Y. (2019). Temporal pattern attention for multivariate time series forecasting. *Machine Learning, 108*, 1421–1441. https://doi.org/10.1007/s10994-019-05815-0

Shin, Y., Kim, M., Cai, C., & Chang, Q. (2019). FARVaR: Functional Autoregressive Value-at-Risk. *Journal of Financial Econometrics, 17*, 284–337. https://doi.org/10.1093/jjfinec/nby031

Song, X., Zhang, W., Ge, Z., Huang, S., Huang, Y., & Xiong, S. (2022). A study of the influencing factors on the carbon emission trading price in China based on the improved gray relational analysis model. *Sustainability, 14*(13), 8002. https://doi.org/10.3390/su14138002

Srinivasan, S., Kumarasamy, S., Andreadakis, Z. E., & Lind, P. G. (2023). Artificial intelligence and mathematical models of power grids driven by renewable energy sources: A survey. *Energies, 16*(14), 5383. https://doi.org/10.3390/en16145383

Tan, K., Wu, Y., Xu, F., Ji, X., & Li, C. (2024). Modeling and analyzing carbon emission market volatility and impact: Evidence from Guangdong province, China. *Systems, 12*(11), 458. https://doi.org/10.3390/systems12110458

Tange, R. I., Rasmussen, M. A., Taira, E., & Bro, R. (2017). Benchmarking support vector regression against partial least squares regression and artificial neural network: Effect of sample size on model performance. *Journal of Near Infrared Spectroscopy, 25*(6), 381–390. https://doi.org/10.1177/0967033517734945

Tsai, I. C. (2024). Fossil energy risk exposure of the UK electricity system: The moderating role of electricity generation mixes and energy source. *Energy Policy, 188*, 114065. https://doi.org/10.1016/j.enpol.2024.114065

Villoria-Sáez, P., Tam, V. W., del Río Merino, M., Arrebola, C. V., & Wang, X. (2016). Effectiveness of greenhouse-gas emission trading schemes implementation: A review on legislations. *Journal of Cleaner Production, 127*, 49–58. https://doi.org/10.1016/j.jclepro.2016.03.148

Wang, W., Yang, Y., Wang, X., Wang, W., & Li, J. (2019). Development of convolutional neural network and its application in image classification: A survey. *Optical Engineering, 58*(4), 040901–040901. https://doi.org/10.1117/1.OE.58.4.040901

Wang, W., Zhao, X., Zhang, Q., Fu, C., & Xie, P. (2022). Auction mechanism design of the Chinese national carbon market for carbon neutralization. *Chinese Journal of Population, Resources and Environment, 20*(2), 115–124. https://doi.org/10.1016/j.cjpre.2022.06.002

Wang, Z., Xing, T., & Wang, X. (2025). Economic uncertainty and stock market asymmetric volatility: Analysis based on the asymmetric GARCH-MIDAS model. *International Journal of Emerging Markets, 20*(8), 3488–3511. https://doi.org/10.1108/IJOEM-05-2023-0841

Zhang, W., & Wu, Z. (2022). Optimal hybrid framework for carbon price forecasting using time series analysis and least squares support vector machine. *Journal of Forecasting, 41*(3), 615–632. https://doi.org/10.1002/for.2831

Zhao, Y., Liu, L., Wang, A., & Liu, M. (2023). A novel deep learning based forecasting model for carbon emissions trading: A comparative analysis of regional markets. *Solar Energy, 262*, 111863. https://doi.org/10.1016/j.solener.2023.111863

Zhou, F., Huang, Z., & Zhang, C. (2022). Carbon price forecasting based on CEEMDAN and LSTM. *Applied Energy, 311*, 118601. https://doi.org/10.1016/j.apenergy.2022.118601

Zhou, K., & Li, Y. (2019). Influencing factors and fluctuation characteristics of China's carbon emission trading price. *Physica A: Statistical Mechanics and its Applications, 524*, 459–474. https://doi.org/10.1016/j.physa.2019.04.249