Journal of Industrial and Management Optimization

JIMO, 22 (5): 2428–2450.
DOI: 10.3934/jimo.2026089
Received: 12 November 2025
Revised: 04 March 2026
Accepted: 14 April 2026
Published: 23 April 2026

https://www.aimspress.com/journal/jimo

Research article

ACVPNet: An air cargo volume prediction model for China based on periodic structure and multi-dimensional feature interaction

Xiaoyuan Cheng*

SWUFE-UD Institute of Data Science at SWUFE, Southwestern University of Finance and Economics, Chengdu, Sichuan, 611130, China

* **Correspondence:** Email: cxycxy0224@163.com; Tel: +86-17686619523.

Abstract: Accurate air cargo volume (ACV) forecasting is critical for operational management and global supply chain decision optimization in aviation logistics. Existing methods present limitations in capturing the multi-period seasonal dependencies of ACV sequences and high-order nonlinear interactions between multi-source influencing factors, resulting in insufficient forecasting accuracy under small-sample monthly data scenarios. To address these gaps, this study proposes ACVPNet, a lightweight multi-layer perceptron (MLP)-based ACV forecasting framework integrating multi-period time dependency modeling and multivariate feature interaction learning. The proposed model adopts a dual-branch prediction architecture that independently models cyclical and trend components. By incorporating specialized modules to jointly capture intra-period and inter-period dependencies, it effectively integrates cross-feature information, thereby enhancing both the flexibility and interpretability of temporal representation learning. Specifically, ACVPNet integrates two core modules: the temporal local multi-layer perceptron (TLM), which extracts nonlinear temporal dependencies within and across cycles through localized patch-based learning, and the feature interaction MLP (FIM), which captures inter-variable correlations among economic, consumption, and energy indicators. Together, these components enable ACVPNet to learn complex periodic patterns and multi-dimensional relationships from heterogeneous time series data. Experiments based on the CEIC historical China ACV dataset demonstrate the superiority of the proposed model over three representative baseline models: LSTM, Informer, and SARIMA. In the one-year-ahead ACV forecasting task, ACVPNet achieved an RMSE of 0.083 million tons, representing reductions of 37.6%, 49.7%, and 39.4% compared with the three baseline models, thereby demonstrating a

significant performance advantage. The results confirm that ACVPNet provides high-precision and temporally consistent ACV forecasts, offering valuable insights for aviation market analysis, capacity planning, and policy formulation.

Keywords: deep learning; periodic feature decomposition; 2D temporal characteristics; feature interaction; air cargo volume prediction
Mathematics Subject Classification: 90B06, 91A35

1. Introduction

The accurate forecasting of air cargo volume (ACV) carries significant operational and strategic importance in the aviation sector [1,2]. Air cargo not only represents a vital component of global trade flows but also serves as a key revenue stream for airlines and freight forwarders, especially in an era increasingly dominated by e-commerce, high-value goods, and time-sensitive logistics [3,4]. Precise ACV predictions enable carriers and logistics providers to calibrate capacity, optimize pricing, allocate handling resources and anticipate seasonal or cyclical shifts in demand. In parallel, regulatory authorities and airport operators utilize forecasted volumes for infrastructure planning, terminal expansions, slot allocation, and environmental impact assessments [5]. Consequently, developing high-accuracy ACV prediction models has become indispensable for enhancing supply-chain resilience, reducing inventory risk, and supporting efficient network management [6,7].

Currently, ACV sequence prediction methods can be broadly categorized into two categories [8]. The first category comprises model-driven, explicit statistical or signal-decomposition algorithms. Classical autoregressive integrated moving average models (ARIMA) and their seasonal variant (SARIMA) remain widely used for ACV and related aviation time series because of their interpretability and low data requirements [8–11]; SARIMA models are commonly applied to capture deterministic trend and seasonal components in cargo throughput. To address nonstationary and multiscale behavior, empirical mode decomposition (EMD) or its recursive variants have been combined with SARIMA/ARIMA to decompose the original series into intrinsic mode functions (IMFs) and model each component separately (EMD–SARIMA), which often reduces forecasting error for complex, seasonal traffic series [12,13]. Additionally, the extended algorithm, which combines the SARIMA model with the volatility module (e.g., SARIMA-EGARCH), can effectively handle sudden fluctuations in freight volume during unexpected events [7,14]. Grey forecasting and its corresponding combination algorithms have also garnered attention in research on predicting air passenger traffic [15]. Although model-driven approaches offer advantages such as high analytical transparency and moderate data requirements, their strict model assumptions often constrain their capacity to characterize the inherent high-order nonlinear features and abrupt state transitions embedded in modern ACV data [8].

The second category comprises data-driven and deep-learning methods that learn nonlinear mappings directly from historical observations and exogenous covariates [8,16,17]. Early neural approaches—multi-layer perceptron (MLP), support vector machine (SVM) and random forest models—have been applied to air transport demand problems with multiple input features, demonstrating improved fit over purely linear baselines when sufficient training data are available [6,18,19]. Recurrent architectures, notably long short-term memory (LSTM) networks [20], have been widely adopted because their gating mechanisms mitigate vanishing gradients and enable the capture of long-range

temporal dependencies in passenger and freight series. Beyond pure RNNs, hybrid architectures such as ARIMA–LSTM models and decomposition-augmented deep frameworks have been developed to integrate linear and nonlinear modeling paradigms [21,22]. These methods leverage the complementary strengths of both approaches: classical statistical models effectively capture both deterministic and stochastic patterns, while deep neural networks excel at representing complex, nonlinear dynamics. By combining these advantages, hybrid models have demonstrated notable improvements in forecasting accuracy across numerous empirical studies [23].

However, it is essential to note that ACV, as a key indicator of macroeconomic activity, is inherently influenced by the interplay and interaction of multiple periodic factors [7]. Such multi-periodic characteristics are widely observed across various domains—for instance, meteorological observations exhibit both diurnal and annual cycles. At the same time, electric load demand often fluctuates with weekly and quarterly alternations. Similarly, in the case of ACV, the cargo volume at any given time is not only closely correlated with adjacent temporal observations (intra-period dependency) but also significantly associated with corresponding positions in neighboring cycles (inter-period dependency). The intertwined nature of these dependencies significantly increases modeling complexity and poses substantial challenges for accurate forecasting. Nevertheless, existing research rarely addresses this limitation explicitly.

Furthermore, most existing multivariate modeling studies on ACV-related factors do not sufficiently capture the nonlinear dependencies and dynamic interactions among influencing variables. Inadequate representation of these characteristic factors restricts the model's capacity to fully characterize the coupled relationships among economic, transportation, and trade indicators, thereby limiting both predictive accuracy and interpretability.

To enhance the accuracy and interpretability of ACV forecasting, this paper proposes a novel deep learning architecture with periodic feature interaction, termed ACVPNet. The main contributions of this study are summarized as follows:

(1) We develop a new ACV prediction framework that explicitly integrates periodic structural information and multi-feature interaction. The proposed ACVPNet comprises three key components: periodic data reconstruction, local temporal feature extraction, and cross-feature interaction modeling.

(2) A periodic reshaping (PR) method is introduced to transform one-dimensional time series into a two-dimensional periodic matrix, based on cycle lengths obtained via the Fast Fourier Transform (FFT). This reconstruction process enables the model to disentangle intra- and inter-period dependencies, facilitating the learning of complex periodic patterns underlying air cargo fluctuations.

(3) To effectively capture temporal dependencies within the reshaped two-dimensional representation, we design a temporal local MLP (TLM) module, which replaces traditional linear convolution kernels with nonlinear MLP-based mappings. This design enhances the model's ability to learn both intra-patch and inter-patch nonlinear correlations, thereby improving its flexibility and robustness.

(4) Furthermore, a feature interaction MLP (FIM) module is proposed to model the dependencies among multiple influencing factors—such as economic indicators, transportation capacity, and trade volumes. This component enables efficient cross-dimensional information fusion, enhancing the overall representational power of the network.

(5) The unique architecture of ACVPNet enables joint modeling of multi-period temporal dynamics and cross-feature interactions, providing a flexible and interpretable framework for high-precision ACV forecasting under complex economic environments.

The remainder of this paper is organized as follows: Section 2 describes the dataset and preprocessing procedures; Section 3 details the proposed ACV prediction model and its

implementation flow; Section 4 presents the experimental results and demonstrates the superiority of the proposed approach; and Section 5 concludes the paper and summarizes the key findings.

2. Dataset

2.1. Experimental dataset

Considering that the temporal dynamics of ACV are influenced by multiple correlated indicators, a comprehensive exploration of multivariate interdependencies is essential to enhance modeling accuracy. To this end, 13 influencing factors were initially selected from three perspectives—macroeconomic conditions, consumption, and energy production—as outlined in Table 1 [1,12]. It should be noted that certain energy indicators do not have a direct one-to-one correlation with freight volumes [24]. However, both are linked to economic activity and oil prices, thereby forming an indirect relationship. Furthermore, high-precision ACV forecasting relies on the effective modeling of historical temporal patterns. Therefore, the historical ACV dataset, spanning from July 1998 to August 2025, was obtained from the CEIC database for model training and evaluation. This period represents the maximum time span available for the raw samples in the CEIC database. The final temporal range of the experimental dataset is subsequently determined based on the overlapping time coverage of the selected feature variables. The CEIC database is a publicly available commercial database, and all raw data used in this study (including China's ACV data, and exogenous indicators from macroeconomic, consumption, and energy production dimensions) can be legally obtained through the official website of CEIC (https://www.ceicdata.com/en) with standard access rights.

Table 1. Explanation of ACV- related metrics.

Variation	Value	Unit
Economic development	Per capita GDP	RMB
	Nominal GDP	Million USD
	Primary industry rate	%
	Secondary industry rate	%
	Tertiary industry rate	%
Resident consumption level	Personal consumption expenditure	RMB
	Consumer index	%
	Urban population	Persons
	Population density	Persons/km^2
Energy production	Crude processing volume	Million tons
	Coal oil volume	Thousand tons
	Total crude volume	Thousand barrels
	Fuel volume	Thousand tons

2.2. Feature variable selection

In multi-source data fusion time series forecasting tasks, a significant challenge arises from the heterogeneous sampling frequencies among different data sources. In this study, the selected

correlated indicators originate primarily from economic and energy domains, whose temporal resolutions are not entirely consistent with that of ACV. Specifically, some indicators are released on a quarterly or annual basis, whereas ACV data are recorded on a monthly basis. To address this mismatch, the Legendre Polynomial Interpolation method (third-order) was employed to perform temporal interpolation and resampling on low-frequency indicators, thereby aligning all features to a unified monthly time scale consistent with the ACV data. The Legendre polynomial offers excellent numerical stability and smoothness, effectively preserving the original variation trends of the indicators while avoiding the oscillation effects often introduced by high-order polynomial interpolation [25].

Furthermore, it is essential to note that in multivariate time series forecasting, the appropriateness of the selected features directly impacts the model's predictive performance. To ensure strong statistical correlations between the input variables and air cargo volume, this study applies the mutual information (MI) method to screen the candidate indicators (see Table 1). Mutual information quantifies nonlinear dependencies between variables, overcoming the limitations of traditional linear correlation coefficients when dealing with complex economic and transportation indicators [26]. The general process is as follows:

For the target variable Y (air cargo volume), its entropy $H(Y)$ measures the degree of uncertainty associated with the variable:

$$H(Y) = -\sum_{y \in \mathcal{Y}} P_Y(y) \log P_Y(y) \tag{1}$$

where $P_Y(y)$ denotes the probability density function corresponding to ACV.

Meanwhile, for the selected correlated variable X_i in Table 1, its joint entropy with Y is defined as $H(X_i, Y)$, which reflects the combined uncertainty of both variables.

$$H(X_i, Y) = -\sum_{x \in \mathcal{X}_i} \sum_{y \in \mathcal{Y}} P_{X_i,Y}(x, y) \log P_{X_i,Y}(x, y) \tag{2}$$

Accordingly, the mutual information $I(X_i; Y)$ between X_i and Y can be expressed equivalently either as the difference between entropy terms or in terms of the probability ratio:

$$I(X_i; Y) = \sum_{x \in \mathcal{X}_i} \sum_{y \in \mathcal{Y}} P_{X_i,Y}(x, y) \log \frac{P_{X_i,Y}(x, y)}{P_{X_i}(x) P_Y(y)} \tag{3}$$

where the probability ratio $\dfrac{P_{X_i,Y}(x, y)}{P_{X_i}(x) P_Y(y)}$ measures the deviation of the joint distribution from complete independence. A larger MI value $I(X_i; Y)$ indicates a stronger dependency between the variables. Moreover, a higher mutual information score suggests that the corresponding indicator possesses greater explanatory power and influence over fluctuations in air cargo volume from an information-theoretic perspective.

Finally, all candidate indicators are ranked in descending order according to their mutual information (MI) values (see Figure 1). Variables with MI scores exceeding the predefined threshold of 0.6 are classified as strongly correlated with ACV. The MI threshold is set to 0.6 based on the empirical distribution of MI values in Figure 1, where a clear gap ("elbow") is observed between strongly informative and weakly informative indicators. This choice helps reduce weak predictors and improves generalization under the small-sample monthly setting.

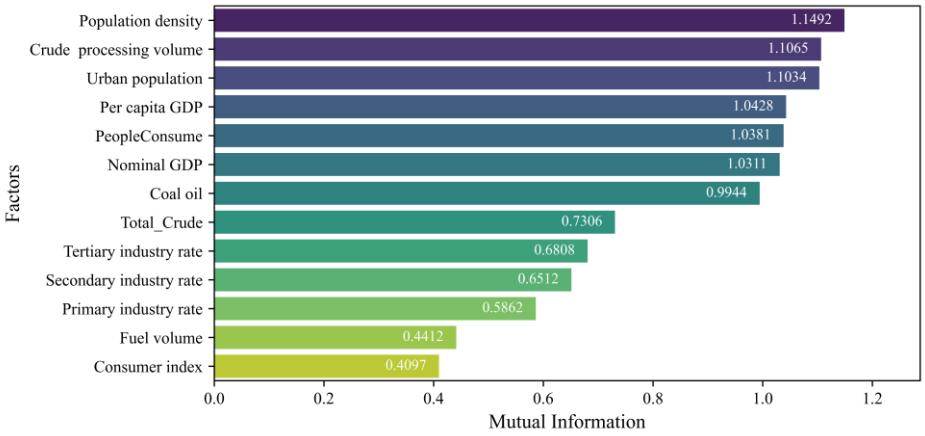

Figure 1. Distribution map of the impact factor MI.

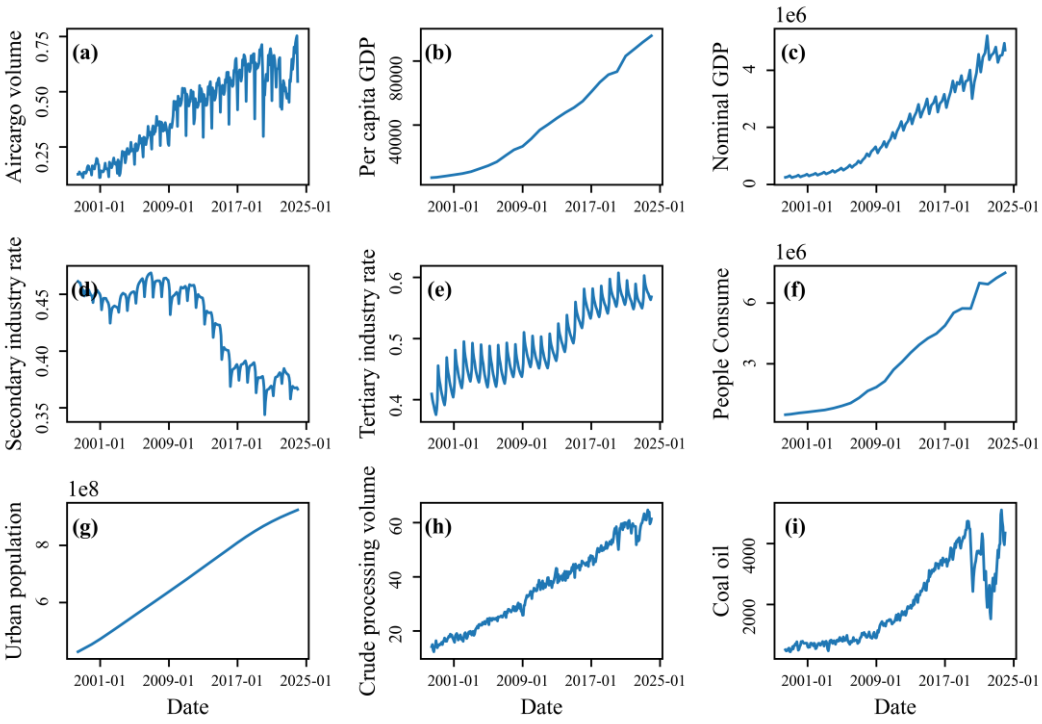

Figure 2. Sequence diagram of multiple variables. (a) represents China's air cargo volume, (b)–(i) are the relevant screening factors.

After the mutual information–based screening, additional refinement was conducted by considering both the physical interpretability of the indicators and their mutual independence. Within the energy production category, crude processing volume and total crude volume exhibit strong collinearity. Given that the MI value of crude processing volume is substantially higher than that of total crude volume, the former provides stronger explanatory power with respect to ACV fluctuations. To avoid redundancy while preserving the most informative variables, only crude processing volume and coal oil volume were retained as representative indicators in the energy production dimension. For the resident consumption level category, although population density demonstrated a relatively high MI value, urban population was considered more directly reflective of potential air logistics demand, as air cargo activities are more closely associated with urban economic concentration and

consumption intensity. Therefore, population density was excluded in favor of urban population to enhance both interpretability and practical relevance of the selected features. The selected feature indicators and the temporal changes in ACV are shown in Figure 2.

3. Methodology

3.1. Sequence feature analysis

As previously discussed, the ACV time series exhibits pronounced temporal–frequency complexity, with its variations arising from the superposition and coupling of multiple periodic components [12]. At shorter time scales, ACV is influenced by factors such as flight scheduling, holiday-driven logistics demand, and seasonal meteorological conditions, displaying distinct high-frequency intra-cycle fluctuations. In contrast, over longer time scales, it is driven by low-frequency factors such as macroeconomic conditions, industrial production cycles, and international trade activities, leading to inter-cycle trends and structural shifts.

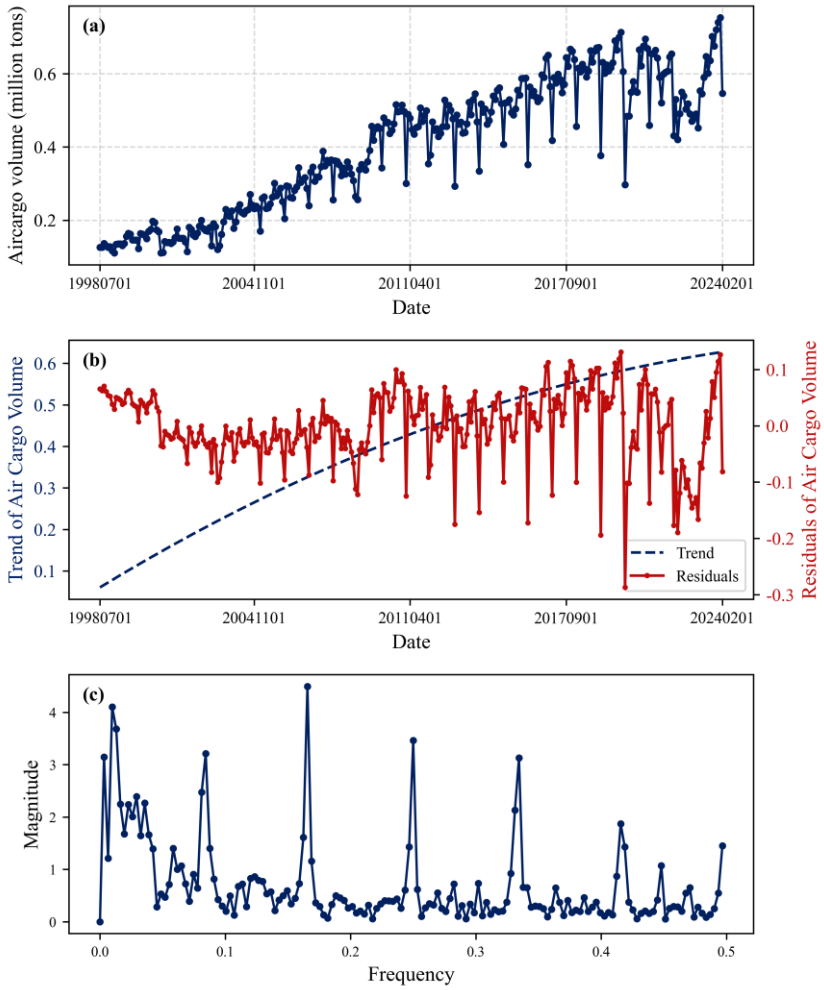

Figure 3. Analysis of sequential characteristics in air cargo volume. (a) Original sequence; (b) trend component and residual periodic component of the freight volume sequence; (c) frequency spectrum of the residual periodic component.

To verify these characteristics, the ACV time series obtained in Section 2.2 was analyzed through trend decomposition and frequency-domain analysis. First, the long-term trend component was extracted using trend decomposition, and the residual component was regarded as the periodic term. Subsequently, a Fast Fourier Transform (FFT) was applied to the residual series, and the resulting spectrum is presented in Figure 3(c). Multiple dominant frequency components with significant amplitudes can be observed, confirming that the dynamic evolution of ACV is jointly governed by multi-periodic influences. This finding substantiates the multi-scale fluctuation property of the ACV sequence.

Based on the above analysis, to fully capture the dynamic variations of ACV across different temporal scales, this study proposes a multi-period feature extraction model capable of simultaneously modeling short-term intra-cycle fluctuations and long-term inter-cycle trends, referred to as ACVPNet. The design aims to achieve collaborative modeling of multi-frequency components, thereby enhancing the prediction accuracy and temporal robustness for complex air cargo volume sequences.

3.2. ACVPNet

First, from a technical perspective, the establishment of ACVPNet aligns with the current mainstream paradigm of time series forecasting [27]. ACVPNet is composed of three primary components: sequence reconstruction, feature extraction, and predictor, with specific components illustrated in Figure 4.

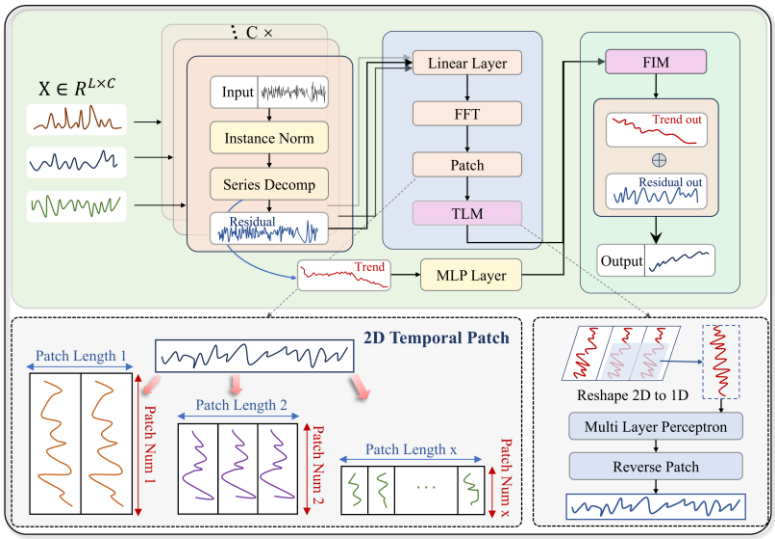

Figure 4. Structure of the proposed model: ACVPNet.

3.3. Sequence reconstruction

3.3.1. Instance norm

Accurate prediction of real-world time series using deep learning models requires careful handling of distributional discrepancies between training and testing data. Failure to address this issue often leads to substantial degradation in both accuracy and robustness. Instance normalization has been extensively employed in recent research , where the mean and standard deviation of each ACV sequence are computed to normalize individual samples. The formula is expressed as follows:

$$\overline{X}_i = \frac{X_i - \mu_i}{\sqrt{\sigma_i^2}} \cdot \eta + \omega \tag{4}$$

where μ_i and σ_i^2 denote the mean and variance, respectively, η and ω represent the learnable affine parameters. Finally, before generating ACV predictions, statistical information from the historical sequence and the optimized affine parameters are utilized to rescale the data.

3.3.2. Series decomposition

As illustrated in Figure 3(b), the ACV sequence comprises both low-frequency trend components and high-frequency periodic components. To achieve a more accurate modeling of ACV, separating these two components and modeling them independently enables the model better to capture the distinct characteristics of each sequence component [28]. Therefore, the proposed model performs sequence decomposition using a sliding window mechanism, followed by a dual-branch prediction architecture that independently forecasts the trend component and the residual periodic component. The detailed procedure of the sequence decomposition process is described as follows:

$$\begin{aligned} X^{trend} &= \text{AvgPool}(\text{Padding}(X)), \\ X^{season} &= X - X^{trend} \end{aligned} \tag{5}$$

where X is the ACV sequence, X^{trend} is the trend item for ACV, and X^{season} is the residual term sequence. To mitigate boundary effects, the ACV sequence is symmetrically extended at both ends through mirror padding, while AvgPool represents average pooling implemented via a one-dimensional convolution kernel.

3.4. Feature extraction

3.4.1. Temporal local MLP

As outlined in Section 3.1, the ACV sequence is characterized by the coexistence of multiple periodic components. Given the intertwined characteristics of periodic fluctuations and local perturbations in ACV sequences, single-scale temporal modeling often struggles to capture both intra-cycle variations and inter-cycle dynamics simultaneously. To address this limitation, this study designs the temporal local MLP (TLM) module, which constructs a locally aware temporal representation structure to enhance the model's ability to learn complex temporal dependencies within ACV sequences.

In practical forecasting tasks, ACV data typically exhibit pronounced periodic and seasonal fluctuations, while also being influenced by disturbances such as policy adjustments, market fluctuations, and unexpected events. The TLM module incorporates a local temporal patch strategy, dividing the time series into segments at both intra-cycle and inter-cycle levels, enabling the model to capture variations occurring within and across cycles simultaneously. Specifically, the module first applies a Fast Fourier Transform (FFT) to the input sequence to extract multiple characteristic periods. Based on these extracted periodic scales, the sequence is segmented along the temporal dimension using a sliding window mechanism, producing a set of local patches, as illustrated in Figure 5. Each patch is then processed through a lightweight multi-layer perceptron (MLP) for feature interaction.

The MLP performs nonlinear mappings within each local window to effectively model dynamic dependencies among temporal fragments.

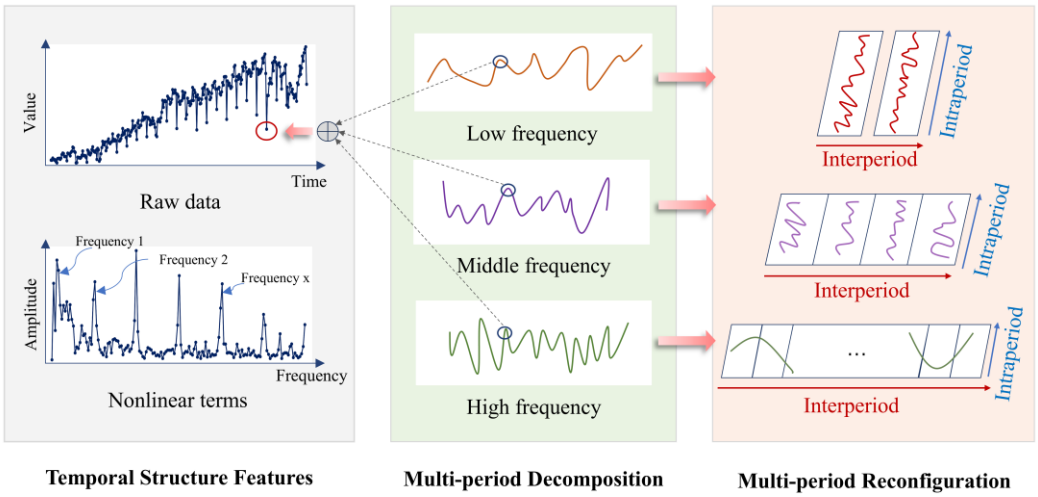

Figure 5. Schematic diagram of patch segmentation under multiple periods.

Specifically, let the input sequence be $X \in \mathbb{R}^{P \times p \times C}$, where P、p、C denote the batch size, number of periods, period length, and feature dimension, respectively. For each predefined patch scale (k_h, k_w), the TLM module extracts local temporal blocks through an unfold operation. These blocks undergo feature transformation via an MLP and are subsequently reconstructed into the original temporal structure:

$$Y_{(k_h,k_w)} = \text{Fold}\{\text{MLP}_{(k_h,k_w)}(\text{Unfold}(X))\} \tag{6}$$

The outputs of multi-scale or multi-period features are fused through weighted aggregation to obtain the final locally enhanced feature representation, thereby achieving a joint modeling of micro-level intra-cycle variations and macro-level inter-cycle trends.

Unlike traditional global temporal convolutions or average pooling operations, the TLM module preserves the complete structural information of temporal segments, thus avoiding the attenuation of local features caused by smoothing operations. Moreover, Instance Normalization is applied along the patch dimension to eliminate distributional discrepancies among different sequence instances, thereby enhancing the model's robustness.

Through this local patch modeling mechanism, the TLM module can flexibly perceive key dynamic features of ACV sequences across different temporal scales, reinforcing the model's responsiveness to local perturbations and seasonal variations, and providing a more stable and high-resolution temporal representation for subsequent variable/factor interactions.

3.4.2. Feature interaction MLP

For the ACV forecasting task, various influencing factors — such as economic level, energy production, and household consumption — directly or indirectly affect air cargo volume. Meanwhile, these factors often exhibit complex nonlinear dependencies and interactions with each other. Such

dynamic coupling among variables makes it difficult for temporal modeling alone (e.g., via TLM) to capture their combined influence on cargo volume fully.

To address this issue, we design the feature interaction MLP (FIM) module, which explicitly models nonlinear inter-variable dependencies along the feature dimension, thereby enhancing the model's capability to represent the synergistic effects of multi-source factors. To effectively capture these cross-variable dependencies, the FIM module performs variable-level feature interaction modeling at each time step, where an MLP is employed to realize nonlinear mixing and remapping of multi-variable features. The structural illustration of this module is shown in Figure 6.

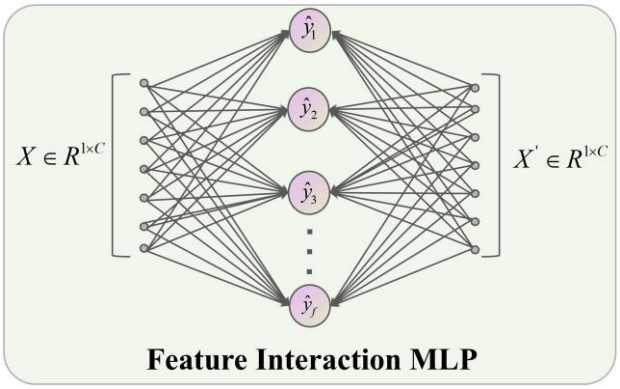

Figure 6. Structure of the feature interaction MLP module.

Specifically, let the multivariate input sequence at a given time be $X \in \mathbb{R}^{1\times C}$, where C denotes the feature dimension/number of variables. The FIM module first performs layer normalization on the feature dimension to eliminate distribution inconsistencies among different variables.

$$\mathbf{X} = LayerNorm(\mathbf{X}) \tag{7}$$

Then, a nonlinear mapping is applied to the feature vector X, projecting it into a latent space. This process enables the model to characterize nonlinear and higher-order dependencies among variables in a higher-dimensional and more abstract feature domain, thereby capturing latent coupling patterns and implicit interaction features among multiple variables that influence air cargo volume more effectively. As a result, the overall robustness and forecasting performance of the model are improved.

$$Y = \sigma(\mathbf{W}_1 X + \mathbf{b}_1) \tag{8}$$

where, $\sigma(\cdot)$ denotes the GELU activation function, and $\mathbf{W}$ represents the weight matrix. The introduction of GELU in the FIM module enhances the nonlinear expressiveness of feature interactions. Compared to conventional ReLU-like functions, GELU adaptively modulates the input amplitude according to a Gaussian distribution, resulting in a smoother activation transition [29].

Finally, the feature variables are subjected to another nonlinear transformation and mapped back to the original dimensionality, allowing the subsequent prediction module to fuse the forecasting results of the trend and periodic components.

$$X' = \mathbf{W}_2 \sigma Y + \mathbf{b}_2 \tag{9}$$

It is worth emphasizing that, unlike traditional attention-based or convolution-based feature fusion methods, the FIM module performs efficient interaction learning directly along the feature

dimension through a lightweight MLP structure. This design not only avoids the issue of high computational complexity but also effectively captures nonlinear mappings among variables. In addition, a dropout mechanism is incorporated within the module to prevent overfitting. At the same time, a residual connection establishes a robust information pathway between the fused features and the original input, ensuring stability during feature interaction.

Through this feature interaction modeling mechanism, the FIM module can dynamically perceive the coupling patterns among multiple variables at each time step, thereby reinforcing the model's ability to understand the synergistic effects among correlated factors.

3.5. Predictor

As described in Section 3.2, ACVPNet adopts a dual-branch prediction architecture to effectively model the distinct temporal patterns present in the ACV sequence. After the decomposed periodic component sequence passes through the TLM module for intra- and inter-period feature perception, it is subsequently fed into the FIM module for multivariate coupling interaction. At this stage, the model needs to fuse the extracted periodic information with the trend component information to generate the complete ACV forecasting output.

To this end, a simple yet efficient MLP predictor is employed as the trend component predictor. For the trend component separated from the ACV sequence, complex feature extraction mechanisms such as multi-head attention may introduce unnecessary computational overhead and structural complexity, while a lightweight MLP projection layer can achieve effective modeling [30]. This design not only preserves the essential representational capacity of the trend information but also enhances the efficiency of the overall model.

Finally, the predicted periodic and trend components are summed to obtain the complete ACV sequence prediction.

3.6. Air cargo volume forecasting process

In practical application scenarios, the proposed deep learning model ACVPNet achieves high-precision forecasting of ACV through three main stages: data preprocessing, model training, and prediction generation, as illustrated in Figure 7.

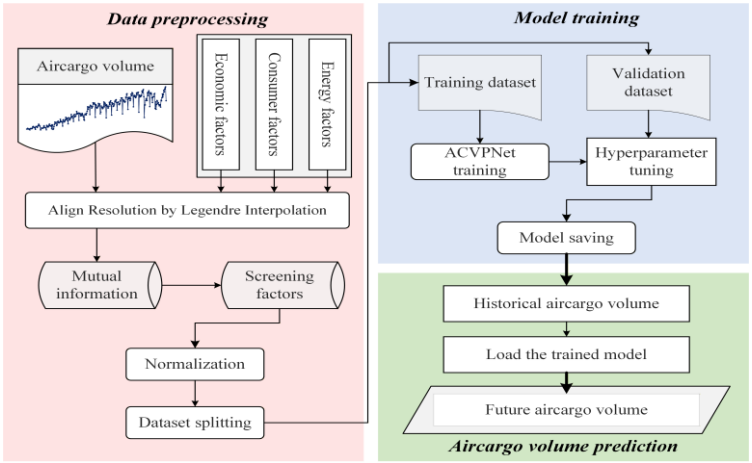

Figure 7. The flowchart for ACV sequence prediction.

The overall workflow can be summarized as follows:

(1) Data preprocessing: All selected influencing factors are first temporally aligned with the ACV series, followed by correlation-based feature screening. Subsequently, the experimental samples are standardized to eliminate dimensional inconsistencies among different variables.

(2) Model training: The dataset is divided into training and validation subsets to iteratively optimize the parameters of the proposed model and prevent overfitting.

(3) Prediction generation: Once the model is trained, a segment of historical data is fed into the network to produce accurate ACV forecasts directly.

This workflow ensures that the ACVPNet model effectively captures both the temporal dependencies and multi-factor interactions underlying ACV dynamics, thereby supporting robust and interpretable forecasting performance.

4. Experiment and result

4.1. Experimental setup

4.1.1. Experimental details

All experiments were conducted in a computational environment configured with PyTorch 2.9.0 and CUDA 12.6. The hardware platform consisted of an AMD Ryzen 9 7940HS CPU (4.0 GHz) and an NVIDIA GeForce RTX 4070 Laptop GPU, equipped with 8 GB of video memory. Given the overall size of the sample dataset, we selected the last 12 months of ACV data as the test set, with the remaining data divided into an 80% training set and a 20% validation set. To construct supervised learning samples, we adopt a rolling window strategy with a look-back length of 36 months and a forecasting horizon of 12 months (stride = 1 month). The dataset is split chronologically to prevent information leakage: earlier samples are used for training and the later part (before the test horizon) is used for validation. To ensure stable and efficient training, the hyperparameters of the proposed ACVPNet were carefully tuned through traversal optimization, and the final settings are summarized in Table 2.

Table 2. The hyperparameters selected for the proposed model.

Hyperparameter	Value
Batch size	16
Learning rate	0.0001
Loss function	MSE
Look-back window size	36
Number of Periods	2
Patch kernel sizes	[2, 3]
FFN hidden dimension	64
Encoder depth	2
Dropout	0.4
Early stopping patience	5

During the training phase, several optimization strategies were employed to enhance convergence stability and model generalization. For instance, a learning rate warm-up scheme gradually increases

the learning rate during the initial training phase to stabilize optimization and prevent early gradient instability. Early stopping monitors the validation RMSE (equivalently MSE loss) after each epoch. Training terminates if the validation metric does not improve for a predefined number of epochs (see Table 2) and the model parameters are restored to the best checkpoint observed during training.

4.1.2. Comparative methods

In recent years, the application of artificial intelligence techniques for high-precision time series forecasting has become a mainstream trend. In the field of cargo volume prediction, several researchers have developed and adapted advanced deep learning models based on the unique characteristics of cargo volume sequences [1,16,18,31]. To demonstrate the effectiveness and superiority of the proposed model, two representative AI-based forecasting models commonly used in cargo volume prediction tasks were selected for comparison. Moreover, the classical SARIMA model was also included as a baseline to highlight the performance gap between traditional statistical approaches and deep learning methods. A brief introduction to the comparative method is provided below.

(1) LSTM

The Long Short-Term Memory (LSTM) network [32] is a recurrent neural model capable of capturing long-term temporal dependencies through gated mechanisms. Its ability to mitigate gradient vanishing makes it suitable for modeling sequential dynamics, enabling effective representation of historical dependencies in ACV forecasting.

(2) Informer

The Informer model utilizes a self-attention mechanism with ProbSparse attention to capture long-range temporal dependencies in long sequences efficiently [33]. Its high computational efficiency and ability to model complex temporal patterns make it advantageous for large-scale ACV sequence forecasting.

(3) SARIMA

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model extends the traditional ARIMA by incorporating seasonal components, enabling it to capture both trend and periodic fluctuations. Its clear statistical structure makes it suitable for ACV forecasting tasks characterized by strong seasonality and long-term trend variations [7].

4.1.3. Evaluation metrics

To comprehensively assess the predictive capability of the predictive performance of the proposed model on the ACV time series, five quantitative metrics are employed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Square Percentage Error (MSPE), and Symmetric Mean Absolute Percentage Error (SMAPE). These indicators jointly capture different aspects of prediction deviation, reflecting the model's overall accuracy, robustness, and proportional error characteristics in forecasting ACV variations. It should be noted that all data used in calculating the indicators are presented in their original units and scales. The corresponding mathematical definitions are presented below:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|L_i - \hat{L}_i\right| \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(L_i - \hat{L}_i\right)^2} \tag{11}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{L_i - \hat{L}_i}{L_i}\right| \tag{12}$$

$$\mathrm{MSPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left(\frac{L_i - \hat{L}_i}{L_i}\right)^2 \tag{13}$$

$$\mathrm{SMAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{\left|L_i - \hat{L}_i\right|}{\left(\left|L_i\right| + \left|\hat{L}_i\right|\right)/2} \tag{14}$$

where L_i represents the ground truth and $\hat{L}_i$ represents the predicted ACV value.

4.2. Experimental results

4.2.1. Assessment of the consistency of prediction results

To evaluate the effectiveness of the proposed model in forecasting China's air cargo volume, the experiment was conducted using the dataset described in Section 2.1 and the hyperparameter settings listed in Table 2. The true ACV values from March 2023 to February 2024 were used as the reference sequence, against which the model's predictions were compared. The corresponding time series curves are illustrated in Figure 8.

As shown in the figure, the predicted ACV sequence aligns closely with the ground truth in both overall trend and periodic structure, demonstrating strong temporal consistency and pattern fidelity. Notably, while the target sequence exhibits a declining trend from June 2022 to February 2023, the model successfully captures the subsequent upward tendency in the forecast period. This indicates that the proposed model effectively leverages long-term historical dependencies without being misled by short-term fluctuations, thereby achieving stable and reliable forecasting performance for ACV time series.

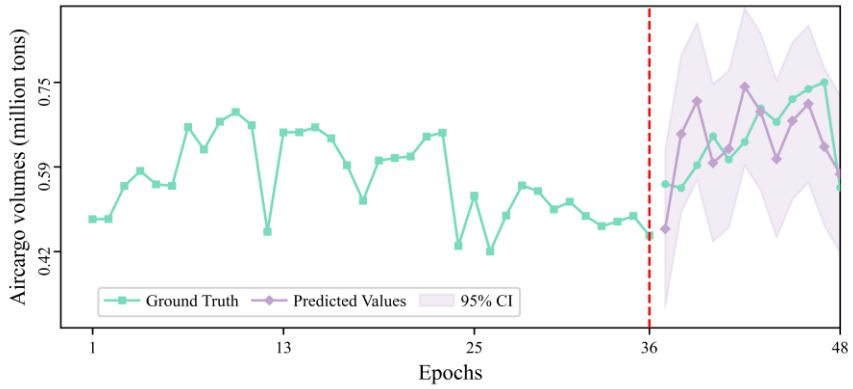

Figure 8. Prediction bias in ACV sequences.

4.2.2. Model performance assessment

To further demonstrate the effectiveness and superiority of the proposed ACVPNet model, three representative ACV forecasting models introduced in Section 4.1.2 were selected as comparative baselines. For a fair comparison, the hidden layer dimensions of the LSTM and Informer models were set identically to the "FFN hidden dimension" listed in Table 2. The Informer model's factor parameter was set to 3, its embedding dimension to 128, and the look-back window size was uniformly defined as 36 across all AI-based models. The evaluation metrics described in Section 4.1.3 were employed to quantitatively assess the prediction accuracy of each method. It should be noted that to ensure the reliability of experimental results, we randomly set five random seeds when training each AI-based model and calculated the mean and standard deviation of their test results. The mean evaluation results for all models are illustrated in Figure 9, while the corresponding numerical values of each evaluation indicator are provided in Table 3. Additionally, all baseline training datasets employ the same data preprocessing methods and feature factors as ACVPNet, with training strategies consistent with those outlined in Section 4.1.1.

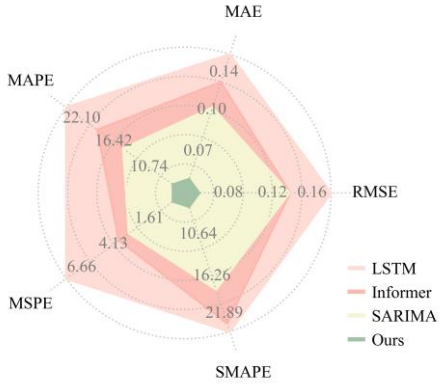

Figure 9. Model comparison results across different evaluation metrics.

Table 3. Accuracy statistics for different methods for all indicators.

Methods\Metrics	RMSE	MAE	MAPE	MSPE	SMAPE
SARIMA	0.137	0.124	18.784	4.159	21.142
Informer	0.165 ± 0.031	0.142 ± 0.029	21.954 ± 3.975	6.579 ± 1.591	21.015 ± 4.921
LSTM	0.133 ± 0.024	0.111 ± 0.023	16.369 ± 3.473	3.714 ± 1.338	18.489 ± 4.201
ACVPNet	0.083 ± 0.005	0.068 ± 0.005	10.667 ± 0.988	1.670 ± 0.296	10.907 ± 1.088

*Note: The evaluation results of AI-based models are recorded as mean ± standard deviation.

As shown in Figure 9, the proposed ACVPNet consistently outperforms the benchmark models across all evaluation metrics, exhibiting clear advantages in both error control and prediction stability. Specifically, ACVPNet achieves significantly lower MAE and RMSE values, indicating its superior capability in capturing fine-grained variations within the ACV sequence. Moreover, the model attains better performance in MAPE and MSPE, demonstrating robust performance across ACV sequences with different magnitudes. In terms of SMAPE, ACVPNet also outperforms the competing models,

reflecting its enhanced resilience to symmetric errors. These results collectively confirm that ACVPNet possesses strong adaptability and precision in handling real-world ACV forecasting tasks.

To provide a more intuitive comparison of the prediction performance and historical feature extraction capabilities of different models, the trajectories of the experimental dataset used for modeling and the corresponding prediction sequences from each model were visualized (see Figure 10).

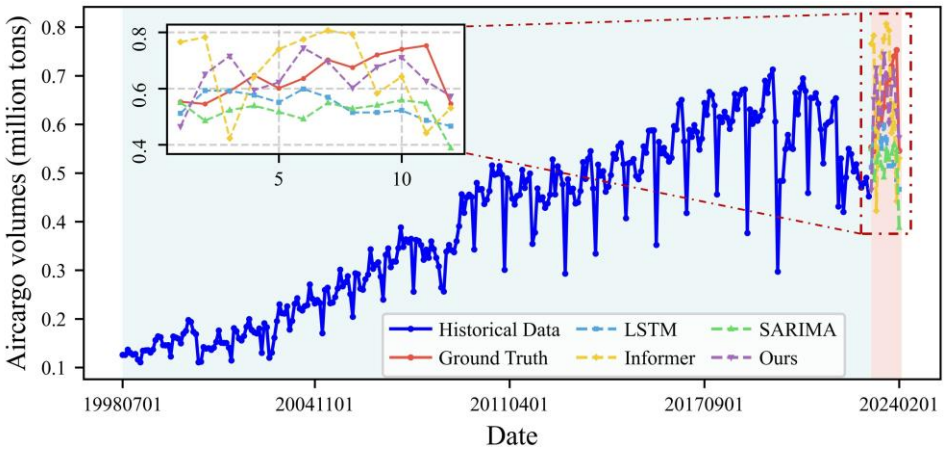

Figure 10. Comparison of predictive performance across different models.

As shown in the results, the Informer model exhibits noticeable deviations from the ground truth, with pronounced fluctuations in its predicted sequence. This weakness is primarily attributed to the model's reliance on large-scale training data to capture long-term temporal dependencies effectively. Under limited sample conditions, Informer struggles to learn stable temporal patterns, and its attention mechanism becomes highly sensitive to noise [34], ultimately leading to degraded and unstable predictive performance. In contrast, the LSTM model demonstrates a relatively smooth prediction at the early stage but shows a noticeable downward trend in the latter part of the sequence. This behavior likely results from RNN's inherent limitation in capturing long-term dependencies, causing it to underfit long-range periodic patterns [35]. The SARIMA model generates predictions whose fluctuation patterns are generally consistent with the actual series; however, it still exhibits a noticeable systematic bias. This limitation primarily stems from its relatively simple functional structure and its inability to account for the complex interactions among multiple variables, which constrains its capacity to capture the nonlinear and multi-periodic characteristics inherent in ACV sequences.

Overall, the proposed ACVPNet model achieves superior performance across multiple aspects. Its prediction series closely aligns with the true ACV trend in both global and local variations, effectively capturing periodic fluctuations and dynamic transitions. These results confirm that ACVPNet possesses a stronger capability for learning temporal dependencies and extracting cross-variable interactions, thereby enabling more accurate and robust forecasting of air cargo volume dynamics.

4.3. Ablation study

As presented in Table 4, the effectiveness of each component within ACVPNet is systematically evaluated through a set of ablation experiments. In these experiments, individual modules are removed or replaced sequentially to assess their respective contributions to the overall model

performance. Specifically, Case 1 models the relationship between historical ACV sequences and future predictions using only a basic multi-layer perceptron. Case 2 incorporates solely the FIM module for feature interaction modeling, while Case 3 adopts only the TLM module for multi-period temporal representation learning. To ensure experimental rigor, the parameter configurations and training strategies in the ablation study were kept consistent with those described in Section 4.1.1. Each model was independently trained five times with random initialization, and the reported accuracy metrics correspond to the average results across all runs. The forecasting horizon was uniformly set to 12 months.

Table 4. Ablations of avnet.

		Modules				Prediction duration
		TLM	FIM	MLP		12 months
Case	1	✗	✗	✓	RMSE	0.145
	2	✗	✓	✓	(million tons)	0.115
	3	✓	✗	✓		0.121
	4	✓	✓	✓		0.085

*Note: The check mark ✓ and cross mark ✗ indicate whether or not a component is included.

The comparison between Case 1 and Case 2 indicates that the proposed FIM module effectively captures the dependencies between ACV and its associated multivariate feature sequences. In addition, the performance gains observed when comparing Case 1 and Case 3 further demonstrate that the TLM module successfully learns multi-period interaction patterns within the temporal sequence. Overall, the ablation results confirm the individual contribution of each module in Avnet, thereby supporting the soundness and robustness of the proposed architectural design.

4.4. Interpretability analysis

To improve the transparency of ACVPNet beyond pure predictive accuracy, we provide an interpretability-oriented analysis focusing on the cycle scales extracted by the FFT-driven periodic modeling in the TLM, and the economic consistency of cross-variable coupling captured by the FIM.

4.4.1. Period-scale interpretation (TLM)

As discussed in Section 3.1, the ACV time series exhibits multi-periodic characteristics in the frequency domain. After decomposing ACV into a long-term trend and a residual periodic component, we apply FFT to the residual series and observe multiple dominant spectral components with salient amplitudes (Figure 3(c)), indicating that ACV dynamics are governed by the superposition of multi-frequency influences rather than a single seasonal rhythm.

In the proposed model, these frequency-domain cues are operationalized as data-driven period scales that guide periodic reshaping and patch-based local modeling. Specifically, the FFT-based period extraction yields two representative cycle lengths, 2 months and 3 months, which are used as the multi-period patch kernel sizes in TLM (see Table 2). Rather than being manually imposed, these scales are obtained from the intrinsic spectral structure of the ACV series and thus reflect recurrent sub-seasonal regularities embedded in historical air cargo fluctuations.

From an economic and operational perspective, the 3-month component is consistent with widely observed quarterly rhythms in macroeconomic activity and supply-chain decision cycles (e.g., quarterly production planning), which can indirectly manifest in air cargo demand through recurring quarter-level logistics scheduling and shipment consolidation. The 2-month component can be interpreted as a sub-quarter (bi-monthly) fluctuation, capturing shorter-term adjustments such as order replenishment and capacity deployment responses that are typically realized at a several-week horizon.

Technically, TLM leverages these scales by segmenting the periodic component into local temporal patches at each period length and applying lightweight nonlinear mappings within patches, followed by a weighted aggregation across scales. This design enables ACVPNet to jointly represent micro-level intra-cycle variations and macro-level inter-cycle evolution, aligning with the earlier observation that ACV is simultaneously affected by short-term disturbances (e.g., scheduling and seasonal conditions) and longer-term economic drivers.

4.4.2. Interaction map (FIM)

To visualize the nonlinear coupling patterns learned by the FIM, we derive a symmetric interaction-strength matrix from the trained FIM parameters. Let $W \in R^{d\times c}$ denote the first-layer weight matrix of the FIM mapping (after normalization). We define the interaction strength between variables i and j as:

$$S_{ij} = \sum_{k=1}^{d} \left|W_{k,i}\right| \bullet \left|W_{k,j}\right| \tag{15}$$

In which, d denotes the dimension size of the hidden layer in the FIM module. The resulting matrix was normalized and visualized as a heatmap in Figure 11.

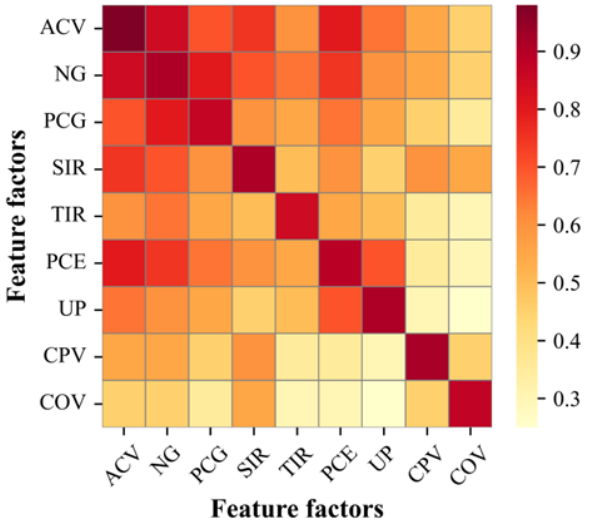

Figure 11. Interaction strength heatmap among feature variables. The abbreviated variable names in the labels correspond to: ACV, Nominal GDP, Per capita GDP, Secondary industry rate, Tertiary industry rate, Personal consumption expenditure, Urban population, Crude processing volume, and Coal oil volume.

We can see that, the ACV-related coupling concentrates on Nominal GDP, Secondary industry rate, and Personal consumption expenditure, indicating that the model mainly relies on a compact set of economically meaningful drivers when forming cross-variable representations. Second, the interaction map exhibits a structured organization broadly aligned with the three feature groups (macroeconomic–consumption–energy). In particular, strong cross-group couplings between GDP-related variables and consumption-related variables reflect the linkage between aggregate economic activity and resident consumption demand, which is consistent with the fact that air cargo volumes are jointly shaped by production/trade intensity and consumption-driven high-value logistics.

5. Conclusions

Accurate forecasting of ACV is of great significance for optimizing logistics management, supporting strategic planning in the aviation industry, and enhancing supply chain resilience. This study proposes a novel deep learning framework, ACVPNet, which effectively models the complex periodic variations inherent in ACV time series, aiming to enhance the accuracy and stability of ACV forecasting. All modeling and testing data used in this work were obtained from the historical air cargo volume dataset published in the CEIC database. In the task of predicting China's ACV for the following year, ACVPNet achieved a Root Mean Square Error (RMSE) of 0.083 million tons. In contrast, the predicted sequence demonstrated a trend evolution and short-term fluctuation pattern closely aligned with the ground truth. This high level of predictive accuracy and temporal consistency provides valuable insights for aviation market forecasting, capacity planning, and policy decision-making.

To demonstrate the superiority of the proposed model, three representative ACV forecasting models—LSTM, Informer, and SARIMA—were selected for comparison. Across five quantitative evaluation metrics, ACVPNet consistently outperformed these baselines. For example, in the one-year prediction task, ACVPNet reduced the RMSE by 37.6%, 49.7%, and 39.4% compared with LSTM, Informer, and SARIMA, respectively. Moreover, through the analysis of both the historical modeling dataset and the predicted temporal trajectories, ACVPNet was shown to effectively capture periodic patterns within and across cycles by leveraging its multi-granularity temporal learning mechanism. Furthermore, the multivariate feature interaction module enabled the model to remain robust against short-term perturbations near the prediction horizon, thereby achieving an accurate and comprehensive understanding of future ACV trends.

Nevertheless, it is worth noting that although the proposed model exhibits outstanding predictive performance, this study lacks an in-depth validation of the physical interpretability behind the model's perception of periodic features. The absence of such interpretability limits the broader development of ACV forecasting research. Therefore, future work will focus on enhancing the explainability of deep learning-based forecasting models by incorporating interpretable temporal feature extraction and domain-informed priors, ultimately promoting a deeper understanding of the underlying mechanisms governing air cargo dynamics. Additionally, validating the model's generalization capabilities by incorporating data from more countries and regions into testing represents one of the future research objectives.

Author contributions

Xiaoyuan Cheng: Conceptualization, methodology, software, writing–original draft, writing–review & editing.

Use of Generative-AI tools declaration

The author declares that no generative AI tools were used to create any figures, images or graphical elements in this work. Only assistive AI tools were employed to polish sentences, correct grammar and improve language clarity. All content, including text and images, is original and verified by the author, who takes full responsibility for its accuracy and integrity.

Acknowledgments

Thanks to the CEIC Database for providing the historical dataset on China's air cargo volume, as well as other data utilized in this paper. This support has been of great help to our research. The website is https://www.ceicdata.com/en. We don't have funding. All the raw data used in this study (including China's ACV data, and exogenous indicators from macroeconomic, consumption, and energy production dimensions) can be legally obtained through the official website of CEIC (https://www.ceicdata.com/en) with standard access rights.

Conflict of interest

The authors declare no competing interests.

References

1. J. G. M. Anguita, O. D. Olariaga, Air cargo transport demand forecasting using ConvLSTM2D, an artificial neural network architecture approach, *Case Stud. Transp. Policy*, **12** (2023), 101009. https://doi.org/10.1016/j.cstp.2023.101009
2. C. C. Hwang, G. C. Shiao, Analyzing air cargo flows of international routes: an empirical study of Taiwan Taoyuan International Airport, *J. Transp. Geogr.*, **19** (2011), 738–744. https://doi.org/10.1016/j.jtrangeo.2010.09.001
3. F. Kupfer, H. Meersman, E. Onghena, E. V. Voorde, The underlying drivers and future development of air cargo, *J. Air Transp. Manag.,* **61** (2017), 6–14. https://doi.org/10.1016/j.jairtraman.2016.07.002
4. W. W. L. Lo, Y. Wan, A. Zhang, Empirical estimation of price and income elasticities of air cargo demand: The case of Hong Kong, *Transp. res., Part A Policy pract.*, **78** (2015), 309–324. https://doi.org/10.1016/j.tra.2015.05.014
5. Y. Rodríguez, O. D. Olariaga, Air traffic demand forecasting with a bayesian structural time series approach, *Period. Polytech. Transp. Eng.,* **52** (2024), 75–85. https://doi.org/10.3311/PPtr.20973
6. D. C. Han, Y. Y. Peng, Prediction of air freight volume based on BP neural network, *Int. Conf. Electron. Inf. Technol. Comput. Eng.,* **4** (2024), 904–907. https://doi.org/10.1145/3650400.3650553
7. Q. H. Nguyen, Modeling the volatility of international air freight: A case study of Singapore using the SARIMAX-EGARCH model, *J. Air Transp. Manag.*, **117** (2024), 102593. https://doi.org/10.1016/j.jairtraman.2024.102593

8. J. M. Liu, L. N. Ding, X. Y. Guan, J. Gui, J. B. Xu, Comparative analysis of forecasting for air cargo volume: Statistical techniques vs. machine learning, *J. Data Inf. Manag.*, **2** (2020), 243–255. https://doi.org/10.1007/s42488-020-00031-1
9. Q. H. Nguyen, P. Q. Tran, P. D. Ngo, Air cargo traffic forecasting model: An empirical study in Vietnam using the SARIMA-X/(E)GARCH model, *Res. Transp. Bus. Manag.*, **59**, (2025), 101268. https://doi.org/10.1016/j.rtbm.2024.101268
10. K. C. Min, H. K. Ha, Forecasting the daily demand of air cargo using data mining with CHAID approach, *J. Korean Soc. Transp.*, **38** (2020), 190–207. https://doi.org/10.7470/jkst.2020.38.3.190
11. H. Shin, G. Lee, Factors Affecting Air Cargo Demand: Focus on Trade Volumes between South Korea and Intra-Asia, *J. Korean Acad. Int. Commer.*, **34** (2019), 191–214. https://doi.org/10.18104/kaic.2019.34.3.191
12. H. T. Li, J. C. Bai, X. Cui, Y. W. Li, S. L. Sun, A new secondary decomposition-ensemble approach with cuckoo search optimization for air cargo forecasting, *Appl. Soft Comput.*, **90** (2020), 106161. https://doi.org/10.1016/j.asoc.2020.106161
13. B. Z. Niu, Z. P. Dai, X. P. Zhuo, Co-opetition effect of promised-delivery-time sensitive demand on air cargo carriers' big data investment and demand signal sharing decisions, *Transp. Res. E: Logist. Transp. Rev.,* **123** (2019), 29–44. https://doi.org/10.1016/j.tre.2019.01.011
14. X. Q. Hu, R. M. Jia, Y. Q. Wang, Research on Chengdu air cargo forecast based on improved ARIMA-GARCH, *Int. J. Model. Oper. Manag.*, **8** (2021), 299–312. https://doi.org/10.1504/IJMOM.2021.116802
15. T. Y. Chou, G. S. Liang, T. C. Han, Application of fuzzy regression on air cargo volume forecast, *Qual. Quant.,* **47**, (2013), 897–908. https://doi.org/10.1007/s11135-011-9572-4
16. C. Çatuk, Forecasting Turkey's air cargo tonnage: A comparative analysis of statistical techniques and machine learning methods, *J. Article*, **9** (2025), 109–117. https://doi.org/10.30518/jav.1582814
17. J. X. Che, W. X. Xia, Y. F. Xu, K. Hu, Multivariate wind speed forecasting with genetic algorithm-based feature selection and oppositional learning sparrow search, *Inf. Sci.*, **695** (2025), 121736. https://doi.org/10.1016/j.ins.2024.121736
18. M. S. Alam, J. B. Deb, A. A. Amin, S. Chowdhury, An artificial neural network for predicting air traffic demand based on socio-economic parameters, *Decis. Anal. J.*, **10** (2024), 100382. https://doi.org/10.1016/j.dajour.2023.100382
19. T. Y. Du, Q. W. Yin, Study on freight volume prediction of routes based on random forest model, *TCSISR*, **5** (2024), 1733–1739. https://doi.org/10.62051/fedtxr69
20. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
21. A. Neagoe, E. I. Tică, L. I. Vuță, O. Nedelcu, G. E. Dumitran, B. Popa, Hybrid LSTM-ARIMA model for improving multi-step inflow forecasting in a reservoir, *Water*, **17** (2025), 3051. https://doi.org/10.3390/w17213051
22. Y. W. Wei, W. Y. Tao, H. R. Zhang, D. B. Kou, Research on cargo volume forecasting based on the ARIMA-LSTM hybrid model, *ICDACAI*, **4** (2024) 896–900. https://doi.org/10.1109/ICDACAI65086.2024.00169
23. S. Q. Luo, T. H. Jiao, X. Y. Zhang, Y. T. Liu, R. R. Li, M. H. Guo, Logistics cargo volume prediction model based on combined ARIMA-LSTM prediction methods, *IACIS*, **4** (2024), 1–8. https://doi.org/10.1109/IACIS61494.2024.10721663

24. F. Kupfer, H. Meersman, E. Onghena, E. V. Voorde, The underlying drivers and future development of air cargo, *J. Air Transp. Manag.,* **61** (2017), 6–14. https://doi.org/10.1016/j.jairtraman.2016.07.002
25. M. S. Devi, F. Garestier, S. Hosseini, Legendre polynomials for nonlinear modeling in InSAR time series, *IEEE Geosci. Remote Sens. Lett.*, **22** (2025), 1–5. https://doi.org/10.1109/LGRS.2025.3570030
26. G. P. Zeng, A unified definition of mutual information with applications in machine learning, *Math. Probl. Eng.*, **2** (2015), 201874. https://doi.org/10.1155/2015/201874
27. Z. Li, S. Y. Qi, Y. D. Li, Z. L. Xu, Revisiting long-term time series forecasting: An investigation on linear mapping, *arXiv preprint arXiv:2305.10721,* 2023. https://doi.org/10.48550/arXiv.2305.10721
28. Z. P. Cao, Z. M. Sha, D. Lv, P. Z. Wei, B. W. Xiong, S. R. Ye, MGAMNet: a multi-granularity aware and hierarchically mixed network for BDS-3 satellite clock bias prediction, *GPS Solut.*, **30** (2026), 13. https://doi.org/10.1007/s10291-025-01984-9
29. D. Hendrycks, K. Gimpel, Gaussian error linear units (GELUs), *arXiv preprint arXiv:1606.08415,* 2023. https://doi.org/10.48550/arXiv.1606.08415
30. A. L. Zeng, M. X. Chen, L. Zhang, Q. Xu, Are transformers effective for time series forecasting? *arXiv preprint arXiv:2205.13504,* 2022. https://doi.org/10.48550/arXiv.2205.13504
31. C. S. Fiskin, O. Turgut, S. Westgaard, A. G. Cerit, Time series forecasting of domestic shipping market: comparison of SARIMAX, ANN-based models and SARIMAX-ANN hybrid model, *Int. J. Shipp. Transp. Logist.*, **14** (2022), 193–221. https://doi.org/10.1504/IJSTL.2022.122409
32. C. C. Hu, Cargo volume forecast and personnel scheduling model of logistics network based on ARIMA, LSTM and MOP, *ICEACE*, **4** (2024), 1250–1254. https://doi.org/10.1109/ICEACE63551.2024.10898553
33. H. Y. Zhou, S. H. Zhang, J. Q. Peng, S. Zhang, J. Li, H. Xiong, et al., Informer: Beyond efficient transformer for long sequence time-series forecasting, *arXiv preprint arXiv:2012.07436,* 2021. https://doi.org/10.48550/arXiv.2012.07436
34. Y. Q. Nie, N. H. Nguyen, P. Sinthong, J. Kalagnanam, A time series is worth 64 words: Long-term forecasting with transformers, *arXiv preprint arXiv:2211.14730,* 2023. https://doi.org/10.48550/arXiv.2211.14730
35. S. J. Bai, J. Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, *arXiv preprint arXiv:1803.01271,* 2018. https://doi.org/10.48550/arXiv.1803.01271