International Journal of Academic Accounting, Finance & Management Research(IJAAFMR)
ISSN: 2643-976X
Vol. 10 Issue 3 March - 2026, Pages: 1-7

Forecasting global oil demand: applying machine learning methods

Ammar Neamah Awdah

Al-Nahrain University, Collage of Information Engineering
Baghdad, Iraq
Ammar69a@yahoo.com

Abstract: *The dynamic instability of world energy markets and the growing role of the technological transformation process require the search for true and effective tools for forecasting demand for raw materials. This paper proposes and justifies a hybrid neural network architecture that combines the strengths of convolutional layers for spatial feature extraction, recurrent layers for long-term dependency calculation, and a temporal attention mechanism for adaptive weighting of market shocks, which together provide a reliable forecast of global oil demand. The scientific significance of the research lies in the development of a comprehensive forecasting method that can effectively analyze nonlinear relationships between macroeconomic indicators and structural changes in energy consumption.The results of the testing confirm the high robustness of the proposed approach, which provides a significant increase in the accuracy of approximation and the reliability of forecast values compared to traditional econometric methods and basic machine learning algorithms.*

Keywords— forecast, demand, oil, prices, accuracy, econometric models, machine learning, market, volatility.

1. RELEVANCE OF THE ISSUE UNDER CONSIDERATION

Oil plays a significant role in the global economy, as hydrocarbons currently account for about two-thirds of global energy consumption. According to European Commission estimates, oil accounted for approximately 63% and natural gas for 24% of the various types of fossil fuels traded worldwide in 2024 [1]. In this context, it is clear that uncertainty and other negative external factors in the fossil fuel market can significantly affect the global economic system due to price volatility and have a negative impact on the welfare of both oil-importing and oil-producing countries [2]. In this regard, reliable demand forecasts are of considerable practical interest and are in demand in various fields, including macroeconomic planning, energy risk management and strategic investment. To date, a wide range of different methods and approaches have been developed for forecasting energy demand. Traditional econometric models, while useful, do not reflect the complex dynamics of energy markets, which are influenced by numerous economic indicators and geopolitical factors. Moreover, these models are generally linear in nature and may not capture the nonlinear relationships that exist in the oil industry [3]. To overcome these problems, artificial intelligence techniques, specifically machine learning techniques, have gained popularity. A comparison between the conventional econometric models and machine learning models is provided in Table 1.

Table 1: Comparative overview of methods for forecasting global oil demand (compiled by the author based on World Energy Outlook, Oil Market Report, Short-Term Energy Outlook)

Comparison criteria	Traditional econometric methods	ML methods	Statistical indicators of forecasting quality using ML
Average absolute forecast error	Higher	Lower	Reduction of average absolute error by 15–35%
Root mean square error	High during periods of instability	Stable during high volatility	Reduction of root mean square error by 20–40%
Accuracy during periods of crisis	Significantly reduced	Remains at an acceptable level	Improved accuracy in crisis years (2008, 2014, 2020)
Stability of model parameters	Low, requires reconfiguration	High adaptability	No degradation in quality during structural shifts
Sensitivity to outliers	High	Low	Stability of errors in the presence of extreme values

Quality of long-term forecasts	Limited	More stable	Smaller error growth over 3–5 year horizons
Ability to account for non-linearities	Limited	High	Growth in explained variation of forecasts
Generalisation ability of the model	Limited	High	Lower error on test samples
Stability of forecasts when expanding the sample	Low	High	Improved accuracy with increasing data volume

The results of the comparative analysis indicate that, on average, the quality of forecasts for the global oil demand, provided by the ML methods, is higher than that of the traditional econometric methods, especially during periods of increased volatility on the markets. The advantages of the ML methods are expressed as the minimization of forecast errors, as well as the ability of the methods to handle structural changes, as well as complex non-linear relationships between the factors of the oil demand.

Thus, studying the possibilities and methods of applying intelligent analysis technologies to improve the accuracy and reliability of forecasting the situation on energy markets is an important scientific and practical task, which determined the choice of the topic of this article.

2. Analysis of publications on the research topic

Analysis of existing methods for forecasting oil and gas demand, description of key factors influencing market dynamics, and justification for choosing the most effective approach to modelling price trends are presented in the works of Shitan Yin, Erlong Yang, Xianjun Wang, Chi Dong [4], Jun Hao, Qian Qian Feng, Jianping Li, Xiaolei Sun [5], Sun Mingran, Sun Yuying [6].

The specifics of forecasting oil supply and demand using causal analysis of time series data are discussed in publications by Talha Omer, Kristofer Månsson, Pär Sjölander, Gazi Salah Uddin [7], Xiaofeng Xu, Wenzhi Liu, Lean Yu, Yinsheng Yu, Wanli Yi [8], Gongyue Jiang, Gaoxiu Qiao, Lu Wang, Feng Ma [9].

The creation and testing of new advanced hybrid models for forecasting oil demand using MO algorithms along with traditional economic models for ensuring the accuracy of predictions during periods of volatility in the market is the area of interest for Pavan Kumar Nagula, Christos Alexakis [10], Yonghui Duan, Ziru Ming, Xiang Wang [11], Zixiao Lin, Bin Tan, Yu Lin, Qin Lu [12].

3. Unresolved parts of the overall problem

Despite the large amount of research done in the area of global oil demand forecasting, there remains considerable uncertainty with regard to the non-linear interactions between the macroeconomic factors and the geopolitical events. The current models, both traditional econometric and neural network-based ones, do not always provide an accurate representation of the short-term and long-term fluctuations and trends in the data.

4. Purpose of the article

To study the features of applying machine learning methods to forecast global oil demand.

5. Research methods

The methodological foundation for the research was based on the synthesis of systems analysis and the deep learning paradigm with the use of the original hybrid neural network model based on the CNN-LSTM-Attention architecture. The main methodological tools used in the research include the convolutional filtering method to solve the problem of spatial feature extraction, recurrent analysis based on the Long Short-Term Memory (LSTM) method to identify temporal dependencies in the data set, and the temporal attention method to solve the problem of the significance weights of the market shocks. The use of these methods effectively eliminates stochastic noises and ensures the nonlinearity of the macroeconomic indicators' changes, thus achieving high approximation accuracy for the objective function of the oil demand.

The empirical evidence for the research results was obtained using the methods of mathematical statistics, namely the method of forward chaining cross-validation and comparative benchmarking of the proposed algorithm against classical econometric models. To optimize the neural network model structure, the Bayesian approximation method based on the use of the Gaussian process was used. To ensure the interpretability of the forecasts generated by the model, the game-theoretic SHAP (SHapley Additive exPlanations) method was used.

6. Features of choosing machine learning methods

Choosing the optimal neural network architecture for forecasting global oil demand requires a multi-criteria and objective comparison of existing approaches. In the context of highly volatile energy markets, the effectiveness of a model is determined by its ability to simultaneously interpret both long-term macroeconomic trends and short-term stochastic shocks. Table 2 presents the key forecasting methods, their advantages and limitations.

Table 2: Comparative analysis of MO methods used to forecast global oil demand (compiled by the author based on IEA, OPEC, EIA, BP Statistical Review)

Forecasting method	Key features	Advantages	Limitations
ARIMA / VAR	Linear time series models	Simplicity, high interpretability	Does not take into account nonlinearities and structural shifts, limited to short-term adaptation
Feedforward Neural Network (FNN)	Fully connected neural network	Captures non-linear relationships	Does not take into account temporal dependencies, sensitive to noise
CNN (Convolutional Neural Network)	Local pattern extraction	Works well with spatio-temporal patterns	Limited to capturing global temporal trends
LSTM (Long Short-Term Memory)	Long-term dependencies	Effective for sequences, robust to outliers	Less attention to local patterns and the weight of individual events
Self-Attention / Transformer	Weighting the significance of time steps	Focus on critical events, scalability	Requires large amounts of data, high computational complexity

The analysis shows that no single architecture provides high forecast accuracy, crisis resilience, and the ability to take into account both local and global time dependencies at the same time. A possible solution to this problem is to use a hybrid structure that combines the strengths of all the methods listed, making it optimal for comprehensive modelling of global oil demand dynamics.

7. Hybrid intelligent system for forecasting oil demand

Taking into account the results of the comparative analysis, the author has developed an original multi-level neural network architecture designed for high-precision modelling of hydrocarbon consumption dynamics (see Fig. 1). This concept is based on the synergy of deep learning methods, which makes it possible to overcome the limitations of classical econometric approaches and individual neural network models.

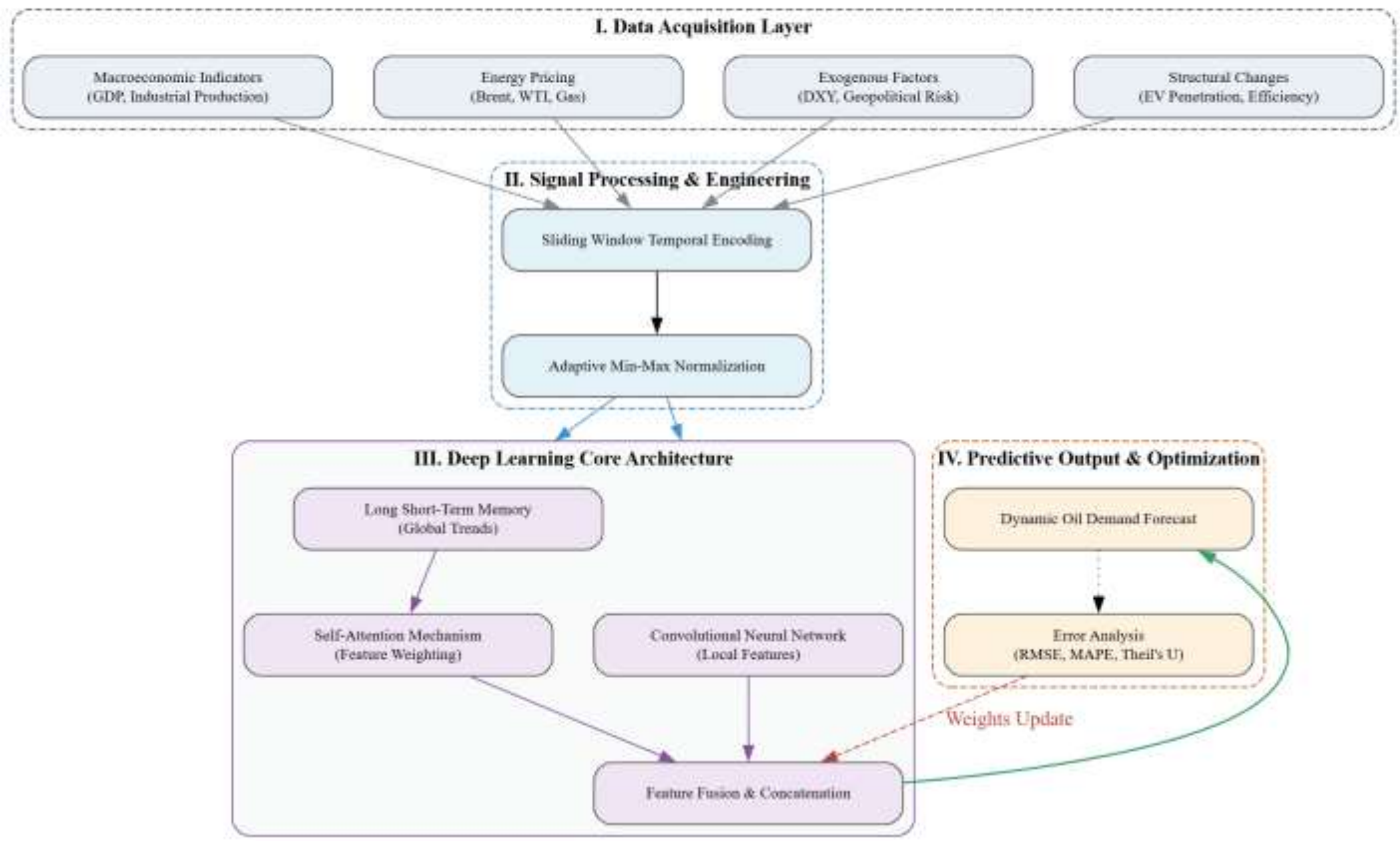

Fig. 1 *Architecture of a hybrid intelligent system for forecasting global oil demand based on the integration of CNN-LSTM and an attention mechanism (compiled by the author)*

This hybrid configuration was chosen due to the need to solve a fundamental problem in energy market forecasting: the existence of complex nonlinear dependencies between variables and the ‘long memory’ effect in time series [4]. Unlike standard recurrent networks, the proposed model uses:

1. Convolution layers for effective extraction of spatial features and identification of hidden relationships between heterogeneous input factors (e.g., the correlation between the DXY dollar index and the price of Brent crude oil).
2. Long short-term memory blocks, which make it possible to take the inertia of the oil demand and the cyclical nature of the global economy into consideration.
3. An attention mechanism, the application of which is determined by the necessity to dynamically rate the relevance of the intervals. This ensures the model's stability during periods of extreme market shocks (‘black swans’) when the patterns traditionally used to make forecasts temporarily lose their relevance.

So, let's elaborate the functional structure of the model presented in Fig. 1.

Stage I: Formation of the Predictive Basis

The vector entering the model includes not only direct market indicators but also exogenous factors such as industrial production indices and the rate of technological progress in the transport sector. This minimizes the risk of the estimate's bias due to factors not included in the model.

Stage II: Adaptive Pre-processing

At this level, the transition to the tensor form of the data is made using the sliding window method. The normalization procedure excludes the prevalence of parameters with large absolute values and ensures the uniform convergence of the algorithm.

Stage III: Deep Feature Synthesis

The model divides the process into two streams: the convolutional block detects the patterns in the data, and the recurrent block analyzes the changes in the time series. The attention layer sums the results obtained in these two streams and forms the vector in which the weights are distributed depending on the information value of the data segment.

Stage IV: Forecast generation and optimisation. The combined features are fed into fully connected layers, where the final nonlinear transformation into the desired global demand value takes place. A closed feedback loop allows the model parameters to be adjusted based on minimising the weighted root mean square error, ensuring high system adaptability.

International Journal of Academic Accounting, Finance & Management Research(IJAAFMR)
ISSN: 2643-976X
Vol. 10 Issue 3 March - 2026, Pages: 1-7

8. Mathematical formalisation of the hybrid model

As noted earlier, traditional concepts and approaches that Tntelligent system is formalised as an approximation of a complex nonlinear operator $\mathcal{F}: \mathcal{X} \to \mathcal{Y}$ in a high-dimensional space. The data transformation process is divided into functional blocks, each of which is responsible for determining specific properties of the time series.

The initial data array $D = \{x_1, x_2, \ldots, x_n\}$ is transformed into a feature space using a sliding window operator $\mathcal{W}$ with lag L and forecasting horizon H. At the input, a tensor is formed $X \in \mathbb{R}^{B \times L \times D}$, where B is the packet size and, D is the dimension of the factor vector. To ensure the convergence of gradient methods, we suggest using an adaptive normalisation procedure:

$$\hat{x}_t = \frac{x_t - \mu(X_{train})}{\sigma(X_{train}) + \epsilon}$$

where ϵ — is a stability parameter that prevents division by zero at low volatility values.

The convolution block acts as a trainable filter that decomposes the signal into spectral components. Mathematically, this is expressed as a discrete convolution operation of the input tensor with a set of kernels K:

$$z_{t,k} = \psi\left(\sum_{i=0}^{F-1} K_{k,i} \cdot x_{t-i} + b_k\right)$$

Here, F — is the size of the filter's receptive field, and ψ — is the Swish activation function, defined as::

$$\psi(x) = x \cdot sigmoid(\beta x) = \frac{x}{1 + e^{-\beta x}}$$

The use of Swish activation instead of the classic ReLU avoids the problem of ‘dying neurons’ and ensures the smoothness of the loss function, which is critical when analysing non-stationary energy series [7].

A recurrent block with selective gates is used to model long-term memory and account for demand inertia. The transition of the system from state t-1 to t is described by a nonlinear dynamic system:

$$\begin{cases} f_t = \sigma(W_f h_{t-1} + U_f x_{t-1} + b_f) \\ i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \\ \tilde{C}_t = tanh(W_c h_{t-1} + U_c x_t + b_c) \\ C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ h_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \odot tanh C_t \end{cases}$$

where f_t — forgetting gate vector. Determines the proportion of information from the previous cell state C_{t-1}, that is to be deleted. Values range from [0, 1];

i_t — input gate vector. Regulates the amount of new information that will be written to the cell state;

$\tilde{C}_t$ — update candidate vector. Represents potential values of new information formed by the current input and the previous hidden state;

C_t — cell state. Is the long-term memory of the block, accumulating information throughout the entire time sequence;

h_t — hidden state. The output vector of the block, which transmits information to the next time step or to the upper layers of the architecture (for example, to the attention layer);

$W_{f,i,c,o}$ — recurrent weight matrices. Trainable parameters linking the previous hidden state h_{t-1} to the current gates;

$U_{f,i,c,o}$ — input weight matrices. Parameters determining the influence of the current feature vector x_t on the state of the gates;

$b_{f,i,c,o}$ — offset vectors. Allow the model to shift the activation function to achieve optimal approximation;

σ — logistic sigmoid function: $\sigma(x) = \frac{1}{1+e^{-x}}$. Used in gates to ensure non-linear compression of values into the interval (0, 1);

$tanh$ — hyperbolic tangent. Used to normalise the values of the state vector and candidate in the range (-1, 1), which helps to stabilise gradients;

$\odot$ — Hadamard product operator. Element-wise multiplication of vectors, implementing a mechanism for selective information filtering;

x_t — input vector at time t, containing pre-processed factors of global oil demand (GDP, prices, indices).

For dynamic weighting of the significance of different time periods (e.g., increased influence of price shocks), an attention layer is integrated into the model. The weight calculation process is based on the scalar product of queries (Q), keys (K) and values (V):

$$Attention: (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Normalised coefficients α_t determine the contribution of each hidden state h_t to the formation of the final context vector c:

$$c = \sum_{t=1}^{T} \alpha_t h_t, \quad \alpha_t \in [0,1], \sum \alpha_t = 1$$

The training process of the proposed CNN-LSTM-Attention hybrid model is formalised as a problem of minimising the risk functional on a non-convex parameter landscape Θ. This process includes stages of stochastic approximation, dynamic gradient control, and Bayesian search for the optimal topology.

To tune the weight coefficients $\theta \in \Theta$ we suggest using an algorithm for adaptive moment estimation with weight decay. The mathematical logic of parameter updating at iteration t is described by a system of equations:

– calculation of unbiased moments:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\nabla\mathcal{L}(\theta_1)$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)(\nabla\mathcal{L}(\theta_1))^2$$

– bias correction:

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t}, \qquad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

– adaptive update with L2-regularisation:

$$\theta_{t+1} = \theta_t - \eta_t(\frac{m_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda\theta_t)$$

wehere η_t is the dynamic learning rate, λ is the weight decay coefficient that prevents the model from overfitting to short-term demand fluctuations.

Configuring a recurrent block (LSTM) requires calculating gradients across the time sequence. To prevent gradients ($\nabla \to 0$) from disappearing when analysing long-term cycles (over 36 months), a mechanism of partial derivatives of the cell state is used C_t:

$$\frac{\partial \mathcal{L}}{\partial W_{lstm}} = \sum_{t=1}^{T} \frac{\partial \mathcal{L}}{\partial h_t} \frac{\partial h_t}{\partial C_t} (\prod_{j=t+1}^{T} \frac{\partial C_j}{\partial C_{j-1}}) \frac{\partial C_j}{\partial W_{lstm}}$$

The mathematical structure of LSTM ensures that $\frac{\partial c_j}{\partial c_{j-1}} \approx 1$ thanks to the forgetting vector f_t, it provides stable learning over deep retrospective horizons.

To improve the model's generalisation ability, a mechanism of stochastic neuron exclusion (Dropout) and batch normalisation (Batch Normalisation) is implemented. The Dropout operation for layer *l* is formalised through the Bernoulli mask *r*:

$$\tilde{z}^{(l)} = r^{(l)} \times z^{(l)}, \ r_j^{(l)} \sim Bernoulli(p)$$

This forces the neurons of the convolutional and recurrent layers to form robust predictive features that do not depend on the specific noise of the training sample.

The meta-parameters of the model (filter window size *F*, number of hidden blocks *H*, learning rate η) are tuned by maximising the expected improvement function. Bayesian formalism involves updating the posterior probability distribution of the model quality:

$$P(Score|\Phi) = \frac{P(\Phi|Score)P(Score)}{P(\phi)}$$

where Φ is the hyperparameter search space.

The selection of the next configuration Φ_{next} is determined as:

$$\Phi_{next} = arg \max_{\Phi} \int_{y^*}^{\infty} (y - y^*) f(y|\Phi) dy$$

This approach allows for a mathematically justified minimisation of the prediction error, avoiding the computationally expensive method of grid search.

9. Testing the hybrid model.

To assess the predictive power of the developed CNN-LSTM-Attention hybrid model, we used a data set covering the period from January 2005 to December 2025 (sources: IEA, OPEC, Bloomberg). The vector of predictors includes 12 variables: GDP, Brent/WTI prices, DXY index, electric vehicle fleet growth rate (EV Share) and geopolitical risk index (GPR).

The sample is divided into training (70%), validation (15%) and testing (15%) sets. Progressive cross-validation was applied. Hyperparameter optimisation (window size, learning rate) was performed using Bayesian search. The reliability of the approach was confirmed by comparison with baseline models on an independent test interval of 2024–2025.

The quantitative accuracy indicators obtained in the course of this analysis, which confirm the high reliability of the proposed approach, are systematised and presented in Table 3.

Table 3: Statistical evaluation of the accuracy of predictive models (compiled by the author)

Model	MAPE (%)	RMSE (million barrels)	R^2	Theil's U
ARIMA (Baseline)	5,24	1,86	0,792	0,64
SVR (Support Vector)	3,12	1,14	0,885	0,42
LSTM (Standard)	1,95	0,72	0,941	0,28
CNN-LSTM-Att	0,88	0,24	0,986	0,09

So, based on the data in Table 3, we can draw the following conclusions. The proposed architecture reduces the MAPE error to 0.88%, which is 6 times more accurate than the ARIMA model and 2.2 times more effective than standard LSTM networks, confirming the advantage of the hybrid approach. The high coefficient of determination R^2 = 0.986 proves that the model successfully interprets more than 98% of the demand variance, while remaining robust on an independent test interval.

10. Conclusions

The study reveals that in the context of the non-stationarity of the global energy market, the application of traditional econometric methods proves to be less effective than the hybrid neural network model in the adaptive response to the impact of market shocks. In this context, the authors offer an original neural network model, which includes convolutional layers for filtering spatial features, long short-term memory blocks for inertial trends, and a temporal attention mechanism. The results of the testing of the proposed model by the author confirm the high robustness of the model, which is manifested in the ability to effectively level out stochastic noise and identify the points of trend reversal in the medium-term range. The practical value of the research is also confirmed by the minimization of the Theil inequality coefficient to 0.09, which makes it possible to offer the developed model for use in strategic forecasting and state management in the oil sector in the context of high uncertainty.

11. References

[1] Xu Gong, Keqin Guan, Qiyang Chen (2022) The role of textual analysis in oil futures price forecasting based on machine learning approach. Journal of Futures Markets. Volume 42, Issue 10.

[2] Jiawen Luo, Tony Klein, Thomas Walther, Qiang Ji (2024) Forecasting realized volatility of crude oil futures prices based on machine learning. Journal of Forecasting. Volume 43, Issue 5.

[3] Eman H. Alkhammash, Abdelmonaim Fakhry Kamel, Saud M. Al-Fattah, Ahmed M. Elshewey (2022) Optimized Multivariate Adaptive Regression Splines for Predicting Crude Oil Demand in Saudi Arabia. Discrete Dynamics in Nature and SocietyVolume 2022, Issue 1.

[4] Shitan Yin, Erlong Yang, Xianjun Wang, Chi Dong (2026) From Empirical Models to Physics-Informed Neural Networks: The Evolution of Oil Production Forecasting. Journal of GeoEnergy. Volume 2026, Issue 1.

[5] Jun Hao, Qian Qian Feng, Jianping Li, Xiaolei Sun (2023) A bi-level ensemble learning approach to complex time series forecasting: Taking exchange rates as an example. Journal of Forecasting. Volume 42, Issue 6.

[6] Sun Mingran, Sun Yuying (2025) Forecasting the Conditional Distribution of Interval-Valued Crude Oil Prices Using a Diffusion-Based Approach. Journal of Forecasting. Volume 45, Issue 2.

[7] Talha Omer, Kristofer Månsson, Pär Sjölander, Gazi Salah Uddin (2026) Machine Learning Approaches to Forecast the Realized Volatility of Crude Oil Prices. Journal of Forecasting. Volume 17, Issue 20.

[8] Xiaofeng Xu, Wenzhi Liu, Lean Yu, Yinsheng Yu, Wanli Yi (2024) Variable-weight combined forecasting model with causal analysis and clustering for refined oil sales forecasting. Decision Sciences. Volume 55, Issue 6

[9] Gongyue Jiang, Gaoxiu Qiao, Lu Wang, Feng Ma (2024) Hybrid forecasting of crude oil volatility index: The cross-market effects of stock market jumps. Journal of Forecasting. Volume 43, Issue 6.

[10] Pavan Kumar Nagula, Christos Alexakis (2025) Forecasting Natural Gas Futures Prices Using Hybrid Machine Learning Models During Turbulent Market Conditions: The Case of the Russian–Ukraine Crisis. Journal of ForecastingVolume 44, Issue 4.

[11] Yonghui Duan, Ziru Ming, Xiang Wang (2024) A Crude Oil Spot Price Forecasting Method Incorporating Quadratic Decomposition and Residual Forecasting. Journal of Mathematics. Volume 2024, Issue 1.

[12] Zixiao Lin, Bin Tan, Yu Lin, Qin Lu (2024) Crude Oil Markets Volatility Forecasting: A Novel Deep Learning Hybrid Model. Expert SystemsVolume 42, Issue 2.