Research in Education, Technology, and Multiculture
ISSN: 3025-6763 | Journal's DOI: 10.61436/rietm
Homepage: https://imrecsjournal.com/journals/index.php/rietm

Hybrid LSTM–XGBoost Model with Residual Error Correction for Multivariate Gold Price Forecasting Using Macroeconomic Indicators

Andros Juan Brando Nainggolan*, **Yogi Austin Hizkia Hutajulu, Kevin,**
Marlince Novita Karoseri Nababan
Department of Information Systems, Universitas Prima Indonesia, Indonesia

ARTICLE INFO
Article History:
Received: 07 May 2026
Accepted: 10 June 2026
Published: 16 June 2026

Keywords:
Gold price forecasting;
Hybrid LSTM–XGBoost;
Multivariate forecasting;
Residual learning;
Time series forecasting;
Walk-forward validation.

Corresponding Author:
androsjuanbn@gmail.com

ABSTRACT

Gold plays a critical role in financial markets and is widely regarded as a hedge and safe-haven asset during periods of economic uncertainty. Accurate gold price forecasting is therefore essential for investment strategy, portfolio allocation, and risk management. This study proposes a hybrid forecasting framework that integrates Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) through explicit residual error correction for multivariate gold price prediction. Monthly gold prices and selected macroeconomic indicators, including CPI, DXY, US10Y, WTI, and S&P 500, covering the period from 2010 to 2025, are employed. The dataset consists of 192 monthly observations. Prior to modeling, logarithmic transformation, stationarity testing using the Augmented Dickey–Fuller (ADF) test, first-order differencing, and Min–Max normalization are applied to ensure statistical validity and numerical stability. The LSTM component captures temporal dependencies in sequential data, while the XGBoost model nonlinearly models residual structures to enhance predictive performance. A 12-month sliding-window mechanism is employed to capture annual temporal dependencies, and the XGBoost component is trained to learn residual errors not explained by the LSTM forecasts. Empirical results demonstrate that the proposed hybrid model achieves superior accuracy compared to standalone LSTM, XGBoost, and ARIMA baselines, obtaining an RMSE of 0.0989, an MAE of 0.0691, and a MAPE of 0.8503 in the testing period. Diebold–Mariano tests confirm that the hybrid model significantly outperforms both LSTM and ARIMA ($p < 0.01$). Walk-forward validation further indicates stable forecasting performance across rolling evaluation windows. The validation results demonstrate consistent predictive accuracy across multiple rolling windows, supporting the robustness and generalizability of the proposed framework. These findings suggest that integrating temporal learning with structured residual correction provides a robust, statistically grounded approach to multivariate gold price forecasting.

To cite this article:
Nainggolan et al. (2026). Hybrid LSTM–XGBoost model with residual error correction for multivariate gold price forecasting using macroeconomic indicators. *Research in Education, Technology, and Multiculture, 5*(1), 60-74. doi: https://doi.org/10.61436/rietm/v5i1.pp60-74

■ INTRODUCTION

Gold has long been regarded as a strategic financial asset and is frequently considered a hedge and safe-haven instrument during periods of economic instability and financial turbulence (Baur & McDermott, 2010; Pendaraki & Charda, 2025). Empirical evidence continues to confirm the role of gold in mitigating portfolio risk and preserving value during market stress and macroeconomic uncertainty (Echaust & Just, 2022; Ameur et al., 2024). Given its sensitivity to inflation, exchange rate movements, interest rate dynamics, and global commodity fluctuations, gold price behavior reflects complex interactions among macroeconomic and financial variables. Accurate gold price forecasting is therefore essential for investment decision-making, portfolio diversification, and financial risk management, as well as for supporting institutional and policy-related economic assessments (Taneva-Angelova et al., 2025). Previous empirical studies have modeled gold price movements using multiple macroeconomic indicators, including inflation measures, currency indices, crude oil prices, and stock market indices (Yi et al., 2023; Zangana & Obeyd, 2024; Quang & Thang, 2025). These studies highlight the nonlinear dependencies and temporal structures inherent in multivariate financial time series. Such characteristics make forecasting gold prices challenging, particularly as interactions among explanatory variables evolve.

Recent evidence suggests that gold continues to play an important role as both a hedge and a safe-haven asset amid increasing economic uncertainty and geopolitical tensions. Echaust and Just (2022) show that gold remains effective in reducing extreme downside risk during major financial crises and continues to offer diversification benefits in periods of severe market stress. Similarly, Ameur et al. (2024) find that gold retains significant hedge and safe-haven characteristics during episodes of heightened financial and geopolitical uncertainty, supporting its role as a defensive asset within diversified investment portfolios. More recently, Pendaraki and Charda (2025) highlight the dynamic relationship between gold and stock markets during crises, emphasizing its relevance to portfolio protection and risk management. Although Faraj et al. (2025) suggest that rising volatility in the gold market may weaken its traditional safe-haven function under certain conditions, their findings nevertheless underline the continued importance of gold in modern financial systems. Taken together, these studies provide contemporary evidence that gold remains a valuable instrument for preserving value and mitigating risk across changing macroeconomic environments, further reinforcing the importance of accurate gold price forecasting.

Traditional statistical models, most notably the Autoregressive Integrated Moving Average (ARIMA), have been widely applied in time series forecasting due to their solid theoretical foundation and effectiveness in modeling linear temporal structures (Box et al., 2016; Madhika et al., 2023). In gold price prediction, ARIMA has often served as a benchmark model for comparison with more advanced approaches (Madhika et al., 2023). However, financial time series frequently exhibit nonlinear dynamics and complex cross-variable interactions that may not be fully captured by purely linear models (Makala & Li, 2021). Recent comparative studies suggest that machine learning and deep learning approaches can provide improved predictive performance when nonlinear relationships dominate the data-generating process (Kontopoulou et al., 2023). Among deep learning architectures, Long Short-Term Memory (LSTM) networks have gained significant attention for financial time series forecasting (Hochreiter & Schmidhuber, 1997; Nensi et al., 2025). LSTM models are designed to capture long-term temporal dependencies and mitigate vanishing-gradient issues in recurrent neural networks, making them well-suited for modeling sequential financial data. Several studies have demonstrated the effectiveness of LSTM-based models in gold price and stock index forecasting (Nensi et al., 2025; Mun et al., 2025). Nevertheless, despite their strength in modeling sequential patterns, LSTM models may still leave residual prediction errors when complex nonlinear feature interactions are not fully represented through temporal learning alone (Hewamalage et al., 2021).

Recent developments in gold price forecasting have increasingly focused on hybrid deep-learning approaches to better capture the nonlinear and highly dynamic nature of financial time series (Saini et al., 2025; Zhao et al., 2025). Saini et al. (2025) demonstrate that hybrid neural-network architectures can achieve superior forecasting performance compared to standalone models by combining complementary learning mechanisms within a unified framework. Their findings suggest that integrating different modeling approaches can improve predictive stability and better capture complex interactions among commodity prices and macroeconomic variables. These developments indicate a growing interest in hybrid forecasting frameworks as a practical approach to addressing the challenges of multivariate financial time-series prediction.

Extreme Gradient Boosting (XGBoost), an ensemble learning algorithm based on gradient-boosted decision trees, has also demonstrated strong predictive performance across various financial applications (Chen & Guestrin, 2016; Tan et al., 2023). Tree-based boosting models are particularly effective at capturing nonlinear interactions among features and heterogeneous effects across explanatory variables. However, unlike recurrent neural networks, XGBoost does not inherently model temporal dependencies when applied directly to time series data. To leverage complementary strengths, hybrid forecasting approaches integrating statistical, machine learning, and deep learning models have been increasingly explored (Qiu et al., 2024; Zhou, 2025). Hybrid frameworks often aim to combine temporal dependency learning with nonlinear feature modeling, resulting in improved predictive stability compared to standalone approaches (Nasir et al., 2025). In particular, residual-based hybrid strategies in which one model captures the primary structure, and another learns the residual error component, have shown promising performance improvements in several forecasting contexts (Bechere et al., 2025).

Although recent studies have reported encouraging results from hybrid forecasting models, several methodological challenges remain. Existing hybrid frameworks primarily focus on improving forecasting accuracy. At the same time, the residual learning process is often incorporated into broader model architectures without clearly specifying how residual errors are generated, modeled, and reintegrated into the final forecasting output (Qiu et al., 2024; Saini et al., 2025). As a result, the specific contribution of residual learning to forecasting improvement is not always clearly demonstrated. In addition, model evaluation in many studies relies mainly on conventional accuracy metrics, whereas formal statistical significance testing and robustness assessment under changing market conditions remain relatively limited. Consequently, there remains a need for a forecasting framework that not only improves predictive accuracy but also provides clearer methodological transparency and more comprehensive evaluation procedures.

Building on these methodological concerns, several limitations remain particularly relevant to research on multivariate gold price forecasting. First, explicit residual-error correction mechanisms within LSTM–XGBoost frameworks are not consistently implemented in a structured, transparent manner. Second, many forecasting studies rely solely on conventional accuracy metrics and do not conduct formal statistical significance testing to validate performance differences, even though robust forecast evaluation methodologies have been emphasized in recent literature (Cerqueira et al., 2020). Third, robustness analysis via rolling walk-forward validation is often limited, providing limited insight into model stability under evolving market conditions. Motivated by these gaps, this study aims to develop and rigorously evaluate a structured hybrid LSTM–XGBoost framework for multivariate gold price forecasting. The proposed approach applies a logarithmic transformation to stabilize the variance of gold prices, employs an LSTM to capture temporal dynamics, and uses XGBoost to model residual errors via nonlinear correction explicitly. Unlike many prior studies, model performance is assessed not only using multiple forecasting accuracy metrics but also through Diebold–Mariano statistical testing (Diebold & Mariano, 1995) and rolling walk-forward validation to ensure both predictive superiority and temporal robustness. By integrating sequential learning, structured residual correction, and rigorous statistical evaluation, this study contributes a methodologically grounded framework for multivariate gold price forecasting. It provides empirical evidence on the effectiveness of hybrid modeling strategies in complex financial time-series environments.

To address these identified limitations, the proposed framework combines LSTM-based temporal learning with XGBoost-based residual correction. It evaluates forecasting performance using both Diebold–Mariano statistical testing and rolling walk-forward validation. Through this approach, the study provides a more transparent, robust, and empirically grounded framework for multivariate gold price forecasting.

■ METHOD

Research Design and Procedures

This study proposes a structured hybrid forecasting framework integrating Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) for multivariate gold price prediction. The framework is designed to combine temporal dependency modeling with nonlinear residual correction while maintaining rigorous evaluation procedures.

As illustrated in Figure 1, the proposed research framework consists of several sequential stages, beginning with data collection and preprocessing, followed by model development and comprehensive validation.

The overall research procedure consists of five main stages. First, monthly gold price and macroeconomic data are collected from publicly available financial databases. Second, the data are preprocessed through logarithmic transformation, stationarity testing, normalization, and sequence construction. Third, ARIMA, LSTM, XGBoost, and hybrid LSTM–XGBoost models are developed and trained using the training dataset. Fourth, forecasting performance is evaluated using RMSE, MAE, and MAPE metrics. Finally, statistical and robustness validation are conducted using the Diebold–Mariano test and rolling walk-forward validation.

The framework utilizes gold price as the target variable. It incorporates several macroeconomic indicators, namely the Consumer Price Index (CPI), the US Dollar Index (DXY), the US Treasury Yield (US10Y), crude oil prices (WTI), and the S&P 500 index. CPI represents inflationary pressure, DXY reflects fluctuations in the value of the US dollar, US10Y captures long-term interest rate expectations, WTI serves as an indicator of global energy market conditions, and the S&P 500 index represents stock market performance and investor sentiment. These variables are

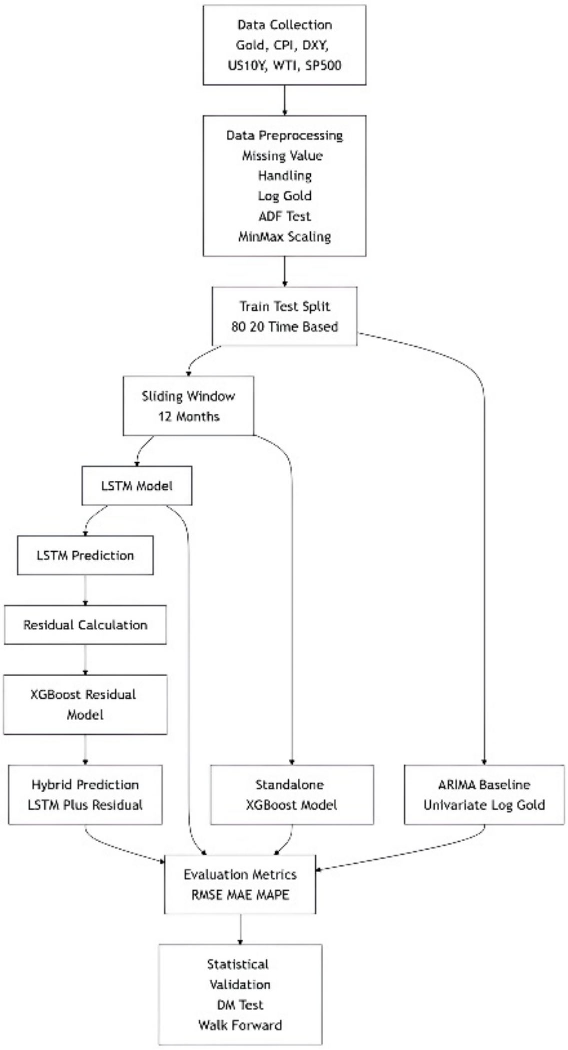

Figure 1. Research Framework of the Proposed Hybrid LSTM-XGBoost Model

included to capture multiple economic channels influencing gold price dynamics within a multivariate forecasting framework. The process starts with the collection of monthly gold price data and the selected explanatory variables described above. These variables were selected based on prior empirical findings identifying macroeconomic drivers of gold price dynamics (Yi et al., 2023; Zangana & Obeyd, 2024).

During preprocessing, missing values are handled, and a logarithmic transformation is applied to the gold price series to stabilize variance and reduce heteroskedasticity, which is commonly observed in financial time series (Box & Cox, 1964). Stationarity characteristics of the log-transformed series are examined using the Augmented Dickey–Fuller (ADF) test (Dickey & Fuller, 1979). Subsequently, Min–Max normalization is applied to the training data and extended to the test set to prevent information leakage and ensure numerical stability during model training (Lim & Zohren, 2021). The dataset is divided chronologically into training and testing sets using an 80:20 time-based split.

For deep learning and machine learning models, a 12-month sliding window is used to generate multivariate input sequences (Zhan & Kim, 2024). Three modeling strategies are implemented for comparative analysis:

1. A univariate ARIMA model applied to the log-transformed gold price series as a classical statistical baseline.
2. Standalone LSTM and standalone XGBoost models for nonlinear multivariate forecasting.
3. A hybrid LSTM–XGBoost model employing structured residual error correction.

In the hybrid architecture, the LSTM model first generates baseline predictions. The residual errors between actual and predicted values are then modeled using XGBoost to capture nonlinear residual patterns (Bechere et al., 2025). The final hybrid forecast is obtained by summing the LSTM output and the predicted residual component.

Forecasting performance is evaluated using RMSE, MAE, and MAPE metrics. To assess the statistical significance of performance differences, the Diebold–Mariano (DM) test is conducted (Diebold & Mariano, 1995). Furthermore, rolling walk-forward validation with iterative retraining is implemented to evaluate temporal robustness under dynamic market conditions. This structured framework enables systematic comparison between linear statistical modeling, standalone nonlinear approaches, and residual-based hybrid integration.

Data collection and preprocessing

The dataset employed in this study comprises monthly observations from January 2010 to December 2025, totaling 192 observations. Although the dataset is relatively small for deep learning applications, the use of a compact single-layer LSTM architecture and rolling walk-forward validation mitigates the risk of overfitting. The multivariate framework includes gold price as the target variable, along with selected macroeconomic indicators: the Consumer Price Index (CPI), the US Dollar Index (DXY), the US Treasury Yield (US10Y), the crude oil price (WTI), and the S&P 500 index. These variables were selected based on prior empirical evidence highlighting their influence on gold price dynamics and financial market interactions (Yi et al., 2023; Zangana & Obeyd, 2024).

The selected explanatory variables represent key macroeconomic and financial indicators associated with gold price movements. CPI is used as a proxy for inflation; DXY represents the strength of the US dollar; US10Y reflects long-term interest-rate expectations; WTI captures global energy market conditions; and the S&P 500 index represents stock-market performance and investor sentiment. These variables were selected to capture multiple economic channels that may influence gold price dynamics.

Gold price, DXY, WTI, and S&P 500 data were obtained from Investing.com, while CPI and US10Y were retrieved from the Federal Reserve Economic Data (FRED) database. All variables were aligned to a monthly frequency and chronologically ordered to preserve temporal consistency. To stabilize variance and reduce heteroskedasticity commonly observed in financial time series, a logarithmic transformation was applied to the gold price series. Log transformation is widely used in time series analysis, as a special case of the Box–Cox transformation, to improve variance stability and to interpret proportional changes more effectively (Box & Cox, 1964). The forecasting models in this study are therefore developed using log-transformed gold prices.

The selected data sources are widely used in financial and economic research due to their reliability and consistency. Investing.com aggregates historical market data from major financial exchanges and data providers. At the same time, the Federal Reserve Economic Data (FRED) database, maintained by the Federal Reserve Bank of St. Louis, serves as an authoritative source of macroeconomic and financial indicators. The use of these established databases helps ensure data validity, consistency, and reproducibility throughout the forecasting process.

Stationarity was formally examined using the Augmented Dickey–Fuller (ADF) test (Dickey & Fuller, 1979). The results indicate that the log-level gold price series is non-stationary ($p > 0.05$), suggesting the presence of a unit root. After applying first-order differencing, the ADF test confirms stationarity ($p < 0.01$). These findings justify using first differencing ($d = 1$) in the ARIMA specification and ensure compliance with the linear time-series modeling assumptions. Prior to model training, all variables were normalized using Min–Max scaling. Importantly, the scaling parameters were fitted exclusively on the training set and subsequently applied to the testing set to prevent information leakage. It is important to note that observations in the testing set that fall outside the minimum–maximum range observed in the training data are not removed or clipped. Instead, the scaling transformation is applied using the parameters estimated from the training set, allowing scaled values to exceed the conventional [0,1] interval when necessary. This approach preserves the original information contained in previously unseen observations and reflects realistic out-of-sample forecasting conditions. This procedure ensures unbiased out-of-sample evaluation and numerical stability during model optimization. The dataset was divided chronologically using an 80:20 time-based split, with 153 observations allocated to training and 38 to testing. Random shuffling was avoided to preserve the sequential structure.

To construct supervised learning inputs for the LSTM model, a sliding-window mechanism with 12 time steps was applied, enabling the model to capture one-year temporal dependencies in the multivariate series. The selection of a 12-month sliding window is supported by the autocorrelation structure of the log-transformed gold price series. As shown in Figure 2, the Autocorrelation Function (ACF) analysis indicates persistent positive temporal dependence across multiple lags, with strong autocorrelation observed during the first 12 months. Therefore, a 12-step input window was adopted to capture one-year temporal dependencies while maintaining a compact model architecture suitable for the available sample size.

All computational procedures were conducted in a Python-based environment using Google Colaboratory to ensure reproducibility and numerical consistency. The LSTM model was implemented using the TensorFlow Keras API, while gradient boosting regression was performed using the XGBoost Python library.

Equipment and Parameters

This study was conducted using a Python-based computational environment implemented in Google Colaboratory to ensure reproducibility and numerical consistency. The LSTM model was developed using the TensorFlow Keras API, while the residual learning component was implemented with the XGBoost Python library. Time-series preprocessing and statistical analyses were performed using Pandas, NumPy, and Statsmodels libraries.

The main model parameters consisted of a 12-month sliding-window configuration, 64 LSTM memory units, a batch size of 16, and 50 training epochs. For the XGBoost residual

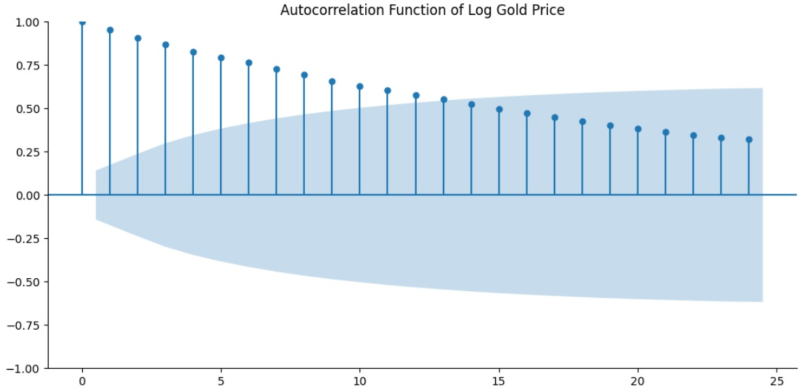

Figure 2. Autocorrelation Function (ACF) Plot of the Log-Transformed Gold Price Series

model, the hyperparameters were set to 200 trees (*n_estimators* = 200), a learning rate of 0.05, a maximum tree depth of 3, and a fixed random state of 42 to ensure reproducibility. These parameter settings were selected based on preliminary experimentation and prior empirical studies to balance predictive accuracy, computational efficiency, and model generalization capability. These evaluation procedures were implemented to assess predictive accuracy, statistical significance, and temporal robustness of the forecasting models.

Data Analysis

ARIMA model

The Autoregressive Integrated Moving Average (ARIMA) model was used as a classical statistical benchmark for evaluating the performance of machine learning and hybrid approaches. ARIMA integrated autoregressive (AR), differencing (I), and moving average (MA) components to model linear temporal dependencies in time series data (Box et al., 2016).The general ARIMA (*p*, *d*, *q*) formulation is expressed as:

$$\boldsymbol{\phi_p(B)(1-B)^d y_t = \theta_q(B)\varepsilon_t}$$

where $\phi_p(\mathcal{B})$ denotes the autoregressive (AR) polynomial, $\theta_q(\mathcal{B})$ represents the moving-average (MA) polynomial, B is the backshift operator, d is the order of differencing, and εt is the random error term.

As discussed in the data collection section, the Augmented Dickey–Fuller (ADF) test indicates that the log-transformed gold price series is nonstationary in levels but becomes stationary after first differencing. Accordingly, first-order differencing (d = 1) is applied to satisfy the stationarity requirement of the ARIMA framework.

Model order selection was conducted using the Akaike Information Criterion (AIC), and the optimal specification obtained in this study is ARIMA (0,1,0). This configuration corresponds to a first-difference random walk model, implying that the current value of the log gold price is primarily explained by its immediately preceding value plus a stochastic disturbance. It is important to note that the ARIMA model in this study is implemented in a univariate setting using only the log-transformed gold price series. Consequently, it does not incorporate macroeconomic predictors such as the CPI, DXY, US10Y, WTI, or the S&P 500. The ARIMA forecasts therefore represent a purely linear historical baseline against which the multivariate LSTM and hybrid LSTM–XGBoost models are compared.

The use of a univariate ARIMA benchmark is intended to provide a classical statistical baseline that reflects forecasting performance based solely on historical gold price information. The objective of this comparison is not to establish feature parity with the multivariate models, but rather to evaluate the extent to which additional macroeconomic information and nonlinear learning mechanisms improve predictive performance beyond conventional time-series forecasting. Therefore, ARIMA is employed as a historical linear benchmark. At the same time, the multivariate LSTM and hybrid LSTM–XGBoost models are designed to assess the added value of exogenous information and nonlinear learning.

Long Short-Term Memory (LSTM) model

Long Short-Term Memory (LSTM) is a variant of Recurrent Neural Network (RNN) specifically designed to capture long-term dependencies in sequential data while mitigating the vanishing gradient problem commonly encountered in conventional RNN architectures (Hochreiter & Schmidhuber, 1997). Through its gated memory structure, the LSTM can selectively retain, update, and forget information across time steps, making it well-suited for modeling nonlinear temporal dynamics in financial time series.

At each time step t, the LSTM unit updates its hidden representation based on the current input vector and previous internal states. The general hidden state update can be expressed as:

$$h_t = LSTM(x_t, h_{t-1}, c_{t-1})$$

where x_t denotes the input vector at time t, h_t represents the hidden state, and c_t is the cell state that carries long-term memory information.

The final one-step-ahead prediction generated by the LSTM model is obtained through a linear transformation of the hidden state:

$$\hat{y}_{LSTM,t} = W_y h_t + b_y$$

where W_y and b_y are trainable output parameters.

In this study, the LSTM model is implemented within a multivariate forecasting framework. The input sequences consist of log-transformed gold prices and selected macroeconomic indicators (CPI, DXY, US10Y, WTI, and S&P 500). A sliding-window approach with 12 time steps is applied to construct supervised learning sequences, enabling the model to capture one-year temporal dependencies before forecasting the next monthly observation. Because sequence generation requires 12 prior observations as model inputs, the 39 observations allocated to the testing set produced 27 valid testing sequences used for final model evaluation. The implemented architecture comprises a single LSTM layer with 64 memory units, followed by a fully connected output layer for one-step-ahead regression. The model is trained using the Adam optimizer with mean squared error (MSE) as the loss function. Training is conducted for 50 epochs with a batch size of 16. During training, 10% of the training data were reserved as an internal validation set (validation_split = 0.1) to monitor model convergence and potential overfitting by evaluating validation loss. To ensure unbiased evaluation, model fitting is performed exclusively on the training set, and predictions are generated on the hold-out testing set.

Unlike the ARIMA benchmark, the LSTM model incorporates multivariate inputs and nonlinear representations, enabling it to capture complex interactions among macroeconomic variables and the temporal dynamics of gold price movements. Hyperparameters were selected based on preliminary experimentation and prior empirical studies to balance predictive accuracy and generalization capability.

The selection of 64 LSTM memory units was based on preliminary experimentation and prior empirical studies. A larger number of memory units may increase model complexity and the risk of overfitting given the relatively limited sample size, whereas smaller architectures may reduce predictive capability. Therefore, 64 units were adopted as a compromise between representational capacity, computational efficiency, and model generalization performance.

Extreme gradient boosting (xgboost)

Extreme Gradient Boosting (XGBoost) is an ensemble learning algorithm based on gradient boosting decision trees that sequentially constructs regression trees to minimize prediction errors (Chen & Guestrin, 2016). The model builds additive tree-based learners in a stagewise manner, with each subsequent tree attempting to correct the errors produced by the preceding ensemble.The prediction function of XGBoost can be expressed as:

$$\hat{\boldsymbol{y}}_{\text{XGB},\boldsymbol{t}} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{f_k}(\boldsymbol{x_t})$$

where f_k denotes an individual regression tree, K represents the total number of trees, and x_t corresponds to the input feature vector at time t.. Through iterative error minimization, XGBoost effectively captures nonlinear feature interactions and complex residual patterns (Chen & Guestrin, 2016; Tan et al., 2023).

In this study, the XGBoost component is implemented as a regression model (XGBRegressor) with 200 trees (n_estimators = 200), a learning rate of 0.05, and a maximum tree depth of 3. A fixed random state of 42 is used to ensure reproducibility. These hyperparameters are selected to balance predictive performance and model stability without excessive complexity. The selection of 200 trees and a maximum tree depth of 3 was based on preliminary experimentation and prior empirical evidence. A larger number of trees and deeper tree structures may improve training performance but increase computational complexity and the risk of overfitting, particularly when modeling residual series with a limited number of observations. Therefore, the selected configuration was adopted to balance predictive capability, model stability, and generalization performance. It is important to note that XGBoost does not inherently model sequential dependencies in time series data. Therefore, in the proposed framework, XGBoost is not used as the primary forecasting model. Instead, it is employed to learn the residual errors generated by the LSTM model. The residual term is defined as the difference between the actual log gold price and the LSTM prediction. By modeling this residual structure, XGBoost performs nonlinear error correction on top of the temporal patterns captured by LSTM.

To be compatible with the tree-based learning structure, the multivariate input sequences constructed for the LSTM are reshaped from three-dimensional tensors into two-dimensional feature matrices before being provided to XGBoost. This transformation preserves the temporal window information while enabling training of tree-based regression models. Through this residual learning mechanism, the XGBoost component enhances the hybrid architecture by addressing nonlinear discrepancies that the sequential LSTM model does not fully capture. These hyperparameters were determined through exploratory tuning to achieve stable performance without excessive model complexity. Hyperparameter tuning was conducted exclusively on the training dataset, using the internal validation split specified in the LSTM training configuration (validation_split = 0.1). The testing dataset was not used during model selection or hyperparameter adjustment and was reserved solely for final out-of-sample performance evaluation.

Hybrid LSTM-XGBoost model

Hybrid forecasting models that combine deep learning and machine learning techniques have demonstrated improved predictive performance by leveraging complementary strengths (Qiu et al., 2024; Zhou, 2025). In particular, sequential neural networks are effective in capturing temporal dependencies, while tree-based ensemble methods are well-suited for modeling nonlinear feature interactions and structured residual patterns.In the proposed framework, the hybrid architecture employs a residual-error correction strategy. First, the LSTM model is trained using multivariate input sequences to generate baseline one-step-ahead forecasts of the log-transformed gold price. The residual error at time *t* is defined as:

$$e_t = y_t - \hat{y}_{LSTM,t}$$

where y_t denotes the actual log gold price and $\hat{y}_{LSTM,t}$ represents the LSTM model's prediction.

Subsequently, the XGBoost regressor is trained using the same input feature representation (after reshaping) to learn the residual structure. Rather than forecasting the gold price directly, XGBoost models the discrepancy between the observed values and the LSTM predictions. During the testing phase, actual residual values are unavailable because future observations have not yet been observed. Therefore, the trained XGBoost model generates residual estimates using the testing input features. These predicted residuals are subsequently added to the corresponding LSTM forecasts to generate the final hybrid predictions. This procedure enables residual correction without requiring future observations during out-of-sample forecasting.

The final hybrid forecast is obtained by combining the baseline LSTM prediction and the predicted residual:

$$\hat{y}_{hybrid,t}=\hat{y}_{LSTM,t}+\hat{y}_{XGB,t}$$

where $\hat{y}_{XGB,t}$ corresponds to the residual correction generated by XGBoost.

This two-stage learning mechanism enables the hybrid model to preserve the temporal representation learned by the LSTM while refining predictions through nonlinear residual adjustment. By explicitly modeling residual errors, the hybrid framework aims to reduce systematic forecasting discrepancies that remain after sequential learning.

Evaluation Metrics

Model performance is evaluated using three widely adopted forecasting accuracy metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These metrics are computed on the testing dataset to ensure out-of-sample evaluation. The RMSE is defined as:

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

The MAE is defined as:

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

The MAPE is defined as:

$$\boldsymbol{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|$$

where y_i represents the actual log-transformed gold price, $\hat{y}_i$ denotes the predicted value, and n is the number of observations in the testing set.

RMSE and MAE measure absolute forecasting errors, with RMSE penalizing larger deviations more heavily due to the squared term. In contrast, MAPE provides a scale-independent measure of relative forecasting accuracy expressed as a percentage. Since this study employs log-transformed gold prices, MAPE reflects proportional deviations in the log scale, enabling consistent comparative evaluation across competing models. The combination of these three metrics provides a comprehensive assessment of forecasting performance from both absolute and relative error perspectives.

Statistical and Validation Tests

To evaluate whether differences in forecasting accuracy are statistically significant, the Diebold–Mariano (DM) test is employed (Diebold & Mariano, 1995). The DM test examines the null hypothesis that two competing forecasting models have equal predictive accuracy. The test is conducted using forecast-error differentials computed with squared-error loss. A statistically significant p-value indicates that the observed difference in predictive performance is unlikely to have occurred by random variation. In this study, the DM test is applied to compare the hybrid model with both the standalone LSTM model and the ARIMA benchmark. These comparisons provide formal statistical validation of whether the residual correction mechanism contributes meaningful predictive improvement beyond both linear univariate modeling and standalone sequential learning.

Beyond statistical significance testing, walk-forward validation is used to assess model robustness under rolling-forecasting conditions. In this procedure, the training window is progressively expanded, and the model is retrained before generating subsequent out-of-sample forecasts. This iterative approach more closely reflects real-world forecasting environments, in which models are updated as new data become available. By incorporating both formal statistical testing and rolling validation, the evaluation framework ensures that model performance is not only accurate in a single train–test split but also stable across evolving temporal segments. All evaluation metrics are computed based on the log-transformed gold price series to ensure full consistency with the modeling framework.

■ RESULT AND DISCUSSION

Forecasting Performance Comparison

This section presents the comparative forecasting performance of the ARIMA, LSTM, XGBoost, and hybrid LSTM–XGBoost models. Evaluation is conducted on the testing dataset using RMSE, MAE, and MAPE as defined in (7)–(9), computed on the log-transformed gold price series.

As shown in Table 1, the hybrid LSTM–XGBoost model achieves the lowest forecasting errors across all evaluation metrics (RMSE = 0.0989, MAE = 0.0691, MAPE = 0.8503%), indicating that residual correction substantially enhances predictive accuracy. The superior performance of the hybrid model suggests that gold price dynamics contain both temporal dependencies and nonlinear residual structures that cannot be fully captured by a single modeling approach. The LSTM component effectively learns long-term sequential patterns and interactions among macroeconomic variables, while the XGBoost residual learner captures remaining nonlinear errors not explained by the recurrent network. This complementary learning mechanism enables the hybrid framework to reduce systematic forecasting discrepancies and improve overall predictive accuracy. The results indicate that residual correction is particularly beneficial for modeling the complex and heterogeneous behavior commonly observed in commodity price movements.

The standalone LSTM model demonstrates competitive performance (RMSE = 0.1160; MAE = 0.0836; MAPE = 1.0284%), yet remains slightly inferior to the hybrid approach, suggesting nonlinear residual patterns not fully captured by sequential learning alone. In contrast, the ARIMA (0,1,0) and standalone XGBoost models produce considerably higher forecasting errors, reflecting limitations of purely linear univariate modeling and non-sequential tree-based regression when applied directly to multivariate time-series data. Overall, these results confirm that integrating temporal representation learning with structured residual correction provides superior predictive performance compared to classical statistical and standalone machine learning approaches.

The empirical results demonstrate that the hybrid LSTM–XGBoost framework consistently outperforms the classical ARIMA model and the standalone LSTM and XGBoost models across all evaluation metrics. The hybrid model achieves the lowest RMSE, MAE, and MAPE, indicating superior performance in both absolute and relative error

Table 1. Forecasting Performance Comparison

Model Type	Model	RMSE	MAE	MAPE (%)
Classical	ARIMA (0,1,0)	0.5653	0.5147	6.4084
Single	LSTM	0.1160	0.0836	1.0284
Single	XGBoost	0.5192	0.4652	5.7850
Hybrid	Hybrid LSTM-XGBoost	0.0989	0.0691	0.8503

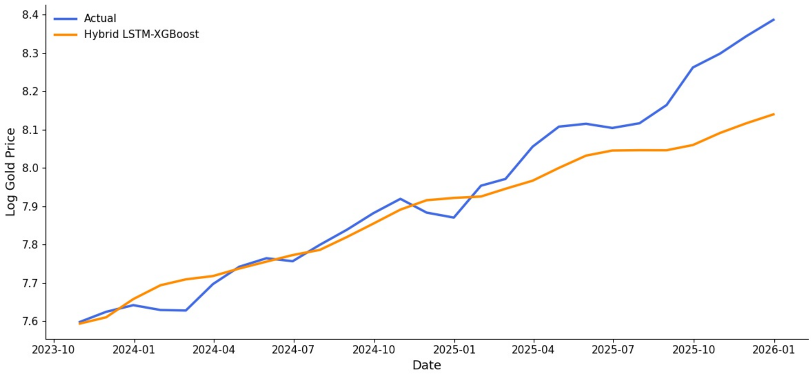

Figure 3. Actual vs. Hybrid LSTM–XGBoost Prediction (Log Scale)

measures. These findings suggest that integrating sequential temporal modeling with structured residual correction provides meaningful predictive advantages in multivariate gold price forecasting. The comparatively weaker performance of ARIMA can be attributed to its univariate, linear formulation. Although first differencing ensures stationarity and statistical validity, the ARIMA (0,1,0) specification essentially represents a first-difference random walk model. Such a structure may be insufficient for capturing nonlinear interactions and cross-variable dependencies embedded in macroeconomically driven gold price dynamics. In multivariate financial environments characterized by structural complexity, purely linear historical modeling may not fully exploit the informational content available in the data.

The standalone LSTM model demonstrates strong predictive performance compared with ARIMA and XGBoost. This result is consistent with the theoretical strengths of LSTM networks originally introduced by Hochreiter and Schmidhuber (1997) in modeling nonlinear temporal dependencies and long-term sequential patterns. However, the observed performance gap between LSTM and the hybrid framework indicates that certain structured residual patterns remain unmodeled when relying solely on recurrent architectures.

Figure 3 compares the actual log-transformed gold prices with the forecasts generated by the hybrid LSTM–XGBoost model during the testing period. The figure is constructed using out-of-sample testing data. As depicted in Figure 3, the hybrid model closely tracks the gold price series over the evaluation horizon. The predicted values replicate the overall upward trajectory and intermediate fluctuations observed in the actual data, indicating that the model successfully captures both temporal continuity and short-term variations. Unlike the standalone models, the hybrid forecasts exhibit smoother adjustments in response to trend changes without producing excessive volatility or abrupt divergence. Although minor deviations are observable at certain points, particularly during sharper movements in the series, the predicted curve remains consistently aligned with the actual trajectory. The visual correspondence between predicted and observed values is consistent with the quantitative results presented in Table 1, where the hybrid model achieves the lowest RMSE (0.0989), MAE (0.0691), and MAPE (0.8503%). The relatively low MAPE value indicates strong proportional forecasting accuracy within the log-transformed framework. Overall, the graphical analysis reinforces the numerical findings, suggesting that the residual correction mechanism improves stability and accuracy in multivariate gold price forecasting.

Figure 4 presents the relationship between the actual and predicted values generated by the hybrid LSTM–XGBoost model on the testing dataset. Using a 12-month sliding window, the 39 testing observations yielded 27 valid testing sequences, all of which are shown in the scatter plot. Most observations are concentrated near the regression line, indicating strong agreement between predicted and observed values. The coefficient of determination ($R^2 = 0.8212$) indicates a strong linear agreement between the observed and predicted values. The relatively limited dispersion around the regression line further supports the hybrid framework's predictive consistency. These results complement the RMSE, MAE, and MAPE findings presented in Table 1, providing additional evidence that the proposed model achieves reliable forecasting performance.

The superior performance of the hybrid model can be explained by the complementary modeling characteristics of LSTM and XGBoost. While LSTM captures temporal continuity and dynamic dependencies, gradient

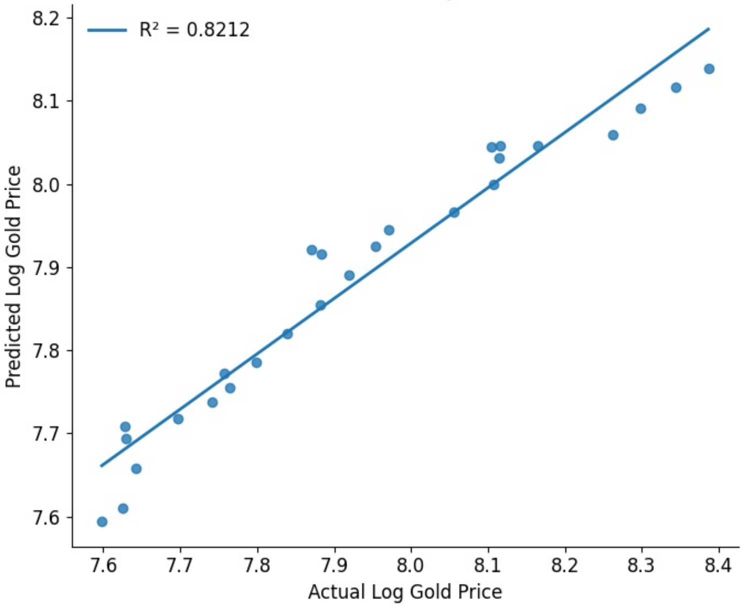

Figure 4. Actual vs. Predicted Values of the Hybrid LSTM–XGBoost Model

boosting methods such as XGBoost are effective in modeling nonlinear feature interactions and complex residual structures (Chen & Guestrin, 2016). By explicitly learning the residual term rather than directly re-estimating the target variable, the hybrid approach refines the baseline prediction and reduces systematic forecasting discrepancies. Similar hybrid strategies integrating sequential networks and boosting mechanisms have been reported to enhance forecasting stability in financial prediction tasks (Qiu et al., 2024; Zhou, 2025). The standalone XGBoost model yields higher forecast errors than the LSTM and the hybrid framework. Although XGBoost has demonstrated strong predictive performance across various financial applications (Chen & Guestrin, 2016; Tan et al., 2023), it does not inherently encode sequential memory when directly applied to time-series data. This limitation underscores the importance of incorporating explicit temporal modeling mechanisms in financial forecasting.

Figure 5 presents the feature importance obtained from the XGBoost residual correction model. Feature importance values were computed using the gain-based importance metric provided by XGBoost, where larger values indicate a greater contribution of a feature to reducing prediction error during residual correction. The results indicate that lagged WTI and US10Y features are among the most influential predictors in the residual learning process, suggesting that energy market conditions and long-term interest-rate expectations contain valuable information not fully captured by the baseline LSTM model. Several lagged gold-price features also appear among the top-ranked predictors, confirming the persistence of temporal dependencies in gold price movements. In addition, the presence of DXY and CPI among the important features highlights the relevance of macroeconomic indicators in explaining the remaining forecasting errors.

From an economic perspective, the dominance of WTI-related features may reflect the influence of global energy market conditions on inflation expectations and commodity price dynamics. Likewise, the importance of US10Y variables suggests that long-term interest-rate expectations play a meaningful role in gold price formation by influencing investment allocation and opportunity costs. These findings provide additional evidence that the superior performance of the hybrid framework is driven not only by sequential learning with an LSTM but also by the incorporation of macroeconomic information through an

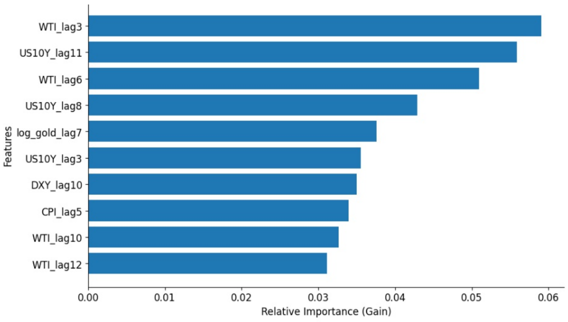

Figure 5. Top 10 Important Features in the XGBoost Residual Correction Model

XGBoost-based residual correction. A formal statistical validation of the hybrid model's performance is presented in the following subsection.

Figure 6 presents the training and validation loss curves obtained during LSTM training. Both loss values decrease substantially during the early training epochs, indicating effective model learning and convergence. The validation loss reaches a relatively stable minimum during the middle stages of training and remains close to the training loss throughout most epochs. Although a slight increase in validation loss is observed in the final epochs, the gap between training and validation loss remains small, indicating only minor overfitting and acceptable model generalization. This behavior suggests that the model does not suffer from severe overfitting and maintains reasonable generalization capability. The convergence behavior observed in Figure 6 is consistent with the robustness demonstrated by the subsequent walk-forward validation results.

Statistical Significance Analysis

To formally assess whether the observed differences in forecasting performance are statistically significant, the Diebold–Mariano (DM) test is conducted using squared error loss (Diebold & Mariano, 1995). Because the forecasting task is formulated as a one-step-ahead prediction problem, the forecast horizon used in the Diebold–Mariano test is h = 1. The loss differential is defined as the difference between the squared forecast errors of the competing models, and its variance is estimated using the sample variance with Bessel's correction (ddof = 1). The test evaluates the null hypothesis that two competing models possess equal predictive accuracy. First, the hybrid LSTM–XGBoost model is compared with the standalone LSTM model. The computed DM statistic is 3.5347 with a p-value of 0.0016. Since the p-value is well below the 0.05 significance threshold, the null hypothesis of equal predictive accuracy is rejected. This result indicates that the hybrid framework provides a statistically significant improvement over the standalone LSTM model.

Second, the hybrid model is compared with the ARIMA (0,1,0) benchmark. The DM statistic equals 6.3327 with a p-value below 0.001. The null hypothesis is again rejected, confirming that the hybrid model significantly outperforms the classical linear specification. These findings provide statistical evidence that the residual error correction mechanism contributes meaningful predictive gains beyond both linear univariate modeling and standalone sequential learning.

The Diebold–Mariano test results further confirm that the hybrid model's performance improvement is statistically significant relative to both the standalone LSTM and the ARIMA benchmark. The rejection of the null hypothesis indicates that the observed reduction in forecast errors is unlikely to be driven by random variation.

Walk-Forward Validation and Robustness Analysis

Walk-forward validation is conducted to assess the robustness of the hybrid LSTM–XGBoost model under rolling-forecasting conditions. In this procedure, the training window is progressively updated, and the model is retrained before generating each subsequent out-of-sample prediction. This approach more closely reflects practical forecasting environments in which new information is continuously available. Across multiple rolling iterations, the hybrid model achieves an average RMSE of 0.0977, MAE of

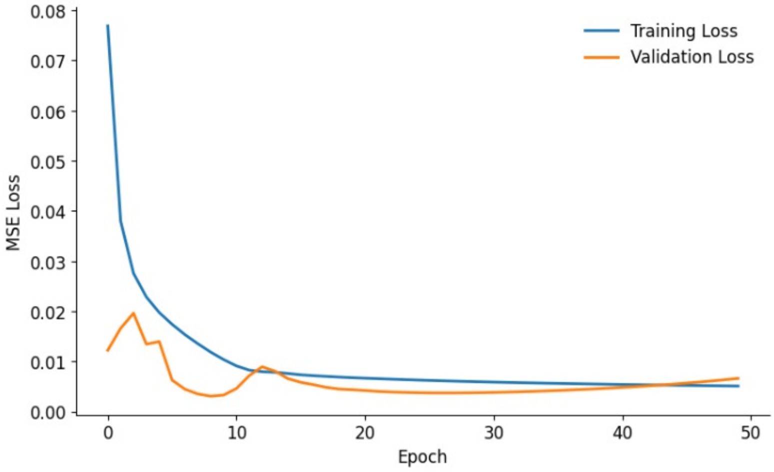

Figure 6. Training and Validation Loss Curves During LSTM Training

0.0846, and MAPE of 1.1347%. The corresponding standard deviations are 0.0306 for RMSE, 0.0284 for MAE, and 0.3830 for MAPE. Although the average MAPE obtained from walk-forward validation is slightly higher than that reported in the single train–test split (0.8503%), the difference remains within a reasonable range. Notably, the average RMSE obtained from walk-forward validation (0.0977) is slightly lower than that reported in the single train–test split (0.0989), whereas the MAE and MAPE values are higher. This difference can be attributed to the distinct sensitivity of the evaluation metrics. RMSE places greater emphasis on larger forecasting errors due to the squared-error formulation, while MAE and MAPE reflect average absolute deviations across all observations. The results suggest that the rolling evaluation yielded fewer extreme forecasting errors, resulting in a marginal reduction in RMSE. However, the average magnitude of prediction errors increased slightly across forecasting windows. The relatively moderate standard deviation values indicate that forecasting errors do not fluctuate excessively across rolling windows, suggesting stable predictive behavior over time. Overall, the walk-forward evaluation supports the robustness of the proposed hybrid framework and demonstrates that its predictive performance is not limited to a single dataset partition.

Walk-forward validation reinforces the robustness of the proposed framework. Although the rolling evaluation yields a slightly higher average MAPE than the single train–test split, the difference remains moderate, and the standard deviation values indicate stable predictive behavior across temporal segments. This stability suggests that the hybrid model does not rely on a favorable data partition but maintains consistent forecasting performance under dynamic evaluation settings. The relatively small discrepancy between single-split and rolling validation results also indicates limited overfitting. From a practical perspective, these findings suggest that multivariate forecasting frameworks that combine sequential representation learning and nonlinear residual correction are better suited to modeling gold price movements influenced by macroeconomic factors. The proposed hybrid structure provides a balanced and adaptable modeling strategy for complex financial time series.

The walk-forward validation procedure was initialized with a training window of 107 observations (70% of the training dataset). A rolling forecast horizon of 12 observations was employed, and the training window was expanded by one observation at each iteration (step size = 1). This configuration enabled repeated out-of-sample evaluation across multiple temporal segments while preserving the chronological structure of the time-series data.

Despite these encouraging results, several limitations should be acknowledged. First, the dataset consists of monthly observations, which may not capture high-frequency market fluctuations. Second, the analysis focuses on one-step-ahead forecasting, leaving multi-step predictive evaluation for future research. Third, only a selected set of macroeconomic indicators is included; incorporating additional financial, geopolitical, or sentiment-based variables may further enhance predictive performance. Finally, while reasonable hyperparameter configurations are employed, more advanced optimization techniques could potentially improve model generalization. Future research may extend this framework by exploring multi-horizon forecasting, richer feature representations, and broader comparative statistical testing across alternative hybrid architectures.

■ CONCLUSION

This study developed and evaluated a hybrid LSTM–XGBoost framework for multivariate gold price forecasting using selected macroeconomic indicators. Empirical results demonstrate that the proposed hybrid model consistently outperforms the ARIMA benchmark and standalone machine learning models across RMSE, MAE, and MAPE metrics. Based on RMSE, the hybrid framework reduces forecasting error by approximately 14.7% relative to the standalone LSTM model, 82.5% relative to ARIMA (0,1,0), and 81.0% relative to the standalone XGBoost model, highlighting the effectiveness of residual correction in improving predictive accuracy. These results demonstrate that residual correction yields meaningful predictive gains across both classical statistical and machine-learning forecasting approaches. The integration of temporal dependency learning and nonlinear residual correction provides measurable improvements in both absolute and proportional forecasting accuracy. The Diebold–Mariano test confirms that the predictive gains of the hybrid model are statistically significant relative to both the standalone LSTM and ARIMA specifications. Furthermore, walk-forward validation indicates that the model maintains stable performance across rolling evaluation windows, supporting its robustness under dynamic forecasting conditions.

These findings underscore the value of structured residual learning within hybrid deep learning frameworks for modeling complex multivariate financial time series. By combining sequential modeling with gradient-boosting–based error refinement, the proposed approach offers a balanced, statistically grounded forecasting strategy for gold price prediction. Future research may extend this framework to multi-step forecasting horizons, incorporate additional macroeconomic or sentiment-based indicators, and explore more advanced hyperparameter optimization techniques to further enhance predictive generalization.

■ DECLARATION OF GENERATIVE AI USAGE IN THE WRITING PROCESS

During the preparation of this manuscript, the authors utilized ChatGPT to assist with language refinement, grammar improvement, and manuscript editing. Following the use of this tool, the authors carefully reviewed and revised the content as necessary and accept full responsibility for the final content of the article.

■ REFERENCES

Baur, D. G., & McDermott, T. K. (2010*). Is gold a safe haven? International evidence. Journal of Banking and Finance, 34*(8), 1886–1898. https://doi.org/10.1016/j.jbankfin.2009.12.008

Bechere, M., Barkat, A., Ghenabzia, A., & Yiltas-Kaplan, D. (2025). Urban traffic prediction using hybrid XGBoost–LSTM model. *International Journal of Computing and Digital Systems, 18*(1), 1–15. https://doi.org/10.12785/ijcds/1571136041

Ben Ameur, H., Jamaani, F., & Abu Alfoul, M. N. (2024). Examining the safe-haven and hedge capabilities of gold and cryptocurrencies: A GARCH and regression quantiles approach in geopolitical and market extremes. *Heliyon, 10*(22), e40400. https://doi.org/10.1016/j.heliyon.2024.e40400

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological), 26*(2), 211–252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control (5th ed.).* John Wiley & Sons.

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning, 109*, 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, 785–794. https://doi.org/10.1145/2939672.2939785

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366), 427–431. https://doi.org/10.1080/01621459.1979.10482531

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Echaust, K., & Just, M. (2022). Is gold still a safe haven for stock markets? New insights through the tail thickness of portfolio return distributions. *International Review of Financial Analysis, 84*, 102388. https://doi.org/10.1016/j.irfa.2022.102388

Faraj, H., McMillan, D., & Al-Sabah, M. (2025). The diminishing luster: Gold’s market volatility and the fading safe haven effect. *Global Finance Journal, 65,* 101145. https://doi.org/10.1016/j.gfj.2025.101145

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting, 37*(1), 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008

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

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data-driven networks. *Future Internet, 15*(8), Article 255. https://doi.org/10.3390/fi15080255

Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A, 379*(2194), Article 20200209. https://doi.org/10.1098/rsta.2020.0209

Madhika, Y. R., Kusrini, K., & Hidayat, T. (2023). Gold price prediction using the ARIMA and LSTM models. *Sinkron: Jurnal dan Penelitian Teknik Informatika, 8*(3), 1255–1264. https://doi.org/10.33395/sinkron.v8i3.12461

Makala, D., & Li, Z. (2021). Prediction of gold price with ARIMA and SVM. *Journal of Physics: Conference Series, 1767*(1), Article 012022. https://doi.org/10.1088/1742-6596/1767/1/012022

Mun, W. K., Sufahani, S. F., Kamil, A. A., & Nawawi, M. K. M. (2025). Prediction of gold prices using hybrid model ARIMA-LSTM. *Advances and Applications in Statistics, 92*(5), 749–766. https://doi.org/10.17654/0972361725031

Nasir, J., Iftikhar, H., Aamir, M., Iftikhar, H., Rodrigues, P. C., & Rehman, M. Z. (2025). A hybrid LMD–ARIMA–machine learning framework for enhanced forecasting of financial time series: Evidence from the NASDAQ composite index. *Mathematics, 13*(15), Article 2389. https://doi.org/10.3390/math13152389

Nensi, A. I. E., Al Maida, M., Notodiputro, K. A., Angraini, Y., & Mualifah, L. N. A. (2025). Performance analysis of ARIMA, LSTM, and hybrid ARIMA-LSTM in forecasting the composite stock price index. *CAUCHY: Jurnal Matematika Murni dan Aplikasi, 10*(2), 588–604. https://doi.org/10.18860/cauchy.v10i2.33379

Pendaraki, K., & Charda, M. (2025). Investigating the dynamic connection between gold and stock markets during crises. *Journal of Risk and Financial Management, 18*(12), Article 694. https://doi.org/10.3390/jrfm18120694

Qiu, C., Zhang, Y., Qian, X., Wu, C., Lou, J., Chen, Y., et al. (2024). A two-stage deep fusion integration framework based on feature fusion and residual correction for gold price forecasting. *IEEE Access, 12*, 85565–85579. https://doi.org/10.1109/ACCESS.2024.3408837

Quang, P. D., & Thang, T. Q. (2025). Analysis and forecasting of daily global gold price: An SARIMA-LSTM approach with Random Forest technique. *Cogent Economics & Finance, 13*(1), 2568969. https://doi.org/10.1080/23322039.2025.2568969

Saini, A., Singh, R. K., & Sinha, P. (2025). Forecasting gold price using hybrid deep neural network LSTM-autoencoder. *Discover Artificial Intelligence, 5,* 281. https://doi.org/10.1007/s44163-025-00464-w

Tan, B., Gan, Z., & Wu, Y. (2023). The measurement and early warning of daily financial stability index based on XGBoost and SHAP: Evidence from China. *Expert Systems with Applications, 227*, 120375. https://doi.org/10.1016/j.eswa.2023.120375

Taneva-Angelova, G., Raychev, S., & Ilieva, G. (2025). A framework for gold price prediction combining classical and intelligent methods with financial, economic, and sentiment data fusion. *International Journal of Financial Studies, 13*(2), Article 102. https://doi.org/10.3390/ijfs13020102

Yi, S. N. C., Chew, L. M., & Yeng, O. L. (2023). Gold prices forecasting using bidirectional LSTM model based on SPX500 index, USD index, crude oil prices and CPI. *Proceedings of the 2023 IEEE 11th International Conference on Information and Communication Technology (ICoICT)*, 539–544. https://doi.org/10.1109/ICoICT58202.2023.10262481

Zangana, H. M., & Obeyd, S. R. (2024). Deep learning-based gold price prediction: A novel approach using time series analysis. *Sistemasi: Jurnal Sistem Informasi, 13*(6), 2581–2591. https://doi.org/10.32520/stmsi.v13i6.4651

Zhan, Z., & Kim, S.-K. (2024). Versatile time-window sliding machine learning techniques for stock market forecasting. *Artificial Intelligence Review, 57*, Article 209. https://doi.org/10.1007/s10462-024-10851-x

Zhao, Y., Guo, Y., & Wang, X. (2025). Hybrid LSTM–Transformer architecture with multi-scale feature fusion for high-accuracy gold futures price forecasting. *Mathematics, 13*(10), 1551. https://doi.org/10.3390/math13101551

Zhou, J. (2025). A dynamic weighted fusion-based hybrid XGBoost-LSTM model for financial distress prediction. *Proceedings of the 2nd Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence (DEAI 2025)*. https://doi.org/10.1145/3745238.3745276