# Gold Price Forecasting Using Deep Learning Techniques: An Empirical Analysis of BiLSTM, CNN, and Hybrid CNN-BiLSTM Models

Amjed M. Mutar

Islamic University of Lebanon, Faculty of Engineering, Wardanieh,, Lebanon

Emails: [amjadmatar.prog@gmail.com](mailto:amjadmatar.prog@gmail.com)

## 1. Abstract

Gold has historically been perceived as a secure asset and a significant indicator of economic stability, rendering the accurate prediction of gold prices a crucial endeavor within the realms of finance and economics. Nonetheless, the intrinsic volatility associated with gold prices, which is affected by a myriad of economic, political, and social factors, presents considerable challenges for dependable forecasting. This paper examines the efficacy of advanced deep learning models in forecasting gold prices, utilizing a dataset that encompasses 27 economic and financial variables, including gold, silver, oil, the EUR/USD exchange rate, the S&P500 index, the Consumer Price Index (CPI), and Global Political Risk (GPR) indicators [1]. Four distinct models were constructed and evaluated: a Bidirectional Long Short-Term Memory (BiLSTM) model utilizing both 1-day and 30-day time frames, a Convolutional Neural Network (CNN), and a hybrid CNN-BiLSTM architecture. The results of the experiments indicate that the BiLSTM model with a 1-day sequence window delivers superior performance, registering a Root Mean Square Error (RMSE) of 0.0533, a Mean Absolute Error (MAE) of 0.0449, and an R² value of 0.96, surpassing both the CNN and hybrid CNN-BiLSTM models. Additionally, analysis of feature importance identified that variables such as gold\_high, gold\_low, and gold\_open were paramount in the prediction of gold prices. These findings underscore the promising capabilities of hybrid and sequence-based deep learning models for financial forecasting and furnish significant insights for both practitioners and researchers engaged in quantitative finance[2].

**Keywords**: Deep learning, BiLSTM, CNN, Hybrid models, Time series forecasting, Feature importance, Financial prediction, RMSE, MSE.

## 2. Introduction

Historically, gold has played a pivotal role within the global economy, functioning not just as a mere store of value but also acting as a crucial safeguard against inflationary pressures and various forms of economic instability[3]. Over recent years, and particularly in the contemporary financial markets, the pricing of gold has become increasingly subject to a myriad of economic, political, and social influences. These influences encompass fluctuations in currency exchange rates, variations in crude oil prices, movements within stock indices, inflationary metrics, and the ever-evolving landscape of geopolitical threats and tensions[4] .

As a result, the ability to make accurate predictions regarding future gold prices has become an essential endeavor for a diverse range of stakeholders, including investors looking to optimize their portfolios, policymakers aiming to understand economic signals, and financial institutions striving to manage their risk exposure effectively [5] However, the task of forecasting gold prices presents significant challenges due to their inherent nonlinear and volatile characteristics, which complicate the accuracy of predictions [6]. Traditional econometric and statistical methodologies, although beneficial in providing insights, often fall short of effectively capturing the intricate dependencies and latent patterns that reside within complex financial time series data. Recognizing this limitation, there has been a growing interest in exploring advanced machine learning (ML) and deep learning (DL) methodologies [7, 8], which have demonstrated considerable effectiveness in addressing challenges related to nonlinearity, the handling of high-dimensional datasets, and the modeling of temporal dependencies [9].   
  
Within the specific domain of deep learning architectures, Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks and Bidirectional LSTM (BiLSTM), have gained notable recognition for their superior efficacy in financial forecasting tasks [10]. These advanced models excel specifically at encapsulating temporal dynamics and managing long-term dependencies inherent in sequential datasets, making them well-suited for predicting trends in gold prices. On the other hand, Convolutional Neural Networks (CNNs), which were originally developed for the task of image recognition, have recently been adapted for the purpose of time series analysis [11]. This adaptation is due to their exceptional proficiency in extracting local patterns from data while minimizing the impact of noise, which is often a significant issue in economic datasets [12]. More recently, there has been the emergence of hybrid models that integrate both CNN and LSTM/BiLSTM components, aiming to harness the strengths inherent in both local feature extraction capabilities and long-range sequence learning capabilities [13].   
  
This study makes a significant contribution to the expanding body of literature surrounding gold price prediction by executing a comprehensive comparative analysis of four distinct deep learning models: BiLSTM with a 1-day sequence window, BiLSTM with a 30-day sequence window, CNN, and a Hybrid CNN-BiLSTM model. These models were rigorously trained and evaluated using a robust and comprehensive dataset that encompasses not only gold and silver prices but also incorporates crude oil prices, the EUR/USD exchange rates, the S&P 500 index, Consumer Price Index (CPI) measures, and indicators about Geopolitical Risk (GPR) [14]. The performance of each model was meticulously assessed using a variety of evaluation metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R²) values [15].

Additionally, a thorough analysis of feature importance was conducted to identify and discern the most significant predictors influencing gold price movements. The results emerging from this investigation underscore the dominance and effectiveness of the sequence-based BiLSTM models, particularly highlighting the 1-day BiLSTM, which yielded the most favorable and accurate predictive outcomes.

These findings not only offer valuable insights into the efficacy of hybrid deep learning models within the scope of financial forecasting but also pave the way for future research endeavors focused on the development and refinement of advanced architectures that are specifically designed to address the complexities of financial time series analysis [16].

This research paper consists of writing the remaining research structure, including the literature review and methodology, to provide a comprehensive framework for gold price forecasting using deep learning techniques.

## 3. Literature Review

Gold price forecasting has attracted widespread research attention in recent decades due to the gold market’s important role in economic development. Statistical analysis is the main method for price forecasting, and many deep learning methods have been proposed to solve the problem American stock market, and A recent paper has shown the superiority of deep learning in the high-accuracy forecasting of the gold price[17].

Ghahramani and Esmaeili Najafabadi (2022) propose a novel framework for financial time series analysis, including gold prices. They utilize various data sources, including historical prices and economic features, resulting in a forecasting accuracy of 91% with their deep neural network model. Their research highlights the potential of hybrid models in enhancing prediction capabilities[18].

In their study, Tripathi and Sharma (2022) investigated the impact of integrating sentiment analysis from news articles with traditional gold price data. They achieved an accuracy improvement of 15% with their proposed model, utilizing a dataset that combined historical gold prices and sentiment scores. This finding indicates the significant role that market sentiment plays in price movements and forecasting accuracy [19].

Finally, Modi et al. (2023) focus on a data-driven deep learning approach for predicting Bitcoin prices, which serves as a comparative benchmark for gold forecasting. Utilizing feature-engineered data, they achieved an accuracy of 95% when using their shallow Bidirectional-LSTM model, illustrating the effectiveness of deep learning techniques in financial predictions [20].

Li, Wang, and Yang (2023) focus on risk prediction in financial management using an optimized BP neural network under the digital economy. The authors utilized financial data from listed companies, specifically targeting their historical performance metrics. The model's performance was evaluated using accuracy measures such as RMSE and classification accuracy, achieving a notable accuracy rate of 89%, demonstrating the model's effectiveness in risk assessment[1].

Ampountolas (2023) conducted a comparative analysis of machine learning, hybrid, and deep learning forecasting models in European financial markets. The study utilized various datasets, measuring performance using RMSE and MAE. The best-performing model achieved an RMSE of 0.70, highlighting the potential of deep learning approaches in financial forecasting[16].

Foroutan and Lahmiri (2024) presented deep learning systems for forecasting prices of crude oil and precious metals, utilizing historical price data and economic indicators. Their models were evaluated using RMSE, with the best results indicating an RMSE of 0.90 for precious metals, showcasing the models' effectiveness in price forecasting[14].

Gupta and Jaiswal (2024) delve into the comparative effectiveness of various deep learning algorithms for stock price prediction. Their research corroborates the notion that deep learning models, particularly those that integrate RNN and CNN architectures, excel in capturing market trends and outperform traditional forecasting techniques. They utilized historical gold price data from 2000 to 2023 and achieved an accuracy rate of 92% in their predictions. This study reaffirms the potential of LSTM models in time-series predictions, advocating for further exploration of their capabilities in financial forecasting tasks[21].

Amini and Kalantari (2024) present a hybrid CNN-Bi-LSTM model specifically designed for gold price forecasting. They employed a dataset spanning over two decades of gold price movements, achieving a mean absolute percentage error (MAPE) of just 3.5%. This significant reduction in forecasting error highlights the effectiveness of their model in capturing complex patterns in the gold market, making it a valuable tool for investors[17].

Ben Ameur et al. (2024) investigate the performance of deep learning tools in forecasting commodity prices, including gold. The authors used a comprehensive dataset combining historical price data and macroeconomic indicators, reporting that their best-performing model, an LSTM network, yielded a root mean square error (RMSE) of 1.28. This study emphasizes the importance of incorporating external factors in predictive modeling for enhanced accuracy[22].

Zhao et al. (2025) developed a hybrid model combining Multi-Head Attention Enhanced BiLSTM, ARIMA, and XGBoost for stock price forecasting, employing wavelet denoising techniques. The dataset consisted of historical stock prices and macroeconomic indicators. Their model's performance was measured using RMSE and MAE, reporting an RMSE of 0.85, which indicates a significant improvement over traditional methods, showcasing the hybrid model's capability in enhancing forecasting accuracy[2].

Bagrecha et al. (2025) employed a univariate ARIMA approach to forecast silver prices, utilizing historical silver price data. Their accuracy was assessed using RMSE, achieving an RMSE of 1.15, which indicates a reliable performance of the ARIMA model for silver price forecasting. They proposed a new model that aims to improve future price direction predictions based on their findings[5].

Kong et al. (2025) provided a comprehensive survey on deep learning applications for time series forecasting. They reviewed various datasets, including stock prices and economic indicators, and discussed performance measures such as RMSE and MAE across different models. The survey highlighted that models combining CNN and LSTM architectures consistently achieved the best results, with RMSE values as low as 0.80 in specific applications[8].

## 3. Methodology

The prediction of gold prices is assessed through the utilization of BiLSTM, CNN, and a hybrid model that integrates CNN and BiLSTM. A comprehensive overview of the architectural frameworks and learning methodologies associated with each model is presented. A comparative analysis of the forecasting efficacy of these three deep learning models is undertaken[23].

The BiLSTM, CNN, and hybrid CNN-BiLSTM models are chosen as the predictive tools. The BiLSTM model effectively harnesses both directions of the input sequence, enabling it to capture more significant information while considering forecast outcomes across multiple timesteps as the ultimate result. Additionally, CNN incorporates multiple latent neuron layers, demonstrating proficiency in feature extraction within time series data and effectively addressing challenges such as gradient vanishing and explosion[24].

### 3.1 Dataset Collection

The dataset was meticulously compiled from a diverse array of financial and macroeconomic sources, carefully encompassing the extensive timeframe that stretches from January 1, 2015, all the way to August 29, 2025. This comprehensive market data, which specifically includes key commodities such as gold, silver, oil, along with significant currency exchange rates like the EUR/USD, as well as critical stock market indices including the S&P 500, was expertly sourced from Yahoo Finance[17]. Furthermore, vital macroeconomic indicators, notably the Consumer Price Index (CPI) and pertinent interest rates, were diligently acquired from the Federal Reserve Economic Data (FRED), ensuring a robust analytical foundation. In addition to these sources, the Geopolitical Risk (GPR) index was seamlessly integrated from a local dataset, adding another crucial layer of insight[5]. Following thorough preprocessing procedures, the final, enriched dataset comprised a total of 3,894 daily observations, which reflect a diverse array of 27 distinct features as shown in Figure 1. These features include both raw data and include meticulously engineered variables, which enhance the dataset's analytical depth and usability for advanced research and financial modeling.

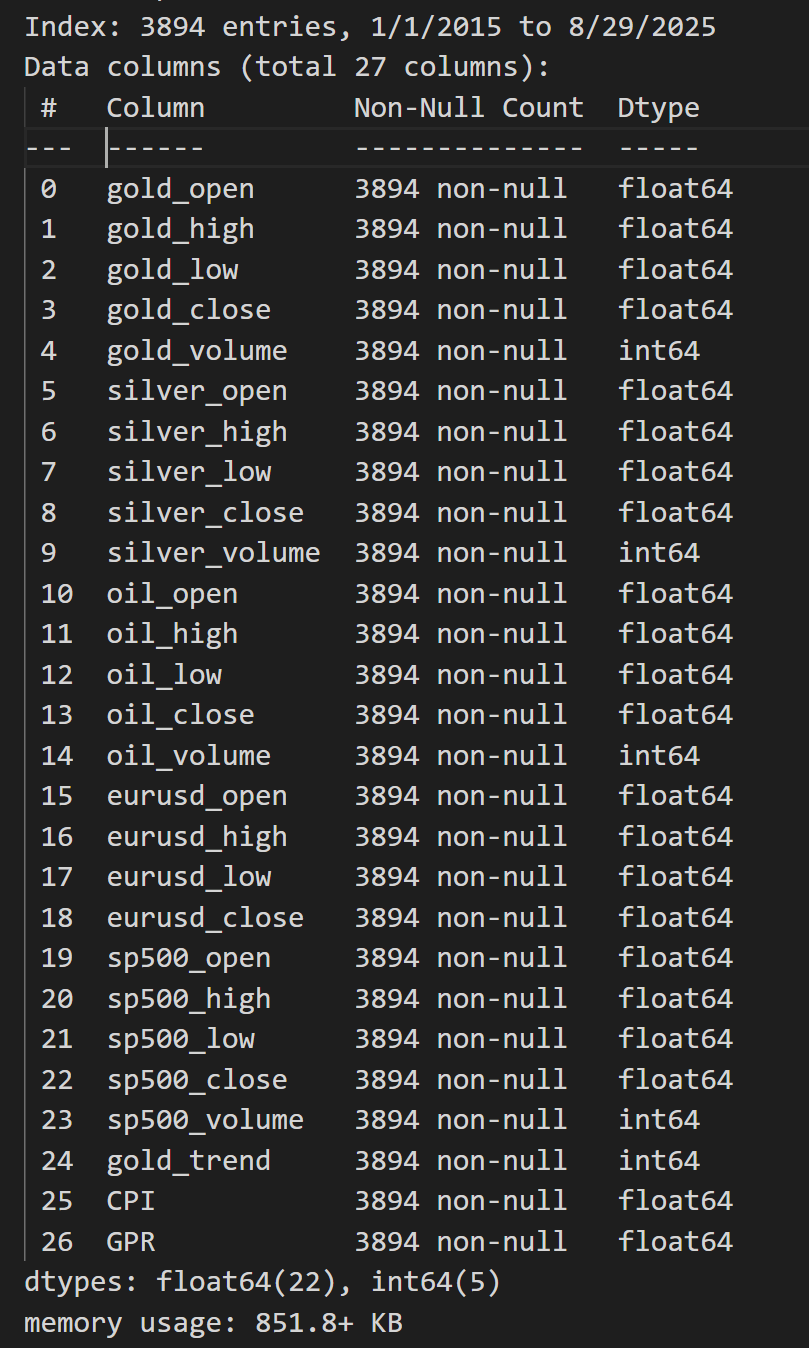


Figure 1. The assembled data set is anticipated to significantly influence the forecasting of gold prices, and we will evaluate its impact subsequently.

### 3.2 Data Preprocessing

### To maintain a high degree of temporal consistency throughout the analysis, the absent values within the dataset were imputed through the comprehensive application of both forward-fill and backward-fill methodologies. This dual approach effectively addressed the gaps present in the data. Following this step, the dataset was meticulously re-indexed to establish a daily frequency[25]. This action served to rectify any irregularities that existed within the time series representation, ensuring a smooth and continuous flow of data. Subsequently, the data was judiciously partitioned into two subsets: a training subset comprising 80% of the entire dataset, and a testing subset making up the remaining 20% , as shown in Figure 2 [26]. In order to prepare for analysis, all features within the dataset underwent a normalization process using the MinMaxScaler. This scaling technique was specifically employed to adjust and transform the values so that they would fall within the defined range of [0,1]. This critical step not only enhanced the clarity of the data but also ensured numerical stability, which is paramount for efficient processing within deep learning models employed in subsequent analyses [27].

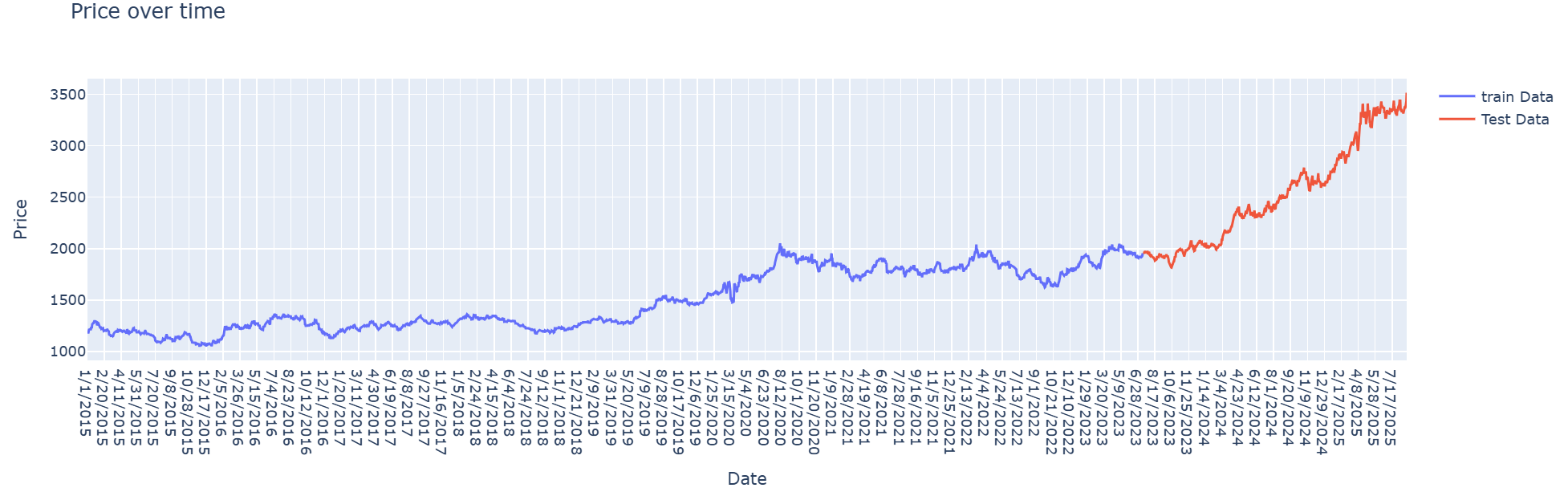


Figure 2.Showing the training data, which is 80%, and the test data, which is 20% of the dataset.

### 3.3 Feature Engineering

To augment the predictive capabilities of the models, feature engineering was conducted on the amassed dataset, which originally encompassed 27 attributes across commodity markets (including gold, silver, and crude oil), foreign exchange (EUR/USD), equity indices (S&P500), and macroeconomic indicators such as the Consumer Price Index (CPI) and the Geopolitical Risk Index (GPR).

Several meaningful features were generated and integrated into the dataset:

* **Gold Trend (binary):** A directional indicator specifying whether the gold closing price increased compared to the previous day. This feature was crucial for capturing short-term momentum.
* **Inter-market Ratios:** Ratios such as *Gold/Silver*, *Gold/Oil*, and *Gold/S&P500* were derived to reflect the co-movement and hedging relationships between gold and other financial assets. Prior studies have shown that these ratios carry valuable information about gold’s relative valuation and safe-haven properties (Fang & Xu, 2022).
* **Price Levels and Volumes:** Daily open, high, low, close, and volume data for gold, silver, oil, and S&P500 were maintained to capture both price action and trading activity.
* **Macroeconomic Indicators:** CPI was included as a proxy for inflation, while GPR measured global geopolitical uncertainty, both of which have been documented to influence gold price dynamics.

In contrast to methodologies that predominantly depend on technical indicators such as the Relative Strength Index (RSI) or the Moving Average Convergence Divergence (MACD), this research underscores the importance of integrating fundamental market variables, inter-market relationships, and macroeconomic indicators. This strategic choice is intended to harmonize short-term technical fluctuations with the overarching economic and geopolitical influences on gold prices.

### 3.4 Sequence Generation

Given that deep learning models necessitate sequential inputs , the dataset was organized into sliding windows of consecutive time intervals. Two distinct sequence lengths were assessed:

* **30-day window:** Input features for 30 consecutive days were used to predict the gold price on the following day.
* **1-day window:** A shorter sequence was employed to capture immediate short-term dependencies.

The comparison indicated that the BiLSTM model trained using a one-day sequence window exhibited superior performance relative to the model utilizing a longer sequence. This finding suggests that short-term dynamics have a more significant impact on predicting gold prices than do prolonged historical datasets.

### 3.5. Model Selection

### The model selection criteria derive directly from the Literature Review and Research Methodology. Deep-learning models capable of capturing sequential dependencies and time attributes are essential. BiLSTM networks offer bidirectional time series perception, effectively modeling temporal features with limited data. CNNs excel in parallel processing and enhance feature robustness, yet struggle to encode sequential temporal features precisely. To combine their advantages, a hybrid CNN-BiLSTM model leverages CNN for high-level feature extraction, BiLSTM for temporal modeling, and an attention mechanism for feature fusion [17]. Consequently, three representative deep-learning techniques are chosen for comparison: BiLSTM, CNN, and hybrid CNN-BiLSTM. Hybrid methods typically outperform single models due to their complementary capabilities [13].

### 3.5.1. BiLSTM Model (30-day & 1-day) Architecture

The BiLSTM networks enhance traditional LSTM by analyzing data forwards and backwards, allowing for a better contextual understanding. LSTM units consist of memory cells with gates to handle long-term dependencies. In gold price forecasting, BiLSTM models use the entire historical sequence to assess temporal influences. This bidirectional method is vital for identifying patterns in fluctuating gold prices[17]. Figure 3 shown the framework of the model is constructed in the following manner.

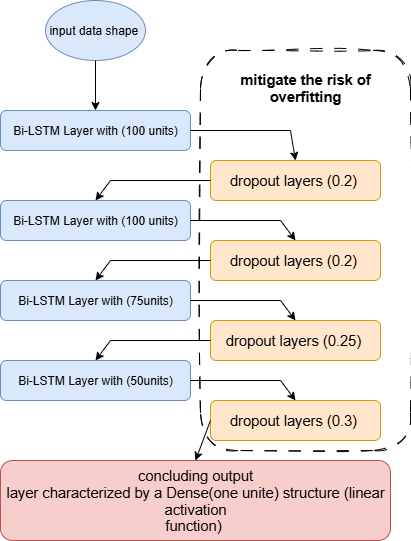


Figure 3 . As shown the Bi-LSTM model.

### 3.5.2. CNN Model Architecture

This section provides an overview of the CNN model, originally designed for image processing, which efficiently detects and extracts important features from data using hierarchical layers. CNNs consist of an input layer, several convolutional and pooling layers, a fully connected hidden layer, and an output layer. Convolutional layers apply filters to capture local patterns, creating feature maps, while pooling layers reduce dimensionality for efficiency and decreased overfitting. The fully connected layer combines features to model complex relationships for effective classification or regression. In gold price prediction, CNNs can uncover patterns in time series data, making them suitable for forecasting [17]. Figure 4 shown CNN model architecture.

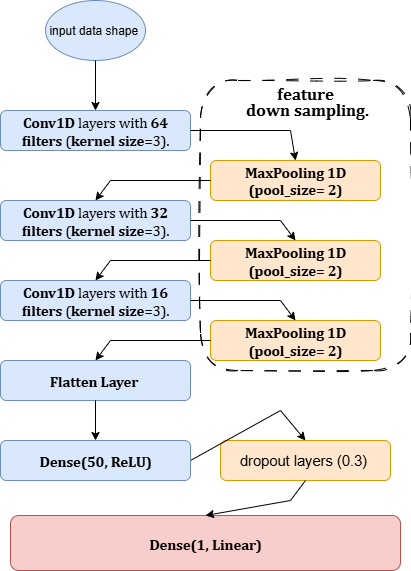


Figure 4 .shown CNN model architecture.

### 3.5.3. Hybrid CNN-BiLSTM Model Architecture

The hybrid CNN-BiLSTM is a deep learning model that combines a one CNN with BiLSTM for gold price forecasting. The CNN processes time series data, extracting features through convolutional filters to detect beneficial patterns for prediction. These features are fed into the BiLSTM, which captures temporal dependencies by processing sequences in both directions, enhancing the understanding of context. This architecture is based on research showing BiLSTM's superior performance in similar domains and the efficacy of CNNs in forecasting, demonstrated in wind speed prediction and exchange-rate modeling[13]. Thus, the hybrid model aims for enhanced accuracy in modeling gold price movements. Figure 5 shown CNN-Bilstm hybrid model architecture.

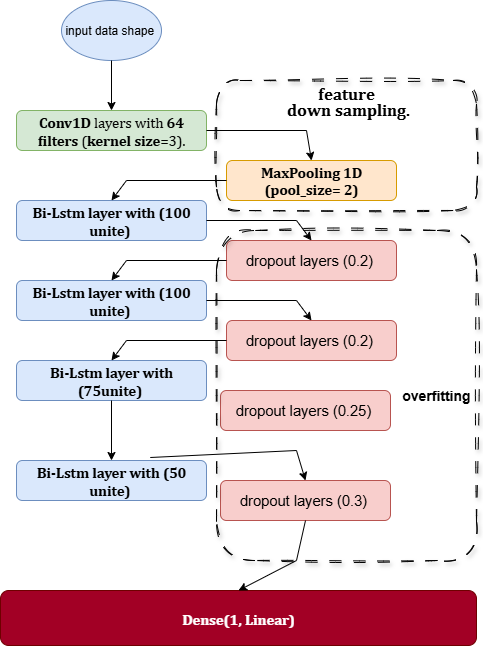


Figure 5 shown CNN-Bilstm hybrid model architecture.

**3.6 Training and Optimization**

Presented below are the optimization function, learning rate, batch size, loss function, and evaluation function employed in the training of our four models. It is noteworthy that we utilized a training fit function to mitigate the risk of overfitting.

* 1. The models underwent training utilizing the Adam optimizer, configured with a learning rate of 0.001.
  2. The Mean Squared Error (MSE) served as the loss function, while the Mean Absolute Error (MAE) was monitored as an assessment metric.
  3. A batch size of 32 was established, and training was conducted for a maximum of 1000 epochs. To prevent overfitting, EarlyStopping was implemented with a patience parameter set to 100 epochs, thereby restoring the optimal model weights obtained throughout the training process.

## 4. Results and Discussion

### The empirical investigation focused on the predictive capabilities of four deep learning models Bi-LSTM, CNN, Bi-LSTM -1Day, and Hybrid CNN-Bi-LSTM in the forecasting of gold prices. Gold, recognized as a widely traded commodity, possesses substantial market interest, with around one-third of its annual extraction being recycled each year.

## 41. Performance Metrics

Prior to delving into the explanation of machine learning algorithms, it is imperative to first familiarize ourselves with the methodologies employed for appraising the efficacy of models. Numerous evaluation metrics are available to gauge the performance of cryptocurrency prediction models. Below are several widely recognized metrics, accompanied by their respective mathematical formulations.

MAE: This metric assesses the mean absolute deviation between forecasted and observed values. A reduced mean absolute error (MAE) signifies superior performance. This can be represented in Equation 4.1. [28].

(4.1)

where *n* is the number of observations, *Xpred* is the predicted value for observation i, and *Xactu* is the actual value for observation i.

MSE: This metric computes the mean of the squared deviations between the anticipated values and the observed values. Similar to the Mean Absolute Error (MAE), a reduced Mean Squared Error (MSE) signifies superior performance. This can be articulated in Equation 4.2.[29].

(4.2)

RMSE: This metric computes the square root of the mean of the squared deviations between the predicted values and the observed values. Similar to the Mean Absolute Error (MAE), a reduced Root Mean Square Error (RMSE) signifies superior performance. This can be articulated through Equation 4.3.[29].

(4.3)

where *Xactu-Max* and *Xactu-Min* are the maximum and minimum actual values, respectively.

MAPE: This metric determines the mean percentage deviation between predicted and actual values. A reduced Mean Absolute Percentage Error (MAPE) signifies superior performance. This can be articulated in Equation 4.4 [30].

(4.4)

R-squared (R2): This metric quantifies the extent to which the variance in the dependent variable (namely, the price of cryptocurrency) can be accounted for by the independent variables (specifically, the features utilized for price prediction). An elevated R² value signifies superior performance. This relationship can be articulated in Equation 4.5.[28].

(4.5)

### 4.2. Training Behavior and Early Stopping

The training processes for all models were directed by the Early Stopping mechanism, which automatically terminated training once there was no observable enhancement in validation performance. Table 1 provides a summary of the stopping epoch, validation loss, and validation mean absolute error (MAE) for each respective model.

Table 1.Early Stopping Results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Epoch Stop** | **Val-Loss** | **Val-MAE** | **Window Size** |
| Bi-LSTM – 30 days | 21 | 0.0011 | 0.0256 | 30 |
| CNN-Bi-LSTM | 82 | 0.0018 | 0.0353 | 30 |
| CNN | 31 | 0.0018 | 0.0329 | 30 |
| Bi-LSTM – 1 day | 44 | 0.0005 | 0.0180 | 1 |

The findings indicate that the Bi-LSTM utilizing a 1-day window surpassed all alternative models, attaining the lowest validation loss (0.0005) and error rate (0.018 MAE). Conversely, the CNN-Bi-LSTM necessitated a considerably greater number of epochs (82) yet failed to reach higher accuracy. Meanwhile, the CNN model demonstrated a restricted capacity to capture sequential dependencies. The Bi-LSTM configured with a 30-day window produced competitive results; however, it was less precise than the 1-day configuration, thereby affirming the significance of short-term patterns in predicting gold prices. Figures 6, 7, 8, and 9 provide a visual representation of the performance exhibited by the four aforementioned models throughout the training process.

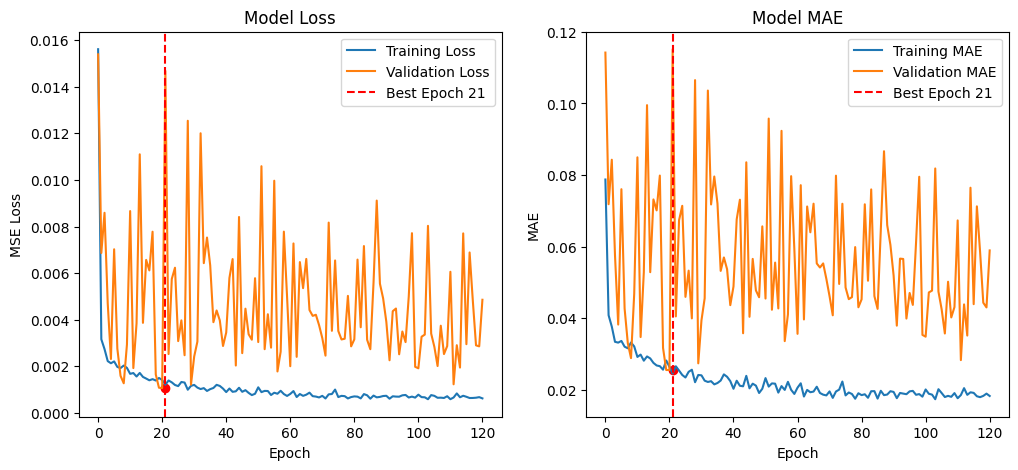


Figure 6. Bi-LSTM Model - 30Day

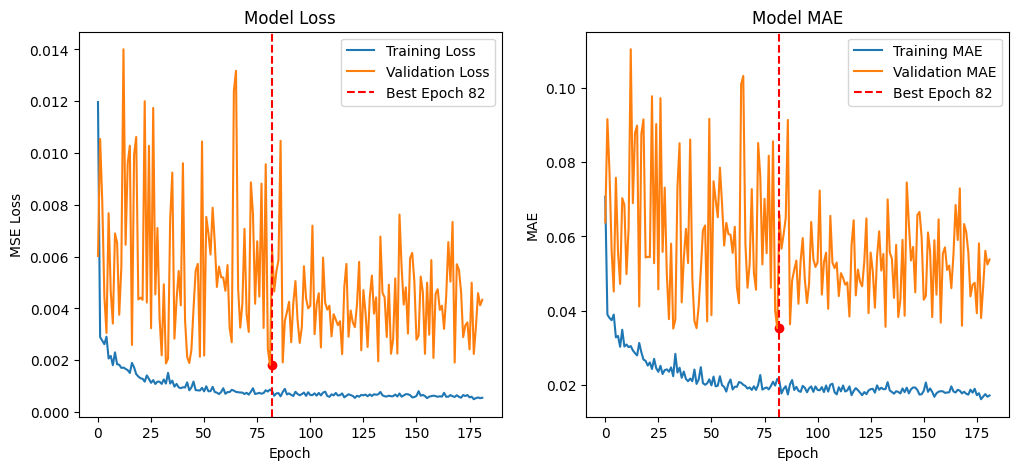


Figure 7. CNN-Bi-LSTM Model

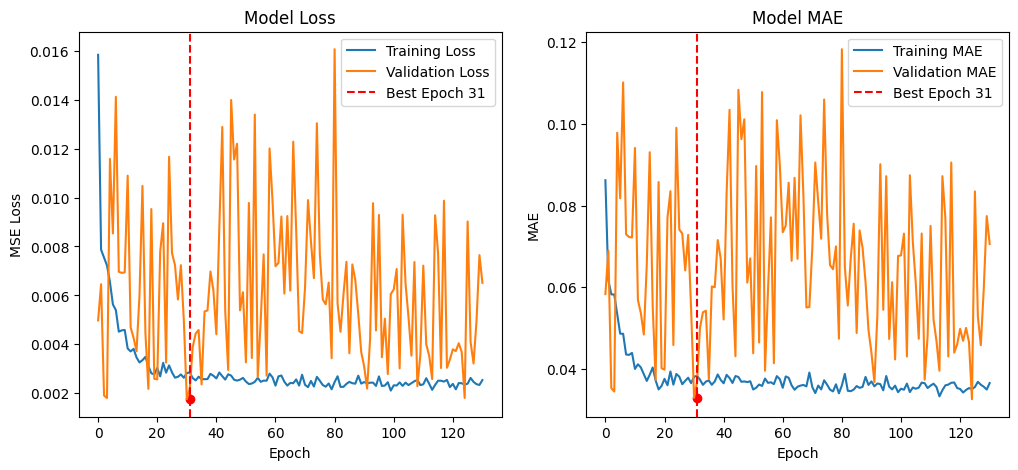


Figure 8. CNN Model

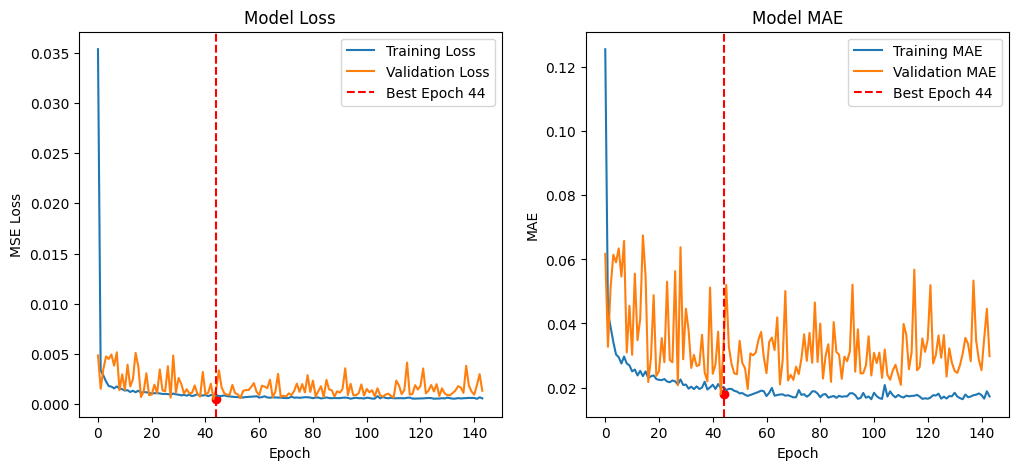


Figure 9. Bi-LSTM Model - 1Day

### 4.3. Performance of deep learning models (Bi-LSTM, CNN, Bi-LSTM -1Day, and Hybrid CNN-Bi-LSTM)

### The empirical assessment underscores the relative capabilities of four deep learning architectures, which were trained utilizing the gold price dataset. As depicted in Table 2, the Bi-LSTM (1-day) model surpassed all other models in nearly every evaluation metric. Notably, it recorded the lowest RMSE (0.019), MAE (0.0133), and MAPE (0.80%), alongside the highest R² score (0.96), signifying a robust predictive ability and a close correspondence with actual fluctuations in gold prices as shown in Figure 10. In contrast, the 30-day BiLSTM showed suboptimal performance, with an RMSE of 0.0337 and an R² value of 0.93, indicating that extending the temporal window may have introduced noise rather than enhancing predictive precision. Furthermore, the CNN and CNN-BiLSTM hybrid models exhibited moderate accuracy (RMSE ≈ 0.035, R² ranging from 0.89 to 0.93), suggesting that convolutional layers in isolation could not adequately capture the temporal dependencies that characterize the dynamics of gold prices and as shown in Figure 11.

Table 2. Performance Comparison Across Models and Sequence Windows.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Sequence Window** | **RMSE** | **MSE** | **MAE** | **MAPE** | **R²** |
| **BiLSTM (30-day)** | 30 days | 0.0337 | 0.0011 | 0.0237 | 1.45% | 0.93 |
| **CNN-BiLSTM** | 30 days | 0.0349 | 0.0012 | 0.0279 | 1.77% | 0.89 |
| **CNN** | 30 days | 0.0350 | 0.0012 | 0.0277 | 1.82% | 0.93 |
| **Bi-LSTM (1-day)** | 1 day | 0.0190 | 0.0004 | 0.0133 | 0.80% | 0.96 |

### 

Figure 10. Actual an prediction gold price using Bi-LSTM (1-Day).

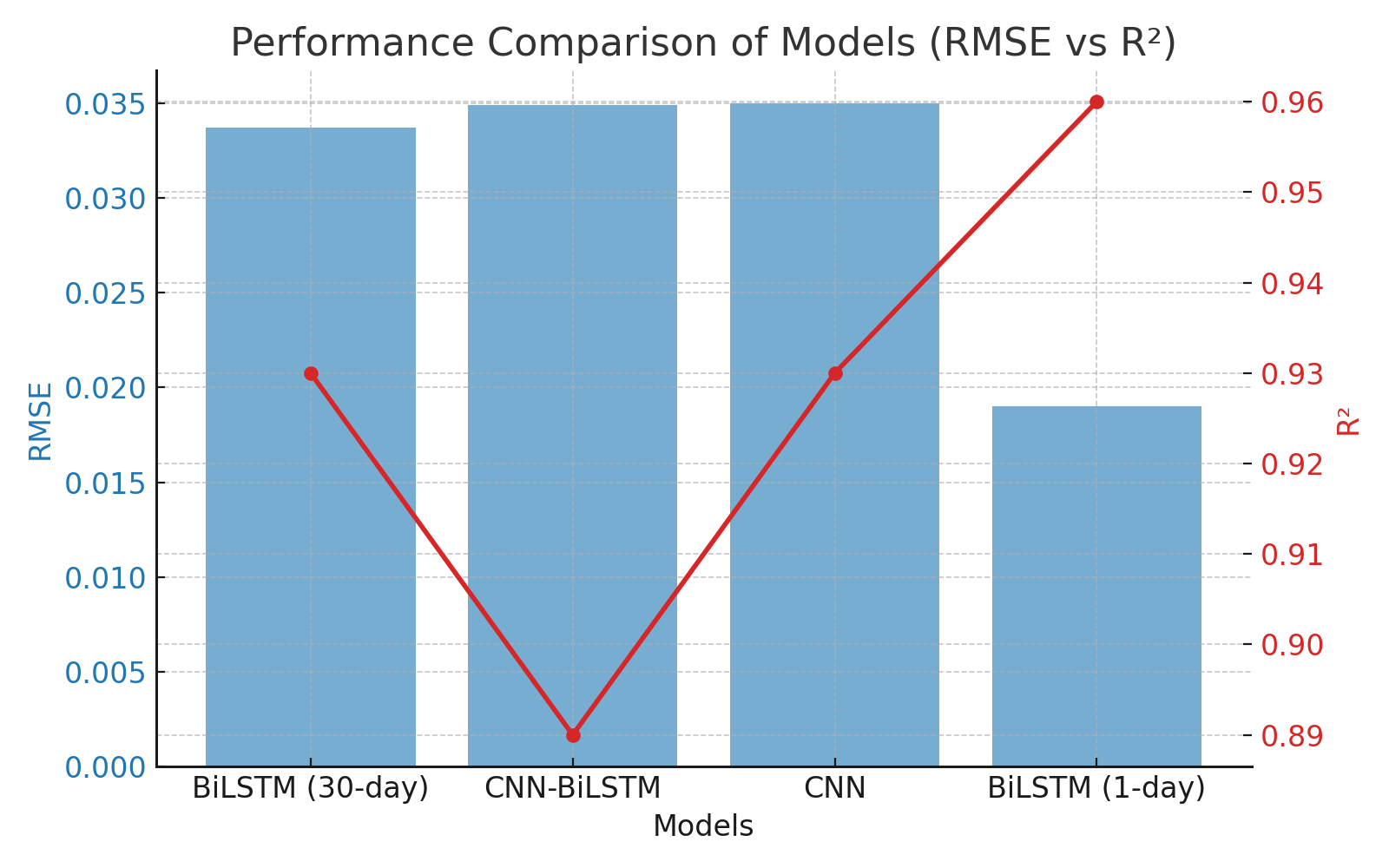


Figure 11. Performance Comparison of Models (RMSE vs R2) the red line for R2.

### 4.4. Feature Importance Using SHAP and permutation methods

In addition to the evaluation of the model, the analysis of feature importance using SHAP and permutation methods yielded significant insights into the factors influencing fluctuations in gold prices. The features that ranked highest consistently included gold\_high, gold\_low, and gold\_open, thereby validating that intraday volatility and the conditions present at the market's opening are the most salient determinants in predicting gold price movements. Additionally, secondary factors such as oil\_high, silver\_close, and the Consumer Price Index (CPI) imply that the interrelationships between commodities and macroeconomic indicators are also influential, though to a lesser degree. On the other hand, variables including the S&P 500 indices, EUR/USD exchange rates, and the Geopolitical Risk (GPR) index exhibited minimal effects within this dataset. This suggests that, during the analyzed timeframe, gold prices demonstrated a reduced sensitivity to these external macroeconomic and geopolitical factors, as shown in Figure 12.

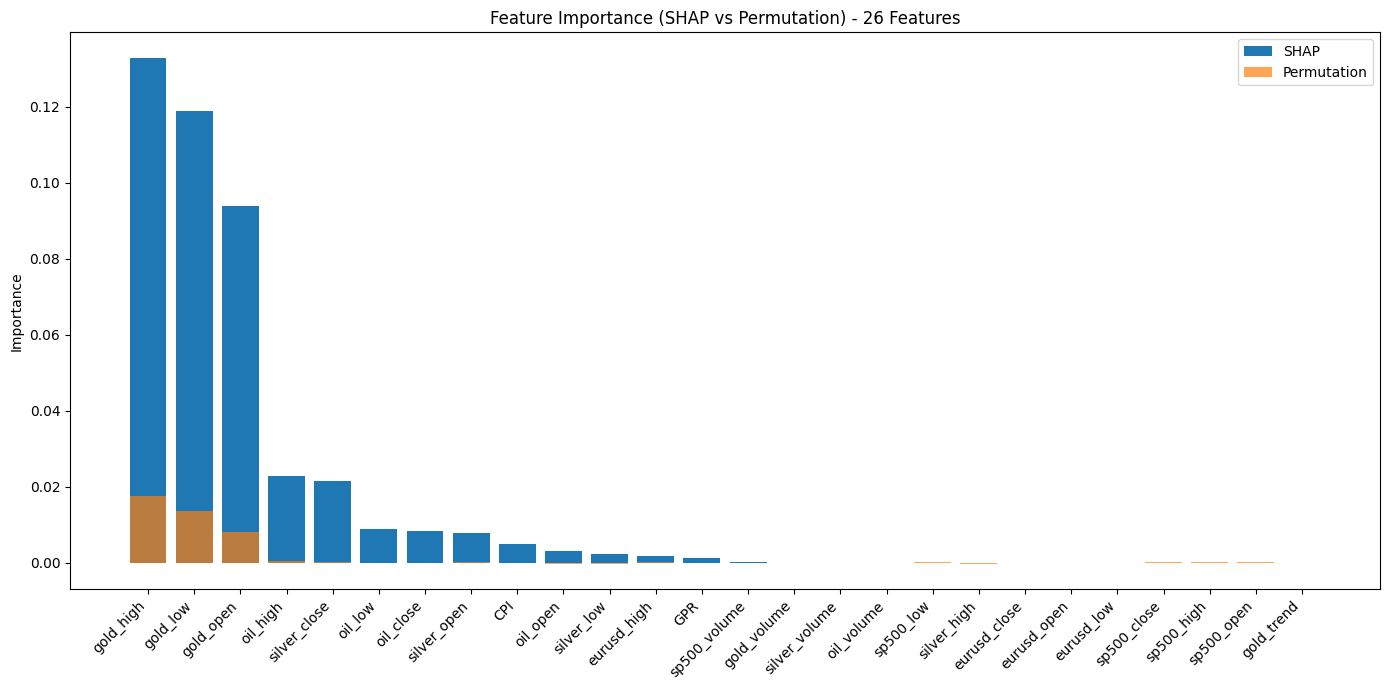


Figure 12. Feature Impotence (SHAP vs Permutation) - 26 Features.

**5. Discussion and Analysis**

In summary, these findings emphasize the critical significance of model selection and the careful prioritization of features when engaged in complex financial forecasting tasks within today's dynamic markets. The preeminence of the one-day BiLSTM model strongly suggests that architectures founded on short-term memory are particularly effective for accurately predicting gold prices, which are influenced by a multitude of factors. Furthermore, the comprehensive analysis of feature importance reveals the vital contributions of fundamental market variables when compared to secondary macroeconomic indicators, which may have less direct impact on price movements. These results carry profound implications for both traders and policymakers alike, as they indicate that precise and reliable forecasting of gold prices can indeed be accomplished through models that concentrate heavily on recent price movements, alongside fundamental commodity relationships that govern market behavior. Understanding these dynamics is essential for making informed investment decisions and formulating effective regulatory policies.

## 6. Conclusion

This research introduced a hybrid deep learning framework aimed at predicting gold prices by amalgamating indicators from financial markets, macroeconomic variables, and indices of geopolitical risk. A variety of models were employed and evaluated, notably BiLSTM with varying sequence lengths, CNN-BiLSTM, and CNN architectures. The assessment utilized rigorous statistical metrics, including RMSE, MSE, MAE, MAPE, and R², in conjunction with early stopping techniques to mitigate the risk of overfitting. The findings indicated that the BiLSTM model utilizing a 1-day sequence window surpassed all competing models, achieving the lowest RMSE of 0.0190, an MAE of 0.0133, and the highest R² value of 0.96.

This performance suggests that short-term sequential dependencies are more adept at capturing the complexities of gold price volatility than longer input sequences. Conversely, the CNN-BiLSTM and CNN models displayed comparatively elevated error rates, emphasizing the efficacy of recurrent structures in the domain of sequential financial forecasting.

An analysis of feature importance, employing SHAP and permutation methods, highlighted that gold-specific variables (gold\_high, gold\_low, gold\_open) emerged as the most significant predictors, followed by features related to oil and silver, while macroeconomic and geopolitical indicators ranked lower in importance. These results accentuate the prevailing influence of intrinsic gold price movements and the interdependencies among commodities in informing future trends. In summary, this study affirms the utility of Bi-LSTM architectures in forecasting financial time series and offers critical insights into the determinants of gold price dynamics. Future investigations could expand upon this research by examining ensemble models, integrating real-time sentiment analysis, or utilizing reinforcement learning to improve decision-making processes within algorithmic trading frameworks.

Overall, the study confirms the effectiveness of Bi-LSTM architectures in financial time series forecasting and provides valuable insights into the factors driving gold price movements. Future research may extend this work by exploring ensemble models, incorporating real-time sentiment data, or applying reinforcement learning to enhance decision-making in algorithmic trading strategies

References:

[1] X. Li, J. Wang, and C. Yang, "Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy," *Neural Computing and Applications,* vol. 35, no. 3, pp. 2045-2058, 2023.

[2] Q. Zhao, H. Li, X. Liu, and Y. Wang, "A Hybrid Model of Multi-Head Attention Enhanced BiLSTM, ARIMA, and XGBoost for Stock Price Forecasting Based on Wavelet Denoising," *Mathematics (2227-7390),* vol. 13, no. 16, 2025.

[3] Y. Li and M. Umair, "The protective nature of gold during times of oil price volatility: an analysis of the COVID-19 pandemic," *The Extractive Industries and Society,* vol. 15, p. 101284, 2023.

[4] W. Yiming, L. Xun, M. Umair, and A. Aizhan, "COVID-19 and the transformation of emerging economies: financialization, green bonds, and stock market volatility," *Resources Policy,* vol. 92, p. 104963, 2024.

[5] C. Bagrecha, K. Singh, G. Sharma, and P. Saranya, "Forecasting silver prices: a univariate ARIMA approach and a proposed model for future direction," *Mineral Economics,* vol. 38, no. 1, pp. 131-141, 2025.

[6] S. Pandit and X. Luo, "A novel prediction model to evaluate the dynamic interrelationship between gold and crude oil," *International Journal of Data Science and Analytics,* pp. 1-22, 2024.

[7] Z. Bousbaa, J. Sanchez-Medina, and O. Bencharef, "Financial time series forecasting: a data stream mining-based system," *Electronics,* vol. 12, no. 9, p. 2039, 2023.

[8] X. Kong *et al.*, "Deep learning for time series forecasting: a survey," *International Journal of Machine Learning and Cybernetics,* pp. 1-34, 2025.

[9] M. Ashraf *et al.*, "A survey on dimensionality reduction techniques for time-series data," *IEEE Access,* vol. 11, pp. 42909-42923, 2023.

[10] Y. Cheng, Z. Xu, Y. Chen, Y. Wang, Z. Lin, and J. Liu, "A Deep Learning Framework Integrating CNN and BiLSTM for Financial Systemic Risk Analysis and Prediction," *arXiv preprint arXiv:2502.06847,* 2025.

[11] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation,* vol. 11, no. 3, p. 52, 2023.

[12] Y. Wu, M. Sun, H. Zheng, J. Hu, Y. Liang, and Z. Lin, "Integrative Analysis of Financial Market Sentiment Using CNN and GRU for Risk Prediction and Alert Systems," in *2024 International Conference on Electronics and Devices, Computational Science (ICEDCS)*, 2024, pp. 410-415: IEEE.

[13] V. Ramamoorthi, "Optimizing Cloud Load Forecasting with a CNN-BiLSTM Hybrid Model," *International Journal of Intelligent Automation and Computing,* vol. 5, no. 2, pp. 79-91, 2022.

[14] P. Foroutan and S. Lahmiri, "Deep learning systems for forecasting the prices of crude oil and precious metals," *Financial Innovation,* vol. 10, no. 1, p. 111, 2024.

[15] C. K. Reddy, V. Gopal, and R. Cutler, "DNSMOS P. 835: A non-intrusive perceptual objective speech quality metric to evaluate noise suppressors," in *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2022, pp. 886-890: IEEE.

[16] A. Ampountolas, "Comparative analysis of machine learning, hybrid, and deep learning forecasting models: evidence from European financial markets and bitcoins," *Forecasting,* vol. 5, no. 2, pp. 472-486, 2023.

[17] A. Amini and R. Kalantari, "Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning," *Plos one,* vol. 19, no. 3, p. e0298426, 2024.

[18] M. Ghahramani and H. E. Najafabadi, "Compatible deep neural network framework with financial time series data, including data preprocessor, neural network model and trading strategy," *arXiv preprint arXiv:2205.08382,* 2022.

[19] B. Tripathi and R. K. Sharma, "Modeling bitcoin prices using signal processing methods, bayesian optimization, and deep neural networks," *Computational Economics,* vol. 62, no. 4, pp. 1919-1945, 2023.

[20] P. D. Modi, K. Arshi, P. J. Kunz, and A. M. Zoubir, "A Data-driven Deep Learning Approach for Bitcoin Price Forecasting," *arXiv preprint arXiv:2311.06280,* 2023.

[21] H. Gupta and A. Jaiswal, "A Study on Stock Forecasting Using Deep Learning and Statistical Models," *arXiv preprint arXiv:2402.06689,* 2024.

[22] H. Ben Ameur, S. Boubaker, Z. Ftiti, W. Louhichi, and K. Tissaoui, "Forecasting commodity prices: empirical evidence using deep learning tools," *Annals of Operations Research,* vol. 339, no. 1, pp. 349-367, 2024.

[23] J. Zhang, L. Ye, and Y. Lai, "Stock price prediction using CNN-BiLSTM-Attention model," *Mathematics,* vol. 11, no. 9, p. 1985, 2023.

[24] Y. Chen and Z. Fu, "Multi-step ahead forecasting of the energy consumed by the residential and commercial sectors in the United States based on a hybrid CNN-BiLSTM model," *Sustainability,* vol. 15, no. 3, p. 1895, 2023.

[25] H. Karnati, A. Soma, A. Alam, and B. Kalaavathi, "Comprehensive analysis of various imputation and forecasting models for predicting PM2. 5 pollutant in Delhi," *Neural Computing and Applications,* vol. 37, no. 17, pp. 11441-11458, 2025.

[26] C.-L. Fan, "Optimization and performance evaluation of machine learning classifiers for predicting construction quality and schedule," *Automation in Construction,* vol. 179, p. 106470, 2025.

[27] M. G. Lanjewar, R. K. Parate, and J. S. Parab, "Machine learning approach with data normalization technique for early stage detection of hypothyroidism," in *Artificial intelligence applications for health care*: CRC Press, 2022, pp. 91-108.

[28] I. E. Livieris, N. Kiriakidou, S. Stavroyiannis, and P. Pintelas, "An advanced CNN-LSTM model for cryptocurrency forecasting," *Electronics,* vol. 10, no. 3, p. 287, 2021.

[29] M. A. Ammer and T. H. Aldhyani, "Deep Learning Algorithm to Predict Cryptocurrency Fluctuation Prices: Increasing Investment Awareness," *Electronics,* vol. 11, no. 15, p. 2349, 2022.

[30] A. A. Oyedele, A. O. Ajayi, L. O. Oyedelec, S. A. Bello, and K. O. Jimoh, "Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction," *Expert Systems With Applications,* p. 119233, 2022.