Republic of Iraq
Ministry of Higher Education &
Scientific Research
Al-Farabi University
Journal of Al-Farabi For
Engineering Sciences

جـمـهـوريــة الـعـراق
وزارة التعليم العالي والبحث العلمي
جامعة الفارابي
مجلـــــة الفارابي للعلوم الهندسية

التاريخ:- ٢٠٢٥/٩/٢٨

العدد:- م ف / ١٥

الى/ الباحث: امجد محمود مطر

م/قبول نشر

تحية طيبة...

تهديكم مجلة الفارابي للعلوم الهندسية اطيب تحية...

نود اعلامكم بأنه تم قبول البحث المقدم من قبلكم والموسوم: _

(Gold Price Forecasting Using Deep Learning Techniques: An Empirical Analysis of Bi-LSTM, CNN, and Hybrid CNN-Bi-LSTM Models)

للنشر في مجلة الفارابي للعلوم الهندسية وتم اجراء الخطوات اللازمة للنشر وتهنئكم المجلة على هذا الإنجاز الرفيع املين التواصل الدائم مع المجلة مستقبلاً ومدها بنتاجكم البحثي.

مع التقدير...

أ.د.متمرس. بهجت رشاد شاهين
رئيس تحرير مجلة الفارابي للعلوم الهندسية
2025/9/28

مرتضى
نسخة منه الى:
المجلة
الباحث

بغداد/ الدورة- شارع المصافي قرب جسر ذو الطابقين
Baghdad/ AL-Dora, AL-Masafi Road, Near AL-Tabkaen Bridge
E-mail: info@alfarabiuc-iq.com P.O. Box: 12083 Baghdad/AL-Dora, IRAQ
Mobile: 07715161939

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

Amjed Mahmood Mutar
Islamic University of Lebanon, Faculty of Engineering, Wardanieh, Lebanon
Ministry of Interior / Directorate of Communications and Information Systems, Baghdad, Iraq
Email: 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 (Bi-LSTM) model utilizing both 1-day and 30-day time frames a Convolutional Neural Network (CNN) and a hybrid “CNN-Bi-LSTM” architecture The results of the experiments indicate that the Bi-LSTM 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 R^2 value of 0.96. surpassing both the CNN and hybrid CNN-Bi-LSTM 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 “Bi-LSTM CNN Hybrid models” Time series forecasting, Feature importance Financial prediction RMSE, MSE.

2. Introduction

Gold has traditionally assumed central importance in the world economy. not simply as an inanimate store of value but as an indispensable hedge against inflation and other forms of economic volatility [3]. In recent times especially in the present day financial markets gold price has been noticed to be influenced more and more by these non classical correlation[other sources that affect the price fluctuations of gold include variations in currency exchange rates, crude oil prices, movements within stock indices, inflationary measures, interest rate pressures and also increasingly perilous political instability [4].

Incidence Up To Date prediction of future gold price has an important role to play which is desired by variety of people/organi- zations including investors who want to get the best portfolio, governments for making better under -standing about economic signals, and financial bodies for risk management effectively [5]. However predicting gold price is a hard problem because gold prices have nonlinear and high volatility nature. that make the prediction problem complicated [6]. Traditional econometric and statistical approaches, while helpful in understanding various aspects, often do not account for complex structure of financial time series data explicitly to model intricate dependencies and under-lying structures. Realizing this a shortcoming there is an ever growing interest in exploring sophisticated machine learning (ML) and deep learning (DL) models [7, 8] . Which have shown significant effectiveness towards dealing with such challenges associated with nonlinearity, modeling high-dimensional input data and capturing temporal dependencies from the data stream [9].

Among the various deep learning architectures, Recurrent Neural Networks (RNNs) and its variations including LSTM networks, bidirectional LSTMs have been widely acknowledged for performing successfully in financial prediction tasks[10]. These models outperform other existing models for the task of predicting trends in gold prices However, recent model, Convolutional Neural Networks (CNNs), which were initially designed for the image recognition problem but have now been successfully modified as an architecture to tackle time series forecasting [11]. This adaptation is attributed to their excellent ability of capturing the local pattern from the data while ignoring noise which might be more influential in economic datasets [12]. Recently, we are seeing a resurgence of hybrid models incorporating both CNN and LSTM/Bi-LSTM components in order to make use of features offered by both local feature capturing capabilities and long-term sequence learning capability [13].

This article is a valuable addition to the growing literature on gold price forecasting, and it is valuable in this study as these models were: Rigorously trai- ned, benchmarked using an

extensive dataset of not only the gold and silver prices but raw oil prices, exchange rates (EUR/USD ratio at daily frequency), s&p 500 index, Consumer Price Index (CPI), Geopolitical asymmetry Risk (GPR) indicators [14]. The models were evaluated using several evaluation methods such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and R-squared values [15]

A comprehensive feature importance assessment was also performed to detect as well as interpret the important predictors influencing gold prices. The findings of this study have emphasised the significance and the considerably acceptable performance level of the sequence-based Bi-LSTM mode particularly 1 day BiLSTM demonstrate remarkable predictive results.

These results not only provide valuable insights on hybrid deep learning models for financial forecasting, but also open the possibility for future research in the design and improvement of new architectures that target to address some of the challenges associated with financial time series [16].

Our research paper is next structured by writing the structure of research, literature review and methodology are being written for designing a full blown framework of deep learning for predicting gold price.

3. Literature Review

Gold price prediction has recently become a research hotspot within the investment With statistical analysis being the primary approach for making predictions and numerous deep learning methods have been put forward to address the issue American stock market, A literature has recently presented the top performance of deep learning in highly accurate prediction [17].

Ghahramani and Esmaeili Najafabadi (2022) introduce a new scenario for financial time series analysis comprising the gold prices. They employ different resources such as historic prices and economical characteristics aimed at a forecasting accuracy of 91% using their deep neural network model. Their results show the promise of hybrid methods in improving prediction performances [18].

In their study Tripathi and Sharma (2022) examined the combined effect of blending sentiment indicators from news articles with traditional gold price series. They reported a 15% accuracy outperform with their PLE model when they combined the historical gold price data with sentiment scores. It shows the importance of the market sentiment in terms of the price and forecasting accuracy [19].

Finally, Modi et al (2023) address a data-driven deep learning method to predict Bitcoin prices, which is used as a competing benchmark for gold prediction. They trained on feature- engineered data and obtained a 95% accuracy when testing with their shallow Bidel model This shows that deep learning algorithms are effective for making financial decisions [20].

Li Wang and Yang (2023) address the issue of risk prediction in financial management Analyses are performed on a BP neural network which is optimized under the digital economy A sample of financial data gathered from the listed companies was considered within their historical performance metrics Evaluation of the model's performance relies upon accuracy measures namely, RMSE and classification accuracy for course corrective decision making with an accuracy level as high as91% proving itself to be efficient for characterization between Business Failure and Non-Failure [1].

Ampountolas (2023) the experiment was based on multiple datasets and performance were measured using RMSE MAE where the best model hit 0.70 in RMSE proving deep learning approach's capability in predicting at financial term [16].

Foroutan and Lahmiri (2024) introduced deep learning models for forecasting the price of crude oil and precious metals using historical prices and economic indicators. Their models, tested based on RMSE loss and the most remarkable result was an RMSE equal 0.90 for precious metals, prove the power of prediction of their models [14].

Gupta and Jaiswal (2024) discuss comparative performance of different deep learning techniques for stock market forecasting. Their study supports our assumption that deep learning models especially the RNN combined with CNN models are very good in recording market trends and can achieve better results compared to traditional forecasting methods They tested data of historic gold price from 2000 to 2023, and achieved an accuracy of 92% in their predictions. This paper confirms that there is still potential for LSTM models in time series prediction and suggests that LSTM has not been well exploited yet in financial forecasting [21].

Amini and Kalantari (2024), a hybrid CNN-Bi-LSTM model is proposed for gold price prediction purposes. They used a dataset of more than 20 years in gold prices, and reported an almost incredibly low mean absolute percentage deviation (MAPE) of 3.5%. The model's RMSE evaluation results of 32.31 indicate that the model predicts gold market trends, providing very useful information for investors.[17].

Ben Ameur et al. (2024)They predicted commodity prices, including gold futures, using deep learning models. They tested a large dataset that included historical prices and macroeconomic features. Their study found that the LSTM network achieved the best

model performance metric, RMSE, of 1.28. This study also found the importance of external factors in improving model training performance.[22].

Zhao et al (2025) proposed a mixed model including Multithread Attention Enhanced Bi-LSTM, ARIMA and XGBoost function for stock price prediction by wavelet demising. Dataset The dataset containing historical stock prices and the widely used macroeconomic indicators that are selected from a variety of different databases. Our model was tested on considering root mean square errors (RMSE) and mean absolute error MAE, having RMSE value 0.85 which means there is significant improvement in comparison to Trade retracement Ratio telling Hybrid mode has capability of enhancing forecast Accuracy[2].

Bagrecha et al. (2025) used a univariate ARIMA method to predict silver price by using the historical data of silver price. Their precision were evaluated with RMSE the results of an RMSE for silver price prediction 1.15 which mean that ARIMA model has good performance in the silver price prediction They proposed a new model to develop future price direction predictions according to their discoveries [5].

Kong et al. (2025) They comprehensively reviewed a set of deep learning models for time series forecasting. They used a dataset consisting of stock prices and economic indicators. They also used performance metrics for deep learning models, including RMSE and MAE. Their study indicated that the hybrid CNN-LSTM model outperformed the other models, achieving the best performance with an RMSE of 0.80 [8].

3. Methodology

We followed a methodology in this paper that relies on collecting diverse historical data, not only for gold prices, but also for silver, oil, economic indicators, consumer prices, and geopolitical risks, to create a robust and diverse dataset.

We will then predict future gold prices using deep learning models, specifically neural networks (Bi-LSTM, CNN, and a hybrid CNN-Bi-LSTM model) [23].

We will provide a comprehensive comparison of the models and their ability to predict future prices. We will also use techniques to identify which features in the dataset have a direct impact on future price predictions, providing a clear picture of how models are trained on the dataset. [24].

3.1 Dataset Collection

We collected the dataset for this study using the Yahoo Finance API, which included historical prices for gold, silver, and oil, as well as exchange rates for major currencies (EUR/USD). We also collected data on economic indicators for the stock market (S&P 500). This dataset was collected for a ten-year period, from January 1, 2015, to August 29,

2025 [17]. To diversify the dataset, we added another economic indicator, the Consumer Price Index (CPI), and the Geopolitical Risk Index (GPR), for the same period to ensure consistency and reliability of the dataset [5]. We then processed the data, standardized it, and filled in any missing data. We now have a dataset consisting of 3,894 days, comprising 27 features or columns, including the target category, the closing price of gold, as shown in Figure 1.

```
Index: 3894 entries, 1/1/2015 to 8/29/2025
Data columns (total 27 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   gold_open      3894 non-null   float64
 1   gold_high      3894 non-null   float64
 2   gold_low       3894 non-null   float64
 3   gold_close     3894 non-null   float64
 4   gold_volume    3894 non-null   int64
 5   silver_open    3894 non-null   float64
 6   silver_high    3894 non-null   float64
 7   silver_low     3894 non-null   float64
 8   silver_close   3894 non-null   float64
 9   silver_volume  3894 non-null   int64
 10  oil_open       3894 non-null   float64
 11  oil_high       3894 non-null   float64
 12  oil_low        3894 non-null   float64
 13  oil_close      3894 non-null   float64
 14  oil_volume     3894 non-null   int64
 15  eurusd_open    3894 non-null   float64
 16  eurusd_high    3894 non-null   float64
 17  eurusd_low     3894 non-null   float64
 18  eurusd_close   3894 non-null   float64
 19  sp500_open     3894 non-null   float64
 20  sp500_high     3894 non-null   float64
 21  sp500_low      3894 non-null   float64
 22  sp500_close    3894 non-null   float64
 23  sp500_volume   3894 non-null   int64
 24  gold_trend     3894 non-null   int64
 25  CPI            3894 non-null   float64
 26  GPR            3894 non-null   float64
dtypes: float64(22), int64(5)
memory usage: 851.8+ KB
```

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

Impute missing values to ensure much time consistency as possible throughout the analysis, missing values within the data were filled forward and backward with a common technique - Forward-Fill/Backward-Fill This two-sta ged approach 'filledin' gaps in the dataset efficiently(protocol). 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 Min Max Scaler 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 Bi-LSTM 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. Bi-LSTM 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-Bi-LSTM model leverages CNN for high - level feature extraction Bi-LSTM for temporal modeling, and an attention mechanism for feature fusion [17]. Consequently, three representative deep-learning techniques are chosen for comparison:- Bi-LSTM CNN and hybrid CNN-Bi-LSTM Hybrid methods typically outperform single models due to their complementary capabilities [13].

3.5.1. Bi-LSTM Model (30-day & 1-day) Architecture

The Bi-LSTM 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, Bi-LSTM 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.

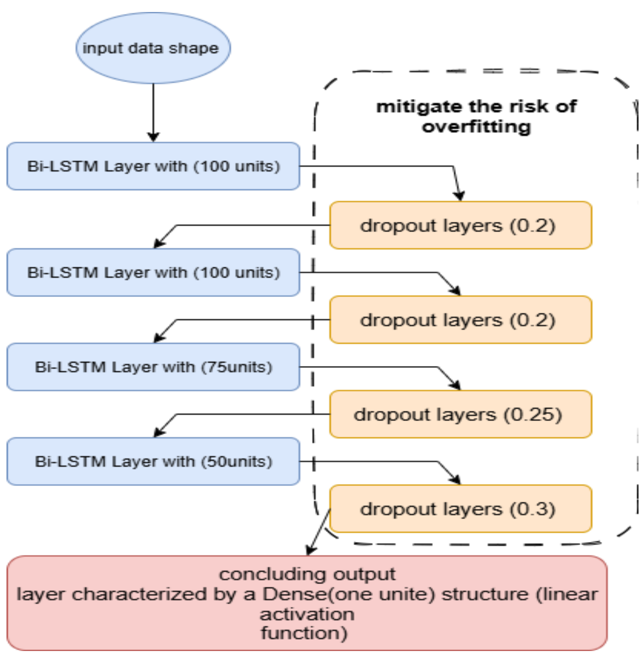

Figure 3 . As shown the Bi-LSTM model

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.

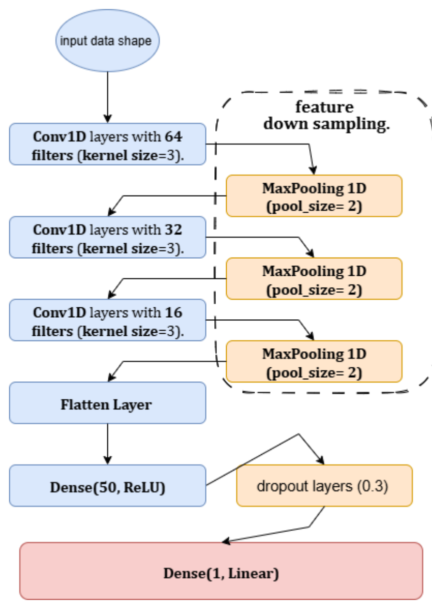

Figure 4 .shown CNN model architecture

3.5.3. Hybrid CNN-Bi-LSTM Model Architecture

The hybrid CNN-Bi-LSTM is a deep learning model that combines a one CNN with Bi_LSTM 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 Bi_LSTM which captures temporal dependencies by processing sequences in both directions, enhancing the understanding of context. This architecture is based on research showing Bi_LSTM'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-Bi-LSTM hybrid model architecture

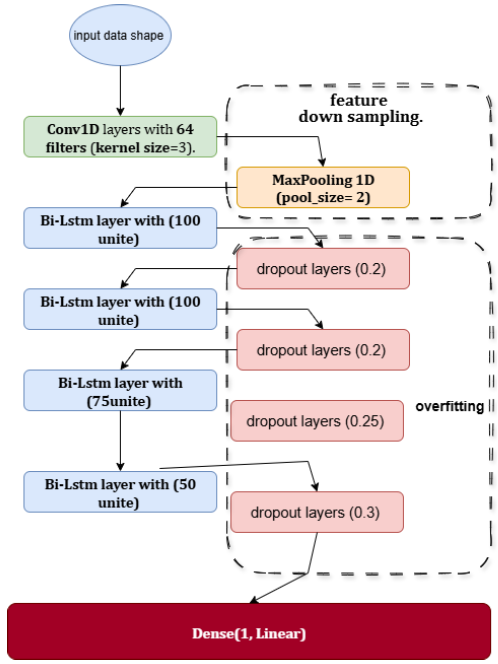

Figure 5 shown CNN-Bi-LSTM hybrid model architecture

3.6 Training and Optimization

Presented below are the optimization function learning rat 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

4.1. 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].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_{actu} - X_{pred}\right| \tag{4.1}$$

where *n* is the number of observations, X_{pred} is the predicted value for observation I and X_{actu} 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].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_{actu} - X_{pred})^2 \tag{4.2}$$

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].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{actu} - X_{pred})^2} \tag{4.3}$$

where $X_{actu\text{-}Max}$ and $X_{actu\text{-}Min}$ 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].

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_{actu} - X_{pred}}{X_{actu}}\right| * 100\% \tag{4.4}$$

R-squared (R^2):- 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 R^2 value signifies superior performance. This relationship can be articulated in Equation 4.5.[28].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_{actu} - X_{pred})^2}{\sum_{i=1}^{n}(X_{actu} - X_{pred})^2} \quad (4.5)$$

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.

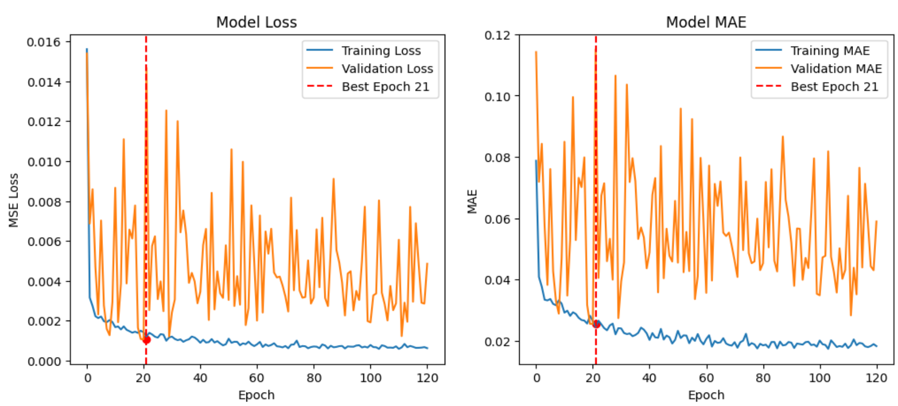

Figure 6. Bi-LSTM Model - 30Day

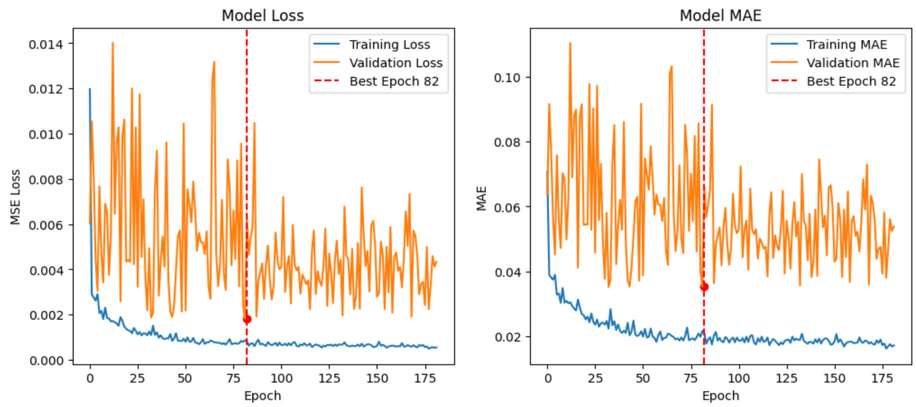

Figure 7. CNN-Bi-LSTM Model

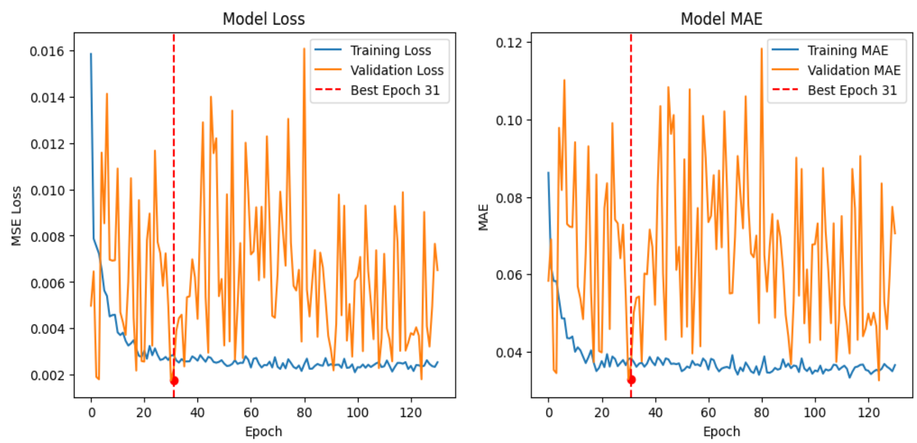

Figure 8. CNN Model

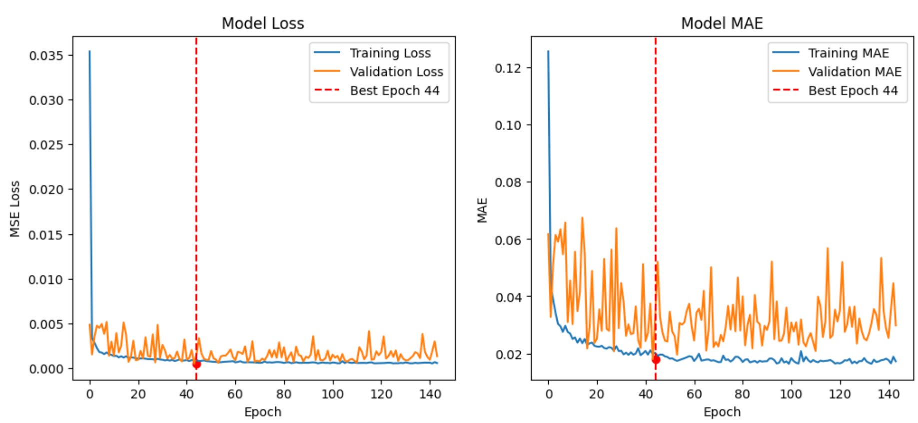

Figure 9. Bi-LSTM Model - 1Day

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 R^2 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 Bi-LSTM showed suboptimal performance, with an RMSE of 0.0337 and an R^2 value of 0.93 indicating that extending the temporal window may have introduced noise rather than enhancing predictive precision. Furthermore the CNN and CNN-Bi-LSTM hybrid models exhibited moderate accuracy (RMSE ≈ 0.035, R^2 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	R^2
Bi-LSTM (30-day)	30 days	0.0337	0.0011	0.0237	1.45%	0.93
CNN-Bi-LSTM	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

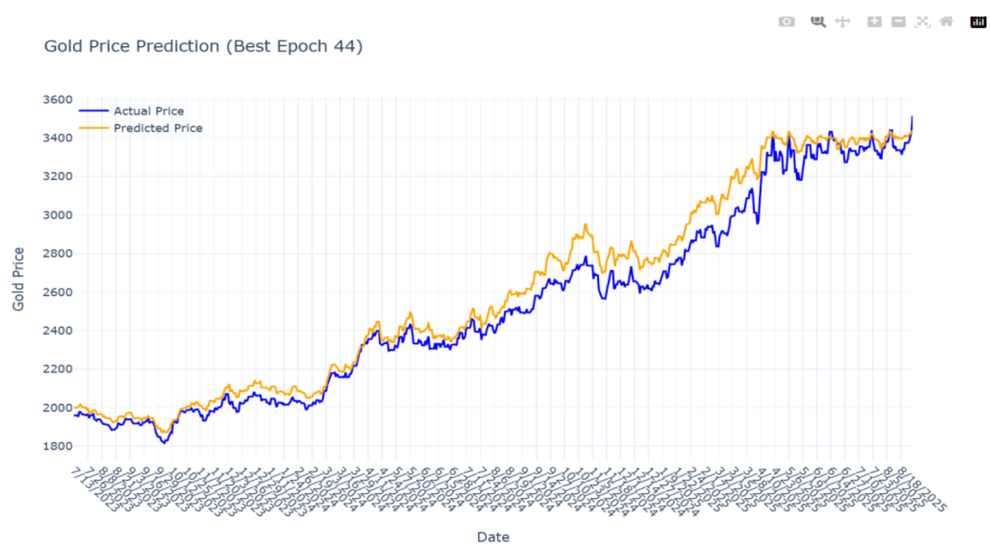

Figure 10. Actual an prediction gold price using Bi-LSTM (1-Day)

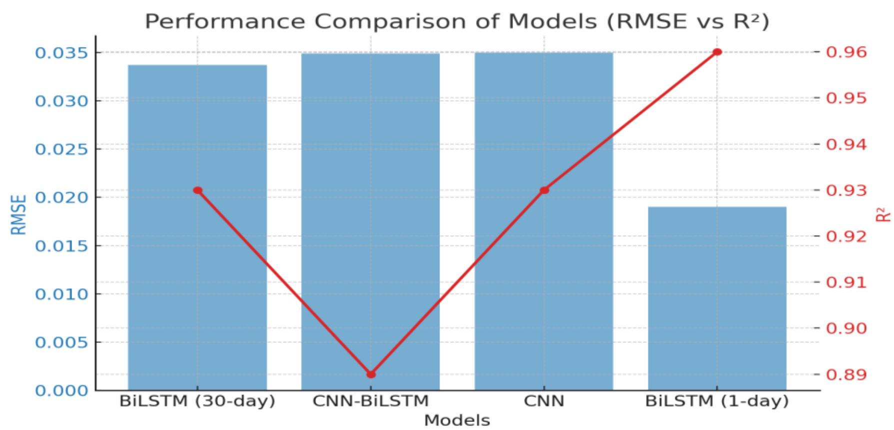

Figure 11. Performance Comparison of Models (RMSE vs R^2) the red line for R^2.

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.

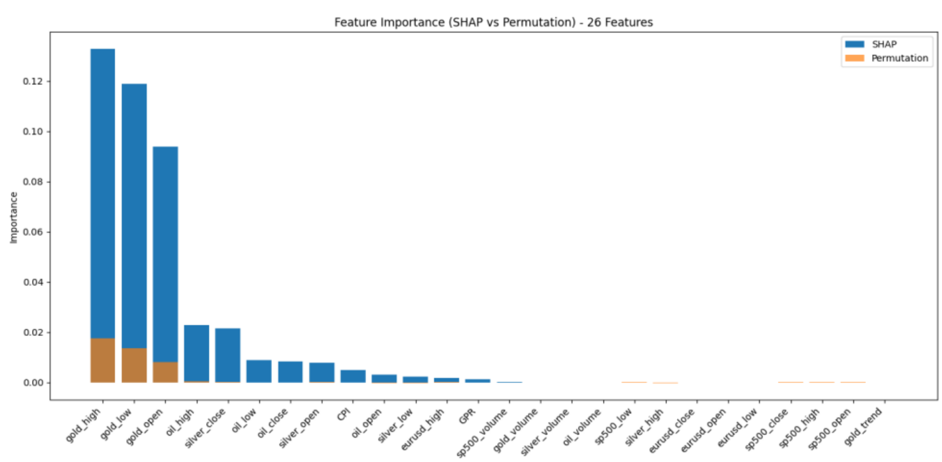

Figure 12. Feature Impotence (SHAP vs Permutation) - 26 Features.

5. Discussion and Analysis

This paper discusses the results of four deep learning models with different time windows for forecasting the next-day gold price. The goal of this study was to find the best model with the best time window and analyze the factors that influence the results from the dataset. The results showed that the Bi-LSTM model with a one-day time window performed best, achieving the lowest RMSE value of 0.0190 and the highest R2 of 0.96. By analyzing the features of our dataset, the analysis revealed that the three most influential years in gold price prediction were historical gold prices (gold_high, gold_low, gold_open). However, other economic years also had a smaller impact on gold price predictions. These results demonstrate the potential of deep learning models in predicting future gold prices, and it is essential to further develop these models to achieve adequate accuracy to support decision-makers and traders in the financial markets in making sound financial decisions.

6. Conclusion

In this study, we used deep learning models to predict gold prices by training them on a dataset combining financial market indicators, macroeconomic drivers, geopolitical risk indicators, and historical gold prices. We selected Bi-LSTM, CNN, and hybrid CNN-Bi-LSTM models. We used evaluation criteria of RMS, MSE, MAE, MAPE, and R^2, and employed early stopping procedures to prevent overfitting. The results showed that the Bi-LSTM sequence window outperformed all the models above, with the lowest RMSE of 0.0190, the lowest MAE of 0.0133, and the highest R^2 of 0.96, respectively.

This experiment demonstrated that using short-term time windows improves accuracy in predicting future gold prices. However, the hybrid model performed well over a 30-day time window.

Feature importance analysis using the SHAP method was of great importance in understanding which features in the dataset influence gold price prediction. We found that gold_high, gold_low, and gold_open were the three most important features. This suggests that training models on historical gold price data will yield accurate results.

In conclusion, this paper finds the Bi-LSTM model to be the most effective in predicting gold prices. Researchers could expand future work by training deep learning models to analyze social media users' sentiment, or applying reinforcement learning to improve decision-making processes.

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 Bi-LSTM, 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 Bi-LSTM 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-Bi-LSTM 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-Bi-LSTM-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-Bi-LSTM 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.

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

Amjed M. Mutar
Email: 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 (Bi-LSTM) model utilizing both 1-day and 30-day time frames a Convolutional Neural Network (CNN) and a hybrid "CNN-Bi-LSTM" architecture The results of the experiments indicate that the Bi-LSTM 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 R^2 value of 0.96. surpassing both the CNN and hybrid CNN-Bi-LSTM 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 "Bi-LSTM CNN Hybrid models" Time series forecasting, Feature importance Financial prediction RMSE, MSE.

2. Introduction

Gold has traditionally assumed central importance in the world economy. not simply as an inanimate store of value but as an indispensable hedge against inflation and other forms of economic volatility [3]. In recent times especially in the present day financial markets gold price has been noticed to be influenced more and more by these non classical correlation[other sources that affect the price fluctuations of gold include variations in currency exchange rates, crude oil prices, movements within stock indices, inflationary measures, interest rate pressures and also increasingly perilous political instability [4].

Incidence Up To Date prediction of future gold price has an important role to play which is desired by variety of people/organi- zations including investors who want to get the best portfolio, governments for making better under -standing about economic signals, and financial bodies for risk management effectively [5]. However predicting gold price is a hard problem because gold prices have nonlinear and high volatility nature. that make the prediction problem complicated [6]. Traditional econometric and statistical approaches, while helpful in understanding various aspects, often do not account for complex structure of financial time series data explicitly to model intricate dependencies and under-lying structures. Realizing this a shortcoming there is an ever growing interest in exploring sophisticated machine learning (ML) and deep learning (DL) models [7, 8] . Which have shown significant effectiveness towards dealing with such challenges associated with nonlinearity, modeling high-dimensional input data and capturing temporal dependencies from the data stream [9].

Among the various deep learning architectures, Recurrent Neural Networks (RNNs) and its variations including LSTM networks, bidirectional LSTMs have been widely acknowledged for performing successfully in financial prediction tasks[10]. These models outperform other existing models for the task of predicting trends in gold prices However, recent model, Convolutional Neural Networks (CNNs), which were initially designed for the image recognition problem but have now been successfully modified as an architecture to tackle time series forecasting [11]. This adaptation is attributed to their excellent ability of capturing the local pattern from the data while ignoring noise which might be more influential in economic datasets [12]. Recently, we are seeing a resurgence of hybrid models incorporating both CNN and LSTM/Bi-LSTM components in order to make use of features offered by both local feature capturing capabilities and long-term sequence learning capability [13].

This article is a valuable addition to the growing literature on gold price forecasting, and it is valuable in this study as these models were: Rigorously trai- ned, benchmarked using an

extensive dataset of not only the gold and silver prices but raw oil prices, exchange rates (EUR/USD ratio at daily frequency), s&p 500 index, Consumer Price Index (CPI), Geopolitical asymmetry Risk (GPR) indicators [14]. The models were evaluated using several evaluation methods such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and R-squared values [15]

A comprehensive feature importance assessment was also performed to detect as well as interpret the important predictors influencing gold prices. The findings of this study have emphasised the significance and the considerably acceptable performance level of the sequence-based Bi-LSTM mode particularly 1 day BiLSTM demonstrate remarkable predictive results.

These results not only provide valuable insights on hybrid deep learning models for financial forecasting, but also open the possibility for future research in the design and improvement of new architectures that target to address some of the challenges associated with financial time series [16].

Our research paper is next structured by writing the structure of research, literature review and methodology are being written for designing a full blown framework of deep learning for predicting gold price.

3. Literature Review

Gold price prediction has recently become a research hotspot within the investment With statistical analysis being the primary approach for making predictions and numerous deep learning methods have been put forward to address the issue American stock market, A literature has recently presented the top performance of deep learning in highly accurate prediction [17].

Ghahramani and Esmaeili Najafabadi (2022) introduce a new scenario for financial time series analysis comprising the gold prices. They employ different resources such as historic prices and economical characteristics aimed at a forecasting accuracy of 91% using their deep neural network model. Their results show the promise of hybrid methods in improving prediction performances [18].

In their study Tripathi and Sharma (2022) examined the combined effect of blending sentiment indicators from news articles with traditional gold price series. They reported a 15% accuracy outperform with their PLE model when they combined the historical gold price data with sentiment scores. It shows the importance of the market sentiment in terms of the price and forecasting accuracy [19].

Finally, Modi et al (2023) address a data-driven deep learning method to predict Bitcoin prices, which is used as a competing benchmark for gold prediction. They trained on feature- engineered data and obtained a 95% accuracy when testing with their shallow Bidel model This shows that deep learning algorithms are effective for making financial decisions [20].

Li Wang and Yang (2023) address the issue of risk prediction in financial management Analyses are performed on a BP neural network which is optimized under the digital economy A sample of financial data gathered from the listed companies was considered within their historical performance metrics Evaluation of the model's performance relies upon accuracy measures namely, RMSE and classification accuracy for course corrective decision making with an accuracy level as high as91% proving itself to be efficient for characterization between Business Failure and Non-Failure [1].

Ampountolas (2023) the experiment was based on multiple datasets and performance were measured using RMSE MAE where the best model hit 0.70 in RMSE proving deep learning approach's capability in predicting at financial term [16].

Foroutan and Lahmiri (2024) introduced deep learning models for forecasting the price of crude oil and precious metals using historical prices and economic indicators. Their models, tested based on RMSE loss and the most remarkable result was an RMSE equal 0.90 for precious metals, prove the power of prediction of their models [14].

Gupta and Jaiswal (2024) discuss comparative performance of different deep learning techniques for stock market forecasting. Their study supports our assumption that deep learning models especially the RNN combined with CNN models are very good in recording market trends and can achieve better results compared to traditional forecasting methods They tested data of historic gold price from 2000 to 2023, and achieved an accuracy of 92% in their predictions. This paper confirms that there is still potential for LSTM models in time series prediction and suggests that LSTM has not been well exploited yet in financial forecasting [21].

Amini and Kalantari (2024), a hybrid CNN-Bi-LSTM model is proposed for gold price prediction purposes. They used a dataset of more than 20 years in gold prices, and reported an almost incredibly low mean absolute percentage deviation (MAPE) of 3.5%. The model's RMSE evaluation results of 32.31 indicate that the model predicts gold market trends, providing very useful information for investors.[17].

Ben Ameur et al. (2024)They predicted commodity prices, including gold futures, using deep learning models. They tested a large dataset that included historical prices and macroeconomic features. Their study found that the LSTM network achieved the best

model performance metric, RMSE, of 1.28. This study also found the importance of external factors in improving model training performance.[22].

Zhao et al (2025) proposed a mixed model including Multithread Attention Enhanced Bi-LSTM, ARIMA and XGBoost function for stock price prediction by wavelet demising. Dataset The dataset containing historical stock prices and the widely used macroeconomic indicators that are selected from a variety of different databases. Our model was tested on considering root mean square errors (RMSE) and mean absolute error MAE, having RMSE value 0.85 which means there is significant improvement in comparison to Trade retracement Ratio telling Hybrid mode has capability of enhancing forecast Accuracy[2].

Bagrecha et al. (2025) used a univariate ARIMA method to predict silver price by using the historical data of silver price. Their precision were evaluated with RMSE the results of an RMSE for silver price prediction 1.15 which mean that ARIMA model has good performance in the silver price prediction They proposed a new model to develop future price direction predictions according to their discoveries [5].

Kong et al. (2025) They comprehensively reviewed a set of deep learning models for time series forecasting. They used a dataset consisting of stock prices and economic indicators. They also used performance metrics for deep learning models, including RMSE and MAE. Their study indicated that the hybrid CNN-LSTM model outperformed the other models, achieving the best performance with an RMSE of 0.80 [8].

3. Methodology

We followed a methodology in this paper that relies on collecting diverse historical data, not only for gold prices, but also for silver, oil, economic indicators, consumer prices, and geopolitical risks, to create a robust and diverse dataset.

We will then predict future gold prices using deep learning models, specifically neural networks (Bi-LSTM, CNN, and a hybrid CNN-Bi-LSTM model) [23].

We will provide a comprehensive comparison of the models and their ability to predict future prices. We will also use techniques to identify which features in the dataset have a direct impact on future price predictions, providing a clear picture of how models are trained on the dataset. [24].

3.1 Dataset Collection

We collected the dataset for this study using the Yahoo Finance API, which included historical prices for gold, silver, and oil, as well as exchange rates for major currencies (EUR/USD). We also collected data on economic indicators for the stock market (S&P 500). This dataset was collected for a ten-year period, from January 1, 2015, to August 29,

2025 [17]. To diversify the dataset, we added another economic indicator, the Consumer Price Index (CPI), and the Geopolitical Risk Index (GPR), for the same period to ensure consistency and reliability of the dataset [5]. We then processed the data, standardized it, and filled in any missing data. We now have a dataset consisting of 3,894 days, comprising 27 features or columns, including the target category, the closing price of gold, as shown in Figure 1.

```
Index: 3894 entries, 1/1/2015 to 8/29/2025
Data columns (total 27 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   gold_open      3894 non-null   float64
 1   gold_high      3894 non-null   float64
 2   gold_low       3894 non-null   float64
 3   gold_close     3894 non-null   float64
 4   gold_volume    3894 non-null   int64
 5   silver_open    3894 non-null   float64
 6   silver_high    3894 non-null   float64
 7   silver_low     3894 non-null   float64
 8   silver_close   3894 non-null   float64
 9   silver_volume  3894 non-null   int64
 10  oil_open       3894 non-null   float64
 11  oil_high       3894 non-null   float64
 12  oil_low        3894 non-null   float64
 13  oil_close      3894 non-null   float64
 14  oil_volume     3894 non-null   int64
 15  eurusd_open    3894 non-null   float64
 16  eurusd_high    3894 non-null   float64
 17  eurusd_low     3894 non-null   float64
 18  eurusd_close   3894 non-null   float64
 19  sp500_open     3894 non-null   float64
 20  sp500_high     3894 non-null   float64
 21  sp500_low      3894 non-null   float64
 22  sp500_close    3894 non-null   float64
 23  sp500_volume   3894 non-null   int64
 24  gold_trend     3894 non-null   int64
 25  CPI            3894 non-null   float64
 26  GPR            3894 non-null   float64
dtypes: float64(22), int64(5)
memory usage: 851.8+ KB
```

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

Impute missing values to ensure much time consistency as possible throughout the analysis, missing values within the data were filled forward and backward with a common technique - Forward-Fill/Backward-Fill This two-sta ged approach 'filledin' gaps in the dataset efficiently(protocol). 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 Min Max Scaler 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].

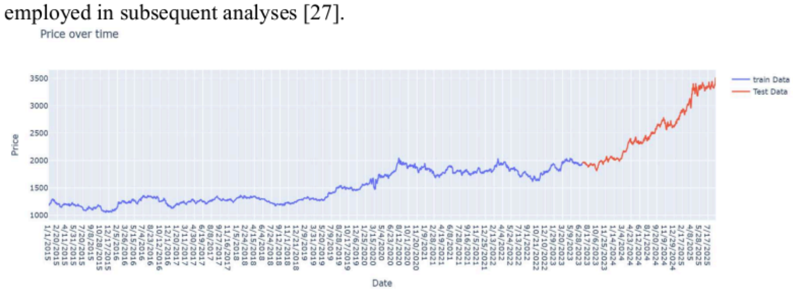

Figure 2.Showing the training data which is 80% and the test data which is 20% of the dataset**.**

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

14

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 Bi-LSTM 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. Bi-LSTM 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-Bi-LSTM model leverages CNN for high - level feature extraction Bi-LSTM for temporal modeling, and an attention mechanism for feature fusion [17]. Consequently, three representative deep-learning techniques are chosen for comparison:- Bi-LSTM CNN and hybrid CNN-Bi-LSTM Hybrid methods typically outperform single models due to their complementary capabilities [13].

3.5.1. Bi-LSTM Model (30-day & 1-day) Architecture

The Bi-LSTM 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, Bi-LSTM 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.

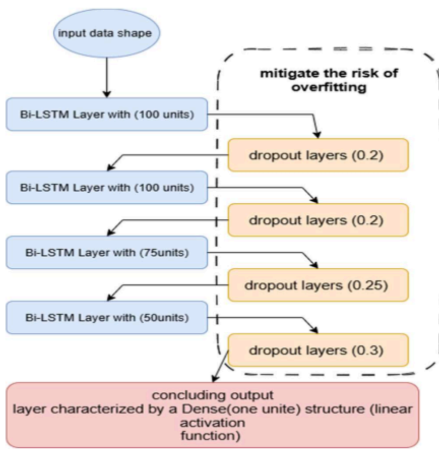

Figure 3 . As shown the Bi-LSTM model

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.

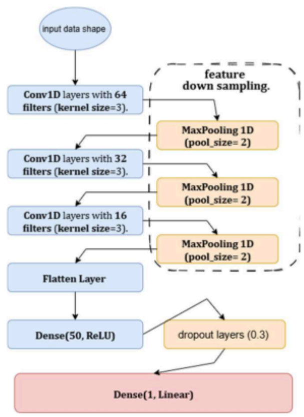

Figure 4 .shown CNN model architecture

3.5.3. Hybrid CNN-Bi-LSTM Model Architecture

The hybrid CNN-Bi-LSTM is a deep learning model that combines a one CNN with Bi_LSTM 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 Bi_LSTM which captures temporal dependencies by processing sequences in both directions, enhancing the understanding of context. This architecture is based on research showing Bi_LSTM'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-Bi-LSTM hybrid model architecture

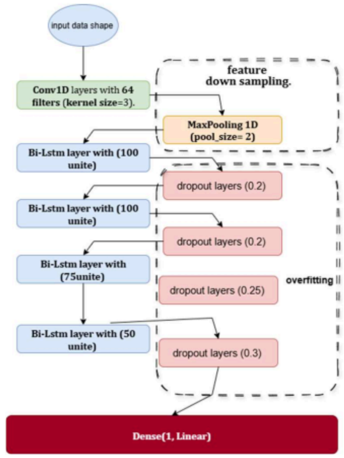

Figure 5 shown CNN-Bi-LSTM hybrid model architecture

3.6 Training and Optimization

Presented below are the optimization function learning rat 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

4.1. 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].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_{actu} - X_{pred}\right| \tag{4.1}$$

where n is the number of observations, X_{pred} is the predicted value for observation I and X_{actu} 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].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_{actu} - X_{pred})^2 \tag{4.2}$$

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].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{actu} - X_{pred})^2} \tag{4.3}$$

where $X_{actu\text{-}Max}$ and $X_{actu\text{-}Min}$ 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].

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_{actu} - X_{pred}}{X_{actu}}\right| * 100\% \tag{4.4}$$

R-squared (R^2):- 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 R^2 value signifies superior performance. This relationship can be articulated in Equation 4.5.[28].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_{actu} - X_{pred})^2}{\sum_{i=1}^{n}(X_{actu} - X_{pred})^2} \quad (4.5)$$

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.

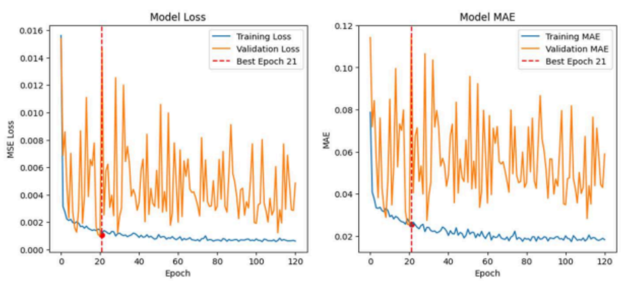

Figure 6. Bi-LSTM Model - 30Day

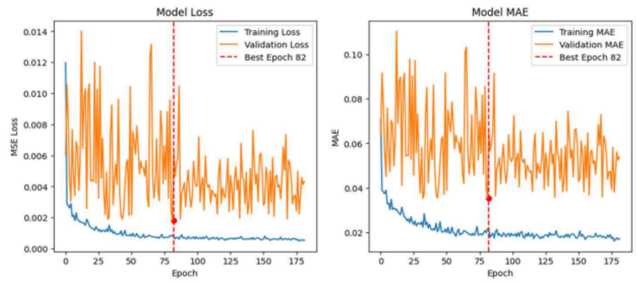

Figure 7. CNN-Bi-LSTM Model

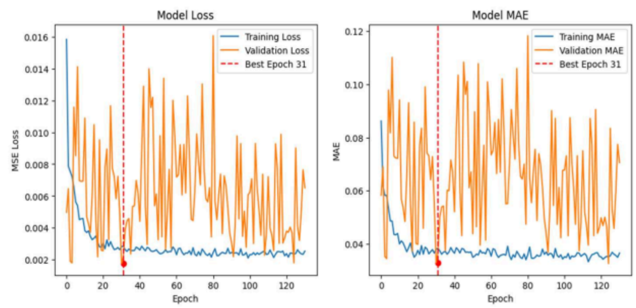

Figure 8. CNN Model

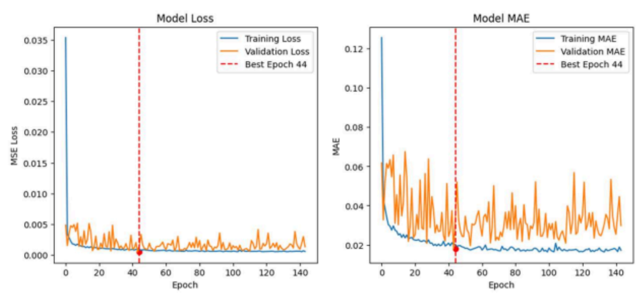

Figure 9. Bi-LSTM Model - 1Day

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 R^2 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 Bi-LSTM showed suboptimal performance, with an RMSE of 0.0337 and an R^2 value of 0.93 indicating that extending the temporal window may have introduced noise rather than enhancing predictive precision. Furthermore the CNN and CNN-Bi-LSTM hybrid models exhibited moderate accuracy (RMSE ≈ 0.035, R^2 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	R^2
Bi-LSTM (30-day)	30 days	0.0337	0.0011	0.0237	1.45%	0.93
CNN-Bi-LSTM	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)

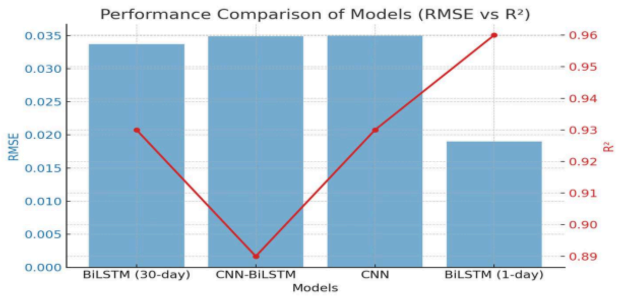

Figure 11. Performance Comparison of Models (RMSE vs R^2) the red line for R^2.

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.

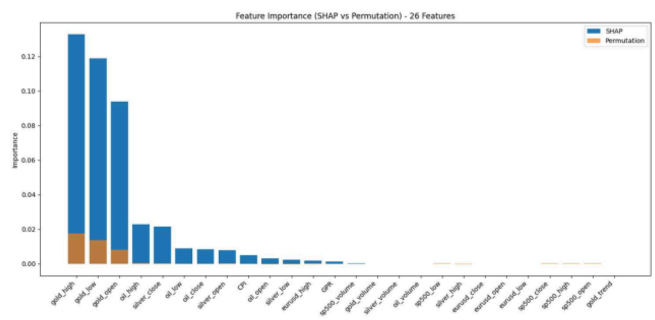

Figure 12. Feature Impotence (SHAP vs Permutation) - 26 Features.

5. Discussion and Analysis

This paper discusses the results of four deep learning models with different time windows for forecasting the next-day gold price. The goal of this study was to find the best model with the best time window and analyze the factors that influence the results from the dataset. The results showed that the Bi-LSTM model with a one-day time window performed best, achieving the lowest RMSE value of 0.0190 and the highest R2 of 0.96. By analyzing the features of our dataset, the analysis revealed that the three most influential years in gold price prediction were historical gold prices (gold_high, gold_low, gold_open). However, other economic years also had a smaller impact on gold price predictions. These results demonstrate the potential of deep learning models in predicting future gold prices, and it is essential to further develop these models to achieve adequate accuracy to support decision-makers and traders in the financial markets in making sound financial decisions.

6. Conclusion

In this study, we used deep learning models to predict gold prices by training them on a dataset combining financial market indicators, macroeconomic drivers, geopolitical risk indicators, and historical gold prices. We selected Bi-LSTM, CNN, and hybrid CNN-Bi-LSTM models. We used evaluation criteria of RMS, MSE, MAE, MAPE, and R^2, and employed early stopping procedures to prevent overfitting. The results showed that the Bi-LSTM sequence window outperformed all the models above, with the lowest RMSE of 0.0190, the lowest MAE of 0.0133, and the highest R^2 of 0.96, respectively.

This experiment demonstrated that using short-term time windows improves accuracy in predicting future gold prices. However, the hybrid model performed well over a 30-day time window.

Feature importance analysis using the SHAP method was of great importance in understanding which features in the dataset influence gold price prediction. We found that gold_high, gold_low, and gold_open were the three most important features. This suggests that training models on historical gold price data will yield accurate results.

In conclusion, this paper finds the Bi-LSTM model to be the most effective in predicting gold prices. Researchers could expand future work by training deep learning models to analyze social media users' sentiment, or applying reinforcement learning to improve decision-making processes.

paper_7

ORIGINALITY REPORT

11%	6%	9%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1%
2	Poonam Nandal, Mamta Dahiya, Meeta Singh, Arvind Dagur, Brijesh Kumar. "Progressive Computational Intelligence, Information Technology and Networking", CRC Press, 2025 Publication	1%
3	Submitted to University of Reading Student Paper	1%
4	Submitted to University of Bahrain Student Paper	<1%
5	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
6	Cristiana Tudor, Margareta Florescu, Persefoni Polychronidou, Pavlos Stamatiou, Vasileios Vlachos, Konstadina Kasabali. "Cloud Adoption in the Digital Era: An Interpretable Machine Learning Analysis of National Readiness and Structural Disparities Across the EU", Applied Sciences, 2025 Publication	<1%
7	journal.su.edu.ly Internet Source	<1%
8	www-emerald-com-443.webvpn.sxu.edu.cn Internet Source	<1%

9	www.ncbi.nlm.nih.gov Internet Source	<1 %
10	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	<1 %
11	Xiuwei Yan, Sijia Liu, Songlei Wang, Jiarui Cui et al. "Predictive Analysis of Linoleic Acid in Red Meat Employing Advanced Ensemble Models of Bayesian and CNN-Bi-LSTM Decision Layer Fusion Based Hyperspectral Imaging", Foods, 2024 Publication	<1 %
12	"Artificial Intelligence Technologies and Applications", IOS Press, 2024 Publication	<1 %
13	Submitted to Institute of Research & Postgraduate Studies, Universiti Kuala Lumpur Student Paper	<1 %
14	bitcoinethereumnews.com Internet Source	<1 %
15	Markapurapu John Dana Ebinezer, Bondalapu Chaitanya Krishna. "Life Insurance Fraud Detection: A Data-Driven Approach Utilizing Ensemble Learning, CVAE, and Bi-LSTM", Applied Sciences, 2025 Publication	<1 %
16	Noor M. Alkudah, Norisma Idris, Aznul Qalid Md Sa, Mohammad A. M. Abushariah. "An Empirical Comparison of Ensemble model and Deep Learning Models for Multi-Level Arabic Fake News Classification using	<1 %

JoNewsFake Dataset", Springer Science and Business Media LLC, 2025
Publication

17	Zhentao Huang, Yahong Ma, Rongrong Wang, Weisu Li, Yongsheng Dai. "A Model for EEG-Based Emotion Recognition: CNN-Bi-LSTM with Attention Mechanism", Electronics, 2023 Publication	<1 %
18	Balakumar Palaniyappan, Senthil Kumar R, Vinopraba T. "Dynamic pricing for load shifting: Reducing electric vehicle charging impacts on the grid through machine learning-based demand response", Sustainable Cities and Society, 2024 Publication	<1 %
19	Mariya Ouaissa, Mariyam Ouaissa, Zakaria Boulouard, Abhishek Kumar, Vandana Sharma, Keshav Kaushik. "Artificial Intelligence for Blockchain and Cybersecurity Powered IoT Applications", Routledge, 2024 Publication	<1 %
20	scholar.lib.vt.edu Internet Source	<1 %
21	umpir.ump.edu.my Internet Source	<1 %
22	dlabi.org Internet Source	<1 %
23	Arwa A. Al Shamsi, Sherief Abdallah. "Ensemble Stacking Model for Sentiment Analysis of Emirati and Arabic Dialects", Journal of King Saud University - Computer and Information Sciences, 2023 Publication	<1 %

24	Ghazi Abdalla, Fatih Özyurt. "Sentiment Analysis of Fast Food Companies With Deep Learning Models", The Computer Journal, 2020 Publication	<1 %
25	Submitted to University of Kelaniya Student Paper	<1 %
26	anserpress.org Internet Source	<1 %
27	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
28	C H Ayishathahira, C Sreejith, C Raseek. "Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing", 2018 International CET Conference on Control, Communication, and Computing (IC4), 2018 Publication	<1 %
29	Chun Jin, Bo Li, Yanli Yang, Xiaodong Yuan, Rang Tu, Linbin Qiu, Xu Chen. "Remaining Useful Life Prediction of Rolling Bearings Based on Empirical Mode Decomposition and Transformer Bi-LSTM Network", Applied Sciences, 2025 Publication	<1 %
30	Hassan Raza, Zafar Akhtar. "Predicting stock prices in the Pakistan market using machine learning and technical indicators", Modern Finance, 2024 Publication	<1 %
31	export.arxiv.org Internet Source	<1 %
32	www.informatica.si Internet Source	

<1%

33	Abhishek Kumar Tripathi, Mangalpady Aruna, P.V. Elumalai, Krishnasamy Karthik et al. "Advancing Solar PV Panel Power Prediction: A Comparative Machine Learning Approach in Fluctuating Environmental Conditions", Case Studies in Thermal Engineering, 2024 Publication	<1%
34	Submitted to British University In Dubai Student Paper	<1%
35	GOPICHAND BANDARUPALLI. "Advancing Smart Transportation via AI for Sustainable Traffic Solutions in Saudi Arabia", Springer Science and Business Media LLC, 2024 Publication	<1%
36	Sewon Lim, Minji Park, Hajin Kim, Seong-Hyeon Kang, Youngjin Lee. "Quantitative analysis of fast non-local means algorithm smoothing factor impact on lung segmentation accuracy in computed tomography images", Radiation Physics and Chemistry, 2024 Publication	<1%
37	Sujithra Kanmani, Surendiran Balasubramanian. "Leveraging Readability and Sentiment in Spam Review Filtering Using Transformer Models", Computer Systems Science and Engineering, 2023 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

Sadik Gharghan

paper_7

Quick Submit

Quick Submit

Association of Educators - 2

Document Details

Submission ID

trn:oid:::1:3352433561

Submission Date

Sep 26, 2025, 5:24 PM GMT+3

Download Date

Sep 26, 2025, 5:27 PM GMT+3

File Name

__v7.docx

File Size

1.2 MB

18 Pages

4,443 Words

25,867 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

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

Amjed M. Mutar
Email: 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 (Bi-LSTM) model utilizing both 1-day and 30-day time frames a Convolutional Neural Network (CNN) and a hybrid "CNN-Bi-LSTM" architecture The results of the experiments indicate that the Bi-LSTM 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 R^2 value of 0.96. surpassing both the CNN and hybrid CNN-Bi-LSTM 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 "Bi-LSTM CNN Hybrid models" Time series forecasting, Feature importance Financial prediction RMSE, MSE.

2. Introduction

Gold has traditionally assumed central importance in the world economy. not simply as an inanimate store of value but as an indispensable hedge against inflation and other forms of economic volatility [3]. In recent times especially in the present day financial markets gold price has been noticed to be influenced more and more by these non classical correlation[other sources that affect the price fluctuations of gold include variations in currency exchange rates, crude oil prices, movements within stock indices, inflationary measures, interest rate pressures and also increasingly perilous political instability [4].

Incidence Up To Date prediction of future gold price has an important role to play which is desired by variety of people/organi- zations including investors who want to get the best portfolio, governments for making better under -standing about economic signals, and financial bodies for risk management effectively [5]. However predicting gold price is a hard problem because gold prices have nonlinear and high volatility nature. that make the prediction problem complicated [6]. Traditional econometric and statistical approaches, while helpful in understanding various aspects, often do not account for complex structure of financial time series data explicitly to model intricate dependencies and under-lying structures. Realizing this a shortcoming there is an ever growing interest in exploring sophisticated machine learning (ML) and deep learning (DL) models [7, 8] . Which have shown significant effectiveness towards dealing with such challenges associated with nonlinearity, modeling high-dimensional input data and capturing temporal dependencies from the data stream [9].

Among the various deep learning architectures, Recurrent Neural Networks (RNNs) and its variations including LSTM networks, bidirectional LSTMs have been widely acknowledged for performing successfully in financial prediction tasks[10]. These models outperform other existing models for the task of predicting trends in gold prices However, recent model, Convolutional Neural Networks (CNNs), which were initially designed for the image recognition problem but have now been successfully modified as an architecture to tackle time series forecasting [11]. This adaptation is attributed to their excellent ability of capturing the local pattern from the data while ignoring noise which might be more influential in economic datasets [12]. Recently, we are seeing a resurgence of hybrid models incorporating both CNN and LSTM/Bi-LSTM components in order to make use of features offered by both local feature capturing capabilities and long-term sequence learning capability [13].

This article is a valuable addition to the growing literature on gold price forecasting, and it is valuable in this study as these models were: Rigorously trai- ned, benchmarked using an

extensive dataset of not only the gold and silver prices but raw oil prices, exchange rates (EUR/USD ratio at daily frequency), s&p 500 index, Consumer Price Index (CPI), Geopolitical asymmetry Risk (GPR) indicators [14]. The models were evaluated using several evaluation methods such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and R-squared values [15]

A comprehensive feature importance assessment was also performed to detect as well as interpret the important predictors influencing gold prices. The findings of this study have emphasised the significance and the considerably acceptable performance level of the sequence-based Bi-LSTM mode particularly 1 day BiLSTM demonstrate remarkable predictive results.

These results not only provide valuable insights on hybrid deep learning models for financial forecasting, but also open the possibility for future research in the design and improvement of new architectures that target to address some of the challenges associated with financial time series [16].

Our research paper is next structured by writing the structure of research, literature review and methodology are being written for designing a full blown framework of deep learning for predicting gold price.

3. Literature Review

Gold price prediction has recently become a research hotspot within the investment With statistical analysis being the primary approach for making predictions and numerous deep learning methods have been put forward to address the issue American stock market, A literature has recently presented the top performance of deep learning in highly accurate prediction [17].

Ghahramani and Esmaeili Najafabadi (2022) introduce a new scenario for financial time series analysis comprising the gold prices. They employ different resources such as historic prices and economical characteristics aimed at a forecasting accuracy of 91% using their deep neural network model. Their results show the promise of hybrid methods in improving prediction performances [18].

In their study Tripathi and Sharma (2022) examined the combined effect of blending sentiment indicators from news articles with traditional gold price series. They reported a 15% accuracy outperform with their PLE model when they combined the historical gold price data with sentiment scores. It shows the importance of the market sentiment in terms of the price and forecasting accuracy [19].

Finally, Modi et al (2023) address a data-driven deep learning method to predict Bitcoin prices, which is used as a competing benchmark for gold prediction. They trained on feature- engineered data and obtained a 95% accuracy when testing with their shallow Bidel model This shows that deep learning algorithms are effective for making financial decisions [20].

Li Wang and Yang (2023) address the issue of risk prediction in financial management Analyses are performed on a BP neural network which is optimized under the digital economy A sample of financial data gathered from the listed companies was considered within their historical performance metrics Evaluation of the model's performance relies upon accuracy measures namely, RMSE and classification accuracy for course corrective decision making with an accuracy level as high as91% proving itself to be efficient for characterization between Business Failure and Non-Failure [1].

Ampountolas (2023) the experiment was based on multiple datasets and performance were measured using RMSE MAE where the best model hit 0.70 in RMSE proving deep learning approach's capability in predicting at financial term [16].

Foroutan and Lahmiri (2024) introduced deep learning models for forecasting the price of crude oil and precious metals using historical prices and economic indicators. Their models, tested based on RMSE loss and the most remarkable result was an RMSE equal 0.90 for precious metals, prove the power of prediction of their models [14].

Gupta and Jaiswal (2024) discuss comparative performance of different deep learning techniques for stock market forecasting. Their study supports our assumption that deep learning models especially the RNN combined with CNN models are very good in recording market trends and can achieve better results compared to traditional forecasting methods They tested data of historic gold price from 2000 to 2023, and achieved an accuracy of 92% in their predictions. This paper confirms that there is still potential for LSTM models in time series prediction and suggests that LSTM has not been well exploited yet in financial forecasting [21].

Amini and Kalantari (2024), a hybrid CNN-Bi-LSTM model is proposed for gold price prediction purposes. They used a dataset of more than 20 years in gold prices, and reported an almost incredibly low mean absolute percentage deviation (MAPE) of 3.5%. The model's RMSE evaluation results of 32.31 indicate that the model predicts gold market trends, providing very useful information for investors.[17].

Ben Ameur et al. (2024)They predicted commodity prices, including gold futures, using deep learning models. They tested a large dataset that included historical prices and macroeconomic features. Their study found that the LSTM network achieved the best

model performance metric, RMSE, of 1.28. This study also found the importance of external factors in improving model training performance.[22].

Zhao et al (2025) proposed a mixed model including Multithread Attention Enhanced Bi-LSTM, ARIMA and XGBoost function for stock price prediction by wavelet demising. Dataset The dataset containing historical stock prices and the widely used macroeconomic indicators that are selected from a variety of different databases. Our model was tested on considering root mean square errors (RMSE) and mean absolute error MAE, having RMSE value 0.85 which means there is significant improvement in comparison to Trade retracement Ratio telling Hybrid mode has capability of enhancing forecast Accuracy[2].

Bagrecha et al. (2025) used a univariate ARIMA method to predict silver price by using the historical data of silver price. Their precision were evaluated with RMSE the results of an RMSE for silver price prediction 1.15 which mean that ARIMA model has good performance in the silver price prediction They proposed a new model to develop future price direction predictions according to their discoveries [5].

Kong et al. (2025) They comprehensively reviewed a set of deep learning models for time series forecasting. They used a dataset consisting of stock prices and economic indicators. They also used performance metrics for deep learning models, including RMSE and MAE. Their study indicated that the hybrid CNN-LSTM model outperformed the other models, achieving the best performance with an RMSE of 0.80 [8].

3. Methodology

We followed a methodology in this paper that relies on collecting diverse historical data, not only for gold prices, but also for silver, oil, economic indicators, consumer prices, and geopolitical risks, to create a robust and diverse dataset.

We will then predict future gold prices using deep learning models, specifically neural networks (Bi-LSTM, CNN, and a hybrid CNN-Bi-LSTM model) [23].

We will provide a comprehensive comparison of the models and their ability to predict future prices. We will also use techniques to identify which features in the dataset have a direct impact on future price predictions, providing a clear picture of how models are trained on the dataset. [24].

3.1 Dataset Collection

We collected the dataset for this study using the Yahoo Finance API, which included historical prices for gold, silver, and oil, as well as exchange rates for major currencies (EUR/USD). We also collected data on economic indicators for the stock market (S&P 500). This dataset was collected for a ten-year period, from January 1, 2015, to August 29,

2025 [17]. To diversify the dataset, we added another economic indicator, the Consumer Price Index (CPI), and the Geopolitical Risk Index (GPR), for the same period to ensure consistency and reliability of the dataset [5]. We then processed the data, standardized it, and filled in any missing data. We now have a dataset consisting of 3,894 days, comprising 27 features or columns, including the target category, the closing price of gold, as shown in Figure 1.

```
Index: 3894 entries, 1/1/2015 to 8/29/2025
Data columns (total 27 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   gold_open      3894 non-null   float64
 1   gold_high      3894 non-null   float64
 2   gold_low       3894 non-null   float64
 3   gold_close     3894 non-null   float64
 4   gold_volume    3894 non-null   int64
 5   silver_open    3894 non-null   float64
 6   silver_high    3894 non-null   float64
 7   silver_low     3894 non-null   float64
 8   silver_close   3894 non-null   float64
 9   silver_volume  3894 non-null   int64
 10  oil_open       3894 non-null   float64
 11  oil_high       3894 non-null   float64
 12  oil_low        3894 non-null   float64
 13  oil_close      3894 non-null   float64
 14  oil_volume     3894 non-null   int64
 15  eurusd_open    3894 non-null   float64
 16  eurusd_high    3894 non-null   float64
 17  eurusd_low     3894 non-null   float64
 18  eurusd_close   3894 non-null   float64
 19  sp500_open     3894 non-null   float64
 20  sp500_high     3894 non-null   float64
 21  sp500_low      3894 non-null   float64
 22  sp500_close    3894 non-null   float64
 23  sp500_volume   3894 non-null   int64
 24  gold_trend     3894 non-null   int64
 25  CPI            3894 non-null   float64
 26  GPR            3894 non-null   float64
dtypes: float64(22), int64(5)
memory usage: 851.8+ KB
```

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

Impute missing values to ensure much time consistency as possible throughout the analysis, missing values within the data were filled forward and backward with a common technique - Forward-Fill/Backward-Fill This two-sta ged approach 'filledin' gaps in the dataset efficiently(protocol). 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 Min Max Scaler 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 Bi-LSTM 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. Bi-LSTM 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-Bi-LSTM model leverages CNN for high - level feature extraction Bi-LSTM for temporal modeling, and an attention mechanism for feature fusion [17]. Consequently, three representative deep-learning techniques are chosen for comparison:- Bi-LSTM CNN and hybrid CNN-Bi-LSTM Hybrid methods typically outperform single models due to their complementary capabilities [13].

3.5.1. Bi-LSTM Model (30-day & 1-day) **Architecture**
The Bi-LSTM 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, Bi-LSTM 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.

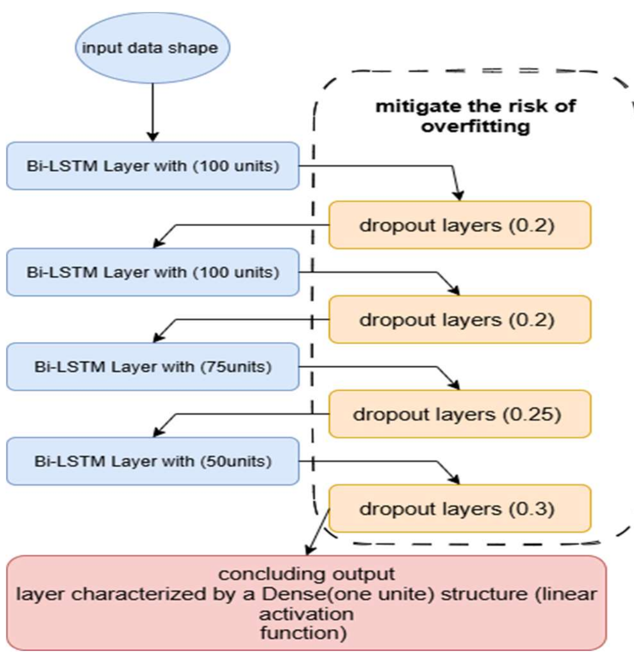

Figure 3 . As shown the Bi-LSTM model

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.

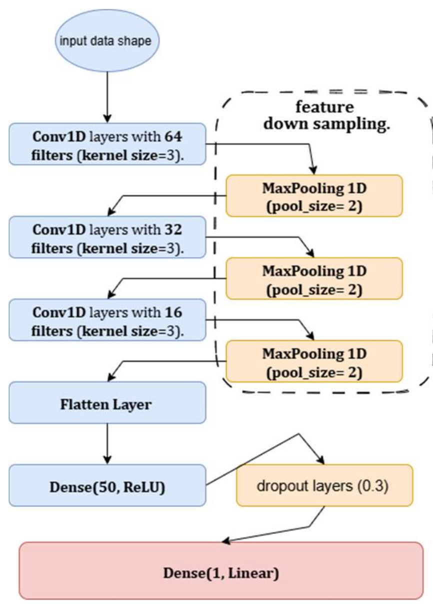

Figure 4 .shown CNN model architecture

3.5.3. Hybrid CNN-Bi-LSTM Model Architecture

The hybrid CNN-Bi-LSTM is a deep learning model that combines a one CNN with Bi_LSTM 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 Bi_LSTM which captures temporal dependencies by processing sequences in both directions, enhancing the understanding of context. This architecture is based on research showing Bi_LSTM'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-Bi-LSTM hybrid model architecture

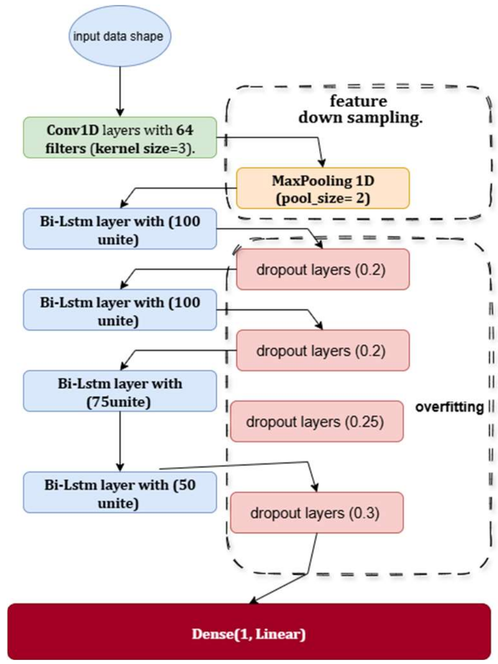

Figure 5 shown CNN-Bi-LSTM hybrid model architecture

3.6 Training and Optimization

Presented below are the optimization function learning rat 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

4.1. 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].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_{actu} - X_{pred}\right| \tag{4.1}$$

where n is the number of observations, X_{pred} is the predicted value for observation I and X_{actu} 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].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_{actu} - X_{pred})^2 \tag{4.2}$$

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].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{actu} - X_{pred})^2} \tag{4.3}$$

where $X_{actu\text{-}Max}$ and $X_{actu\text{-}Min}$ 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].

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_{actu} - X_{pred}}{X_{actu}}\right| * 100\% \tag{4.4}$$

R-squared (R^2):- 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 R^2 value signifies superior performance. This relationship can be articulated in Equation 4.5.[28].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_{actu} - X_{pred})^2}{\sum_{i=1}^{n}(X_{actu} - X_{pred})^2} \quad (4.5)$$

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.

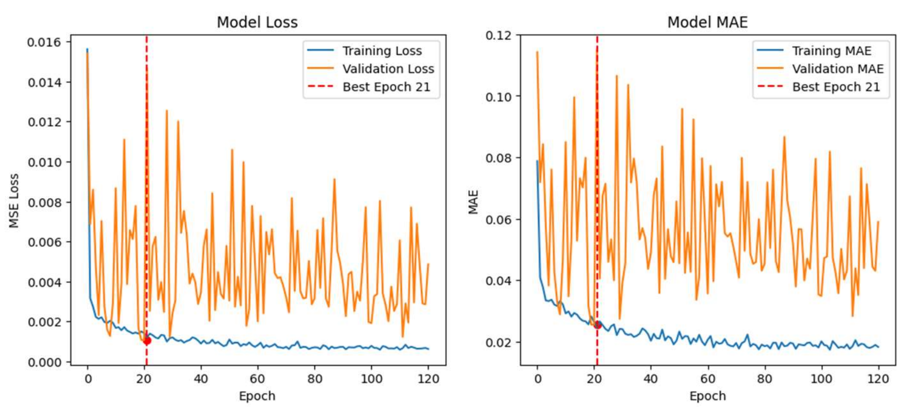

Figure 6. Bi-LSTM Model - 30Day

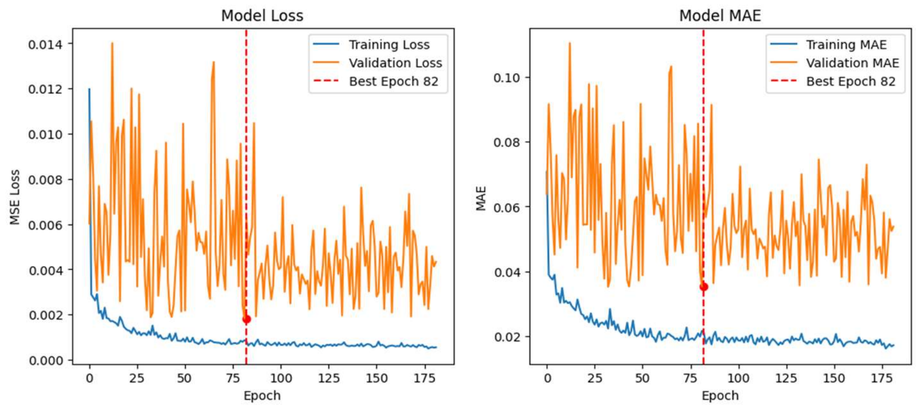

Figure 7. CNN-Bi-LSTM Model

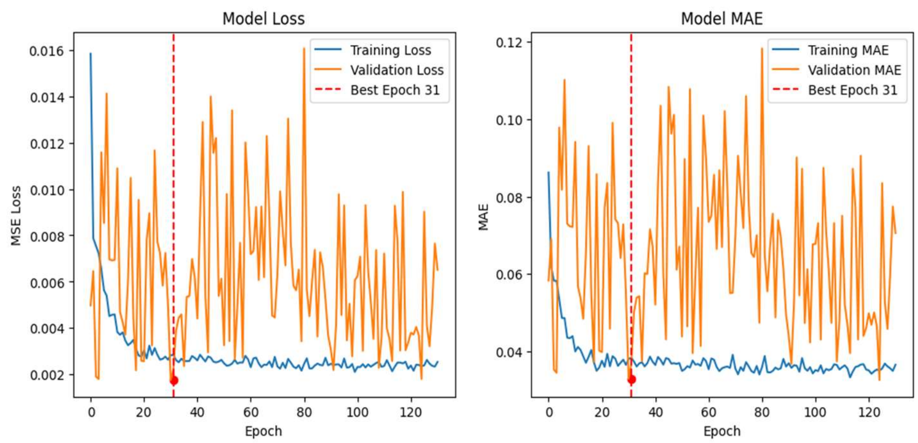

Figure 8. CNN Model

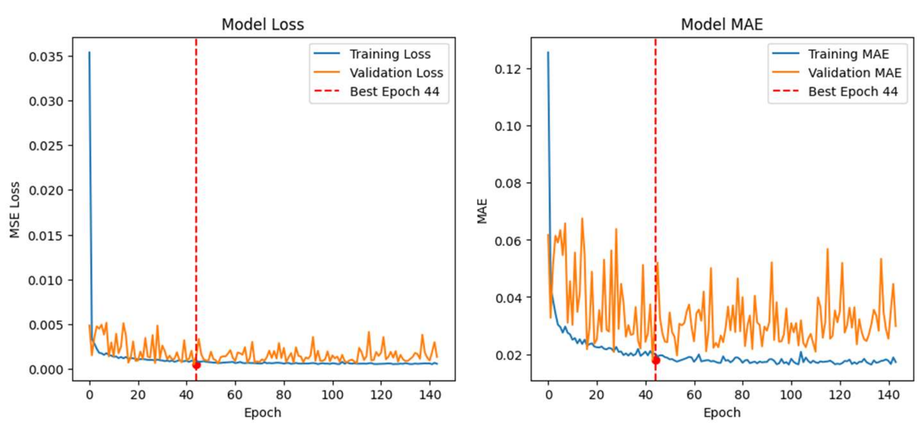

Figure 9. Bi-LSTM Model - 1Day

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 R^2 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 Bi-LSTM showed suboptimal performance, with an RMSE of 0.0337 and an R^2 value of 0.93 indicating that extending the temporal window may have introduced noise rather than enhancing predictive precision. Furthermore the CNN and CNN-Bi-LSTM hybrid models exhibited moderate accuracy (RMSE $\approx$ 0.035, R^2 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	R^2
Bi-LSTM (30-day)	30 days	0.0337	0.0011	0.0237	1.45%	0.93
CNN-Bi-LSTM	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)

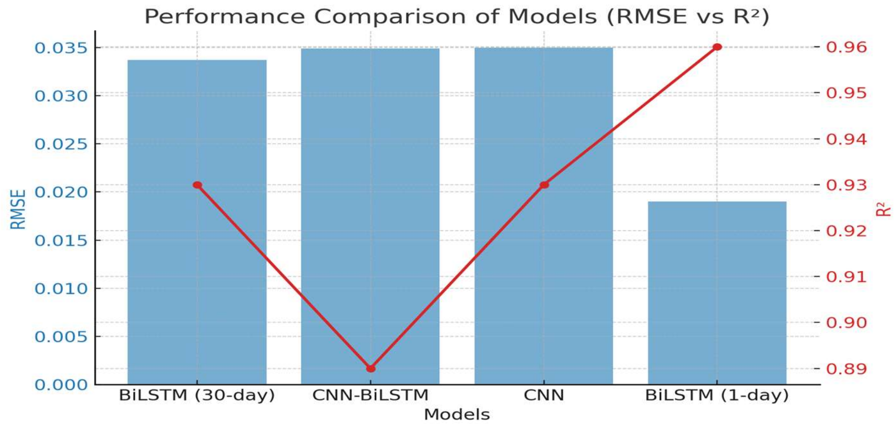

Figure 11. Performance Comparison of Models (RMSE vs R^2) the red line for R^2.

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.

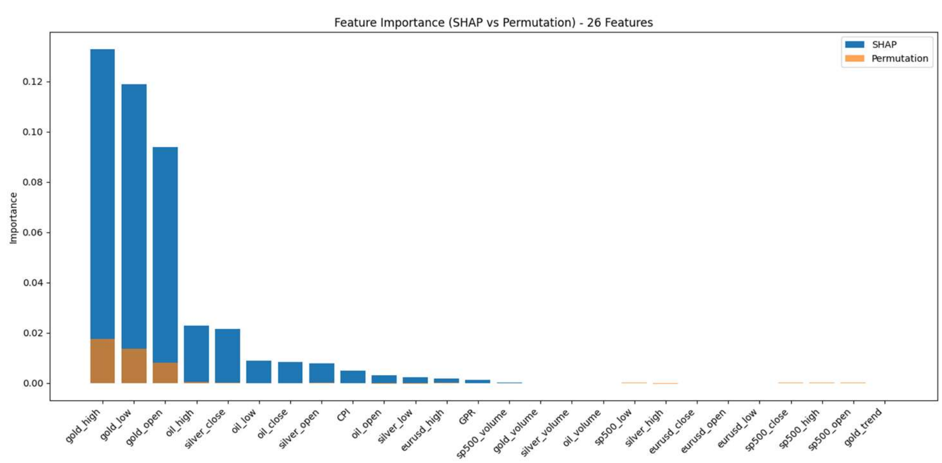

Figure 12. Feature Impotence (SHAP vs Permutation) - 26 Features.

5. Discussion and Analysis

This paper discusses the results of four deep learning models with different time windows for forecasting the next-day gold price. The goal of this study was to find the best model with the best time window and analyze the factors that influence the results from the dataset. The results showed that the Bi-LSTM model with a one-day time window performed best, achieving the lowest RMSE value of 0.0190 and the highest R2 of 0.96. By analyzing the features of our dataset, the analysis revealed that the three most influential years in gold price prediction were historical gold prices (gold_high, gold_low, gold_open). However, other economic years also had a smaller impact on gold price predictions. These results demonstrate the potential of deep learning models in predicting future gold prices, and it is essential to further develop these models to achieve adequate accuracy to support decision-makers and traders in the financial markets in making sound financial decisions.

6. Conclusion

In this study, we used deep learning models to predict gold prices by training them on a dataset combining financial market indicators, macroeconomic drivers, geopolitical risk indicators, and historical gold prices. We selected Bi-LSTM, CNN, and hybrid CNN-Bi-LSTM models. We used evaluation criteria of RMS, MSE, MAE, MAPE, and R^2, and employed early stopping procedures to prevent overfitting. The results showed that the Bi-LSTM sequence window outperformed all the models above, with the lowest RMSE of 0.0190, the lowest MAE of 0.0133, and the highest R^2 of 0.96, respectively.

This experiment demonstrated that using short-term time windows improves accuracy in predicting future gold prices. However, the hybrid model performed well over a 30-day time window.

Feature importance analysis using the SHAP method was of great importance in understanding which features in the dataset influence gold price prediction. We found that gold_high, gold_low, and gold_open were the three most important features. This suggests that training models on historical gold price data will yield accurate results.

In conclusion, this paper finds the Bi-LSTM model to be the most effective in predicting gold prices. Researchers could expand future work by training deep learning models to analyze social media users' sentiment, or applying reinforcement learning to improve decision-making processes.