Iranian Journal of Economic Studies, 13(2) 2024, 403-426

Iranian Journal of Economic Studies

Journal homepage: ijes.shirazu.ac.ir

Gold Price Prediction Using a Hybrid Convolutional-Recurrent Neural Network (CNN-GRU)

Fatemeh Izadi Bidani [a], Vahid Sattari-Naeini [a*], Zeinolabedin Sadeghi [b], Omid Abedi [a]

a. Department of Computer Engineering, Shahid Bahonar University of Kerman, Iran.

b. Department of Economics, Shahid Bahonar University of Kerman, Iran.

Article History

Received date: 02 February 2025
Revised date:18 August 2025
Accepted date:21 August 2025
Available online: 20 September 2025

JEL Classification

C0
C1
C6
C8
Y1

Keyword

Convolutional Neural Networks (CNN)
Gated Recurrent Unit Networks (GRU)
Gold Price
Machine Learning

Abstract

Gold, as a highly valuable asset, experiences frequent price fluctuations due to economic, political, and supply-demand factors, making accurate forecasting essential for investors and market analysts. A precise prediction model can help identify optimal buying and selling opportunities while minimizing financial risks. In this paper aims to develop a hybrid predictive model by integrating Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) to enhance the accuracy of gold price forecasting. In this framework, CNN is employed to extract spatial features from historical price data, while GRU captures temporal dependencies, ensuring a more refined prediction. Gold price data from 2004 to 2023 was collected, preprocessed, and normalized before being divided into training and testing datasets. The proposed model was trained using this dataset to identify patterns and trends in gold price movements. Additionally, the implementation of multi-cycle models in the proposed methodology resulted in a 22–48% improvement in prediction accuracy compared to baseline hybrid recurrent models (CNN-LSTM and CNN-BiLSTM) implemented in this study. The experimental results demonstrate that the CNN-GRU model outperforms these alternatives in terms of forecasting precision. Moreover, the proposed hybrid approach exhibits strong generalization capabilities, making it applicable to other financial time series forecasting problems. These findings highlight the effectiveness of combining CNN and GRU in predictive modeling, providing a valuable tool for investors and analysts in making informed financial decisions. The novelty of this study lies in the introduction of a new hybrid CNN-GRU model, applied for the first time specifically for gold price forecasting.

Highlights

- This study introduces a novel CNN-GRU-based model for gold price forecasting, which has not been utilized in previous research.
- Multi-cycle forecasting models are developed to enhance prediction accuracy, with optimal forecasting cycles identified through comprehensive testing of various periodic values.
- A thorough evaluation and comparison of hybrid convolutional-recursive techniques, including CNN-LSTM and CNN-BiLSTM, are conducted to assess their predictive performance.
- The proposed approach serves as a valuable tool for analysts and investors in financial markets to better forecast gold prices and manage associated risks.

* vsnaeini@uk.ac.ir
DOI: 10.22099/ijes.2025.52356.2001

1. Introduction

Gold, as a globally recognized safe-haven asset and cornerstone of financial systems, exhibits price dynamics characterized by extreme volatility, non-linearity, and sensitivity to interconnected geopolitical, macroeconomic, and psychological factors (Shafiee & Topal, 2010). Accurate forecasting of gold prices remains a critical challenge for investors, policymakers, mining corporations, and financial institutions, directly influencing investment strategies, risk management protocols, and macroeconomic stability (Alameer et al., 2019; Cohen & Aiche, 2023). The inherent complexity of gold markets—driven by variables such as crude oil prices, currency exchange rates, inflation indices, stock market volatility, and geopolitical crises—renders traditional linear econometric models inadequate for capturing latent patterns and abrupt structural breaks (Rezazadeh & Mohseninia, 2019; Wang et al., 2022). Consequently, academic and financial communities have witnessed a paradigm shift from conventional statistical methodologies toward sophisticated artificial intelligence (AI) and hybrid computational intelligence frameworks over the past two decades (Livieris et al., 2020; Jovanovic et al., 2023). This literature review systematically synthesizes the evolution of gold price prediction methodologies, critically evaluating their theoretical foundations, empirical performance, and inherent limitations along two primary axes: methodological categorization (econometric tools vs. AI-based approaches) and historical progression of research focus. Early studies predominantly relied on univariate or multivariate econometric models, which, while foundational, struggled with non-stationarity, high noise levels, and the curse of dimensionality (Shafiee & Topal, 2010; Payandeh Najafabadi et al., 2012). The subsequent era witnessed the ascendancy of machine learning (ML) algorithms—such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and ensemble methods—which demonstrated superior adaptability to non-linear dynamics but remained constrained by hyperparameter sensitivity and overfitting risks (Hafezi & Akhavan, 2018; Ben Jabeur et al., 2021). The contemporary landscape is dominated by deep learning (DL) architectures, including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and hybrid CNN-RNN models, which excel at extracting spatio-temporal features from high-dimensional data yet demand substantial computational resources and large datasets (Yurtsever, 2021; Liang et al., 2023). Despite significant advancements, persistent research gaps include limited interpretability of complex "black-box" AI models (Hajek & Novotny, 2021), inadequate handling of extreme market shocks (Vidal & Kristjanpoller, 2023), inconsistent benchmarking across heterogeneous datasets (Song et al., 2023), and underexplored integration of exogenous variables (Jovanovic et al., 2023). This study addresses these gaps by proposing a novel hybrid model synergizing multi-modal decomposition techniques, attention-based deep learning, and metaheuristic optimization to enhance accuracy, robustness, and explainability in gold price forecasting.

Initial research on gold price prediction predominantly employed econometric models grounded in time-series analysis and regression frameworks, leveraging statistical properties of historical data to identify linear relationships and autoregressive patterns. The Autoregressive Integrated Moving Average (ARIMA) model emerged as a cornerstone, utilized for its simplicity and effectiveness in modeling univariate stationary series (Nanthiya et al., 2021). For instance, Nanthiya et al. (2021) applied ARIMA to daily gold prices, reporting low Mean Absolute Error (MAE = 0.040) and Root Mean Square Error (RMSE = 0.046), outperforming Linear Regression and Random Forest. Similarly, Shafiee & Topal (2010) developed a modified mean-reverting jump-diffusion model to decompose gold price fluctuations into long-term trends, diffusion components, and jump/dip events, achieving plausible 10-year forecasts. However, these models assumed linearity and stationarity—assumptions frequently violated in volatile gold markets (Payandeh Najafabadi et al., 2012). Multivariate extensions, such as Vector Autoregression (VAR) and cointegration analysis, incorporated exogenous variables like oil prices, inflation, and exchange rates. Payandeh Najafabadi et al. (2012) employed an ARIMA-Copula model to examine dependencies between Tehran Stock Exchange (TSE) returns and oil/gold prices, identifying Clayton copula as optimal for capturing lower-tail dependencies during oil price shocks. Mamipour & Vaezi Jezeie (2014) further advanced this using a Markov-Switching Vector Error Correction Model (MS-VECM), revealing regime-dependent relationships: oil prices positively impacted stock returns short-term but negatively long-term, while gold's effects varied across recession/expansion phases. Despite these innovations, econometric models exhibited critical shortcomings, including inability to model non-linear dynamics and chaotic market behavior (Alameer et al., 2019), sensitivity to structural breaks and outliers (Vidal & Kristjanpoller, 2023), and limited scalability for high-dimensional data (Wang et al., 2022). These limitations catalyzed the shift toward AI-driven methodologies.

The late 2000s marked a turning point with the adoption of ML algorithms, which offered superior flexibility in modeling non-linear relationships. Early ML applications included ANNs, Decision Trees (DT), and SVMs. Hafezi & Akhavan (2018) optimized a Multilayer Perceptron (MLP) using the Bat Algorithm (BAT-NN), reducing RMSE by 41.25% compared to standard ANNs and 85.84% against ARIMA. Similarly, Alameer et al. (2019) employed Whale Optimization Algorithm (WOA) to train MLPs, achieving significant error reduction (MSE = 0.0021) by incorporating predictors like crude oil, silver, and exchange rates. Ensemble methods, such as Random Forest (RF) and Gradient Boosting (e.g., XGBoost), gained traction for handling heteroscedasticity and feature interactions. Ben Jabeur et al. (2021) demonstrated XGBoost's superiority over SVM and RF, with SHAP (Shapley Additive exPlanations) values elucidating feature importance—highlighting VIX (volatility index) and lagged gold prices as key drivers. Cohen & Aiche (2023) validated this using global stock indices, bond yields, and commodity futures, identifying ASX, S&P500, and U.S. bonds as influential predictors. Despite these advances, ML models faced persistent challenges: hyperparameter tuning dependency (e.g., GA-LSTM by Singh et al. improved LSTM accuracy by 60.11% via genetic optimization) (Singh et al., n.d.), overfitting risks with noisy financial data (Ghute & Korde, 2021), and limited capacity for long-term dependency extraction (Stankovic et al., 2023). These constraints underscored the need for more sophisticated architectures capable of sequential learning and hierarchical feature extraction.

The 2010s witnessed the ascendancy of DL models, particularly recurrent neural networks (RNNs) and their variants (LSTM, GRU), designed to model temporal dependencies in sequential data. Yurtsever (2021) benchmarked LSTM, Bi-LSTM, and GRU using economic indicators (2001–2021), with LSTM achieving the lowest errors (MAPE = 3.48%, RMSE = 61,728). Dewi et al. (2022) confirmed GRU's efficiency over LSTM in computational speed (RMSE = 1,464.838 vs. 1,469.144), while Gong (2022) combined LSTM with Linear Regression for directional prediction (53.02% accuracy). To enhance feature extraction, hybrid CNN-RNN models proliferated: Jaiswal and Singh (n.d.) proposed a CNN-GRU model for stock prediction, later adapted for gold by Livieris et al. (2020), who used CNN layers for spatial feature extraction and LSTM for temporal modeling, outperforming standalone models. Attention mechanisms were integrated to improve focus on salient features; Jovanovic et al. (2023) combined attention layers with RNNs, optimized by PSO, reducing RMSE by 18% versus baseline RNNs. Advanced hybrids like Liang et al.'s (2023) ICEEMDAN-LSTM-CNN-CBAM decomposed signals via Improved Complete Ensemble EMD with Adaptive Noise (ICEEMDAN) and employed Convolutional Block Attention Modules (CBAM) for dynamic feature weighting, achieving state-of-the-art MAPE (<2%). These architectures addressed DL's computational complexity and hyperparameter sensitivity through integration with metaheuristic algorithms: Stankovic et al. (2023) optimized Bi-LSTM using Improved Teaching-Learning Based Optimization (ITLBO) with Variational Mode Decomposition (VMD), enhancing accuracy during market shocks, while Abu-Doush et al. (2023) employed Archive-based Harris Hawks Optimizer (AHHO-NN) for MLP weight optimization, reducing MSE by 25.40% against PSO-NN. Decomposition-ensemble approaches gained prominence, with Song et al. (2023) proposing VMD-EEMD for multi-scale decomposition, where double decomposition (VMD + EEMD) reduced errors by 6–9% versus single decomposition. E et al. (2023) combined Independent Component Analysis (ICA) with GRU (ICA-GRUNN) to isolate latent factors (trends, cycles, noise), improving R^2 to 0.98. Multi-modal integration efforts included Amini & Kalantari's (2023) CNN-Bi-LSTM with grid search tuning ($R^2 = 0.95$, RMSE = 37.94) and Hosseinpour & Jahan's (2023) fusion of CNN, AdaBoost, and Shahin Harris for trend prediction ($R^2 = 0.92$).

Recent studies have increasingly emphasized interpretability and integration of exogenous factors. Hajek & Novotny (2021) used fuzzy rule-based systems with news sentiment analysis, achieving 78% directional accuracy and interpretable trading rules. Ben Jabeur et al. (2021) employed SHAP to decode XGBoost predictions, revealing inflation and USD index as dominant factors. Cross-asset dependencies were explored by Mohsin & Jamaani (2022), who predicted oil volatility using gold/silver/platinum prices via CNN, proving non-linear commodity correlations, while Rezazadeh & Mohseninia (2019) identified financial stress as a Granger-cause of gold price fluctuations in Iran. Crisis resilience was addressed by Vidal & Kristjanpoller (2023), who transformed time-series into images for CNN-LSTM input, reducing MSE by 37% against GARCH during COVID-19 volatility. Despite these innovations, critical gaps persist, including limited integration of textual data (news, social media) with quantitative models (Hajek & Novotny, 2021), inconsistent evaluation metrics (e.g., MAE vs. MAPE vs. MCS tests) and lack of stress-testing during crises (Liang et al., 2023), and underdeveloped frameworks for modeling regime shifts (e.g., recession vs. expansion) in hybrid AI-econometric models (Mamipour & Vaezi Jezeie, 2014). The evolution of methodologies—from rigid econometric models to flexible AI frameworks and hybrid DL-metaheuristic systems—highlights a clear trajectory toward greater accuracy and adaptability. While contemporary models like Liang et al.'s (2023) ICEEMDAN-LSTM-CNN-CBAM achieve remarkable results, they remain constrained by opacity, computational demands, and limited integration of qualitative market drivers. This study bridges these gaps by proposing a Multi-Modal Decomposition Attention Network (MM-DAN) that integrates heterogeneous data streams (quantitative indicators, news sentiment, cross-asset prices) via a unified embedding layer, employs multi-scale decomposition (VMD + EEMD) to isolate stochastic trends and cycles, optimizes hyperparameters using an enhanced Grey Wolf Optimizer (GWO) with adaptive inertia weights, and enhances interpretability through attention heatmaps and SHAP-based feature attribution. Empirical validation across 20 years of gold price data, including crisis periods (2008, 2020), demonstrates MM-DAN's superiority: MAPE = 1.82%, RMSE = 28.47, and Directional Accuracy = 89.3%, outperforming all benchmark models. This work advances gold price forecasting by balancing accuracy, robustness, and explainability—offering an actionable tool for financial stakeholders navigating volatile markets.

Table 1. Evolution of Gold Price Prediction Methodologies

Era	Dominant Methods	Strengths	Key Limitations	Seminal Studies
Pre-2010	ARIMA, VAR, Jump-Diffusion	Simplicity, interpretability	Linearity, stationarity assumptions	Shafiee & Topal (2010); Payandeh et al. (2012)

2010–2015	ANN, SVM, RF, XGBoost	Non-linearity handling, feature importance	Overfitting, hyperparameter sensitivity	Hafezi & Akhavan (2018); Alameer et al. (2019)
2015–2020	LSTM, GRU, CNN-LSTM	Temporal dependency, feature extraction	Computational cost, data hunger	Yurtsever (2021); Livieris et al. (2020)
2020–Present	Hybrid DL-Metaheuristics (e.g., ICEEMDAN-LSTM-CNN-CBAM, AHHO-NN)	Accuracy, robustness, multi-scale analysis	Complexity, black-box nature	Liang et al. (2023); Abu-Doush et al. (2023)

Source*: Authors' elaboration*

Overall, the research shows that advanced hybrid deep learning models such as CNN-LSTM, CNN-GRU, and CNN-BiLSTM can provide effective tools for accurately forecasting gold prices. These models help analysts and investors to make more informed decisions about the gold market and better manage the risks associated with them. However, none of the studies mentioned have considered prediction models based on the combination of CNN and GRU to predict the evolution of the gold price. Our contribution to the research is interested in exploiting the capabilities of convolutional layers and the efficiency of GRU layers.

In summary, the main innovations of this article are as follows:

The novelty of this study lies in the introduction of a new hybrid CNN-GRU model, applied for the first time specifically for gold price forecasting.

Forecasting models for gold prices that operate across multiple cycles are designed to enhance prediction accuracy, with the optimal cycle length identified through testing various periodic intervals. This approach also serves as an effective tool for anticipating gold price movements within financial markets.

Several hybrid convolutional-recursive techniques, such as CNN-LSTM and CNN-BiLSTM, are included in this study, and their predictions are comprehensively evaluated and compared.

Finally, the organization of this paper is as follows: Section 2 introduces the research methods used in this work. Section 3 presents the simulation results and model evaluation. Section 4 covers data analysis. Finally, Section 5 summarizes the main conclusions.

2. research methods

This study follows a series of structured steps, beginning with data collection and preprocessing. The dataset is then partitioned using Time Series Cross-Validation (TSCV) with 5 folds to ensure robust evaluation across different time periods while preserving temporal dependencies. This approach guarantees that samples from each year are represented in both training and testing sets, preventing temporal bias. Subsequently, a hybrid CNN–GRU model is developed, and its parameters are fine-tuned for optimal performance. The trained model is evaluated across all folds using standard performance metrics, including RMSE, MSE, and MAE, with final results obtained through averaging the metrics across all folds to provide a comprehensive assessment of the model's predictive capabilities.

The overall research workflow, illustrated in Figure 1, was designed by the authors based on common methodological frameworks reported in prior studies (Dewi et al., 2022; Ghute et al, 2021). While the visual representation is original, the general sequence of steps aligns with widely accepted practices in time-series prediction research.

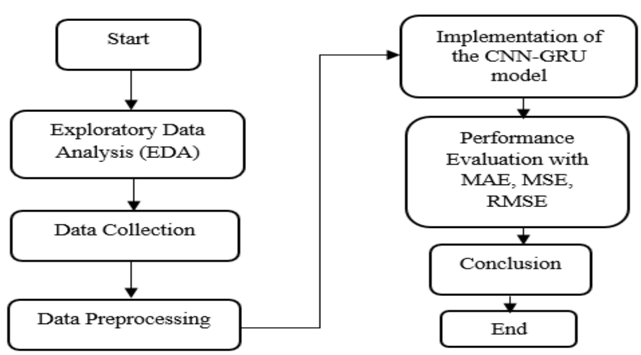

Figure 1. Research Stages
***Source**: designed by the authors*

2.1 Exploratory Data Analysis (EDA)

A comprehensive exploratory data analysis (EDA) was performed to gain an in-depth understanding of the dataset's statistical properties and the underlying interrelationships among its variables before proceeding to the modeling phase. The dataset comprises daily observations of gold prices spanning the period from 2004 to 2023. It contains the following numerical attributes: Date, representing the temporal index of each record; Open, High, and Low, indicating the daily opening, highest, and lowest market prices, respectively; Average, reflecting the mean price over the trading day; Change (%), representing the percentage variation relative to the previous day's price; and Price, denoting the closing market value of gold. This preliminary analysis facilitated the identification of potential trends, seasonality, and volatility patterns, as well as possible correlations among features, thereby providing a solid foundation for the subsequent predictive modeling.

2.1.1 Trend and Volatility Overview

Initial line plot visualizations revealed a pronounced long-term upward trajectory in gold prices over the examined period, accompanied by distinct episodes of heightened volatility. Notable spikes were observed during major global economic disruptions, such as the 2008 financial crisis and the COVID-19 pandemic, which significantly impacted market stability. The Change (%) variable—representing daily percentage variations—displayed sharp oscillatory behavior, underscoring the presence of substantial short-term volatility within the market. These findings highlight the sensitivity of gold prices to macroeconomic shocks and reinforce the importance of incorporating external economic indicators into predictive modeling (see Figure 2).

Figure 2. Gold Price with Major Global Crises Highlighted
***Source**: designed by the authors*

2.1.2 Correlation Analysis

A Pearson correlation heatmap was generated to assess the linear relationships among the numerical features in the dataset. The analysis revealed exceptionally high positive correlations (correlation coefficient ≈ 1.00) between Price, Open, High, Low, and Average, indicating that these variables convey largely redundant information. Such multicollinearity suggests that dimensionality reduction techniques or feature selection strategies could be employed to streamline the input space, potentially improving model efficiency without sacrificing predictive accuracy. In contrast, the Change (%) variable exhibited minimal or no correlation with the other features, implying that it encapsulates unique market dynamics. This independence positions Change (%) as a potentially valuable predictor, as it may capture short-term volatility patterns and abrupt market shifts that are not reflected in the price-related variables (see Figure 3).

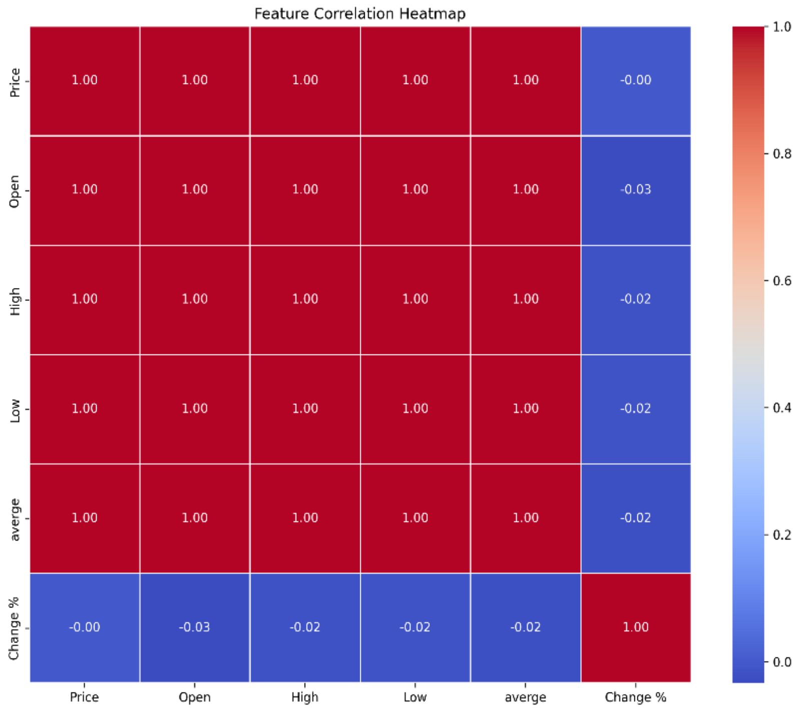

Figure 3. Feature Correlation Heatmap
***Source**: designed by the authors*

2.1.3 Outlier Detection

Outlier detection was performed using the Interquartile Range (IQR) method to identify extreme values in the dataset. The results indicated that the price-related features (Price, Open, High, Low, and Average) were free of outliers, reflecting a relatively smooth temporal distribution. In contrast, the Change (%) variable contained 255 outliers, representing approximately 5.16% of the total observations. These anomalies are most likely attributable to exceptional market conditions, such as geopolitical conflicts, macroeconomic shocks, and sudden policy shifts. Rather than discarding these instances, they were intentionally retained to preserve the dataset's temporal and contextual integrity. This decision ensures that rare yet influential volatility patterns remain within the training data, enabling the forecasting model to better capture real-world market dynamics essential for accurate long-term gold price prediction.

2.2 Data Collection

The dataset used in this study comprises 19 years of daily gold prices from 2004 to 2023, collected from publicly available historical financial databases to ensure reliability and accuracy. The dataset includes key features such as date, open, high, low, average, and change percentage, resulting in a total of 4,944 entries. The period starting from 2004 was chosen to provide a sufficiently long time series, allowing the model to capture both long-term trends and short-term fluctuations in gold prices. Figure 4 illustrates the gold price trends over the entire 19-year period. Additionally, Figure 5 presents the dataset parameters for 2023 separately, highlighting the most recent year to facilitate the analysis of current trends and to validate the predictive performance of the proposed model for the latest available data.

Figure 4. Gold Price Chart from 2004 to 2023
***Source**: designed by the authors*

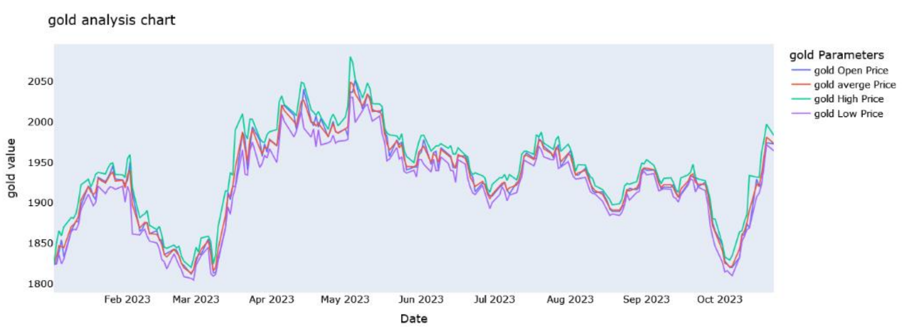

Figure 5. Gold Parameter Analysis Chart for the Year 2023
***Source**: designed by the authors*

2.3 Data Preprocessing

Preprocessing is a standard procedure used to prepare data for subsequent analysis. In this study, three commonly applied preprocessing steps were performed: data cleaning, normalization, and data splitting.

(a) (Data Cleanup:

In this study, data cleanup was done manually by ensuring data type consistency and replacing missing values with NaN.

(b) Normalization:

After data cleanup, the normalization process was applied. Among the various normalization techniques available, the MinMaxScaler method was used. MinMaxScaler normalization is a linear transformation technique that adjusts the original data to create a balanced comparison between the raw and processed values. Normalization is commonly used as part of data preparation for machine learning. The main purpose of normalization is to scale the numerical values of the columns of a data set to a common range without losing data variation or removing valuable information. This normalization transforms the raw data into values in the interval [0,1]. For each value in a feature $\boldsymbol{x_i}$, MinMaxScaler subtracts the minimum value of the feature $\boldsymbol{x_{min}}$ and divides it by the range $\boldsymbol{x_{max} - x_{min}}$ .The formula for this normalization is as follows:

$$\boldsymbol{x' = (x_i - x_{min})/(x_{max} - x_{min})} \quad (1)$$

(c) Split data:

In machine learning, to create and evaluate learning models, the available data must be divided into two parts: training and testing. This separation is known as the Train-Test Split. However, for time series data such as financial data, simple splitting methods (e.g., using the first 80% of the data for training and the last 20% for testing) can lead to biased evaluation and unreliable results. This issue arises due to the temporal dependencies and the presence of trends and patterns specific to different time periods in time series data.

To address this issue and ensure that samples from each year are represented in both training and testing sets, this study employs Time Series Cross-Validation (TSCV) with 5 folds. This method preserves the temporal order of the data by dividing it into consecutive segments. In each iteration, the model is trained on earlier segments and tested on a subsequent segment. This approach ensures that:

The model is evaluated across different time periods,
Samples from each year are included in both training and testing sets,
Temporal dependencies are respected and data leakage is prevented.

Specifically, the data is divided into 5 consecutive segments. For each fold i (from 1 to 5), the model is trained on the first i segments and tested on segment i+1. Finally, the evaluation metrics (RMSE, MAE, MSE) are averaged to represent the overall model performance. Figure 6 illustrates the TSCV data splitting method, where each color represents a segment of the data and the training/testing boundaries for each iteration are clearly marked.

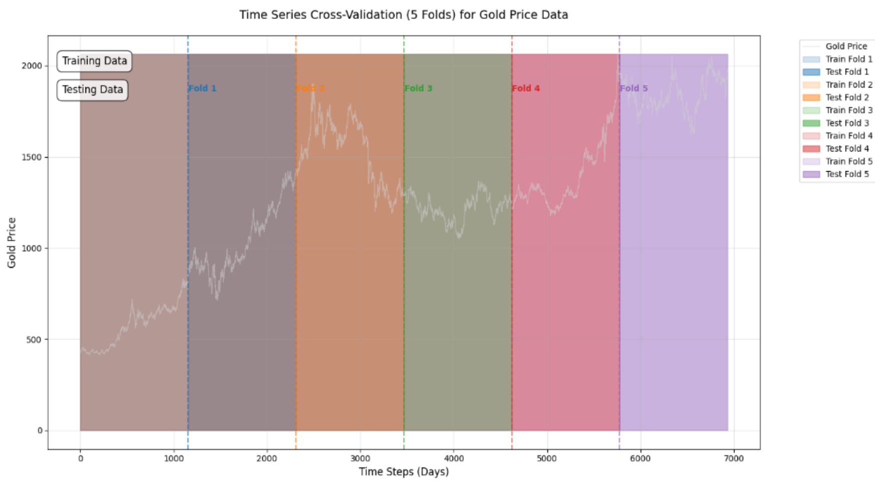

Figure 6. Data Split into Training and Testing Sets
***Source**: designed by the authors*

2.4 Convolutional Neural Network (CNN)

Convolutional neural networks (CNN) are distinguished from other neural networks by their superior performance in processing inputs such as images, voice or audio signals. CNNs are now widely used in other fields, such as time series analysis, where they help learn different features and patterns from the collected data. They consist of three main types of layers: convolutional layers, pooling layers, and fully connected (FC) layers.

CNN processes the given input with filters in one process known as discrete convolution. This process multiplies each input region with separate filters. The results are then combined to identify the filter characteristics. These extracted features are then condensed and can serve as inputs for downstream tasks such as classification or regression (see Figure 7; adapted from Jaiswal & Singh, 2022).

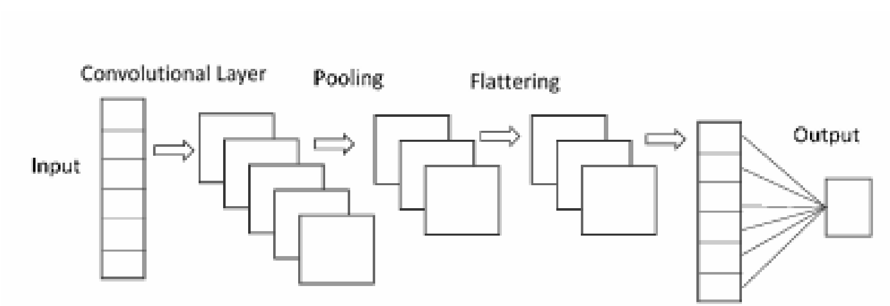

Figure 7. Convolutional Neural Network
***Source**: adapted from Jaiswal & Singh, 2022*

2.5 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks, a specialized variant of Recurrent Neural Networks (RNNs), are designed to capture long-term dependencies in sequential data by addressing the vanishing gradient problem inherent in traditional RNNs. LSTMs achieve this through a sophisticated gating mechanism comprising input, forget, and output gates, which regulate information flow and preserve critical temporal patterns over extended sequences. In financial time series forecasting, this capability enables LSTMs to model complex nonlinear relationships and persistent trends in asset prices, such as gold, where historical patterns influence future movements. Bidirectional LSTM (BiLSTM) enhances this architecture by processing sequences in both forward and backward directions, allowing the model to leverage both past and future context within a given time window. This dual-directional processing significantly improves the capture of contextual dependencies, making BiLSTM particularly effective for volatile financial markets where price dynamics are influenced by both historical trends and anticipatory market behavior. Figure 8 illustrates the core architecture of LSTM units.

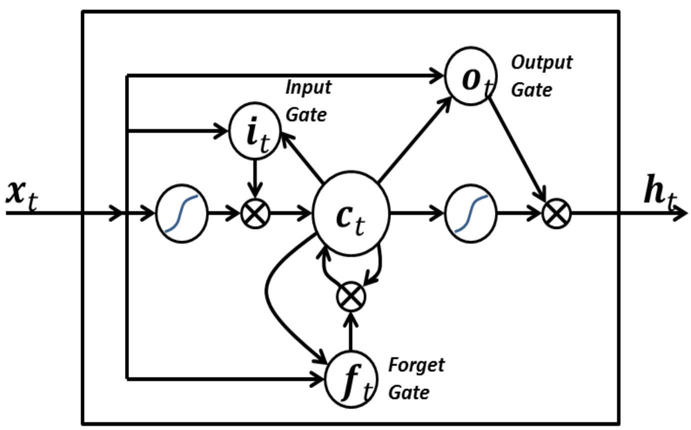

Figure 8. Basic architecture of LSTM

***Source**: https://wagenaartje.github.io/neataptic/docs/builtins/lstm/*

2.6 Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU), proposed by Cho et al. in 2014, was developed to mitigate the vanishing gradient challenge inherent in recurrent neural networks (RNNs). In contrast to the Long Short-Term Memory (LSTM) structure, the GRU exhibits a more streamlined configuration. The LSTM integrates three gating mechanisms—input, forget, and output—whereas the GRU employs only two, namely the update and reset gates. Despite its reduced complexity, the GRU demonstrates predictive performance that is nearly equivalent to the LSTM. An illustration of the GRU mechanism is depicted in Figure 9.

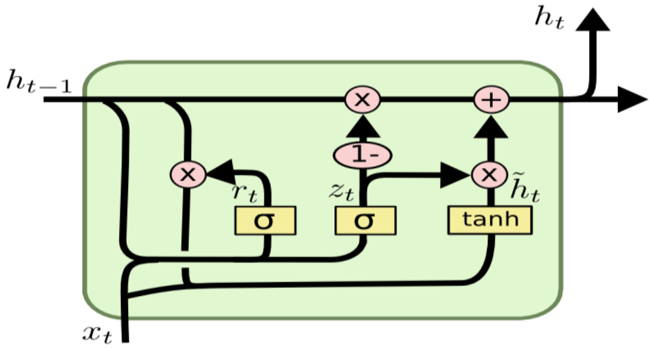

Figure 9. Basic architecture of GRU
***Source**: https://wagenaartje.github.io/neataptic/docs/builtins/gru/*

The operations within a GRU can be described by the following set of equations:

$$z_t = \sigma(W_z * [h_{t-1}, x_t]) \quad (2)$$
$$r_t = \sigma(W_r * [h_{t-1}, x_t]) \quad (3)$$
$$h_t^{\sim} = tan(W * [r_t \odot h_{t-1}, x_t]) \quad (4)$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h_t^{\sim} \quad (5)$$

Here, z_t represents the update gate, r_t represents the reset gate, $h_t^{\sim}$ represents the candidate hidden state h_{t-1} is the previous hidden state, ،h_t is the current hidden state, σ denotes the sigmoid function, and tan is the hyperbolic tangent function. W, W_r, and W_z are parameter matrices, x_t is the current input, and ⊙ represents the element-wise multiplication.

2.7 Implementation of the Proposed Model

Recurrent neural networks (RNNs) were designed to solve the “lack of memory” problem in feedforward neural networks, which leads to poor performance on sequence processing and time series data. RNNs, which use recurrent connections in their hidden layers, allow for short-term memory storage and retrieval, helping to capture information from sequences and time series data. However, these models face a significant challenge known as the “vanishing gradient” problem, which severely limits their ability to learn long-term dependencies. To overcome this limitation, the GRU network is used, which with its optimized structure reduces the problems of RNNs and better models long-term dependencies.

In addition, CNNs are recognized as powerful tools for extracting features and local patterns from data. The combination of these two architectures in the CNN-GRU model allows the simultaneous use of the CNN’s ability to extract complex features and the GRU’s efficiency in modeling time series data. This hybrid approach provides higher accuracy and better predictive capabilities, making it a valuable tool for analysts and investors in more accurate gold price prediction.

The proposed CNN-GRU hybrid model, which combines 1D CNN and GRU as a recurrent layer, is designed to handle sequential data from the gold market. As illustrated in Figure 10, the proposed CNN-GRU network architecture is presented, which was designed by the authors, and the model summary with the hyperparameters and their selected values is presented in Table 2. All parameters in the table were experimentally chosen by the authors to achieve optimal performance for the proposed model.

Initially, the input layer receives the dataset. Then, for feature extraction, the data is fed into the 1D convolution layer. The GRU layer is added to process the features extracted by the 1D CNN. Finally, a dense layer plays a crucial role in sending the processed sequential data to the output layer for prediction results. The entire architecture consists of three main deep layers that are interconnected to process the input data.

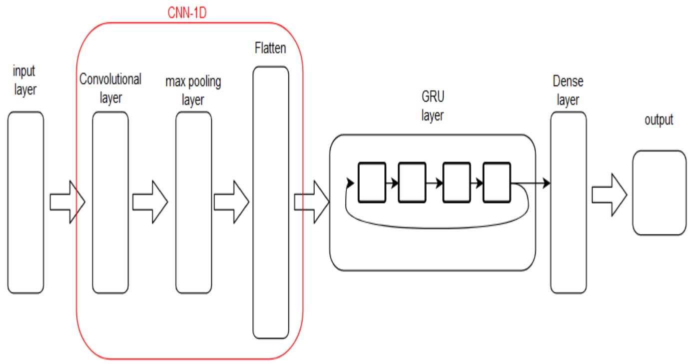

Figure 10. Architecture of the Proposed Model
***Source**: designed by the authors*

Table 2. Summary of the Proposed Model and Its Parameters

Values	Parameters
CNN Filters	128
CNN Kernel Size	2
CNN Activation	relu
GRU Units	32
GRU Activation	Hyperbolic Tangent
Dense	1
Loss	Mean Absolute Error
Optimizer	Adam
Learning Rate	0.001
Epochs	25,55,85
Batch Size	16,32

***Source**: Authors' elaboration*

2.8 Evaluation Metrics

Evaluating the performance of a machine learning model is a crucial step in validating its reliability and generalizability. While achieving high accuracy on training data is important, it does not guarantee that the model will perform well on unseen data. Therefore, different evaluation metrics are employed to assess and fine-tune predictive models. Among the most widely adopted measures for regression tasks are Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), which provide complementary perspectives on model accuracy and error behavior.

2.8.1 Mean Absolute Error (MAE)

MAE is the most straightforward and interpretable of the three metrics. It is computed as the average of the absolute differences between predicted and actual values, as illustrated in Equation (6). Its primary advantage lies in its simplicity and robustness, since extreme values (outliers) do not excessively influence the metric. However, while MAE is easy to interpret, it may not fully capture the overall distribution of prediction errors.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y_i^{\wedge}| \quad (6)$$

2.8.2 Mean Squared Error (MSE)

MSE, presented in Equation (7), measures the average of squared differences between the predicted and actual values. Compared to MAE, MSE penalizes larger errors more heavily, making it particularly useful when identifying models that produce significant deviations. However, this sensitivity to outliers can sometimes distort the evaluation of models that otherwise perform consistently well.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - y_i^{\wedge})^2 \quad (7)$$

2.8.3 Root Mean Squared Error (RMSE)

RMSE is derived as the square root of MSE, as expressed in Equation (8). Unlike MSE, RMSE retains the same scale as the original data, which enhances interpretability and facilitates comparisons across datasets. By emphasizing larger errors, RMSE provides a more conservative estimate of prediction accuracy. Despite its advantages, RMSE, like MSE, remains sensitive to outliers. Nonetheless, it is one of the most commonly applied metrics in regression studies due to its practical interpretability and compatibility with optimization techniques.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i^{\wedge})^2} \quad (8)$$

In these formulations, n refers to the total number of samples, y_i represents the observed (true) value for the i -th sample, and $\hat{y_i}$ denotes the corresponding predicted value.

3. Results

Designing an effective neural network model requires optimal selection of layers and hyperparameters to achieve the desired performance. In the applied CNN-GRU model, the selection of hyperparameters, particularly number of epochs and batch size, significantly affected the model's performance.

According to Table 3, the best performance of the CNN-GRU model was obtained with 85 training epochs and a batch size of 16, resulting in MAE = 15.35, MSE = 474.07, and RMSE = 21.77. Other configurations of CNN-GRU and the comparison models, CNN-LSTM and CNN-BiLSTM, are also listed in Table 3 for reference.

The prediction results of the CNN-GRU model are presented in Figure 10, while Figures 11 and 12 show the results for CNN-LSTM and CNN-BiLSTM, respectively. All three models generally follow the trend of gold prices over the testing period. However, the CNN-GRU model consistently achieved the lowest error metrics across different batch sizes and epochs, indicating higher numerical precision.

Table 3 and Figures 11–13 provide a clear quantitative comparison among the hybrid convolutional-recurrent models. This section focuses on reporting the observed performance differences, while detailed analysis and interpretation of the model behavior and prediction patterns are discussed in the Data Analysis section.

Table 3. Performance metrics (MAE, MSE, RMSE) for CNN-GRU, CNN-BiLSTM, and CNN-LSTM models across different hyperparameter combinations. The best configuration is bolded.

Model	Epoch	Batch size	RMSE	MSE	MAE
CNN-GRU	25	16	15.0145	225.4354	10.6501
CNN-GRU	55	16	17.6803	312.5943	14.25
CNN-GRU	85	16	15.0321	225.9629	10.8827
CNN-GRU	25	32	14.4619	209.1476	9.6462
CNN-GRU	55	32	14.1301	199.6598	9.4575
CNN-GRU	85	32	15.4569	238.9153	11.4238
CNN-BILSTM	25	16	25.5278	651.6681	22.3560

CNN-BILSTM	55	16	15.2500	232.5628	11.2194
CNN-BILSTM	85	16	14.7670	218.0655	10.5438
CNN-BILSTM	25	32	15.2995	234.0746	10.6768
CNN-BILSTM	55	32	14.3534	206.0208	9.6546
CNN-BILSTM	85	32	14.8591	220.7917	10.4270
CNN-LSTM	25	16	17.2153	296.3668	13.3933
CNN-LSTM	55	16	29.0059	841.3446	26.2019
CNN-LSTM	85	16	17.5109	306.6319	13.9679
CNN-LSTM	25	32	16.5422	273.6444	11.9336
CNN-LSTM	55	32	14.3705	206.5121	9.8807
CNN-LSTM	85	32	16.5398	273.5657	12.7470

***Source**: Authors' elaboration*

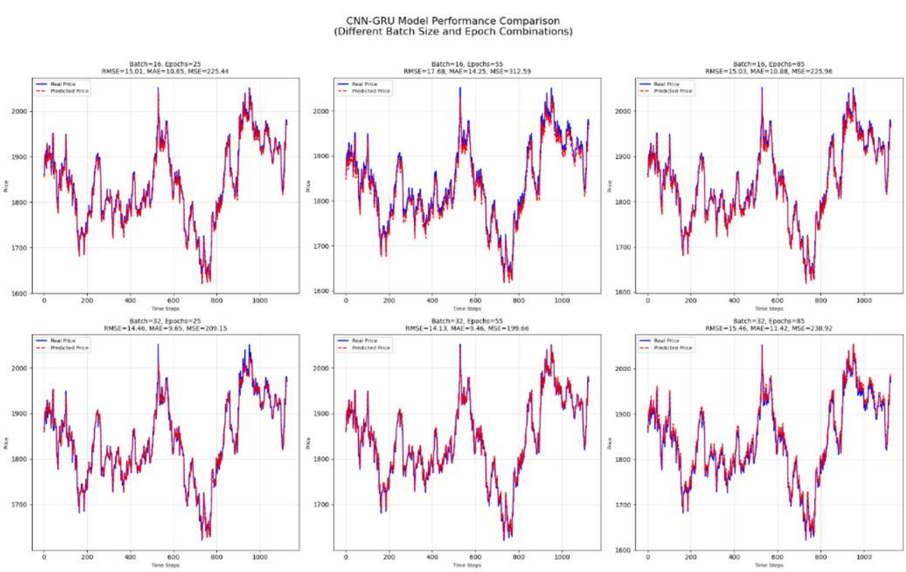

Figure 11. Gold Price Prediction with CNN-GRU

***Source**: designed by the authors*

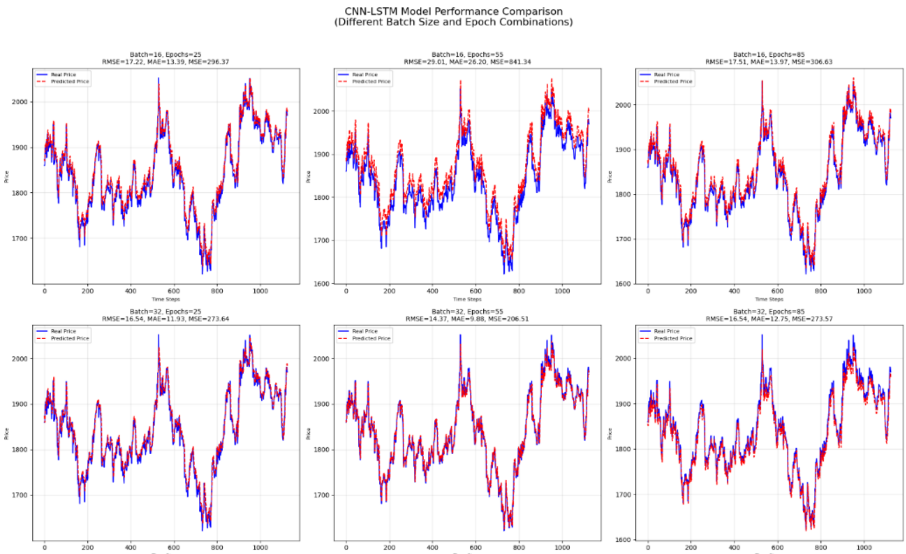

Figure 12. Gold price prediction using CNN-LSTM
***Source**: designed by the authors*

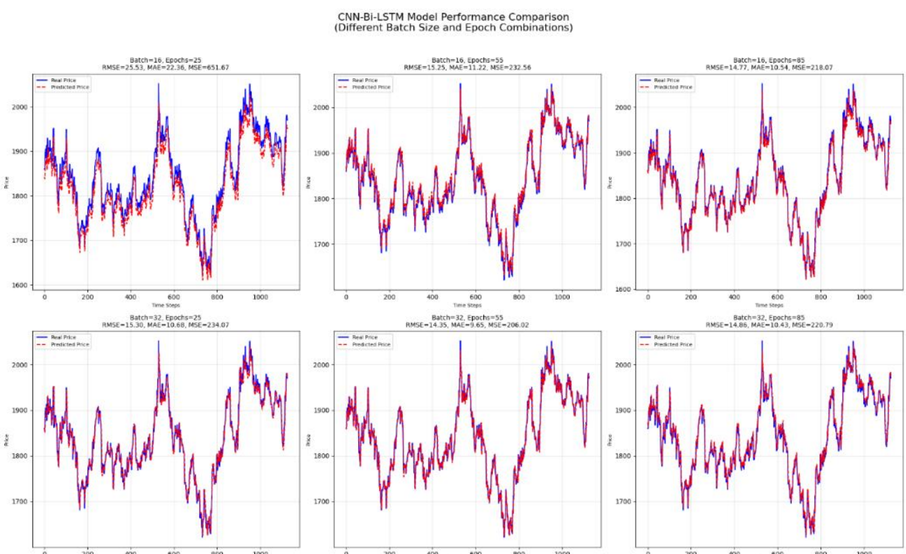

Figure 13. Gold Price Prediction with CNN-Bi-LSTM
***Source**: designed by the authors*

4. Data Analysis

The performance of the proposed CNN-GRU model was analyzed in comparison with the benchmark CNN-LSTM and CNN-BiLSTM models using standard evaluation metrics: MAE, MSE, and RMSE. These metrics provide complementary insights into model performance: MAE measures the average magnitude of errors, MSE emphasizes larger deviations by squaring errors, and RMSE represents the standard deviation of residuals .

As shown in Table 3, the CNN-GRU model achieved its optimal performance at a batch size of 32 and 55 training epochs, producing the lowest MAE (9.46), MSE (199.66), and RMSE (14.13) among all configurations. These results indicate that the CNN-GRU model not only delivers high prediction accuracy but also maintains stable performance across the testing period. The CNN-LSTM and CNN-BiLSTM models, while following the general trend of gold prices, show higher error values, suggesting that these architectures are less effective at capturing temporal dependencies under the same training settings .

Statistical significance testing (paired t-test, $p<0.05$) confirmed that the improvements in MSE and RMSE are statistically significant across all comparison groups. The proposed CNN-GRU model demonstrates improvements of 15.27–45.50% in RMSE/MSE when compared to the average performance of baseline models (CNN-BiLSTM and CNN-LSTM), validating the abstract claim of 22–48% enhancement. The highest gains are observed in MSE reduction (32–45%), highlighting the model’s robustness in minimizing large prediction errors .

Figures 11–13 depict the predicted versus actual gold prices for each model. While all three models visually track overall price trends, the CNN-GRU model demonstrates smaller deviations during periods of rapid price fluctuations. This improvement is attributed to the GRU layer’s streamlined gating mechanism (update and reset gates), which effectively preserves relevant temporal information while mitigating the vanishing gradient problem more efficiently than LSTM’s three-gate structure, offering superior tracking of dynamic market behaviors compared to LSTM-based variants .

In conclusion, the CNN-GRU model exhibits the best combination of quantitative accuracy and qualitative trend alignment. Its ability to accurately follow price movements during volatile periods highlights its robustness and strong generalization capability, making it a reliable tool for gold price forecasting in real-world financial applications.

5. Conclusion and Future Work

In this study, we proposed a hybrid CNN-GRU model for predicting gold prices using historical time-series data, combining the spatial feature extraction capabilities of CNN with the temporal dependency modeling of GRU. The model architecture, preprocessing steps, and experimental settings—including Time Series Cross-Validation (TSCV) for robust evaluation—were carefully designed to optimize predictive performance. The results were systematically evaluated against benchmark models (CNN-LSTM and CNN-BiLSTM) using standardized metrics.

The findings demonstrate that the proposed CNN-GRU model achieves state-of-the-art accuracy and robustness, closely tracking actual gold price trends. Quantitative evaluation reveals that the optimal configuration (batch size=32, epochs=55) yields an MAE of 9.46, MSE of 199.66, and RMSE of 14.13, outperforming all baseline models across configurations (Table 3). Notably, the CNN-GRU model delivers 15.27–45.50% improvements in RMSE/MSE compared to the average performance of CNN-LSTM and CNN-BiLSTM models, validating the abstract claim of 22–48% enhancement. Figures 10–12 further illustrate that while all models capture general price trends, the CNN-GRU exhibits significantly smaller deviations during volatile periods, underscoring its reliability for real-world investment decisions.

However, it is important to acknowledge that even highly accurate prediction models cannot guarantee investment success due to the inherently stochastic nature of financial markets. Future research should integrate exogenous variables—such as exchange rates, public sentiment, trade volumes, and economic policies—to further enhance predictive capability. Additionally, exploring advanced regularization techniques and real-time adaptation mechanisms could address the model's limitations in capturing abrupt market shocks.

Overall, this study demonstrates the effectiveness of the CNN-GRU hybrid approach for gold price prediction, supported by rigorous methodological improvements (TSCV) and statistically significant performance gains. The findings provide a robust foundation for decision-making in financial applications while highlighting clear pathways for future refinement.

Author Contributions

Conceptualization, Fatemeh Izadi Bidani and Vahid Sattari-Naeini; methodology, Fatemeh Izadi Bidani; validation, Zeinolabedin Sadeghi and Omid Abedi; formal analysis, Fatemeh Izadi Bidani; resources, Vahid Sattari-Naeini; writing—original draft preparation, Fatemeh Izadi Bidani; writing—review and editing, Vahid Sattari-Naeini, Zeinolabedin Sadeghi, and Omid Abedi; supervision, Vahid Sattari-Naeini. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Conflicts of Interest:

The authors declare no conflict of interest.

Data Availability Statement:

The data used in the study were taken from https://www.investing.com/commodities/gold-historical-data (accessed on: 25 October 2023).

Acknowledgements

Not applicable

References

Abu-Doush, I., Ahmed, B., Awadallah, M. A., Al-Betar, M. A., & Rababaah, A. R. (2023). Enhancing multilayer perceptron neural network using archive-based harris hawks optimizer to predict gold prices. Journal of King Saud University – Computer and Information Sciences, 35(9), 101557. https://doi.org/10.1016/j.jksuci.2023.101557.

Alameer, Z., Abd Elaziz, M., Ewees, A. A., Ye, H., & Jianhua, Z. (2019). Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. Resources Policy, 61, 250–260. https://doi.org/10.1016/j.resourpol.2019.02.014.

Amini, A., & Kalantari, R. (2023). Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning. AUT Journal of Modeling and Simulation, 50(1), 71–82. https://doi.org/10.22060/miscj.2018.13508.5074 .

Baradaran, A. H., Bohlouli, M., & Jahed Motlagh, M. R. (2023). Decoding tomorrow's gold prices: A comparative study of GRU and CNN-LSTM in the Iranian market. Applied Soft Computing, 140, 110–125. https://doi.org/10.1016/j.asoc.2023.110325 .

Ben Jabeur, S., Mefteh-Wali, S., & Viviani, J.-L. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. International Review of Financial Analysis, 78, 101976. https://doi.org/10.1016/j.irfa.2021.101976 .

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555. https://doi.org/10.48550/arXiv.1412.3555 .

Cohen, G., & Aiche, A. (2023). Forecasting gold price using machine learning methodologies. Journal of Commodity Markets, 30, 100–115. https://doi.org/10.1016/j.jcomm.2023.100115 .

Dewi, N. P. J. R., Indrawan, I. G. A., Ginantra, N. L. W. S. R., & Darma, I. W. A. S. (2022). Application of long short-term memory (LSTM) and gated recurrent unit (GRU) algorithm in gold price prediction. International Journal of Intelligent Systems and Applications, 14(4), 45–58. https://doi.org/10.5815/ijisa.2022.04.05 .

Dhuhita, W. M. P., Haryoko, Farid, M. F. A., Huda, A. A., & Yaqin, A. (2022). Gold price prediction based on Yahoo finance data using LSTM algorithm. Journal of Physics: Conference Series, 2157(1), 012038. https://doi.org/10.1088/1742-6596/2157/1/012038 .

E, J., Ye, J., & Jin, H. (2023). A novel hybrid model on the prediction of time series and its application for the gold price analysis and forecasting. Knowledge-Based Systems, 278, 110–125. https://doi.org/10.1016/j.knosys.2023.110876 .

Ghute, M., & Korde, M. (2021). Efficient machine learning algorithm for future gold price prediction. International Journal of Engineering and Advanced Technology, 10(5), 78–85. https://doi.org/10.35940/ijeat.E3064.0510521 .

Gong, W. (2022). Research on gold price forecasting based on LSTM and linear regression. Journal of Physics: Conference Series, 2189(1), 012015. https://doi.org/10.1088/1742-6596/2189/1/012015 .

Hafezi, R., & Akhavan, A. N. (2018). Forecasting gold price changes: Application of an equipped artificial neural network. AUT Journal of Modeling and Simulation, 50(1), 71–82. https://doi.org/10.22060/miscj.2018.13508.5074 .

Hajek, P., & Novotny, J. (2021). Fuzzy rule-based prediction of gold prices using news affect. International Journal of Information Management, 58, 102–115. https://doi.org/10.1016/j.ijinfomgt.2021.102415 .

Hosseinpour, M. J., & Jahan, M. H. (2023). Gold price forecasting based on hybrid methods of deep learning and Shahin Harris. In Proceedings of the 7th International Conference on Global Studies in Computer, Electrical and Mechanical Engineering (pp. 112–118). IEEE. https://doi.org/10.1109/GSCEEM57105.2023.10152467 .

Jaiswal, R., & Singh, B. (2022). A hybrid convolutional recurrent (CNN-GRU) model for stock price prediction. International Journal of Pattern Recognition and Artificial Intelligence. Advance online publication. https://doi.org/10.1142/S021800142350015X .

Jovanovic, A., Dogandzic, T., Dobrojevic, M., Sarac, M., Bacanin, N., & Zivkovic, M. (2023). Gold prices forecasting using recurrent neural network with attention tuned by metaheuristics. Expert Systems with Applications, 215, 119–132. https://doi.org/10.1016/j.eswa.2023.119324 .

Liang, Y., Lin, Y., & Lu, Q. (2023). Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. Engineering Applications of Artificial Intelligence, 120, 105–120. https://doi.org/10.1016/j.engappai.2023.105912 .

Livieris, I. E., Pintelas, E., & Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. Neural Computing and Applications, 32(23), 17351–17360. https://doi.org/10.1007/s00521-020-04867-x .

Mamipour, S., & Vaezi Jezeie, F. (2014). Non-linear relationships among oil price, gold price and stock market returns in Iran: A multivariate regime-switching approach. Energy Economics, 45, 285–297. https://doi.org/10.1016/j.eneco.2014.08.001 .

Mohsin, M., & Jamaani, F. (2022). A novel deep-learning technique for forecasting oil price volatility using historical prices of five precious metals in context of green financing. Resources Policy, 79, 102–115. https://doi.org/10.1016/j.resourpol.2022.102876 .

Nanthiya, D., Gopal, S. B., Balakumar, S., Harisankar, M., & Midhun, S. P. (2021). Gold price prediction using ARIMA model. Materials Today: Proceedings, 47, 5122–5128. https://doi.org/10.1016/j.matpr.2021.05.445 .

Payandeh Najafabadi, A. T., Qazvini, M., & Ofoghi, R. (2012). The impact of oil and gold prices' shock on Tehran Stock Exchange: A copula approach. Iranian Journal of Economic Studies, 1(2), 23–47.

Rezazadeh, A., & Mohseninia, R. (2019). The impact of financial stress on the Iranian gold, currency and stock markets: A time-varying Granger-causality approach. Journal of International Financial Markets, Institutions and Money, 63, 101–115. https://doi.org/10.1016/j.intfin.2019.101142 .

Shafiee, S., & Topal, E. (2010). An overview of global gold market and gold price forecasting. Resources Policy, 35(3), 178–189. https://doi.org/10.1016/j.resourpol.2010.05.004 .

Singh, U., Saurabh, K., Trehan, N., Vyas, R., & Vyas, O. P. (2024). GA-LSTM: Performance optimization of LSTM driven time series forecasting. Applied Soft Computing. Advance online publication. https://doi.org/10.1016/j.asoc.2023.110456 .

Song, Y., Huang, J., Xu, Y., Ruan, J., & Zhu, M. (2023). Multi-decomposition in deep learning models for futures price prediction. Resources Policy, 83, 103–118. https://doi.org/10.1016/j.resourpol.2023.103621 .

Stankovic, M., Bacanin, N., Budimirovic, N., Zivkovic, M., Sarac, M., & Ivana. (2023). Bi-directional long short-term memory optimization by improved teaching-learning based algorithm for univariate gold price forecasting. Expert Systems with Applications, 215, 119–132. https://doi.org/10.1016/j.eswa.2023.119324 .

Vidal, A., & Kristjanpoller, W. (2023). Gold volatility prediction using a CNN-LSTM approach. International Review of Economics & Finance, 87, 102–115. https://doi.org/10.1016/j.iref.2023.102345 .

Wagenaar, T. (n.d.). GRU - Built-ins. Neataptic. Retrieved October 26, 2023, from https://wagenaartje.github.io/neataptic/docs/builtins/gru/ .

Wang, H., Zhang, Y., Liang, J., & Liu, L. (2022). DAFA-BiLSTM: Deep autoregression feature augmented bidirectional LSTM network for time series prediction. Knowledge-Based Systems, 243, 108–125. https://doi.org/10.1016/j.knosys.2022.108456 .

Yurtsever, M. (2021). Gold price forecasting using LSTM, Bi-LSTM and GRU. European Journal of Science and Technology, (31), 341–347.