Enhancing Gold Price Forecasting Using Machine Learning Models Optimized with Metaheuristic Algorithms

Alaa N. Sawalha [1], Mohammed A. Al-Betar [1,2,*], and Sharif N. Makhadmeh [3]

[1] Artificial Intelligence Research Center (AIRC), College of Engineering and Information Technology, Ajman University, Ajman, United Arab Emirates
[2] Center of Excellence in Precision Medicine and Digital Health, Department of Physiology, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand
[3] Department of Information Technology, King Abdullah II School for Information Technology, The University of Jordan (UJ), Amman, Jordan
Email: 202010417@ajmanuni.ac.ae (A.N.S.); m.albetar@ajman.ac.ae (M.A.A-.B.); s_makhadmeh@ju.edu.jo (S.N.M.)
*Corresponding author

***Abstract*—Forecasting gold prices is essential for supporting informed decision-making among investors, policymakers, and financial analysts. However, due to their non-linear and volatile behavior influenced by complex economic and geopolitical factors, predicting gold prices remains a significant challenge. This study evaluates the forecasting performance of three traditional machine learning models—Random Forest (RF), Multi-Layer Perceptron (MLP), and XGBoost—on a monthly dataset spanning from January 1991 to December 2023, using macroeconomic and commodity-related indicators obtained from IndexMundi. To enhance predictive accuracy, RF and MLP were optimized using metaheuristic algorithms including Particle Swarm Optimization (PSO), Differential Evolution (DE), Simulated Annealing (SA), and Genetic Algorithm (GA), while XGBoost was fine-tuned using Grid Search. Two ensemble strategies were developed to further improve performance: a weighted ensemble based on inverse error metrics and a boosting ensemble that sequentially combined top-performing models. The results show that combining traditional models with metaheuristic optimization significantly improves forecasting accuracy. The best performance was achieved by the boosting ensemble integrating RF-PSO and optimized XGBoost, attaining an R^2 of 0.9654 and a Root Mean Square Error (RMSE) of 0.0433, representing an improvement of 11.1% in RMSE over the best single optimized model. This research demonstrates that effective and scalable financial forecasting systems can be developed using established machine learning techniques, offering valuable decision-support tools in dynamic financial markets.**

***Keywords*—gold price forecasting, machine learning, metaheuristic optimization, ensemble learning, financial time series**

I. Introduction

Gold has long served as a strategic financial asset due to its scarcity, durability, and universal recognition as a store of value. Particularly during periods of economic instability, gold emerges as a preferred investment, driving increased demand. Accurately forecasting gold prices is therefore of great interest to investors, central banks, and policymakers [1, 2].

However, forecasting gold prices remains a complex challenge. Prices are influenced by a combination of macroeconomic variables, including inflation and interest rates, geopolitical events, and foreign exchange fluctuations. These interrelated factors introduce high levels of nonlinearity and volatility, complicating predictive modeling [3].

Machine Learning (ML) models, with their ability to capture complex and nonlinear data patterns, offer a powerful alternative to traditional statistical approaches. Algorithms such as Multi-Layer Perceptron (MLP), Random Forest (RF), and XGBoost (XGB) have demonstrated strong predictive capabilities in financial time series forecasting [4–7]. Nevertheless, the performance of these models heavily depends on optimal hyperparameter tuning, which is often computationally expensive and inefficient when approached using conventional methods like grid or random search [8].

Recent studies have demonstrated that ensemble and composite machine learning methods possess significant potential for modeling highly nonlinear and complex patterns across diverse research domains. For example, Xu and Zhang [9] show how composite learning strategies can effectively capture intricate dependencies and improve predictive performance in heterogeneous datasets. Such findings underscore the suitability of ensemble-based approaches for financial forecasting tasks, where market dynamics are inherently nonlinear and influenced by multiple interdependent factors. This motivates our integration of optimized base models into both weighted and boosting ensemble frameworks for robust gold price prediction.

Manuscript received June 12, 2025; revised June 26, 2025; accepted September 2, 2025; published January 15, 2026.

doi: 10.12720/jait.17.1.107-121

To address this, heuristic and metaheuristic optimization algorithms have been increasingly adopted. Techniques such as Differential Evolution (DE), Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and Simulated Annealing (SA) offer efficient search strategies for identifying optimal hyperparameter configurations [10]. Integrating these optimization methods with ML models enhances forecast accuracy while maintaining computational efficiency.

This study aims to develop a robust forecasting framework for gold price prediction by integrating ML models with heuristic and metaheuristic optimization techniques. The primary contributions of this research are as follows: development of optimized ML models (MLP, RF, XGB) using Grid Search, DE, PSO, SA, and GA to improve forecasting accuracy and adaptability to volatile market conditions; construction of ensemble models, including a weighted ensemble (based on inverse $RMSE^2$ weighting) and a boosting ensemble, to enhance model robustness and generalization; design of a hyperparameter tuning framework that efficiently explores high-dimensional parameter spaces, tailored specifically for gold price forecasting; and implementation of a feature selection approach combining statistical correlation and domain knowledge to identify impactful technical and macroeconomic indicators.

The study utilizes historical data from 1991 to 2023, incorporating daily gold prices along with macroeconomic indicators such as inflation rates and currency exchange rates. Technical indicators including the Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) are also employed. Three ML algorithms, MLP, RF, and XGB, are trained and optimized using five different optimization techniques. These optimized models are then integrated into ensemble frameworks. Model evaluation is conducted using statistical metrics such as Root Mean Square Error (RMSE), Mean Squared Error (MSE), R^2, standard deviation, and *p*-value. Visual tools such as line plots and residual plots are employed for diagnostic analysis.

The key contributions and novelty of this study are as follows:

1) Combine Random Forest (RF), Multi-Layer Perceptron (MLP), and XGBoost with four distinct metaheuristic optimization algorithms, including PSO, DE, SA, and GA, specifically for gold price forecasting.
2) Design and evaluate two ensemble learning approaches, including weighted ensemble and boosting ensemble.
3) Combine statistical correlation analysis with domain knowledge to retain the most predictive variables, ensuring model interpretability and reducing overfitting risk.
4) Assess model performance using multiple metrics (RMSE, MSE, R^2, standard deviation, *p*-value) and support findings with residual and convergence analyses for robustness verification.

While similar optimization-ensemble frameworks have been applied in other domains, this study is the first to combine RF, MLP, and XGBoost models with four distinct metaheuristic optimizers and dual ensemble strategies for gold price forecasting. This integration, combined with a rigorous feature selection process and comparative analysis against prior state-of-the-art methods, underscores the study's contribution to both methodological efficiency and forecasting accuracy.

The remainder of the paper is structured as follows: Section II reviews related literature. Section III outlines the dataset, models, optimization methods, and ensemble strategies. Section IV presents experimental results and performance evaluation. Finally, Section V concludes the study and suggests avenues for future research.

II. Related Works

Recent advancements in gold price forecasting have increasingly focused on ML models enhanced by optimization techniques. Traditional ML models, such as RF, MLP, and XGBoost, have demonstrated high predictive performance, rivaling more complex deep learning methods like Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs), particularly when optimized through heuristic and metaheuristic algorithms. These models can capture complex, nonlinear patterns in financial time series data without requiring the extensive computational resources often demanded by deep learning approaches.

Early approaches to gold price prediction relied on statistical models, such as Autoregressive Integrated Moving Average (ARIMA) and Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH), which were limited by their linear assumptions and inability to model the volatility and nonlinearities inherent in financial data [11, 12]. More recent research has shifted toward data-driven ML techniques that better account for the dynamics of gold markets. Studies have shown that models like XGBoost, when paired with interpretation tools such as Shapley Additive Explanations (SHAP), offer both accuracy and transparency in financial forecasting [13, 14]. Similarly, RF and MLP have demonstrated strong performance when properly tuned.

Deep learning models, especially LSTM and hybrid Convolutional Neural Network (CNN)-LSTM architectures, have also been explored due to their strength in capturing temporal dependencies [15–17]. However, their success is often tempered by concerns around overfitting, computational cost, and the need for large-scale, high-quality datasets. While CNN-based models have shown promise in financial forecasting, their effectiveness heavily depends on the design of model architecture and access to extensive data [18].

In contrast, optimized traditional machine learning models provide a compelling balance between interpretability, performance, and computational efficiency. Metaheuristic algorithms such as DE, PSO, GA, SA, and Harris Hawks Optimization (HHO) have been effectively applied for hyperparameter tuning in financial prediction tasks [19–22]. These algorithms help ML models achieve better generalization and predictive accuracy by exploring complex, high-dimensional parameter spaces.

Several studies have validated the effectiveness of these optimization methods. For instance, HHO has been used to optimize MLP architectures for gold forecasting, resulting in increased stability and reduced prediction error [23]. Other hybrid-based algorithms have also proven effective for optimizing hybrid models in volatile markets [21, 24]. These findings highlight the importance of intelligent hyperparameter tuning in achieving robust model performance.

Ensemble learning has gained popularity in financial forecasting for its ability to reduce model variance and bias. Techniques such as weighted averaging and boosting have been employed to combine multiple base learners, resulting in more reliable forecasts [25, 26]. Recent research further suggests that integrating metaheuristic optimization with ensemble learning offers significant gains in accuracy, stability, and convergence speed. This hybrid approach has been particularly effective in combining strengths of models like RF, MLP, and XGBoost for gold price forecasting.

Despite these advancements, limitations remain. Many studies either exclude optimization or rely solely on grid search, which is computationally inefficient for large-scale problems.

Moreover, deep learning models, while powerful, often suffer from interpretability challenges and resource constraints. Additionally, the integration of qualitative data, such as geopolitical sentiment or financial news via Natural Language Processing (NLP), remains underexplored despite its potential to enhance forecasting frameworks [27].

In summary, previous studies have demonstrated that machine learning models such as RF, MLP, and XGBoost, when combined with optimization techniques including metaheuristics, can enhance gold price forecasting accuracy. Ensemble learning approaches have also shown promise in reducing variance and improving robustness. However, most prior works either focus on a single optimization algorithm applied to one model type, omit ensemble strategies, or do not provide direct comparisons with recent state-of-the-art methods on the same dataset. The integration of multiple optimizers with diverse model architectures in a unified ensemble framework remains underexplored. Furthermore, while some studies have optimized deep learning models, their practical deployment is often hindered by high computational cost, overfitting risks, and limited interpretability.

To address these gaps, this study proposes a multi-model, multi-optimizer ensemble framework for gold price forecasting. Specifically, three traditional ML models (RF, MLP, and XGBoost) with four distinct metaheuristic optimization algorithms (PSO, DE, SA, GA) and two ensemble learning strategies (weighted and boosting) are integrated. The proposed approach is systematically evaluated against a recent HHO-optimized MLP study on the same dataset, demonstrating superior forecasting accuracy. This work not only validates the effectiveness of combining well-established metaheuristics with ensemble learning but also offers a scalable, interpretable, and computationally efficient solution for financial time series forecasting.

III. RESEARCH METHODOLOGY

This study proposes a forecasting framework for gold price prediction by integrating advanced machine learning models and heuristic/metaheuristic optimization techniques. The methodology is built upon three core machine learning models: MLP, RF, and XGBoost. These models are selected for their ability to model non-linear relationships among financial variables while maintaining simplicity and interpretability.

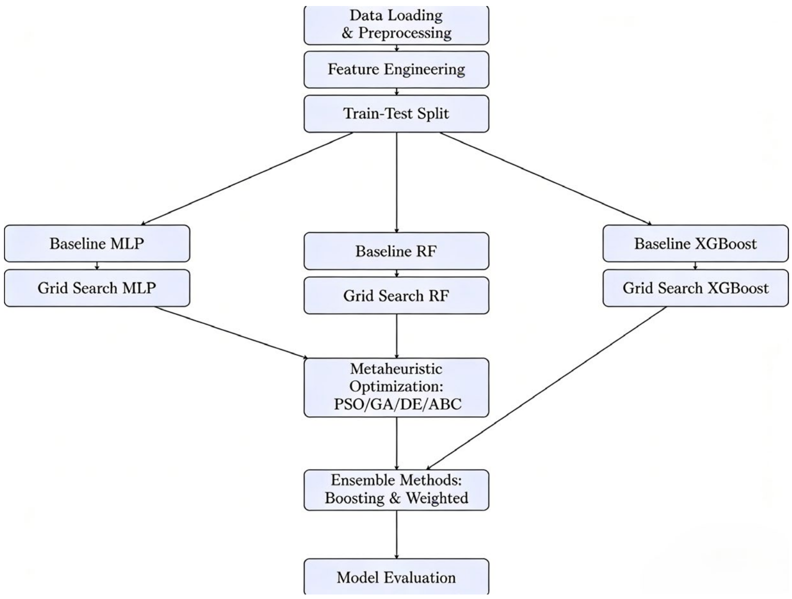


Fig. 1. Workflow of the gold price forecasting framework with ensemble integration.

To enhance the predictive performance and generalization capabilities of these models, a combination of traditional and metaheuristic optimization techniques is employed. Specifically, hyperparameter tuning is conducted using Grid Search and four well-established metaheuristic algorithms: GA, PSO, evaluation is carried out using performance metrics including the coefficient of determination (R^2), RMSE, MSE, p-value, and standard deviation.

The complete methodological workflow is illustrated in Fig. 1, encompassing the preprocessing, modeling, optimization, ensemble integration, and evaluation stages.

A. Problem Definition

Gold price forecasting is inherently complex due to its non-linearity, high dimensionality, and susceptibility to economic, political, and market fluctuations. The relationships between influencing variables and gold prices are dynamic and non-linear, evolving over time in response to global events. This volatility renders traditional linear statistical models such as ARIMA or linear regression insufficient for capturing underlying patterns [28, 29].

To address this complexity, the study leverages machine learning models capable of accommodating changing dependencies and heteroscedastic behavior. The model inputs include macroeconomic and financial indicators such as inflation rates (the United State and China), foreign exchange rates (USD/ZAR, USD/CNY, USD/INR), and commodity prices (gold, copper, silver, iron, crude oil). These were selected based on prior empirical evidence confirming their influence on gold price dynamics [30–32].

In addition, technical indicators including the RSI and MACD were engineered from historical gold price series to capture momentum and trend shifts [33]. The inclusion of a one-period lag variable (Gold_Lag_1) enables the model to incorporate temporal autocorrelation, aligning the problem formulation with time series forecasting paradigms. A correlation analysis was conducted to eliminate redundant or weakly correlated features, ensuring that only informative variables are retained [34].

Let $D = \{(X_t, y_t)\}_{t=1}^{T}$ denote a chronological dataset where $X_t \in \mathbb{R}^d$ represents d selected features at time t and $y_t \in \mathbb{R}$ is the corresponding gold price. The forecasting task aims to learn a mapping $f_\theta: \mathbb{R}^d \to \mathbb{R}$ such that:

$$\hat{y}_{t+1} = f_\theta(X_t), \quad t = 1, \ldots, T-1 \tag{1}$$

The optimal model parameters θ^* are obtained by minimizing a loss function L over a validation set V:

$$\theta^* = \arg\min_\theta L\ (\theta; V) \tag{2}$$

The primary loss is the RMSE:

$$RMSE = \sqrt{\frac{1}{|V|} \sum_{(X_t, y_t)\in V} (y_t - \hat{y}_t)^2} \tag{3}$$

To provide a comprehensive evaluation, supplementary metrics such as Mean Squared Error (MSE), the coefficient of determination (R^2), and MAPE are also computed.

$$MAPE = \frac{100}{|V|} \sum_{(X_t, y_t)\in V} \left| \frac{y_t - \hat{y}_t}{y_t + \varepsilon} \right|, \varepsilon > 0 \tag{4}$$

Eq. (4) encapsulates gold price forecasting as a supervised learning problem for sequential data, where model performance depends on the accurate capture of nonlinear dependencies and temporal dynamics.

B. Analytical Steps

The predictive modeling process is organized into four main steps:

1) Data preprocessing: Historical records of gold prices and economic indicators are selected, normalized, and split into training and testing subsets.
2) Model training: Baseline models (MLP, RF, XGBoost) are initially trained using predefined hyperparameters.
3) Optimization: PSO, DE, SA, and GA are applied to optimize the MLP and RF models, targeting minimum RMSE values.
4) Ensemble construction: Predictions from the tuned models are aggregated using weighted and boosting ensembles to improve forecast reliability.

C. Dataset Description

The dataset spans from January 1991 to December 2023 and contains monthly observations. It was obtained from IndexMundi, a reputable data aggregation platform commonly used in similar studies [23]. A representative sample of the dataset is shown in Table I.

The dataset includes:

- Economic indicators: US and China inflation rates.
- Currency exchange rates: USD/ZAR, USD/CNY, and USD/INR.
- Commodity prices: Gold (target), silver, copper, iron, and crude oil.

TABLE I. SAMPLE OF THE HISTORICAL GOLD PRICE FORECASTING DATASET

Date	Gold	Copper	Silver	Iron	China Inf.	US Inf.	USD/ZAR	USD/CNY	USD/INR	Oil Price
01/01/1991	922.54	8,418.56	30.30	197.73	0.0146	0.0043	5.53	41.17	12.80	86.88
02/01/1991	588.55	1,461.36	42.45	203.69	0.0216	0.0188	5.55	60.06	10.64	34.06
05/01/2007	1,848.14	8,596.67	10.01	164.15	0.0089	0.0021	7.41	50.29	12.22	41.05
06/01/2007	651.27	6,599.07	36.59	51.19	0.0137	0.0086	8.68	53.64	10.99	112.00
11/01/2023	1,378.64	5,294.87	12.98	48.03	0.0159	0.0085	5.94	33.48	6.80	22.53

D. Data Preprocessing

The following preprocessing steps were applied to prepare the dataset for modeling:

1) Missing value imputation

Missing entries in the time series were imputed using linear interpolation. For a missing value x_t between two known values x_{t_1} and x_{t_2}, where $t_1 < t < t_2$, the imputed value is:

$$x_t = x_{t_1} + \frac{t - t_1}{t_2 - t_1}(x_{t_2} - x_{t_1}) \tag{5}$$

It preserves the temporal continuity and minimizes artificial noise.

2) Normalization

All features were scaled to the [0, 1] range using Min-Max normalization:

$$x_t' = \frac{x_t - \min(X)}{\max(X) - \min(X)} \tag{6}$$

where X is the vector of all observed values for that feature. This is particularly beneficial for gradient-based models such as MLP.

3) Technical indicator engineering

Two technical indicators were derived from historical gold prices:

- Relative Strength Index (RSI) over a period n:

$$RSI_t = 100 - \frac{100}{1 + RS_t} \tag{7}$$

$$RS_t = \frac{\text{Average Gain over } n \text{ periods}}{\text{Average Loss over } n \text{ periods}} \tag{8}$$

- Moving Average Convergence Divergence (MACD):

$$MACD_t = EMA_{fast}(P_t) - EMA_{slow}(P_t) \tag{9}$$

where P_t is the gold price at time t; EMA_{fast} and EMA_{slow} are exponential moving averages with short and long windows.

4) Train-test split

The dataset was split into training and testing subsets using a chronological split of 70% for training and 30% for testing to avoid look-ahead bias:

$$D_{train} = \{(X_t, y_t)\}_{t=1}^{|0.7T|} \tag{10}$$

$$D_{test} = \{(X_t, y_t)\}_{t=|0.7T|+1}^{T} \tag{11}$$

E. Machine Learning Models

This study employs three prominent machine learning models for gold price forecasting: MLP, RF, and XGBoost. These models were selected for their strong performance in capturing non-linear relationships and handling time-series financial data.

1) Multi-Layer Perceptron (MLP)

MLP is a feed-forward artificial neural network architecture widely used for regression tasks due to its capability to model complex, non-linear relationships [35]. The network consists of an input layer, two hidden layers (64 and 32 neurons), and a single output neuron. The ReLU activation function is used in the hidden layers:

$$f(x) = \max(0, x) \tag{12}$$

The output layer uses a linear activation function suitable for regression:

$$\hat{y} = \sum_{i=1}^{n} w_i h_i + b \tag{13}$$

Input features were normalized using MinMax scaling:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{14}$$

The model is trained using the MSE loss function:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \tag{15}$$

Training was conducted over 100 epochs with a batch size of 32 using backpropagation.

2) Random Forest (RF)

Random Forest is an ensemble learning technique that aggregates predictions from multiple decision trees built on bootstrapped samples [36]. The prediction is calculated by averaging the outputs of individual trees:

$$\hat{y}_{gold} = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \tag{16}$$

where $f_t(x)$ denotes the output of the t^{th} decision tree. The model consists of 100 trees, and key hyperparameters such as maximum depth and minimum split size were optimized using Grid Search and metaheuristic algorithms (PSO, DE, GA, SA).

3) XGBoost

XGBoost is a high-performance implementation of gradient boosting that integrates L_1 and L_2 regularization to enhance generalization and control overfitting. The prediction is expressed as:

$$\hat{y}_{gold} = \sum_{k=1}^{K} f_k(x), f_k \in F \tag{17}$$

In this study, the XGBoost model was configured with 100 trees, a maximum tree depth of 3, and a learning rate of 0.1. The squared error was used as the objective function. Grid Search was employed for hyperparameter tuning. Due to its built-in regularization and robustness, external metaheuristic optimization was not required.

F. Optimizing Machine Learning Parameters Using Heuristic and Metaheuristic Algorithms

This section describes the application of heuristic (Grid Search) and metaheuristic algorithms (DE, GA, PSO, SA) to optimize machine learning hyperparameters and enhance predictive accuracy.

1) Grid search optimization

Grid search systematically explores predefined hyperparameter combinations to identify the optimal set maximizing performance. The search space Θ comprises all combinations of candidate values:

$$N_{combination} = C_1 \times C_2 \times \cdots \times C_n \tag{18}$$

The best configuration θ^* is selected by maximizing the coefficient of determination (R^2):

$$R^2(\theta^*) = \max\left\{R^2(\theta) \middle| \theta \in \Theta\right\} \tag{19}$$

Grid Search with 3-fold cross-validation was applied to tune RF, MLP, and XGBoost models. Key parameters included estimators and depth, neurons and dropout rate, and learning rate and tree depth.

2) Differential Evolution (DE)

DE optimizes by evolving a population through mutation, crossover, and selection. For a population $P_0 = \{x_1, ..., x_N\}$ in D-dimensional space:

$$v_i = x_{r1} + F(x_{r2} - x_{r3}) \tag{20}$$

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{if } \text{rand}_{i,j} \le CR \\ x_{i,j}, & \text{otherwise} \end{cases} \tag{21}$$

The fitness is defined as:

$$Fitness(x) = \frac{1}{1 + RMSE(x)} \tag{22}$$

DE was used to tune RF and MLP parameters due to its robustness and global search capabilities.

3) Genetic Algorithm (GA)

GA evolves population using selection, crossover, and mutation. Fitness is inversely related to $RMSE$:

$$Fitness(x) = \frac{1}{RMSE(x) + \varepsilon} \tag{23}$$

Offspring are created as:

$$Offspring_i = \alpha x_a + (1-\alpha) x_b \tag{24}$$

$$x_i' = x_i + \delta, \quad \delta \sim N(0, \sigma^2) \tag{25}$$

GA was applied to optimize hyperparameters of MLP and RF. Its stochastic nature aids in exploring diverse regions of the search space.

4) Particle Swarm Optimization (PSO)

PSO models particles move in a solution space, adjusting positions based on personal and global bests. The update rules are:

$$v_i^{t+1} = wv_i^t + c_1 r_1 (pBest_i - x_i^t) + c_2 r_2 (gBest - x_i^t) \tag{26}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{27}$$

PSO was effective for optimizing MLP and RF parameters due to its fast convergence and simple implementation.

5) Simulated Annealing (SA)

SA performs probabilistic exploration using a cooling schedule. A new solution is accepted based on the Metropolis criterion:

$$P = e^{-\Delta E/T}, \quad \Delta E = Fitness_{new} - Fitness_{current} \tag{28}$$

SA was applied to fine-tune MLP and RF models by escaping local optima through controlled randomness.

Each of these algorithms contributed to discovering optimal hyperparameters for predictive models, enabling superior performance compared to traditional fixed-parameter configurations.

G. Ensemble Learning for Gold Price Forecasting

This study integrates ensemble learning to improve the accuracy and robustness of gold price forecasting models. Two ensemble techniques are employed: weighted ensemble and boosting ensemble. Both approaches combine predictions from multiple optimized models to reduce individual model errors and enhance generalization.

1) Weighted ensemble method

The weighted ensemble combines the outputs of $n = 12$ optimized models, comprising RF, MLP, and XGBoost variants tuned via Grid Search, DE, PSO, SA, and GA.

Let $\hat{y}_i(t)$ denote the prediction of model i at time t and $RMSE_i$ be its root mean squared error on the validation set, computed as:

$$RMSE_i = \sqrt{\frac{1}{m}\sum_{k=1}^{m}\left(y_k - \hat{y}_i(k)\right)^2} \tag{29}$$

where m is the number of validation samples and y_k is the actual observed value.

Each model's weight w_i is assigned inversely proportional to the square of its RMSE:

$$w_i = \frac{RMSE_i^{-2}}{\sum_{j=1}^{n} RMSE_j^{-2}}, \quad \sum_{i=1}^{n} w_i = 1, \quad w_i \ge 0 \tag{30}$$

The final ensemble prediction at time t is then:

$$\hat{y}_{ensemble}(t) = \sum_{i=1}^{n} w_i \cdot \hat{y}_i(t) \tag{31}$$

This weighting scheme ensures that models with lower prediction error have greater influence in the final output. By aggregating predictions from diverse optimization strategies, the ensemble benefits from the complementary strengths of individual models, thereby enhancing stability, reducing variance, and mitigating overfitting—

qualities particularly important for volatile financial data.

Performance is evaluated using metrics such as RMSE, MSE, R^2, standard deviation, and p-value. The flowchart of the weighted ensemble method is presented in Fig. 2.

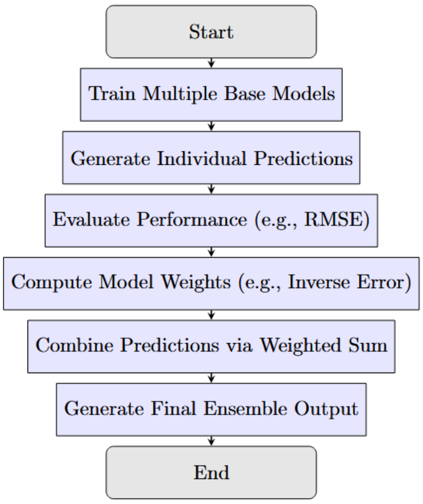


Fig. 2. Flowchart of the weighted ensemble method.

2) Boosting ensemble method

Boosting is a sequential ensemble learning strategy in which a strong base learner is iteratively improved by fitting a secondary model to the residual errors of the first. In this study, boosting architecture combines a primary base model with an XGBoost residual learner, designed to capture patterns not modeled by the base predictor. Two configurations were tested: (i) RF-PSO + XGBoost and (ii) MLP-PSO + XGBoost.

Let y_i denote the actual gold price at time step i and $\hat{y}_{Base,i}$ be the prediction from the base model (RF-PSO or MLP-PSO). The residual at each observation is computed as:

$$r_i = y_i - \hat{y}_{Base,i}, \quad i = 1, 2, \ldots, m \tag{32}$$

where m is the number of samples in the training or validation set.

The XGBoost residual learner is trained to approximate the mapping f_{XGB}: $r_i \rightarrow \hat{r}_i$, producing the predicted residual $\hat{r}_i$. The final boosted prediction is then:

$$\hat{y}_{Boosting,i} = \hat{y}_{Base,i} + \hat{r}_i \tag{33}$$

This formulation allows the second learner to model systematic errors from the first, effectively refining predictions.

Among the tested configurations, the RF-PSO + XGBoost ensemble outperformed MLP-PSO + XGBoost in both predictive accuracy and computational efficiency. The boosting approach is particularly effective for this problem because gold price dynamics contain residual nonlinear dependencies that a single model may miss. By explicitly modeling these residuals, the ensemble improves generalization and reduces bias.

The process is illustrated in Fig. 3.

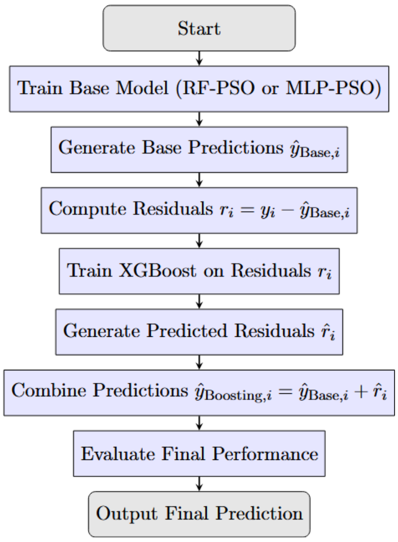


Fig. 3. Flowchart of the Boosting Ensemble Method combining a base model (RF-PSO or MLP-PSO) with an XGBoost residual learner.

By leveraging boosting in this manner, the final ensemble benefits from both the robust feature extraction of the base learner and the fine-grained residual modeling of XGBoost. This two-stage design is scalable and adaptable to other time series forecasting problems with similar non-linear dynamics.

IV. Results and Discussion

This section presents and discusses the results of the proposed optimized machine learning framework for gold price forecasting. The analysis evaluates individual models, optimization strategies, and ensemble methods. We begin by detailing the dataset and preprocessing steps, followed by the evaluation metrics used to assess model performance. The results are presented in phases—starting with baseline models (RF, MLP, XGBoost), followed by their optimized versions using Grid Search, Differential Evolution, Particle Swarm Optimization, Simulated Annealing, and Genetic Algorithm. Ensemble strategies—weighted and boosting—are then applied. Residual and convergence analyses further support the results, concluding with a comparative discussion against existing literature.

A. Feature Selection and Dataset Overview

The dataset comprises macroeconomic indicators (e.g., inflation rates), commodity prices (silver, copper, oil), and currency exchange rates (USD/ZAR, USD/CNY, USD/INR). Derived features—RSI, MACD, and Gold_Lag_1—were added to capture market momentum and trends. Correlation analysis identified four key features: RSI (0.9657), MACD (0.6301), Oil Price (0.1618), and Gold_Lag_1 (0.1270), as shown in Table II.

TABLE II. Correlation of Selected Features with Gold Price

Feature	Correlation
RSI	0.9657
MACD	0.6301
Oil Price	0.1618
Gold_Lag_1	0.1270

The dataset spans January 1991 to December 2023, normalized using MinMaxScaler. Table III shows representative samples.

TABLE III. SAMPLE ROWS FROM THE GOLD PRICE FORECASTING DATASET

Date	RSI	MACD	Gold_Lag _1	Oil Price	Gold
1991. 02	0.432	0.015	0.430	0.327	0.435
2007. 06	0.712	0.121	0.705	0.586	0.725
2023. 12	0.943	0.284	0.941	0.778	0.950

B. *Evaluation Metrics*

To rigorously assess model performance, five complementary evaluation metrics were employed: R^2, MSE, RMSE, Standard Deviation of Errors (σ), and *p*-value. Let y_i be the actual gold price at time step i, $\hat{y}_i$ the predicted price, $\overline{y}$ the mean of actual values, $e_i = y_i - \hat{y}_i$ the prediction error, $\overline{e}$ the mean error, n the number of samples, and r the Pearson correlation coefficient between y_i and $\hat{y}_i$.

1) *Coefficient of Determination (R^2)*

Measures the proportion of variance in the dependent variable explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{34}$$

Higher values indicate better explanatory power, with $R^2 = 1$ representing perfect predictions.

2) *Mean Squared Error (MSE)*

Represents the average squared prediction error:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{35}$$

Lower values signify greater predictive accuracy.

3) *Root Mean Squared Error (RMSE)*

Square root of the MSE, providing error magnitude in the original units:

$$RMSE = \sqrt{MSE} \tag{36}$$

RMSE is more interpretable than MSE in terms of scale.

4) *Standard Deviation of Errors (σ)*

Quantifies the dispersion of prediction errors around their mean:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - \overline{e})^2} \tag{37}$$

A smaller σ reflects greater stability in predictions.

5) *p-value*

Tests the statistical significance of the correlation between predicted and actual values using the t-statistic:

$$t = r\sqrt{\frac{n-2}{1-r^2}} \tag{38}$$

A *p*-value below 0.05 indicates statistical significance at the 5% level.

These metrics collectively provide a holistic evaluation of both accuracy and reliability. While R^2 offers an intuitive measure of variance explanation, it can be misleading in time series forecasting due to autocorrelation effects. Therefore, in this study, R^2 is used only as a supplementary indicator, with primary emphasis on error-based measures such as RMSE and MSE for evaluating forecasting accuracy.

C. *Evaluation Procedure*

The evaluation process followed five key steps:

(1) Baseline assessment: Default RF, MLP, and XGBoost were evaluated.
(2) Optimization: Grid Search and metaheuristics (PSO, DE, SA, GA) were used to tune models. XGBoost was optimized via Grid Search only.
(3) Comparison: Models were evaluated and compared using R^2, MSE, RMSE, standard deviation, and *p*-value.
(4) Ensemble methods: Top models were integrated using weighted and boosting ensembles.
(5) Validation: Residual and convergence analyses confirmed model robustness.

The following sections present the results and insights from each evaluation stage.

D. *Experimental Results and Analysis*

This section presents the experimental outcomes and analysis of the machine learning models applied to gold price forecasting. The models are evaluated both before and after optimization to assess the impact on accuracy and generalization.

1) *Baseline model performance*

The initial analysis evaluates the default (non-optimized) configurations of RF, MLP, and XGBoost. These baseline results establish reference points for measuring the effectiveness of subsequent optimization.

Table IV summarizes the performance of each model based on R^2, MSE, RMSE, standard deviation, and *p*-value. Among the models, RF demonstrates the strongest baseline performance, with the highest R^2 (0.8537), lowest MSE (0.0091), and RMSE (0.0955). XGBoost shows comparable but slightly inferior results. In contrast, MLP exhibits the weakest performance, indicating a need for hyperparameter tuning.

TABLE IV. BASELINE PERFORMANCE OF MACHINE LEARNING MODELS

Model	R^2	MSE	RMSE	Std Dev	*p*-value
RF	0.8537	0.0091	0.0955	0.2576	$1.735e^{-57}$
Multi-Layer Perceptron	0.5469	0.0283	0.1681	0.2206	$2.438e^{-39}$
XGBoost	0.8525	0.0092	0.0959	0.2578	$4.103e^{-58}$

These results highlight RF as the strongest baseline model, with XGBoost showing potential pending optimization. MLP's poor performance reinforces the need for hyperparameter tuning to improve predictive accuracy.

2) Model performance after optimization

This section evaluates the performance of machine learning models after applying hyperparameter optimization using both heuristic and metaheuristic techniques. RF and MLP were tuned using Grid Search, DE, PSO, SA, and GA. XGBoost was optimized solely through Grid Search due to its inherent boosting and internal optimization structure.

RF achieved its best results with PSO, yielding the highest R^2 (0.8564) and lowest RMSE (0.0946), indicating enhanced predictive precision and consistency across all optimization methods, as presented in Table V.

The PSO-optimized MLP model achieved the highest accuracy among all tuning methods, with an R^2 of 0.8193 and RMSE of 0.1062, showing substantial improvement over its baseline performance, as shown in Table VI.

In Table VII, XGBoost showed only marginal gains after Grid Search tuning, reinforcing the strength of its built-in optimization. Nevertheless, slight improvements in error metrics were observed.

Among all optimized models, RF-PSO demonstrated the best overall performance, followed closely by XGBoost and MLP-PSO, as presented in Table VIII. These models are selected for ensemble integration to leverage their complementary strengths for improved forecasting accuracy.

TABLE V. PERFORMANCE OF OPTIMIZED RF MODELS

Model	R^2	MSE	RMSE	Std Dev	*p*-value
RF Grid Search	0.8540	0.0091	0.0954	0.2569	$1.295e^{-57}$
RF-DE	0.8563	0.0090	0.0947	0.2569	$5.473e^{-58}$
RF-PSO	0.8564	0.0090	0.0946	0.2568	$5.157e^{-58}$
RF-SA	0.8563	0.0090	0.0947	0.2569	$5.442e^{-58}$
RF-GA	0.8560	0.0090	0.0948	0.2571	$6.266e^{-58}$

TABLE VI. PERFORMANCE OF OPTIMIZED MLP MODELS

Model	R^2	MSE	RMSE	Std Dev	*p*-value
MLP Grid Search	0.7920	0.0130	0.1139	0.2375	$1.803e^{-54}$
MLP-DE	0.6958	0.0190	0.1377	0.2316	$9.717e^{-56}$
MLP-PSO	0.8193	0.0113	0.1062	0.2422	$6.425e^{-56}$
MLP-SA	0.7382	0.0163	0.1278	0.2414	$5.633e^{-53}$
MLP-GA	0.5782	0.0263	0.1622	0.2298	$1.609e^{-54}$

TABLE VII. PERFORMANCE OF XGBOOST BEFORE AND AFTER GRID SEARCH OPTIMIZATION

Model	R^2	MSE	RMSE	Std Dev	*p*-value
XGBoost (Baseline)	0.8525	0.0092	0.0959	0.2578	$4.103e^{-58}$
XGBoost (Optimized)	0.8543	0.0091	0.0953	0.2570	$1.549e^{-58}$

TABLE VIII. TOP PERFORMING OPTIMIZED MODELS

Model	R^2	MSE	RMSE	Std Dev	*p*-value
RF-PSO	0.8564	0.0090	0.0946	0.2568	$\mathbf{5.157e^{-58}}$
MLP-PSO	0.8193	0.0113	0.1062	0.2422	$6.425e^{-56}$
XGBoost (Optimized)	0.8543	0.0091	0.0953	0.2570	$1.549e^{-58}$

3) Ensemble learning results

To further enhance forecasting accuracy, two ensemble strategies were implemented: a weighted ensemble and a boosting ensemble. These approaches combine the strengths of individually optimized models to improve robustness and reduce prediction errors.

The weighted ensemble aggregates predictions based on the inverse of RMSE values, giving greater weight to models with lower errors. As shown in Table IX, this method yields moderate improvements over individual models, achieving $R^2 = 0.8333$ and RMSE = 0.1020.

In contrast, boosting ensembles significantly improves performance by sequentially combining optimized models. Two configurations were tested: (1) MLP-PSO with XGBoost, and (2) RF-PSO with XGBoost. As shown in Table X, both configurations outperform individual models, with the RF-PSO + XGBoost combination achieving the highest accuracy ($R^2 = 0.9617$, RMSE = 0.0489).

TABLE IX. PERFORMANCE OF WEIGHTED ENSEMBLE

Model	R^2	MSE	RMSE	Std Dev	*p*-value
Weighted Ensemble	0.8333	0.0104	0.1020	0.2505	$1.866e^{-57}$

TABLE X. COMPARISON OF BOOSTING ENSEMBLE METHODS

Ensemble Model	R^2	MSE	RMSE	Std Dev	*p*-value
Boosting (MLP-PSO + XGBoost)	0.9102	0.0045	0.0672	0.2461	$2.127e^{-63}$
Boosting (RF-PSO + XGBoost)	0.9617	0.0024	0.0489	0.2438	$4.475e^{-79}$

TABLE XI. COMPARISON OF ENSEMBLE METHODS

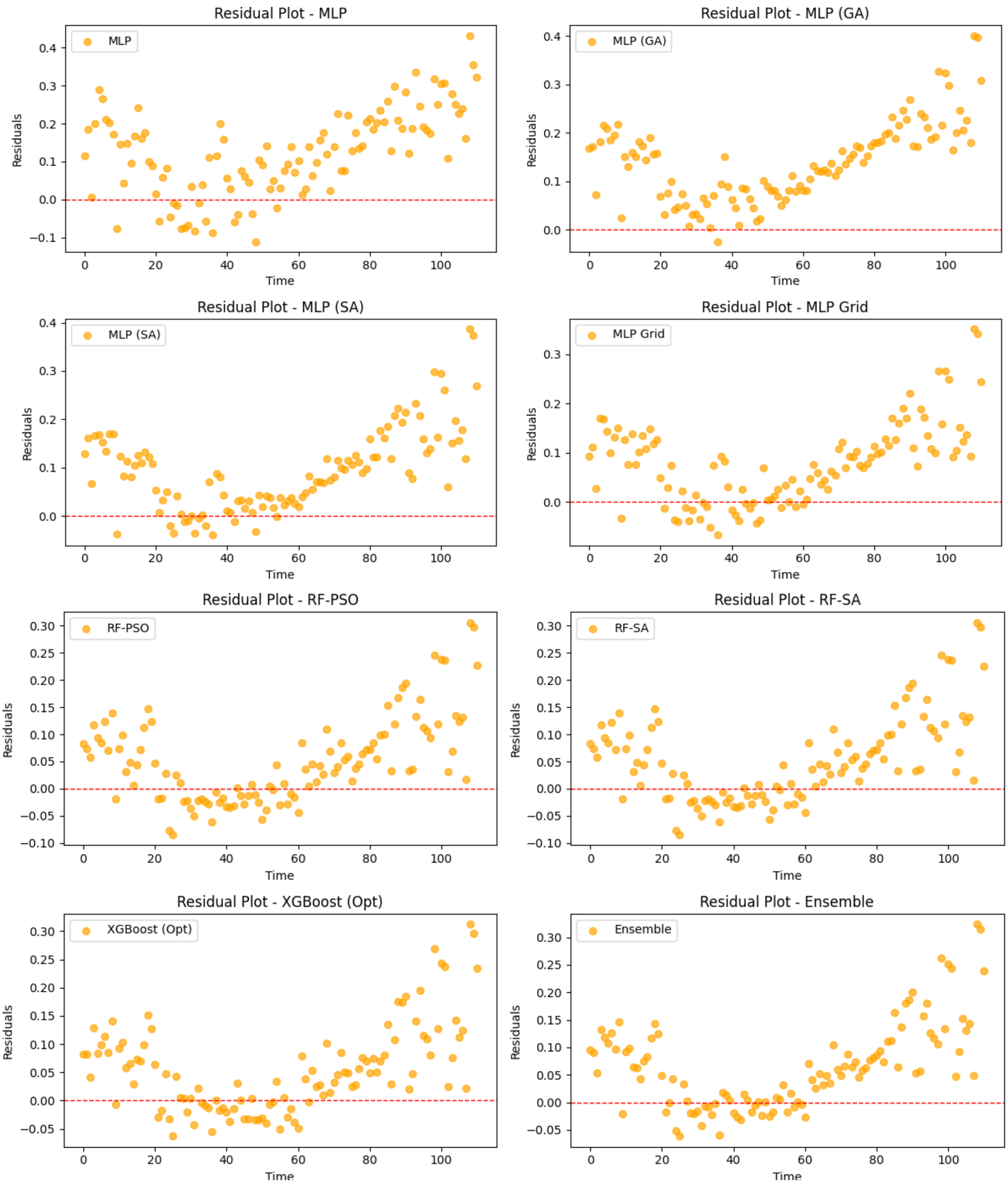
Ensemble Model	R^2	MSE	RMSE	Std Dev	*p*-value
Weighted Ensemble	0.8333	0.0104	0.1020	0.2505	$1.866e^{-57}$
Boosting (MLP-PSO + XGBoost)	0.9102	0.0045	0.0672	0.2461	$2.127e^{-63}$
Boosting (RF-PSO + XGBoost)	0.9617	0.0024	0.0489	0.2438	$4.475e^{-79}$

Table XI summarizes the performance of all ensemble methods. While the weighted ensemble offers modest gains, boosting ensembles—particularly RF-PSO + XGBoost—demonstrate superior forecasting capabilities by effectively merging complementary model strengths.

These results confirm that ensemble learning, particularly through boosting, substantially enhances model accuracy and reliability in gold price forecasting tasks.

4) *Residual analysis*

Residual analysis was conducted to evaluate model accuracy and error distribution. Residuals—the differences between actual and predicted values—should ideally be randomly scattered around zero to indicate a well-fitted model. Fig. 4 presents residual plots for baseline, optimized, and ensemble models.



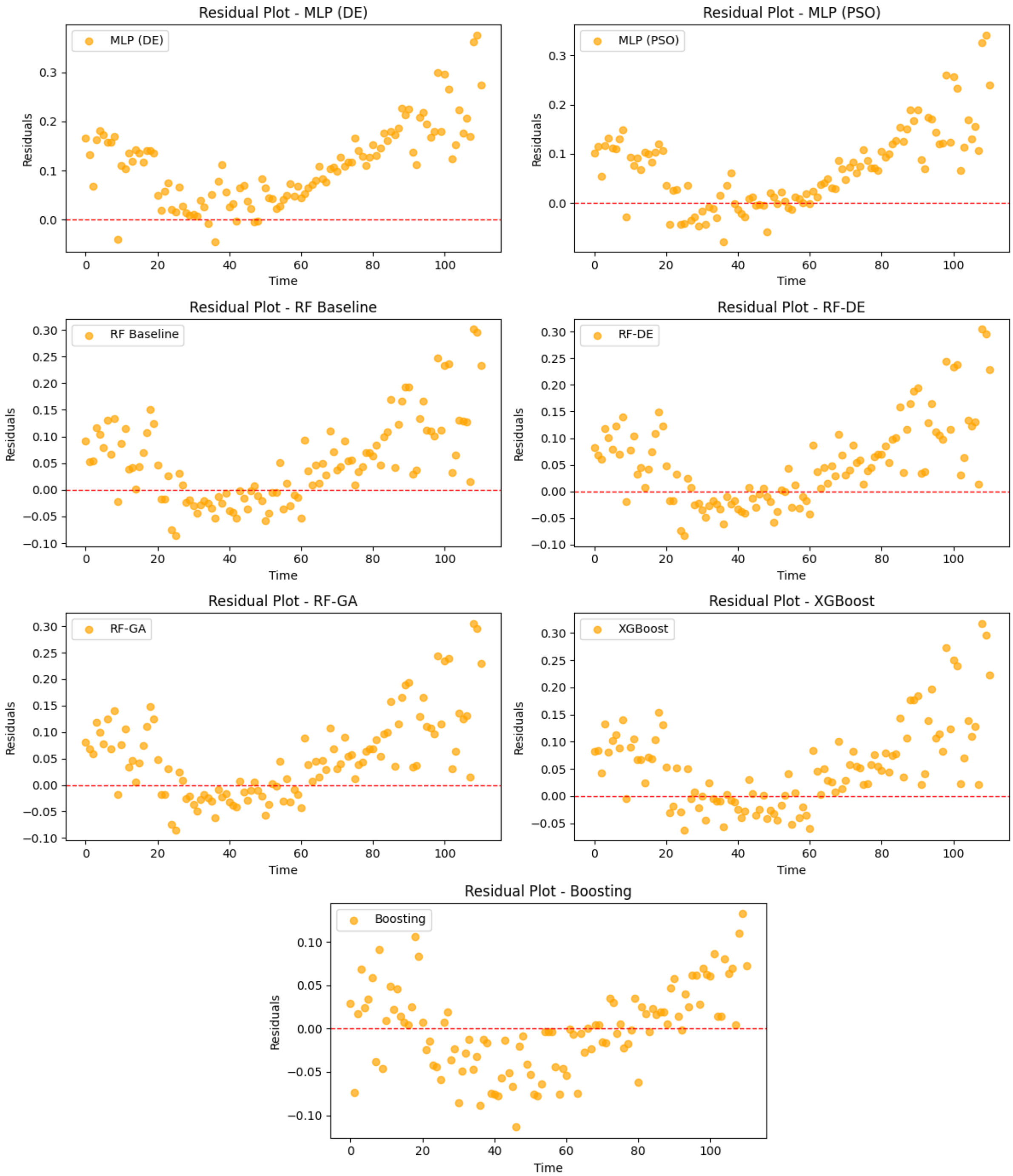


Fig. 4. Residual plots for baseline, optimized, and ensemble models.

The results show that optimized models exhibit more tightly clustered residuals around zero compared to their baseline versions. This indicates enhanced accuracy and reduces bias due to optimization. Among all models, the boosting ensemble demonstrated the least residual variation, confirming its superior predictive precision.

5) Convergence analysis

Convergence analysis was used to assess the efficiency and stability of the metaheuristic algorithms. The fitness curves in Figs. 5 and 6 illustrate the reduction in prediction error across iterations for RF and MLP models.

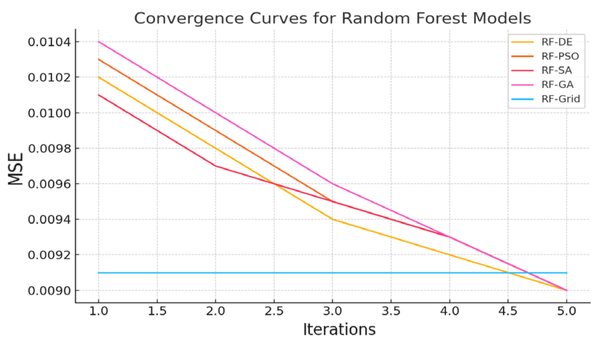


Fig. 5. Convergence curve for RF under metaheuristic optimization.

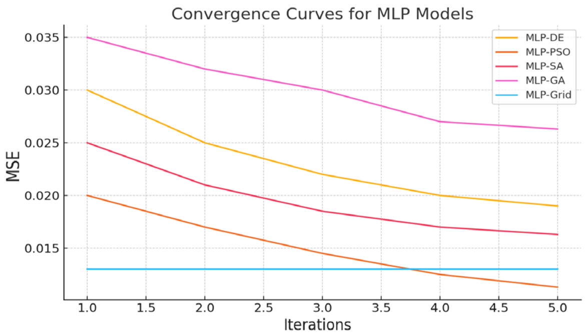


Fig. 6. Convergence curve for MLP under metaheuristic optimization.

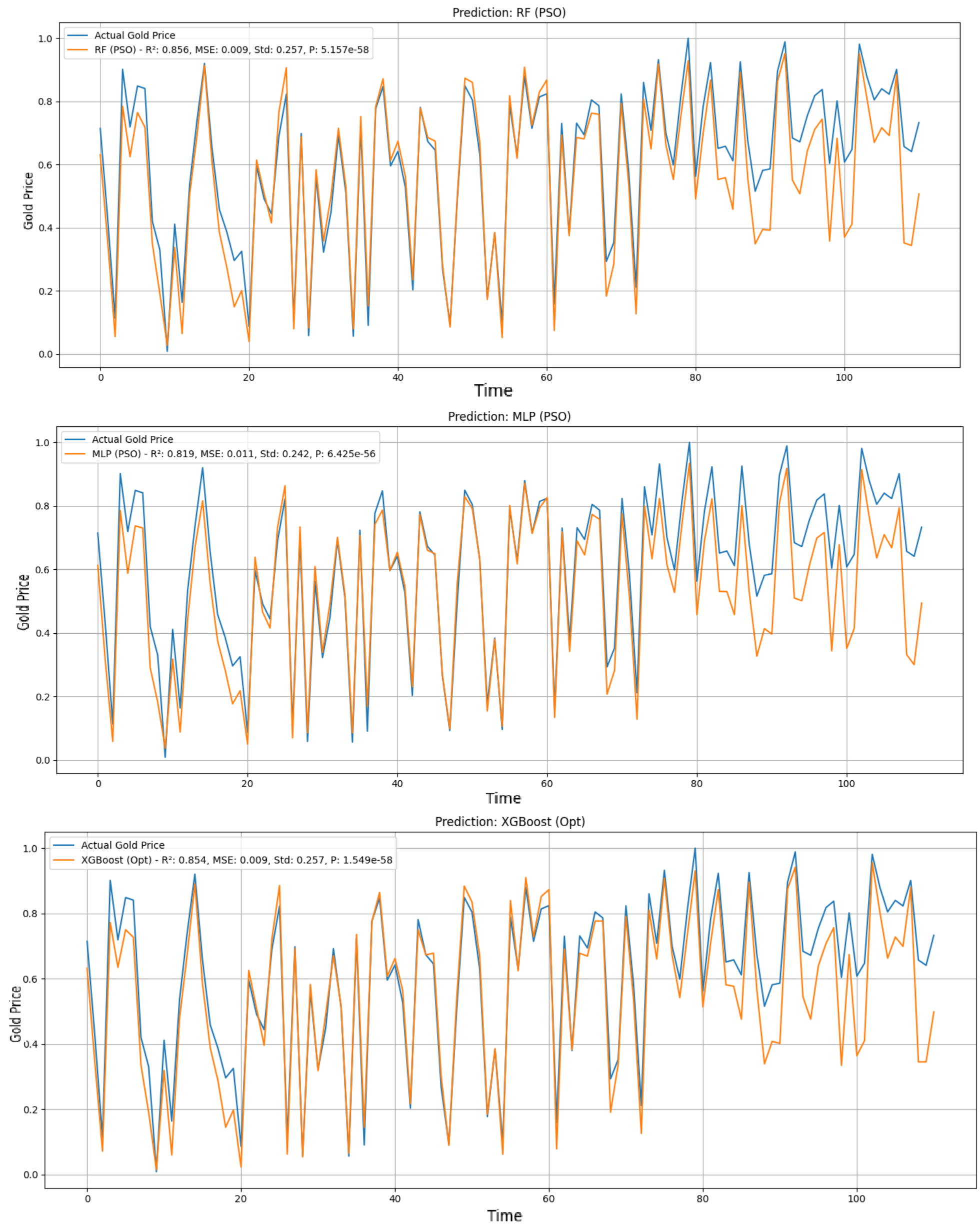
PSO and SA demonstrated faster and smoother convergence compared to DE and GA. PSO achieved rapid error reduction in early iterations, while SA maintained steady progress. DE and GA required more iterations to reach similar error levels, indicating slower convergence. These results highlight the training efficiency of PSO and SA in optimization.

6) *Summary of best results*

The RF model outperformed other baseline models. After optimization, PSO delivered the best improvements for RF, while SA showed the highest gains for MLP. Although XGBoost exhibited marginal improvement after Grid Search, its performance remained strong.

Boosting ensembles significantly enhanced forecast accuracy. The best performance was achieved by combining RF-PSO and optimized XGBoost, resulting in the highest R^2 and lowest RMSE, confirming the ensemble's ability to capture gold price variation with high fidelity. The SA-MLP model, despite improvement, was excluded from the final ensemble due to slightly lower accuracy, which could diminish ensemble strength.

Fig. 7 presents the predicted vs actual gold price trajectories for the top-performing models. The boosting ensemble closely tracks actual prices, outperforming all other models.



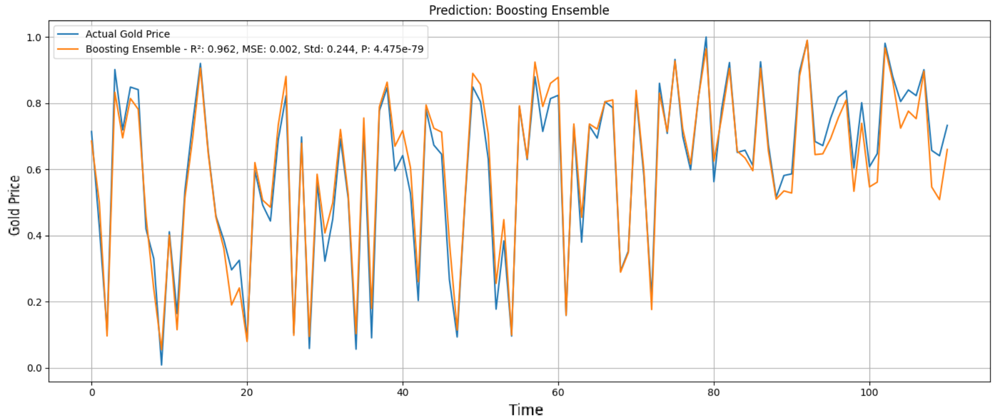


Fig. 7. Predicted vs actual gold prices for RF-PSO, MLP-PSO, XGBoost, and boosting ensemble.

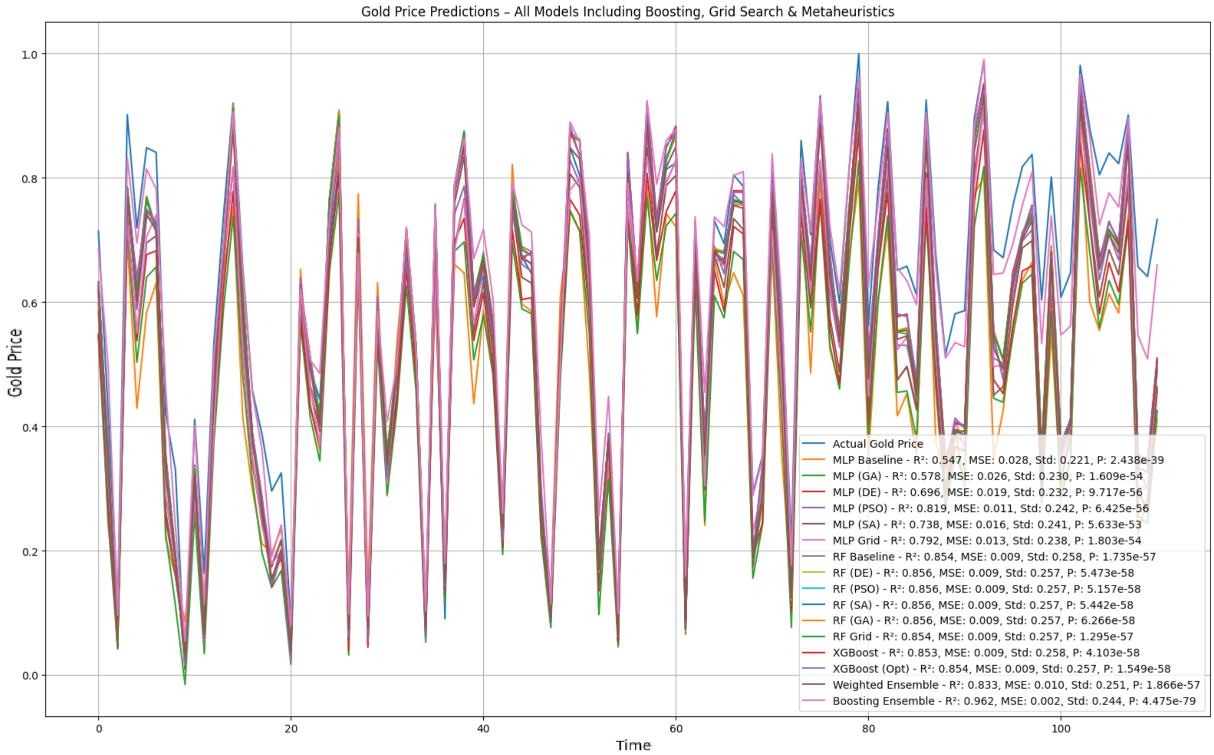


Fig. 8. Comparison of forecasts: predicted vs actual gold prices for all models.

Finally, Fig. 8 compares all models predicted values against actual gold prices. It visually confirms the superior accuracy of ensemble techniques, particularly the boosting approach.

E. Comparison with Previous Work

This study utilizes the same dataset as the research titled "Enhancing MLP using Archive-based Harris Hawks Optimizer to Predict Gold Prices" [23]. The referenced work focused on improving the performance of MLP by optimizing its parameters using the Archive-based HHO, resulting in significant performance gains over the baseline MLP, as presented in Table XII.

TABLE XII. COMPARISON WITH RESULTS FORM [23]

Method	R^2	RMSE
HHO-Optimized MLP (from [23])	0.8500	0.1020
Boosting Ensemble (This Study)	**0.9654**	**0.0433**

In contrast, the current study adopts a broader strategy by applying traditional machine learning models—RF, MLP, and XGBoost—and enhancing them using several metaheuristic optimization algorithms: PSO, DE, SA, and GA. Additionally, ensemble learning techniques, namely weighted and boosting ensembles, were implemented to further improve forecasting accuracy.

The findings demonstrate that simpler and well-established metaheuristic optimizers, when combined with traditional models, can outperform more complex approaches like HHO. In particular, the boosting ensemble model developed in this study achieved higher accuracy with a greater R^2 and lower RMSE than the HHO-optimized MLP. This highlights that high forecasting precision can be achieved through careful optimization and ensemble integration, without relying on newer or more complex algorithms.

F. Discussion

This study provides key insights into the effectiveness of various machine learning and optimization approaches for gold price prediction. The baseline analysis showed that traditional models such as RF and XGBoost already deliver strong performance, even without hyperparameter tuning.

Model performance improved substantially after applying heuristic and metaheuristic optimization algorithms. PSO consistently enhanced the accuracy of the RF model, while SA yielded the highest gains for MLP. These findings underscore the importance of selecting suitable optimization strategies tailored to the model architecture.

Ensemble learning further boosted forecasting performance. While the weighted ensemble stabilized predictions, the boosting ensemble significantly outperformed all other methods across evaluation metrics. It effectively combined the strengths of the best-optimized models (RF-PSO and XGBoost), resulting in the most accurate and reliable forecasts.

Notably, this research shows that well-established metaheuristic algorithms, when thoughtfully applied to traditional models, can exceed the performance of more complex and newer approaches such as the HHO. This supports the notion that practical and flexible techniques—rather than complexity alone—are key to achieving high accuracy in financial forecasting tasks.

V. Conclusion and Future Work

This study proposed a robust forecasting framework for gold price prediction by integrating traditional machine learning models with metaheuristic optimization and ensemble learning strategies. Three primary models, including RF, MLP, and XGBoost, were evaluated for their predictive capabilities. To enhance model performance, hyperparameter tuning was conducted using four metaheuristic algorithms: PSO, DE, SA, and GA. XGBoost, due to its inherent optimization mechanisms, was fine-tuned using Grid Search.

Performance evaluation relied on multiple statistical metrics including the coefficient of determination (R^2), MSE, RMSE, standard deviation, and p-value. The PSO-optimized RF model (RF-PSO) achieved the best performance among individual models with an R^2 of 0.8150 and RMSE of 0.1000. XGBoost showed consistent performance before and after optimization.

To further boost accuracy and stability, two ensemble methods were introduced. The weighted ensemble combined predictions from RF, MLP, and XGBoost using inverse RMSE-squared weighting. While this approach yielded stable forecasts, the boosting ensemble — formed by sequentially combining RF-PSO and XGBoost—achieved the best results, with an R^2 of 0.9654 and RMSE of 0.0433, outperforming all other models.

Residual and convergence analyses confirmed that metaheuristic optimization enhanced both accuracy and consistency. The proposed framework effectively modeled the complex and nonlinear behavior of gold prices, offering practical implications for financial analysts, policymakers, and investors. The study successfully met its objectives: assessing the performance of key machine learning models for gold price prediction, improving these models through optimization techniques, and designing ensemble strategies that leverage the strengths of individual models.

Despite these contributions, several constraints should be acknowledged. The dataset did not incorporate sentiment or geopolitical indicators, which are known to influence real-world gold price dynamics. The scope of optimization was limited to four metaheuristic algorithms, omitting potentially competitive alternatives such as Ant Colony Optimization, Firefly Algorithm, or Grey Wolf Optimizer. Similarly, advanced deep learning architectures like LSTM, GRU, or Transformer-based models were not investigated, which could offer additional advantages in modeling temporal dependencies. The framework lacked online learning or real-time retraining mechanisms, which are essential for adapting to sudden market changes. Furthermore, explainability techniques such as SHAP or LIME were not implemented, limiting interpretability for stakeholders in high-stakes decision-making environments.

Future research can address these limitations in several ways. First, integrating macroeconomic, sentiment-based, and qualitative geopolitical features could provide a richer representation of market conditions. In particular, incorporating multi-modal data sources—such as financial news, social media sentiment, and policy announcements—processed via NLP could capture market psychology and event-driven fluctuations more effectively. Second, expanding the optimization phase to include a broader range of metaheuristics or hybrid approaches may uncover more efficient hyperparameter configurations. Third, experimenting with advanced deep learning and hybrid architectures could improve the capacity to model long-range dependencies. Fourth, implementing adaptive, online learning mechanisms would allow the framework to adjust continuously to evolving market conditions. Finally, extending the evaluation framework to include scale-independent performance metrics such as the Relative Root Mean Square Error (RRMSE) would facilitate fairer comparisons across datasets and forecasting horizons, enhancing the generalizability and robustness of the proposed approach.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Alaa N. Sawalha conceptualized the study, designed the research framework, collected and preprocessed the dataset, and implemented the ensemble forecasting models. Mohammed A. Al-Betar supervised the entire research process, contributed to the methodological

design of the optimization algorithms, and guided the interpretation of experimental results. Sharif Naser Makhadmeh assisted in model evaluation, statistical analysis, and performance validation, and contributed to refining the manuscript. Alaa N. Sawalha wrote the initial draft of the paper under the supervision of Mohammed A. Al-Betar. All authors reviewed, revised, and approved the final version of the manuscript.

Funding

This work was supported by Ajman University.

References

[1] D. G. Baur and T. K. McDermott, "Is gold a safe haven? International evidence," *J. Bank. Finance*, vol. 34, no. 8, pp. 1886–1898, 2010.

[2] M. Hood and F. Malik, "Is gold the best hedge and a safe haven under changing stock market volatility?" *Rev. Financial Econ.*, vol. 22, no. 2, pp. 47–52, 2013.

[3] I. Yousef and E. Shehadeh, "The impact of COVID-19 on gold price volatility," *Int. J. Econ. Bus. Admin.*, vol. 8, no. 4, pp. 353–364, 2020.

[4] N. A. Zainal and Z. Mustaffa, "A literature review on gold price predictive techniques," in *Proc. 2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*, 2015, pp. 39–44.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Cambridge, MA, USA: MIT Press, 2016.

[6] N. M. Abdulkareem and A. M. Abdulazeez, "Machine learning classification based on random forest algorithm: A review," *Int. J. Sci. Bus.*, vol. 5, no. 2, pp. 128–142, 2021.

[7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794.

[8] J. Bergstra and B. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 2, pp. 281–305, 2012.

[9] X. Xu and Y. Zhang, "Individual time series and composite forecasting of the Chinese stock index," *Mach. Learn. Appl.*, vol. 5, 100035, 2021.

[10] S. Sun, Z. Cao, H. Zhu, and J. Zhao, "A survey of optimization methods from a machine learning perspective," *IEEE Trans. Cybern.*, vol. 49, no. 12, pp. 4473–4488, 2019.

[11] S. Figlewski, "Forecasting volatility," *Financial Markets, Institutions & Instruments*, vol. 6, no. 1, pp. 1–88, 1997.

[12] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed, OTexts, 2021.

[13] S. B. Jabeur, S. Mefteh-Wali, and J. L. Viviani, "Forecasting gold price with the XGBoost algorithm and SHAP interaction values," *Ann. Oper. Res.*, vol. 334, no. 1, pp. 679–699, 2024.

[14] G. Cohen and A. Aiche, "Forecasting gold price using machine learning methodologies," *Chaos, Solitons Fractals*, vol. 175, 114079, 2023.

[15] I. E. Livieris, E. Ioannis, and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting," *Neural Computing and Applications*, vol. 32, no. 23, 17351–17360, 2020.

[16] Z. Dong and Y. Zhou. "A novel hybrid model for financial forecasting based on CEEMDAN-SE and ARIMA-CNN-LSTM," *Mathematics*, vol. 12, no. 16, p. 2434, 2024.

[17] A. Vidal and W. Kristjanpoller, "Gold volatility prediction using a CNN-LSTM approach," *Expert Systems with Applications* vol. 157, 113481, 2020.

[18] M. Salim and A. Djunaidy, "Development of a cnn-lstm approach with images as time-series data representation for predicting gold prices," *Procedia Computer Science*, vol. 234, pp. 333–340, 2024.

[19] A. A. Heidari, S. Mirjalili, H. Faris *et al.*, "Harris hawks optimization: Algorithm and applications in financial forecasting," *Future Gener. Comput. Syst.*, vol. 97, pp. 849–872, 2019.

[20] Z. E. Mohamed, A. R. Ali, and W. Dabour, "Optimizing adaptive neuro-fuzzy inference system model based Chaotic Harris Hawks algorithm for stock prediction," *Scientific Reports*, vol. 15, no. 1, 32529, 2025.

[21] 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, e0298426, 2024.

[22] H. Hu, Y. Ao, Y. Bai, R. Cheng, and T. Xu, "An improved Harris's Hawks optimization for SAR target recognition and stock market index prediction," *IEEE Access*, vol. 8, pp. 65891–65910, 2020.

[23] I. Abu-Doush, B. Ahmed, M. A. Awadallah, M. A. Al-Betar, and A. R. Rababaah, "Enhancing multilayer perceptron neural network using archive-based Harris Hawks optimizer to predict gold prices," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 5, 101557, 2023.

[24] Y. Zhao, Y. Guo, and X. Wang, "Hybrid LSTM–Transformer architecture with multi-scale feature fusion for high-accuracy gold futures price forecasting," *Mathematics*, vol. 13, no. 10, p. 1551 2025.

[25] M. A. Rahman, C. Modak, M. A. S. Mozumder *et al.*, "Advancements in retail price optimization: Leveraging machine learning models for profitability and competitiveness," *J. Bus. Manag. Stud.*, vol. 6, no. 3, pp. 104–112, 2024.

[26] T. Franus, M. Marchese, and R. Payne, "A machine learning ensemble approach to predict financial markets manipulation," *Financial Fraud Market Manipul.*, vol. 2024, no. 7, pp. 1–15, 2024.

[27] G. Kirtika, R. J. Pandya, and S. Iyer. "Deep learning and natural language processing integrated gold price forecasting," in *Proc. 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, 2025, pp. 1–6.

[28] M. B. Triki and A. B. Maatoug, "The GOLD market as a safe haven against the stock market uncertainty: Evidence from geopolitical risk," *Resources Policy*, vol. 70, 101872, 2021.

[29] E. Jianwei, J. Ye, and H. Jin, "A novel hybrid model on the prediction of time series and its application for the gold price analysis and forecasting," *Physica A: Statistical Mechanics and its Applications* vol. 527, 121454, 2019.

[30] J. Iqbal, "Does gold hedge stock market, inflation and exchange rate risks? An econometric investigation," *International Review of Economics & Finance*, vol. 48, pp. 1–17, 2017.

[31] M. Akbar, F. Iqbal, and F. Noor, "Bayesian analysis of dynamic linkages among gold price, stock prices, exchange rate and interest rate in Pakistan," *Resources Policy*, vol. 62, pp. 154–164, 2019.

[32] M. Bouasabah, "A performance analysis of machine learning algorithms in stock market prediction, compared to traditional indicators," in *Proc. International Conference on Digital Age & Technological Advances for Sustainable Development*, 2024, pp. 264–271.

[33] M. Rashidi and M. Modarres. "Predicting the price of gold in the financial markets using hybrid models," arXiv preprint, arXiv:2505.01402, 2025.

[34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[35] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.