Multi-Horizon Gold Price Forecasting and Its Implications for Financial Markets

Asifa Iqbal [1,*], Marwa M. Eid[2,3]

[1]School of international languages Zhengzhou University, Henan, China
[2]Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 11152, Egypt
[3]Jadara Research Center, Jadara University, Irbid 21110, Jordan

Emails: asifaiqbal615@gmail.com; mmm@ieee.org

Abstract

Accurate forecasting of gold prices remains a critical challenge in financial markets due to the nonlinear, nonstationary, and regime-dependent nature of commodity price dynamics, particularly for gold quoted against the US dollar (XAU/USD), which plays a central role as a safe-haven asset, inflation hedge, and portfolio diversifier. Motivated by the growing limitations of traditional econometric and manually tuned machine learning approaches in handling long-horizon, multi-timeframe financial data, this study proposes a robust forecasting framework that integrates deep learning with metaheuristic optimization. The main contribution of this work lies in the systematic combination of a Deep Pyramid Recurrent Neural Network (DPRNN) with advanced metaheuristic algorithms for automated hyperparameter optimization, with particular emphasis on Greylag Goose Optimization (GGO), alongside other state-of-the-art optimizers. Using historical XAU/USD data spanning from 2004 to February 2025 across multiple temporal resolutions, baseline model evaluation demonstrates that DPRNN outperforms other deep learning architectures prior to optimization, achieving a Mean Squared Error (MSE) of 0.0589, Root Mean Squared Error (RMSE) of 0.2426, and coefficient of determination (R^2) of 0.79. Following optimization, the proposed GGO-optimized DPRNN framework yields a substantial performance enhancement, reducing the MSE to 2.05×10^{-5} and RMSE to 4.52×10^{-3}, while simultaneously increasing the correlation coefficient to 0.987 and R^2 to 0.983, with near-perfect agreement metrics reflected by a Nash–Sutcliffe Efficiency of 0.986 and Willmott Index of 0.988. These results confirm the effectiveness of GGO in navigating complex hyperparameter search spaces and significantly improving predictive accuracy and stability. From an economic and financial perspective, the findings underscore the practical value of metaheuristic-optimized deep learning models for enhancing gold price forecasting, supporting more informed investment decisions, improved risk management, and greater market efficiency in volatile and uncertain financial environments.

Keywords: Gold price forecasting; Financial time-series modeling; Safe-haven assets; Metaheuristic optimization; Deep learning in financial markets

1 Introduction

Gold, commonly represented in international financial markets through the XAU/USD exchange rate, has long occupied a pivotal position within the global economic and financial system [1], [2], [3]. Unlike conventional financial assets, gold is not directly linked to the creditworthiness of any issuing authority, which grants it a distinctive status as a store of value and a hedge against systemic risk [4], [5]. Historically, gold has served as a monetary anchor, a reserve asset for central banks, and a strategic instrument for wealth preservation [6],

[7]. Its unique economic characteristics have made it particularly attractive during periods of macroeconomic instability, financial crises, geopolitical uncertainty, and heightened inflationary pressures [8], [9]. As a result, gold prices reflect a complex interplay of monetary, financial, and behavioral factors, rendering their accurate prediction a subject of substantial academic and practical importance.

From a monetary and macroeconomic perspective, gold prices are closely associated with inflation expectations, real interest rates, and currency valuation, especially with respect to the United States dollar [10]. In environments characterized by expansionary monetary policy or declining real yields, gold is frequently perceived as a hedge against the erosion of purchasing power. Central banks and institutional investors often interpret gold price movements as an implicit signal of market confidence in fiat currencies and long-term price stability. Furthermore, gold plays a critical role in portfolio diversification strategies due to its historically low correlation with traditional asset classes such as equities and fixed-income securities, particularly during periods of market stress. Consequently, fluctuations in XAU/USD prices have direct implications for asset allocation, risk management, and long-term investment decision-making [11].

The increasing volatility and complexity of global financial markets have further intensified the need for accurate gold price forecasting [12]. Over the past two decades, financial markets have been shaped by repeated episodes of systemic disruption, including global financial crises, sovereign debt instability, pandemics, and abrupt shifts in monetary policy regimes [13]. These events have introduced nonlinear dynamics and regime-dependent behaviors that challenge conventional analytical frameworks. In such settings, price movements often exhibit heterogeneous patterns across different temporal horizons, underscoring the importance of multi-timeframe analysis [14]. Forecasting gold prices across intraday, daily, weekly, and monthly horizons is therefore essential for supporting diverse financial objectives, ranging from short-term trading and hedging to long-term strategic planning and policy analysis [15].

Despite extensive research on gold price modeling, traditional econometric and technical analysis methods exhibit notable limitations when applied to modern financial data. Linear econometric models often rely on restrictive assumptions regarding stationarity and linearity, which may not hold in the presence of structural breaks and evolving market conditions. Similarly, rule-based technical indicators are typically sensitive to noise and parameter selection, limiting their robustness and adaptability. These shortcomings become particularly pronounced when large-scale, high-frequency, and long-horizon datasets are considered simultaneously, as complex interactions across timeframes are difficult to capture using conventional techniques.

In response to these challenges, machine learning has emerged as a powerful paradigm for financial time-series forecasting. By leveraging advanced computational architectures and data-driven learning mechanisms, machine learning models are capable of capturing nonlinear relationships, temporal dependencies, and hidden structures that are difficult to model explicitly. Their ability to process high-dimensional inputs and adapt to evolving patterns has positioned them as promising tools for improving the reliability and efficiency of gold price prediction across multiple temporal scales.

Notwithstanding the potential advantages of machine learning methodologies, forecasting gold prices remains a highly complex task due to several intrinsic characteristics of financial time series. One of the most prominent challenges arises from the high dimensionality associated with multi-timeframe financial data. When Open, High, Low, Close, and Volume information is collected across multiple temporal resolutions, the resulting feature space expands rapidly, increasing computational burden and exacerbating the risk of overfitting. This issue is commonly referred to as the curse of dimensionality and can significantly impair model generalization if not properly addressed.

Feature redundancy constitutes an additional challenge in gold price forecasting. Financial indicators derived from price and volume data often exhibit strong correlations, both within and across timeframes. Such multicollinearity can obscure the true informational contribution of individual features, reduce learning efficiency, and complicate model interpretation. Without effective feature selection mechanisms, machine learning models may allocate excessive capacity to redundant or irrelevant information, thereby limiting their predictive effectiveness and increasing computational cost.

Another fundamental difficulty stems from the nonlinear and nonstationary nature of gold price dynamics. Financial markets evolve continuously in response to macroeconomic developments, monetary policy decisions, and shifts in investor sentiment. Sudden structural changes, such as those induced by financial crises or major policy interventions, can alter the underlying data-generating process. These characteristics challenge learning algorithms, which must balance adaptability to new patterns with stability across long historical horizons.

Deep learning models, while capable of modeling complex dependencies, introduce further challenges related to hyperparameter sensitivity. Model performance is often highly dependent on architectural and training-related parameters, including learning rates, network depth, and regularization strategies. Inappropriate hyperparameter configurations may lead to poor convergence, excessive variance, or biased forecasts. Given the complexity and scale of financial datasets, manual tuning becomes impractical, underscoring the need for automated and systematic optimization strategies.

Overfitting represents a persistent concern, particularly when long-span historical datasets are employed. While extensive historical coverage enables the inclusion of diverse market regimes, it also increases the likelihood that models will learn regime-specific noise rather than generalizable patterns. Ensuring robust performance across heterogeneous economic conditions, including periods of crisis, expansion, and policy transition, remains one of the most critical challenges in XAU/USD price prediction.

The primary objective of this study is to develop and rigorously evaluate machine learning-based forecasting models for predicting XAU/USD prices using long-term, multi-timeframe financial data. By systematically examining multiple learning architectures, the study aims to identify modeling strategies capable of capturing the complex temporal dependencies and nonlinear behaviors inherent in gold price dynamics.

A central focus of the study lies in the integration of metaheuristic optimization algorithms to enhance model performance through automated feature selection and hyperparameter optimization. Feature selection is employed to reduce dimensionality and eliminate redundant information, while hyperparameter optimization seeks to identify configurations that balance model flexibility and generalization. The combined application of these strategies is intended to improve learning efficiency, stability, and scalability within a unified modeling framework.

Through comprehensive model development and optimization, the study seeks to advance methodological practices in financial time-series forecasting and to provide a robust analytical foundation suitable for practical financial decision-making contexts.

This study contributes to the existing literature on financial forecasting by proposing a comprehensive machine learning framework tailored to the prediction of gold prices across multiple temporal resolutions. The framework emphasizes the systematic integration of metaheuristic optimization techniques with advanced learning models to address key challenges related to feature redundancy, hyperparameter sensitivity, and model generalization.

A distinguishing contribution of the study lies in the joint treatment of feature selection and hyperparameter optimization as complementary components of model enhancement rather than as isolated preprocessing steps. By embedding optimization mechanisms directly within the modeling pipeline, the proposed approach aims to improve both predictive stability and computational efficiency. Furthermore, the use of long-term, multi-timeframe financial data enables a thorough examination of model behavior across diverse economic environments, thereby strengthening the empirical relevance and applicability of the framework.

The remainder of this paper is organized to provide a coherent and systematic exposition of the proposed methodology and analytical framework. Following this introduction, the next section describes the dataset, data preprocessing procedures, machine learning models, and optimization techniques employed in the study. Subsequent sections present the experimental design and analytical evaluation, followed by a discussion that situates the findings within the broader context of financial forecasting research. The paper concludes with a summary of key insights and directions for future investigation.

2 Literature Review

Gold price forecasting has been extensively investigated in the literature due to gold's strategic importance as a safe-haven asset, an inflation hedge, and a core component of global financial markets. The XAU/USD exchange rate, in particular, acts as a high-frequency barometer of macro-financial stress because it embeds information about the U.S. dollar's strength, real interest rate expectations, liquidity conditions, and shifts in global risk sentiment. These characteristics make gold price dynamics intrinsically nonlinear, strongly regime-dependent, and sensitive to abrupt shocks, thereby challenging standard assumptions of stationarity and linearity that underlie many classical time-series models. Consequently, the recent forecasting literature increasingly prioritizes robust modeling pipelines that (i) capture long- and short-memory temporal dependencies, (ii) exploit multivariate information from macroeconomic and market indicators, and (iii) incorporate mechanisms for adaptivity and uncertainty quantification to support real-world decision-making under risk.

A major strand of the literature is the systematic advancement of deep learning architectures for daily gold price prediction, often coupled with hyperparameter optimization. One influential study develops a unified forecasting framework for XAU/USD based on multiple deep learning models—LSTM, GRU, CNN, TCN, and RNN—trained on a multi-feature dataset spanning seven years. A key methodological contribution is the explicit examination of input sequence (window) lengths and the use of Bayesian optimization, Genetic Algorithms, and Grey Wolf Optimization to tune model hyperparameters across several window sizes. The empirical results show that the GRU architecture, when optimized via Bayesian methods, delivers the strongest overall performance, suggesting that both the choice of temporal context and the optimization strategy materially affect generalization in gold price prediction [16]. Conceptually, this line of work indicates that gold price movements may be better learned by architectures that balance representational capacity and parameter efficiency, particularly when the forecasting horizon is short (next-day prediction) and the signal-to-noise ratio is modest.

A complementary literature evaluates whether deep learning models genuinely improve upon traditional econometric benchmarks. A detailed comparative study assesses ARIMA against RNN, LSTM, and GRU using daily global gold prices observed during a period of unusually high volatility and macroeconomic disruption. The findings indicate that ARIMA performs weakly, including negative explanatory power under some evaluation criteria, while recurrent deep learning models offer substantially improved accuracy. In particular, GRU achieves the best error statistics and higher R^2 relative to alternatives, reinforcing the argument that recurrent architectures can better accommodate nonlinear dependencies and volatility clustering that are pervasive in gold markets [17]. Methodologically, this comparative evidence supports the view that linear statistical models can serve as informative baselines but are insufficient as standalone solutions when gold price behavior deviates from quasi-linear dynamics due to geopolitical or monetary shocks.

Beyond architecture selection, an important theme concerns feature enrichment and the integration of macro-financial drivers. A study comparing LSTM, Bi-LSTM, and MLP incorporates variables such as economic indicators, interest rates, currency exchange rates, and news emotion, thereby recognizing that gold prices respond to a complex information set rather than to their own lag structure alone. The reported results favor LSTM, indicating that gated recurrent memory mechanisms provide a practical advantage when fusing heterogeneous predictors into a unified temporal representation [18]. Importantly, this work also signals that, although bidirectionality can be beneficial in offline sequence labeling tasks, it may be less advantageous in forecasting settings where causal ordering and forward-only information constraints dominate deployment realities.

Country- and market-specific studies highlight how local economic structures interact with global forces to shape gold price behavior. In Uzbekistan, where gold production and exports are economically pivotal, researchers compare SARIMA, Random Forest, and a hybrid SARIMA–Random Forest model. The hybrid approach demonstrates superior predictability, illustrating how a statistical component can capture structured seasonality and persistence, while the machine learning component models nonlinear interactions among economic indicators and market variables [19]. This suggests that, in emerging-market contexts, hybrid models may be particularly effective because they can represent both policy-driven linearities (e.g., regulated exchange-rate effects) and market-driven nonlinearities (e.g., sudden demand shifts and external shocks).

A further maturation in the literature is the explicit emphasis on interpretability and model transparency, especially given the high-stakes nature of financial forecasting. The Temporal Fusion Transformer (TFT) study applied to the Indian gold market adopts an end-to-end preprocessing and tuning workflow, benchmarks against classical and neural baselines, and then leverages attention mechanisms and variable importance analysis to explain model behavior. The results indicate that TFT not only improves predictive performance relative to ARIMA, linear regression, and LSTM, but also provides an interpretable account of how macroeconomic indicators such as CPI and exchange rates contribute to predictions over time [20]. From an applied standpoint, this interpretability is valuable because it supports scenario analysis, policy assessment, and model governance, enabling stakeholders to scrutinize whether forecasts align with plausible economic narratives rather than merely minimizing error ex post.

Ensemble learning constitutes another major research direction, motivated by the observation that individual models often specialize in different regimes (trend, range-bound, high-volatility). A dynamic weighted ensemble combining deep and hybrid components—LSTM, GRU-LSTM, CNN-LSTM, Attention-LSTM, and XGBoost-LSTM—demonstrates that weighting learners based on reciprocal errors can improve stability and generalization across currency forecasting tasks, which are conceptually close to gold-related forex dynamics [21]. The broader implication for gold forecasting is that ensemble strategies may reduce model risk by diversifying across inductive biases: convolutional modules can exploit local patterns, gradient-boosted components can capture nonlinear feature interactions, and attention mechanisms can adaptively focus on salient timesteps and signals.

Hybrid deep learning architectures are also widely explored in gold-specific settings, where combining convolutional feature extraction with recurrent temporal modeling is often advantageous. In the Iranian market, a comparison between GRU and CNN–LSTM for 18-carat gold price forecasting finds that CNN–LSTM yields lower prediction errors, supporting the hypothesis that convolutional layers can efficiently extract robust short-horizon representations (e.g., local trends and micro-patterns) before recurrent layers model longer dependencies [22]. Relatedly, a CNN–BiGRU hybrid framework is proposed for gold forecasting and reported to outperform competing hybrid and standalone models across a broad error metric set, indicating that bidirectional gated recurrence may further enhance representation learning in contexts where training can exploit richer temporal context, even if deployment remains forward-looking [23]. Collectively, these hybrid deep networks reflect a design pattern in the literature: use convolution for denoising and local pattern extraction, and use gated recurrence to integrate information across time.

Long-horizon and regime-shift resilience is a key focus in settings characterized by persistent instability, as exemplified by research on Turkey's gold and foreign exchange markets. A hybrid approach integrating ARIMA with LSTM/GRU is constructed using a multi-decade dataset and a walk-forward validation strategy, enabling the model to update iteratively as new observations arrive. The reported findings suggest that the hybrid model adapts effectively to abrupt market transitions and offers strong performance in both short- and long-term horizons [24]. This work is conceptually aligned with the argument that forecasting pipelines should not only optimize static train-test accuracy but also implement evaluation schemes that approximate operational deployment, where models face distribution shift and must remain calibrated over time.

A more decision-centric and risk-aware line of work introduces uncertainty estimation into algorithmic trading frameworks. A study integrating Conformal Prediction into an automated trading bot for XAU/USD demonstrates that prediction intervals can inform trade filtering, reducing exposure during uncertain market states and improving the reliability of trading signals [25]. This contribution is significant because it reframes forecasting accuracy as only one objective among several; in practice, the ability to quantify predictive uncertainty can be equally important for controlling downside risk, avoiding overtrading, and improving robustness under sudden volatility spikes.

Another direction expands gold forecasting beyond purely price-based inputs by incorporating multimodal information such as news content and sentiment. A multimodal regression framework uses concept-based news representations, mood signals from specialized groups, and technical indicators, and employs recurrent convolutional and recurrent layers to model temporal dynamics across modalities. By explicitly addressing non-stationarity—via learning how predicted values drift around local averages—the model achieves substantial error reduction relative to baselines [26]. The methodological implication is that gold price movements may be driven not only by lagged price dynamics and macroeconomic variables, but also by information diffusion processes that unfold at different speeds across news topics, thereby requiring architectures designed for heterogeneous temporal response patterns.

Interdisciplinary methods rooted in chaos theory and econophysics further highlight the complexity of gold-related market dynamics and motivate richer temporal representations. Phase Space Reconstruction combined with a Phase Space LSTM framework is used to analyze interactions among Bitcoin, gold, major equity indices, bonds, forex, and commodities, revealing deterministic chaos and hidden dependencies. The resulting improvements in forecasting accuracy and risk-adjusted evaluation emphasize that gold dynamics may be intertwined with broader systemic financial conditions rather than operating as an isolated market [27]. Such findings strengthen the rationale for multivariate forecasting setups that treat gold as part of a coupled financial system.

Even studies not directly centered on gold contribute relevant methodological insights, especially regarding model robustness and feature extraction in financial time series. An ensemble approach for forex trend prediction combining CNN-based feature extraction with bagging and boosting decision trees demonstrates high classification performance on GBP/JPY, underscoring the effectiveness of deep feature learning paired with ensemble aggregation in noisy, high-frequency environments [28]. While the task is directional classification rather than regression, the underlying principle—leveraging CNNs to transform raw market features into more separable representations—has clear parallels in gold-related forecasting and trading systems.

Chaotic dynamics are also incorporated directly into neural network training strategies. One study enhances forecasting performance by initializing GRU and Echo State Network (ESN) parameters with chaotic sequences, improving generalization across multiple financial datasets including XAU/USD. The reported results suggest that chaos-informed initialization can diversify internal dynamics and help models avoid poor local minima or overly smooth representations, which can be particularly beneficial in markets exhibiting sensitive dependence on initial conditions [29]. This connects naturally with broader discussions in the literature about volatility clustering, regime switching, and the limits of purely linear assumptions.

Finally, methodological transfer from other commodity markets underscores the importance of adaptivity, feature relevance, and evaluation under evolving conditions. A soybean price forecasting study integrates external indicators, employs adaptive feature selection (e.g., PCA and L1 regularization), and uses rolling updates to adapt to changing regimes. Although Gradient Boosting outperforms deep learning baselines in that setting, the broader lesson is that no single model family universally dominates across commodities; rather, performance depends on data availability, noise structure, explanatory variables, and the degree of nonlinearity. The emphasis on rolling adaptation and interpretability provides practical guidance for gold forecasting systems operating under distribution shift [30].

Overall, the reviewed literature—spanning 15 studies—demonstrates a clear shift from purely statistical forecasting toward hybrid, ensemble, interpretable, and uncertainty-aware machine learning frameworks. Across diverse markets and datasets, the evidence consistently suggests that gold price prediction benefits from (i) architectures capable of learning nonlinear temporal dependencies (e.g., LSTM/GRU), (ii) hybridization that combines linear structure with nonlinear learning (e.g., ARIMA–LSTM/GRU, SARIMA–RF), (iii) multimodal and macroeconomic feature integration (e.g., news, sentiment, CPI, exchange rates), (iv) interpretability mechanisms (e.g., TFT attention and variable importance), and (v) uncertainty quantification for risk-sensitive deployment (e.g., conformal prediction). Taken together, these directions form a coherent foundation for advanced research in XAU/USD forecasting and the development of decision-support tools for trading, hedging, and macro-financial analysis.

3 Materials and Methods

3.1 Dataset Description

The empirical analysis conducted in this study is based on an extensive historical dataset of XAU/USD prices, which represents the value of gold quoted in United States dollars within international financial markets. The dataset spans from June 2004 to February 2025, providing more than twenty years of continuous market observations. This prolonged temporal coverage is particularly valuable for financial modeling, as it encompasses multiple economic cycles, shifts in monetary policy regimes, and episodes of pronounced market turbulence. Over this period, gold markets have been influenced by a wide range of macroeconomic forces, including inflationary and deflationary environments, financial crises, geopolitical tensions, and changes in global liquidity conditions. As such, the dataset offers a rich empirical foundation for developing forecasting models that are resilient to structural changes and evolving market dynamics.

A key feature of the dataset is its multi-timeframe structure, which captures gold price movements across both high-frequency and low-frequency temporal horizons. Intraday market behavior is represented through observations recorded at 5-minute, 15-minute, 30-minute, 1-hour, and 4-hour intervals, allowing the analysis to incorporate short-term price fluctuations and microstructural effects. In parallel, longer-term dynamics are captured through daily, weekly, and monthly data, which reflect broader macroeconomic trends and strategic investment behavior. This multi-resolution design acknowledges the fact that gold price formation mechanisms differ across time horizons, with short-term prices often driven by liquidity and order flow, while long-term prices are more closely linked to macroeconomic fundamentals and investor expectations.

Across all timeframes, the dataset maintains a consistent set of financial attributes derived from standard market price information. Each observation includes a timestamp and corresponding calendar date, ensuring precise temporal ordering and facilitating alignment across different frequencies. The core price variables consist of the Open, High, Low, and Close prices, expressed in United States dollars, which collectively describe the intraperiod price range and final settlement price. These variables form the foundation of most financial time-series analyses and are widely used in both academic research and professional market practice. In addition to price information, tick volume is included as a proxy measure of trading activity and market participation. Although tick volume does not represent actual traded volume in decentralized markets, it has been shown to capture meaningful information regarding market intensity and liquidity conditions.

The target variable in this study is defined with reference to the gold closing price, which serves as a benchmark indicator in financial valuation, portfolio assessment, and forecasting applications. The closing price is particularly relevant because it reflects the consensus market valuation at the end of each trading interval and is commonly used in the computation of returns, risk measures, and performance metrics. Depending on the modeling configuration, the forecasting task may focus on predicting the future closing price directly or on modeling the directional movement of prices over a specified horizon. Both formulations are economically meaningful, as they support decision-making in contexts such as trading, hedging, and strategic asset allocation.

Importantly, the dataset spans a wide range of economic and financial regimes, including periods of relative stability, systemic crises, and transitional phases associated with changes in monetary policy. This diversity enhances the suitability of the data for training machine learning models that aim to generalize beyond specific historical episodes. By exposing models to heterogeneous market conditions, the study seeks to mitigate the risk of regime-dependent overfitting and to improve robustness across different economic environments.

For model development and evaluation, the dataset is partitioned into distinct training, validation, and testing subsets. The training subset is used to estimate model parameters and learn underlying data patterns, while the validation subset supports model selection and optimization processes. The testing subset is reserved exclusively for out-of-sample assessment, ensuring that model evaluation reflects genuine predictive capability rather than in-sample fitting. This partitioning strategy adheres to established best practices in financial machine learning and is designed to preserve temporal integrity while preventing information leakage across datasets.

Understanding the intraday dynamics of market activity is essential for analyzing liquidity patterns, volatility transmission, and trader behavior across different trading sessions. Market participation is not uniformly distributed throughout the day; instead, it exhibits pronounced temporal heterogeneity driven by overlapping trading hours, information releases, and institutional trading strategies. To provide an empirical illustration of these intraday variations, Figure 1 presents the distribution of trading volume across the 24-hour trading cycle. The visualization highlights substantial differences in both central tendency and dispersion of volume at different hours, offering valuable insights into periods of heightened market engagement and elevated liquidity risk.

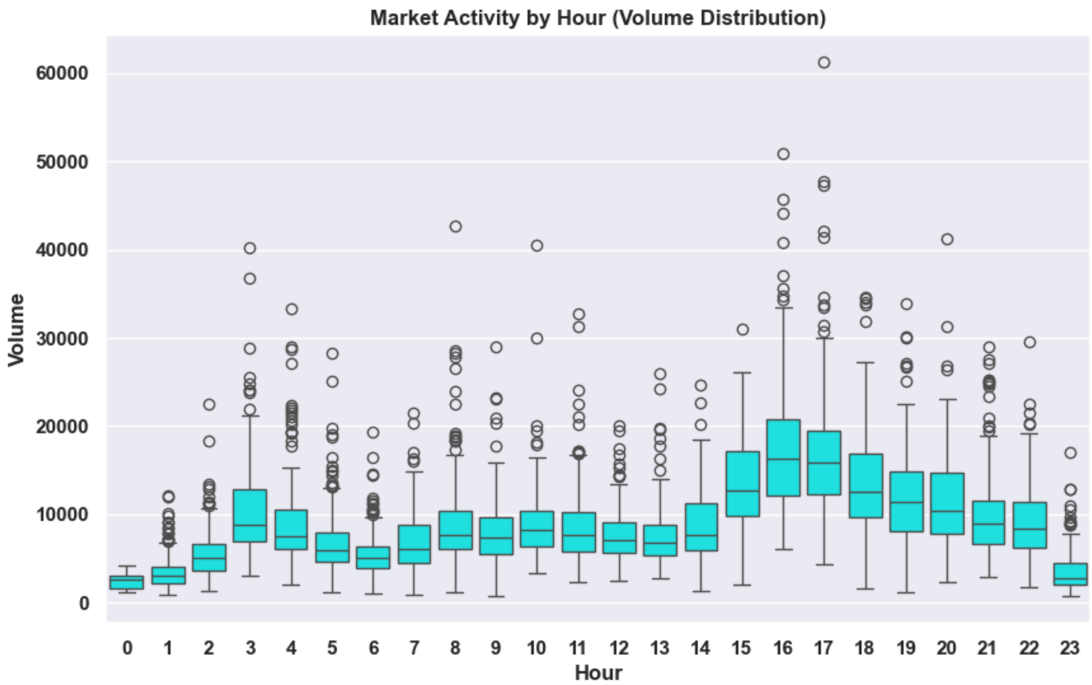

Figure 1: Market Activity by Hour (Volume Distribution).

A comprehensive understanding of asset return behavior requires examining both their distributional characteristics and their temporal variation across trading days. Financial return series are well known for exhibiting non-normal features such as leptokurtosis, asymmetry, and clustering, which have important implications for risk assessment and predictive modeling. To capture these stylized facts from multiple statistical perspectives, Figure 2 provides an integrated visualization framework that combines kernel density estimation, swarm plots, violin plots, and box plots. Together, these graphical representations offer complementary insights into the overall density of returns, cross-day dispersion, central tendency, and the presence of extreme observations, thereby enabling a nuanced assessment of return dynamics and day-of-week effects.

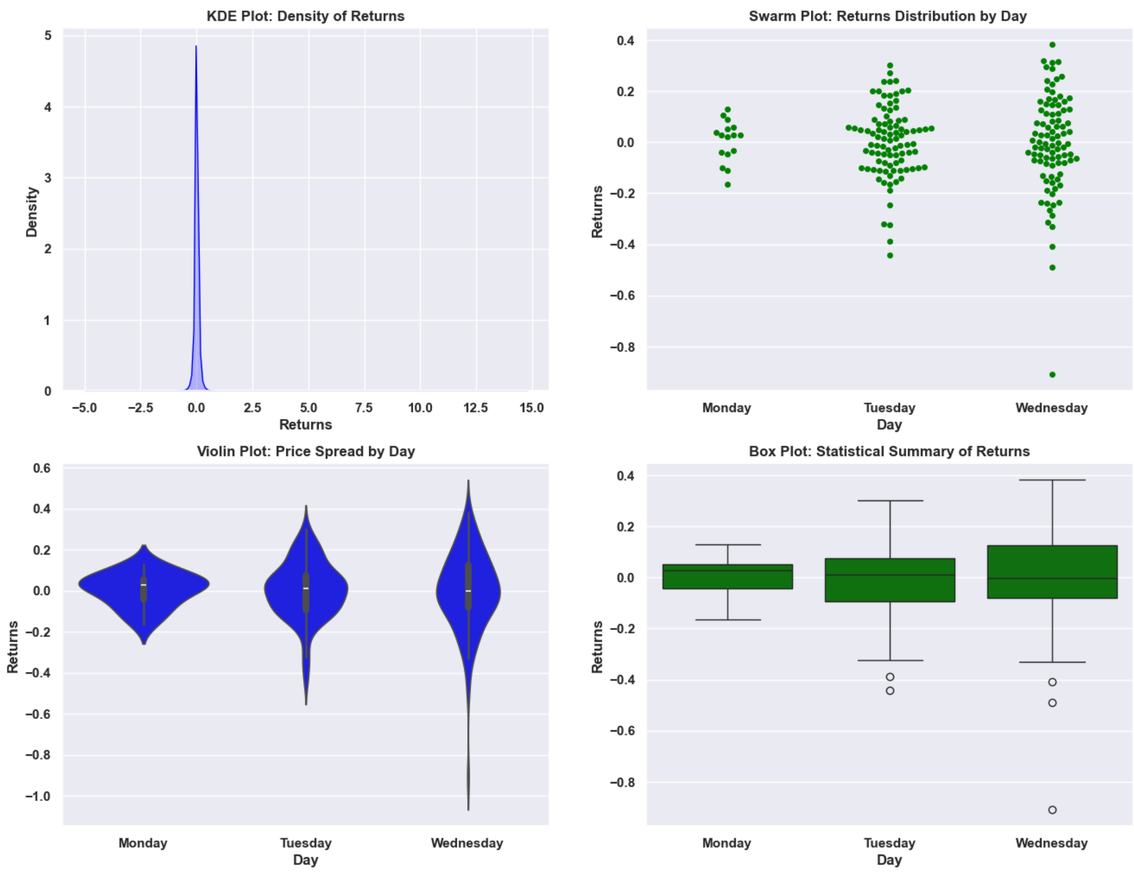

Figure 2: Statistical and Distributional Analysis of Asset Returns.

Examining the relationship between opening and closing prices is fundamental for understanding intraday price formation, market efficiency, and short-term price dynamics in financial markets. The degree of alignment between these two price points reflects the extent to which new information is incorporated during the trading session and provides insights into intraday volatility and directional persistence. As illustrated in Figure 3, the scatter distribution of opening and closing prices reveals a strong linear association, suggesting high price continuity and limited intraday price dislocation over the observed period. Such behavior is indicative of relatively stable trading conditions and supports the use of opening prices as informative predictors in intraday and end-of-day forecasting frameworks.

3.2 Data Preprocessing

Prior to model training, a comprehensive data preprocessing pipeline is implemented to ensure data quality, numerical stability, and compatibility with deep learning architectures. Given the scale and complexity of the multi-timeframe financial dataset, preprocessing plays a critical role in facilitating effective learning and reliable forecasting performance.

The first stage of preprocessing addresses the presence of missing values and irregularities in the time series. Long-horizon financial datasets frequently contain gaps arising from market closures, holidays, or data acquisition inconsistencies. Such irregularities, if left untreated, can distort temporal dependencies and degrade model performance. Appropriate handling strategies are therefore employed to maintain continuity of the time series while preserving its underlying statistical characteristics.

A crucial preprocessing step involves the alignment of timestamps across multiple timeframes. Since observations are recorded at different temporal resolutions, careful synchronization is required to ensure that features derived from various timeframes correspond to the correct chronological context. This alignment

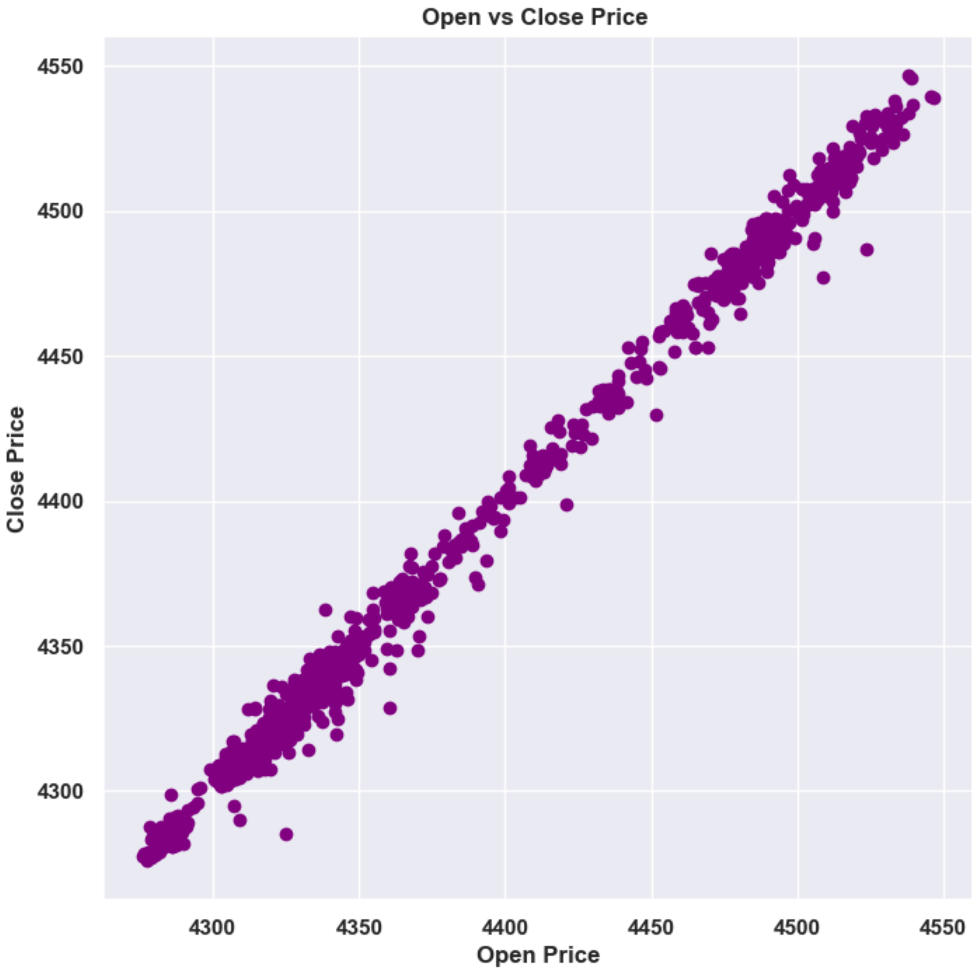

Figure 3: Relationship Between Opening and Closing Prices.

process is essential when integrating information from intraday and longer-term data sources, as misalignment could introduce spurious relationships or lead to information leakage. By enforcing strict temporal consistency, the preprocessing pipeline ensures that all model inputs reflect information available up to the forecasting point.

Feature scaling and normalization are applied to address disparities in the numerical ranges of price and volume variables. Financial time-series data often exhibit large variations in magnitude, particularly when long historical periods are considered. Such variations can hinder gradient-based optimization and slow model convergence. Through appropriate scaling techniques, input features are transformed to comparable ranges, thereby improving numerical stability and enhancing the efficiency of the learning process.

To capture temporal dependencies and dynamic market behavior, lagged features and rolling window representations are constructed from the original price and volume series. These transformations enable the models to incorporate historical context and to learn patterns related to momentum, volatility persistence, and trend evolution. Rolling window features, in particular, allow the extraction of localized statistical information that reflects recent market conditions, which is especially relevant in nonstationary financial environments.

Given the extensive use of overlapping price indicators and multiple timeframes, correlation analysis is conducted to identify redundant and highly correlated features. Excessive multicollinearity can reduce learning efficiency, increase computational burden, and complicate model interpretation. By systematically analyzing feature correlations, the preprocessing stage supports dimensionality reduction and prepares the data for subsequent feature selection procedures.

Finally, the processed data are organized into structured input formats suitable for deep learning models. This includes reshaping time-series data into sequential representations and ensuring consistency with the architectural requirements of the learning algorithms employed in the study. Through this carefully designed preprocessing pipeline, the raw financial data are transformed into a form that supports robust, scalable, and economically meaningful machine learning-based forecasting of XAU/USD prices.

3.3 Deep Learning Models

The selection of deep learning models for financial time-series forecasting requires careful consideration of the unique statistical and structural properties of market data. Financial time series, particularly those associated with commodity markets such as gold, exhibit pronounced nonlinearity, temporal dependence, volatility clustering, and regime shifts driven by macroeconomic and geopolitical factors. These characteristics necessitate modeling approaches that are capable of learning complex sequential patterns while remaining robust to noise and nonstationarity. In the context of XAU/USD price forecasting, the chosen models must effectively process multi-timeframe inputs, accommodate high-dimensional feature spaces, and generalize across heterogeneous market conditions.

Model selection in this study is guided by both theoretical suitability and empirical relevance within the financial forecasting literature. Priority is given to architectures that have demonstrated strong capabilities in sequential learning, representation extraction, and nonlinear function approximation. Rather than relying on a single modeling paradigm, a diverse set of deep learning models is employed to reflect different inductive biases and learning mechanisms. This diversity enables a comprehensive examination of how alternative architectures capture gold price dynamics and provides a robust baseline for subsequent optimization procedures.

The Deep Pyramid Recurrent Neural Network (DPRNN) is included due to its hierarchical recurrent structure, which is designed to capture temporal dependencies at multiple levels of abstraction. By organizing recurrent layers in a pyramidal fashion, the DPRNN progressively reduces temporal resolution while retaining essential sequential information. This hierarchical compression allows the model to integrate short-term fluctuations with longer-term trends, making it particularly suitable for multi-timeframe financial data. In the context of gold price forecasting, such an architecture facilitates the simultaneous modeling of intraday volatility and broader market movements. However, the depth and structural complexity of the DPRNN can increase sensitivity to architectural design choices and training dynamics, highlighting the importance of systematic configuration and optimization.

The Variational Autoencoder (VAE) is employed as a probabilistic learning framework capable of extracting latent representations from financial time series. By modeling the underlying data distribution through a latent variable structure, the VAE enables dimensionality reduction and noise filtering in a principled manner. In financial applications, these latent representations can capture hidden market states and structural patterns that are not immediately observable in raw price data. This characteristic is particularly valuable in gold markets, where prices often reflect unobserved macroeconomic expectations and investor sentiment. Nonetheless, because the primary objective of the VAE is reconstruction rather than direct prediction, its application to forecasting tasks involves a trade-off between representational richness and deterministic forecasting accuracy.

The Encoder–Attention LSTM (EA-LSTM) extends conventional recurrent architectures by integrating an attention mechanism that dynamically weights different segments of the input sequence. This mechanism allows the model to focus selectively on temporally relevant information, thereby enhancing its ability to capture salient market events and structural changes. In the context of XAU/USD forecasting, attention mechanisms are especially useful for identifying periods of heightened informational content, such as rapid price movements or shifts in volatility. Additionally, attention-enhanced models offer improved interpretability by providing insight into which temporal features contribute most significantly to the learning process. However, the increased architectural complexity of EA-LSTM models can lead to higher computational demands and greater sensitivity to hyperparameter settings.

Graph Neural Networks (GNNs) are incorporated to explore the modeling of structured dependencies that extend beyond linear temporal sequences. Although financial time series are inherently ordered in time, relationships among features, timeframes, or derived indicators can exhibit complex interdependencies that are naturally represented in graph form. GNNs enable the propagation and aggregation of information across interconnected nodes, allowing the model to learn relational patterns that may not be captured by purely sequential architectures. In gold price forecasting, such relational modeling can provide a flexible framework for integrating heterogeneous sources of information. However, the effectiveness of GNNs is closely tied to the quality of graph construction, and inappropriate structural assumptions may limit their practical utility.

Convolutional Neural Networks (CNNs) are included due to their demonstrated effectiveness in extracting local patterns and hierarchical representations from structured data. When applied to time-series inputs, convolutional filters can identify recurring short-term motifs, such as local trends, reversals, or volatility bursts. These localized patterns often carry important information for financial forecasting, particularly in high-frequency market environments. CNNs also offer advantages in terms of parameter efficiency and parallel computation, which can be beneficial when processing large-scale financial datasets. Nevertheless, the inherently local receptive fields of CNNs may constrain their ability to capture long-range temporal dependencies unless combined with additional architectural components.

Bidirectional Recurrent Neural Networks (BiRNNs) are employed to leverage contextual information from both past and future directions within a sequence. By processing time-series data in forward and backward order, BiRNNs generate richer sequence representations that incorporate surrounding temporal context. This bidirectional perspective can enhance pattern recognition and improve the modeling of symmetric temporal structures. In financial applications, such representations can be useful for retrospective analysis and sequence modeling. However, from an operational forecasting standpoint, the reliance on future information imposes practical limitations, as future observations are not available at prediction time. Consequently, the applicability of BiRNNs in real-time financial decision-making requires careful consideration.

Taken together, the deep learning models selected in this study represent a broad spectrum of architectural approaches to financial time-series forecasting. Each model exhibits distinct strengths in terms of temporal modeling, feature representation, and computational characteristics, while also presenting specific limitations when applied to the complex dynamics of gold markets. By establishing this diverse set of baseline models, the study lays a rigorous foundation for subsequent enhancements through advanced optimization and feature selection strategies, thereby supporting a systematic investigation of improved forecasting methodologies for XAU/USD prices.

3.4 Metaheuristic Algorithms

The rapid growth in the complexity of deep learning architectures has significantly increased the difficulty of configuring models for financial time-series forecasting tasks. In markets such as gold, where price dynamics are driven by nonlinear interactions among macroeconomic conditions, investor behavior, and liquidity fluctuations, the performance of forecasting models is highly sensitive to hyperparameter settings. These hyperparameters govern essential aspects of the learning process, including convergence speed, stability, and generalization capability. As a result, conventional trial-and-error or grid-based tuning approaches are no longer sufficient for identifying optimal configurations, particularly when large-scale, multi-timeframe datasets are involved. Within this context, metaheuristic optimization algorithms provide a flexible and efficient framework for automating the hyperparameter tuning process in deep learning-based financial forecasting systems.

Metaheuristic algorithms are designed to explore complex and high-dimensional search spaces using stochastic and population-based strategies. Unlike gradient-based optimization methods, metaheuristics do not rely on differentiability assumptions and are therefore well suited for optimizing objective functions that are noisy, nonconvex, or computationally expensive to evaluate. These characteristics are common in financial forecasting problems, where model evaluation often depends on sequential learning processes and where small parameter changes can lead to substantial differences in predictive behavior. By iteratively refining candidate solutions through adaptive search mechanisms, metaheuristics enable the systematic discovery of high-quality hyperparameter configurations that enhance model robustness and learning efficiency.

3.4.1 Role of Metaheuristics in Hyperparameter Optimization

In deep learning-based financial forecasting, hyperparameter optimization plays a critical role in shaping the bias–variance characteristics of predictive models. Hyperparameters such as learning rates, network depth, and regularization controls directly influence the trade-off between model flexibility and generalization. An overly flexible model may closely fit historical gold price movements but fail to generalize to new market conditions, while an overly constrained model may underrepresent the complex dynamics inherent in financial data. Metaheuristic algorithms address this challenge by enabling adaptive exploration of hyperparameter spaces, allowing the learning process to balance bias and variance in a data-driven manner.

Automated hyperparameter search is particularly important in financial contexts where datasets span multiple economic regimes and structural changes. Manual tuning strategies often rely on subjective judgment and limited experimentation, which may lead to suboptimal configurations that perform well only under specific conditions. Metaheuristic optimization mitigates this risk by maintaining a population of candidate solutions and continuously updating them based on performance feedback. Through mechanisms such as selection, mutation, and information sharing, these algorithms can escape local optima and adapt to the underlying structure of the optimization landscape.

Computational efficiency is another crucial consideration in large-scale financial datasets. Deep learning models trained on multi-timeframe XAU/USD data require substantial computational resources, as each training cycle involves processing long sequences and high-dimensional inputs. Exhaustive search methods quickly become infeasible as the number of hyperparameters increases. Metaheuristic algorithms offer a computationally efficient alternative by focusing search efforts on promising regions of the parameter space and reducing redundant evaluations. This efficiency enables the practical deployment of automated optimization strategies even when computational budgets are constrained, thereby supporting scalable and reproducible financial forecasting workflows.

3.5 Greylag Goose Optimization (GGO) Algorithm

The Greylag Goose Optimization (GGO) algorithm is a population-based metaheuristic inspired by the collective foraging, migration, and vigilance behavior of greylag geese. The algorithm models the adaptive balance between exploration and exploitation observed in goose flocks, where individuals dynamically switch between wide-area searching and focused local refinement depending on environmental conditions and collective feedback. This balance is particularly well suited for hyperparameter optimization in deep learning-based financial forecasting, where the optimization landscape is nonlinear, multimodal, and sensitive to small parameter perturbations.

In GGO, each candidate solution represents a search agent, denoted as $\mathbf{X}_i$, within the population of size n. The agents collectively explore the search space over a predefined number of iterations $t_{\max}$ with the objective of minimizing or maximizing a fitness function F_n, depending on the optimization goal. At each iteration, the population is divided into two adaptive groups: an exploration group (n_1) responsible for global search and an exploitation group (n_2) dedicated to local refinement around promising regions. The best-performing agent at iteration t is denoted as $\mathbf{X}^*(t)$ and serves as a reference point for guided movements.

Exploration Phase The exploration mechanism in GGO is designed to encourage diversity and global search during the optimization process. Exploration updates are applied primarily to agents in the exploration group when the iteration index satisfies the condition $t \bmod 2 = 0$. In this phase, agent movements are governed by adaptive coefficients $\mathbf{A}$ and $\mathbf{C}$, which control step size and directional influence.

When the absolute value of $\mathbf{A}$ is less than one, agents are guided toward the current best solution according to

$$\mathbf{X}(t+1) = \mathbf{X}^*(t) - \mathbf{A} \cdot \left| \mathbf{C} \cdot \mathbf{X}^*(t) - \mathbf{X}(t) \right|, \tag{1}$$

which promotes directed exploration around promising regions while maintaining stochastic variability.

When $|\mathbf{A}| \geq 1$, exploration is intensified through interactions among randomly selected agents, denoted as $\mathbf{X}_{Paddle1}$, $\mathbf{X}_{Paddle2}$, and $\mathbf{X}_{Paddle3}$. A time-dependent control parameter $\mathbf{z}$ is computed as

$$\mathbf{z} = 1 - \left(\frac{t}{t_{\max}} \right)^2, \tag{2}$$

which gradually shifts the algorithm from exploration toward exploitation as iterations progress. The agent position is then updated using a weighted combination of neighboring solutions:

$$\mathbf{X}(t+1) = w_1 \mathbf{X}_{Paddle1} + \mathbf{z} w_2 \left(\mathbf{X}_{Paddle2} - \mathbf{X}_{Paddle3} \right) + (1 - \mathbf{z}) w_3 \left(\mathbf{X} - \mathbf{X}_{Paddle1} \right). \tag{3}$$

This formulation enhances population diversity and enables long-range movements across the search space, which is essential for avoiding premature convergence in complex optimization landscapes.

Additionally, when stochastic control parameters satisfy alternative conditions, exploration is performed using a spiral-shaped movement inspired by goose flock dynamics:

$$\mathbf{X}(t+1) = w_4 \left| \mathbf{X}^*(t) - \mathbf{X}(t) \right| e^{bl} \cos(2\pi l) + 2 w_1 (r_4 + r_5) \mathbf{X}^*(t), \tag{4}$$

where b and l regulate the spiral amplitude and frequency. This mechanism allows agents to explore the vicinity of elite solutions while preserving stochastic motion.

Exploitation Phase The exploitation phase focuses on intensifying the search around high-quality solutions to refine convergence. Agents assigned to the exploitation group update their positions using multiple reference points, denoted as sentry agents $\mathbf{X}_{Sentry1}$, $\mathbf{X}_{Sentry2}$, and $\mathbf{X}_{Sentry3}$. Candidate positions are computed as

$$\mathbf{X}_k = \mathbf{X}_{Sentryk} - \mathbf{A}_k \left| \mathbf{C}_k \cdot \mathbf{X}_{Sentryk} - \mathbf{X} \right|, \quad k \in \{1, 2, 3\}, \tag{5}$$

and the final position update is obtained by averaging these candidates:

$$\mathbf{X}(t+1) = \left. \overline{\mathbf{X}_i} \right|_0^3. \tag{6}$$

This strategy enables localized refinement around multiple elite solutions, enhancing stability and convergence accuracy.

In cases where alternative conditions are met, exploitation agents update their positions according to

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \mathbf{D}(1 + \mathbf{z}) w \left(\mathbf{X} - \mathbf{X}_{Flock1} \right), \tag{7}$$

which models coordinated flock movement and reinforces convergence toward optimal regions.

A distinctive feature of GGO is its adaptive adjustment of exploration and exploitation group sizes. When the best fitness value remains unchanged for consecutive iterations, the algorithm increases the number of exploration agents while reducing the exploitation group. This adaptive mechanism enhances global search capability in stagnation scenarios and improves robustness when optimizing deep learning hyperparameters for nonstationary financial datasets.

Through the integration of these exploration and exploitation strategies, GGO provides a flexible and adaptive optimization framework capable of efficiently navigating complex, high-dimensional search spaces. This makes it particularly suitable for tuning deep learning models applied to financial time-series forecasting tasks such as XAU/USD price prediction.

3.5.1 State-of-the-Art Metaheuristic Optimization Algorithms

To ensure a comprehensive evaluation of optimization strategies, this study employs a diverse set of state-of-the-art metaheuristic algorithms, each characterized by distinct search behaviors and adaptive mechanisms. The use of multiple optimizers allows for the assessment of how different exploration and exploitation strategies influence hyperparameter tuning outcomes in financial forecasting applications.

The Genetic Algorithm (GA) is grounded in evolutionary theory and simulates the process of natural selection through iterative cycles of reproduction, crossover, and mutation. Candidate solutions are evaluated based on a fitness function, and higher-performing configurations are more likely to propagate to subsequent generations. In the context of deep learning hyperparameter optimization, GA supports global exploration of the search space while preserving solution diversity. This evolutionary process is particularly effective in avoiding premature convergence and in handling the complex, multimodal optimization landscapes commonly encountered in financial modeling tasks.

Greylag Goose Optimization (GGO) is inspired by the collective migratory behavior and social interactions of greylag geese. The algorithm emphasizes cooperation and information exchange among agents, enabling the population to converge toward high-quality solutions through shared exploration. GGO incorporates adaptive movement strategies that balance exploration of new regions with exploitation of promising areas. In financial forecasting applications, this cooperative behavior supports stable convergence and enhances robustness when optimizing deep learning hyperparameters under noisy and nonstationary conditions.

The Bat Algorithm (BA) draws inspiration from the echolocation capabilities of bats, modeling frequency modulation, loudness adjustment, and pulse emission rates to guide the search process. BA dynamically adapts its exploration and exploitation behavior based on feedback from candidate solutions, allowing it to efficiently navigate complex search spaces. This adaptability makes BA well suited for optimizing deep learning models whose performance landscapes may vary across different regions of the hyperparameter space.

Differential Evolution (DE) is a population-based optimization algorithm that generates new candidate solutions by combining the differences between randomly selected individuals. This vector-based mutation strategy enables efficient exploration of continuous search spaces and has been shown to perform well in high-dimensional optimization problems. In the context of financial forecasting, DE offers a simple yet powerful mechanism for tuning hyperparameters, providing a balance between exploration efficiency and algorithmic simplicity.

The Whale Optimization Algorithm (WOA) is motivated by the bubble-net hunting strategy of humpback whales. It employs adaptive mechanisms that alternate between encircling prey and spiral-shaped movements to explore and exploit the search space. This dual-mode search behavior allows WOA to dynamically adjust its focus based on the optimization stage, making it effective in navigating the nonlinear and multimodal fitness landscapes associated with deep learning hyperparameter tuning.

Multiverse Optimization (MVO) is inspired by cosmological theories and conceptualizes candidate solutions as universes connected through probabilistic exchange mechanisms. White holes, black holes, and wormholes facilitate information transfer among universes, enabling both exploration and exploitation of the search space. In financial forecasting applications, MVO provides a flexible and adaptive framework capable of handling complex objective functions and interacting hyperparameter effects.

Particle Swarm Optimization (PSO) models the collective movement of social organisms, such as bird flocks, through iterative updates of particle velocities and positions. Each particle adjusts its trajectory based on personal experience and shared knowledge from the swarm. PSO is widely recognized for its simplicity and rapid convergence, making it an attractive option for optimizing deep learning models in large-scale financial datasets where computational efficiency is paramount.

Stochastic Fractal Search (SFS) is inspired by fractal growth phenomena observed in natural systems. The algorithm combines diffusion processes with adaptive updating mechanisms to explore the search space in a stochastic yet structured manner. By maintaining diversity and encouraging exploration across multiple scales, SFS is effective in avoiding local optima and sustaining robust search behavior. This property is particularly valuable in financial forecasting, where noisy data and evolving market dynamics can complicate optimization efforts.

Together, these metaheuristic algorithms form a comprehensive optimization framework for enhancing deep learning-based gold price forecasting models. Their complementary search strategies and adaptive behaviors enable systematic and efficient exploration of hyperparameter spaces, supporting the development of robust and scalable financial forecasting systems suitable for complex and dynamic market environments.

4 Experimental Results

4.1 Baseline Model Performance (Before Optimization)

This subsection evaluates the predictive performance of the baseline deep learning models prior to the application of any feature selection or metaheuristic-based hyperparameter optimization. The purpose of this analysis is to establish a quantitative benchmark that reflects the intrinsic forecasting capability of each model architecture when applied to XAU/USD price data under identical experimental conditions. Such a baseline assessment is essential for isolating the true impact of optimization strategies introduced in subsequent stages of the study.

The comparative results, reported in Table 1, reveal noticeable variation in model behavior across all evaluation metrics, underscoring the heterogeneous suitability of different deep learning architectures for gold price forecasting. Error-based metrics, correlation measures, and agreement indices are jointly considered to ensure a comprehensive assessment of predictive accuracy, bias, and robustness.

Among the evaluated baseline models, DPRNN exhibits the strongest overall performance. Specifically, DPRNN achieves an MSE of 0.0589 and an RMSE of 0.2426, indicating relatively low dispersion between predicted and observed gold prices. The MAE value of 0.0802 further confirms the model's ability to limit absolute forecasting deviations, while the MBE of 0.0715 suggests a moderate but controlled systematic bias. From a statistical association perspective, DPRNN records a correlation coefficient of 0.778 and an R^2 value of 0.79, demonstrating a strong linear relationship and high explanatory power. Additionally, the model achieves an NSE of 0.804 and a Willmott Index (WI) of 0.774, reflecting a high level of agreement between predictions and observed values.

The VAE-based model demonstrates competitive but slightly weaker baseline performance relative to DPRNN. While the RMSE of 0.2419 is marginally lower than that of DPRNN, the MSE increases to 0.07195 and the MAE rises to 0.0935, indicating greater sensitivity to absolute prediction errors. The MBE value of 0.0806 suggests a higher degree of systematic deviation. In terms of correlation, the VAE model attains an r value of 0.754 and an R^2 of 0.767, reflecting a moderate reduction in explanatory capability. The NSE and WI values of 0.78 and 0.776, respectively, indicate acceptable agreement, though with reduced stability compared to DPRNN.

The EALSTM model exhibits moderate baseline performance, with an MSE of 0.08074 and an RMSE of 0.2681. The MAE of 0.1067 and MBE of 0.1352 reveal increased absolute and directional errors, suggesting that while the attention mechanism enhances temporal focus, it does not fully mitigate error accumulation over the forecasting horizon. The correlation coefficient of 0.753 and R^2 value of 0.765 indicate a reasonable level of association, though inferior to both DPRNN and VAE. Agreement-based metrics further decline, with an NSE of 0.769 and WI of 0.723, highlighting limitations in predictive consistency.

The GNN-based model shows a further deterioration in baseline forecasting performance. The model records an MSE of 0.16032 and an RMSE of 0.2841, accompanied by an MAE of 0.1458 and an MBE of 0.1521. These values indicate elevated prediction errors and systematic bias. Although the correlation coefficient remains relatively moderate at 0.738 and the R^2 value reaches 0.759, agreement metrics such as NSE (0.754) and WI (0.693) reveal reduced alignment with observed gold price movements. This outcome suggests that relational modeling alone may be insufficient for capturing dominant temporal dependencies in financial price series without extensive structural refinement.

CNN and BiRNN models demonstrate the weakest baseline performance among the evaluated architectures. The CNN model reports an MSE of 0.3489 and an RMSE of 0.5907, with MAE and MBE values increasing to 0.2156 and 0.2913, respectively. These results reflect substantial forecasting deviations and limited robustness. Correspondingly, the correlation coefficient declines to 0.724 and R^2 to 0.733, while NSE and WI fall to 0.733 and 0.659. The BiRNN model performs even less favorably, recording the highest error values with an MSE of 0.5714 and an RMSE of 0.7559. The MAE of 0.2504 and MBE of 0.3841 indicate pronounced absolute and directional errors. Despite achieving an r value of 0.704 and an R^2 of 0.714, the agreement metrics further deteriorate, with NSE and WI reaching only 0.711 and 0.657, respectively.

Table 1: Baseline performance comparison of deep learning models for XAU/USD price forecasting before optimization

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
DPRNN	0.0589	0.2426	0.0802	0.0715	0.778	0.79	6.08	0.804	0.774
VAE	0.07195	0.2419	0.0935	0.0806	0.754	0.767	6.29	0.78	0.776
EALSTM	0.08074	0.2681	0.1067	0.1352	0.753	0.765	6.43	0.769	0.723
GNN	0.16032	0.2841	0.1458	0.1521	0.738	0.759	6.55	0.754	0.693
CNN	0.3489	0.5907	0.2156	0.2913	0.724	0.733	6.86	0.733	0.659
BiRNN	0.5714	0.7559	0.2504	0.3841	0.704	0.714	7.18	0.711	0.657

Overall, the baseline results clearly demonstrate that while certain architectures, particularly DPRNN and VAE, exhibit relatively strong inherent forecasting capability, none of the evaluated models achieve consistently low error and high agreement across all metrics. The observed performance dispersion highlights the sensitivity of deep learning models to architectural design and parameter configuration in financial time-series forecasting. These findings provide a strong empirical justification for the subsequent application of feature selection and metaheuristic-based hyperparameter optimization to enhance predictive accuracy, stability, and generalization in XAU/USD price forecasting.

A rigorous evaluation of predictive model performance requires assessing multiple error- and efficiency-based metrics simultaneously in order to capture both accuracy and robustness. Single-point summary statistics may obscure variability across evaluation runs or validation folds, thereby limiting the interpretability of comparative model performance. To address this, Figure 4 presents the distributional characteristics of key performance metrics using violin plots augmented with jittered data points. This visualization highlights the central tendency, dispersion, and density structure of error measures (such as mean squared error, root mean squared error, mean absolute error, and mean bias error) alongside goodness-of-fit and efficiency indicators, enabling a comprehensive and transparent assessment of model reliability and stability.

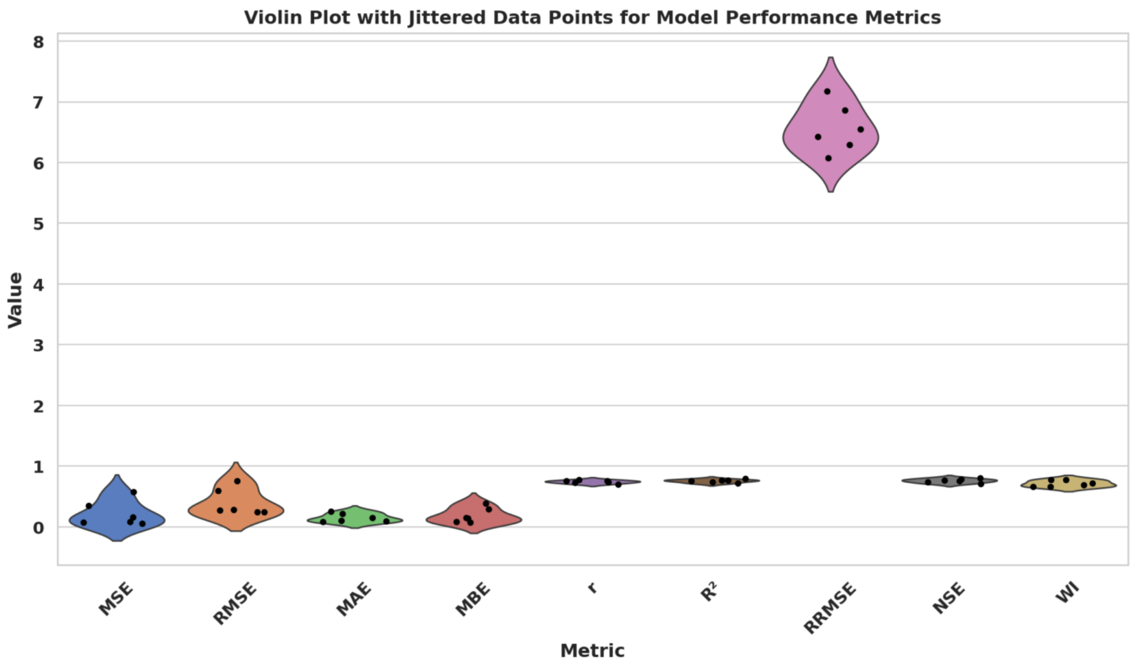

Figure 4: Distribution of Model Performance Metrics Using Violin Plots with Jittered Observations.

Comparative analysis of predictive models benefits from understanding not only their individual performance levels but also the structural similarities in their evaluation outcomes. Hierarchical clustering provides a systematic framework for identifying groups of models that exhibit comparable behavior across multiple performance metrics, thereby revealing latent relationships among modeling approaches. In this context, Figure 5 illustrates a dendrogram resulting from hierarchical clustering applied to the evaluated models.

The branching structure and linkage distances reflect the degree of similarity between models, enabling a clear interpretation of clusters that share analogous predictive characteristics and offering insights into model complementarity and redundancy.

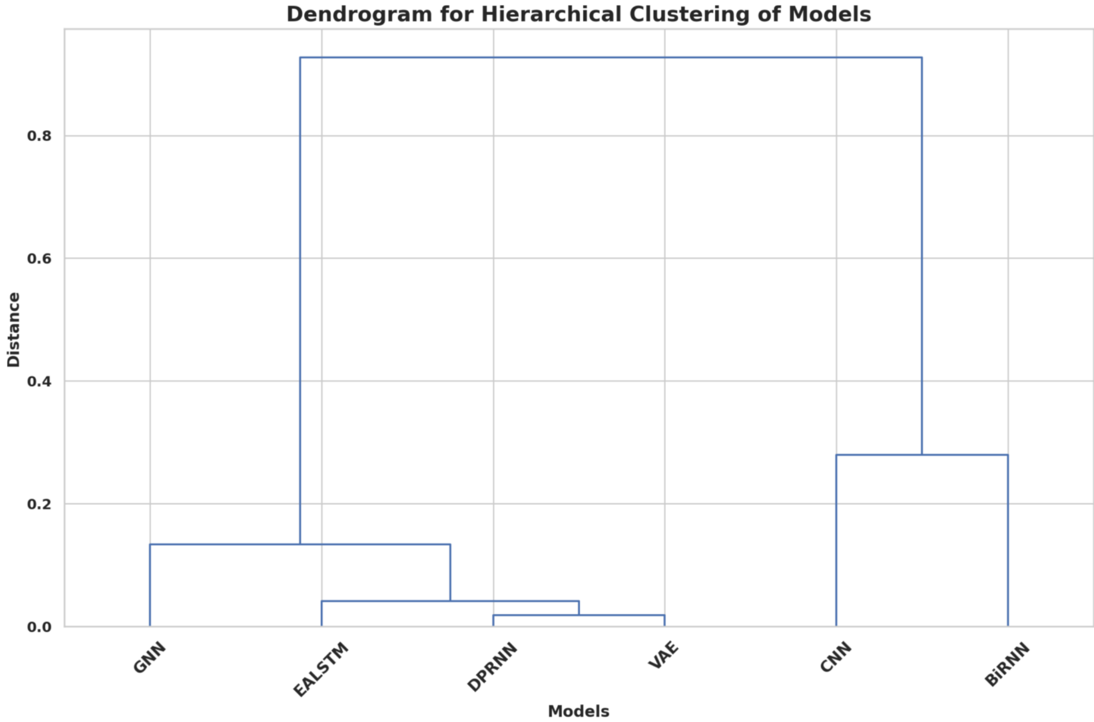

Figure 5: Dendrogram for Hierarchical Clustering of Predictive Models.

A detailed comparison of predictive models requires not only evaluating their average performance but also examining the variability and stability of their evaluation metrics across different configurations or validation instances. Error-based measures and goodness-of-fit indicators may exhibit heterogeneous dispersion patterns that are not fully captured by point estimates alone. To address this issue, Figure 6 presents a combined visualization framework in which box plots are overlaid with swarm plots for individual performance metrics, complemented by explicit annotations of the mean and standard deviation. This integrated representation facilitates a nuanced assessment of central tendency, dispersion, and outlier behavior across models, thereby supporting a more transparent interpretation of model accuracy, robustness, and comparative reliability.

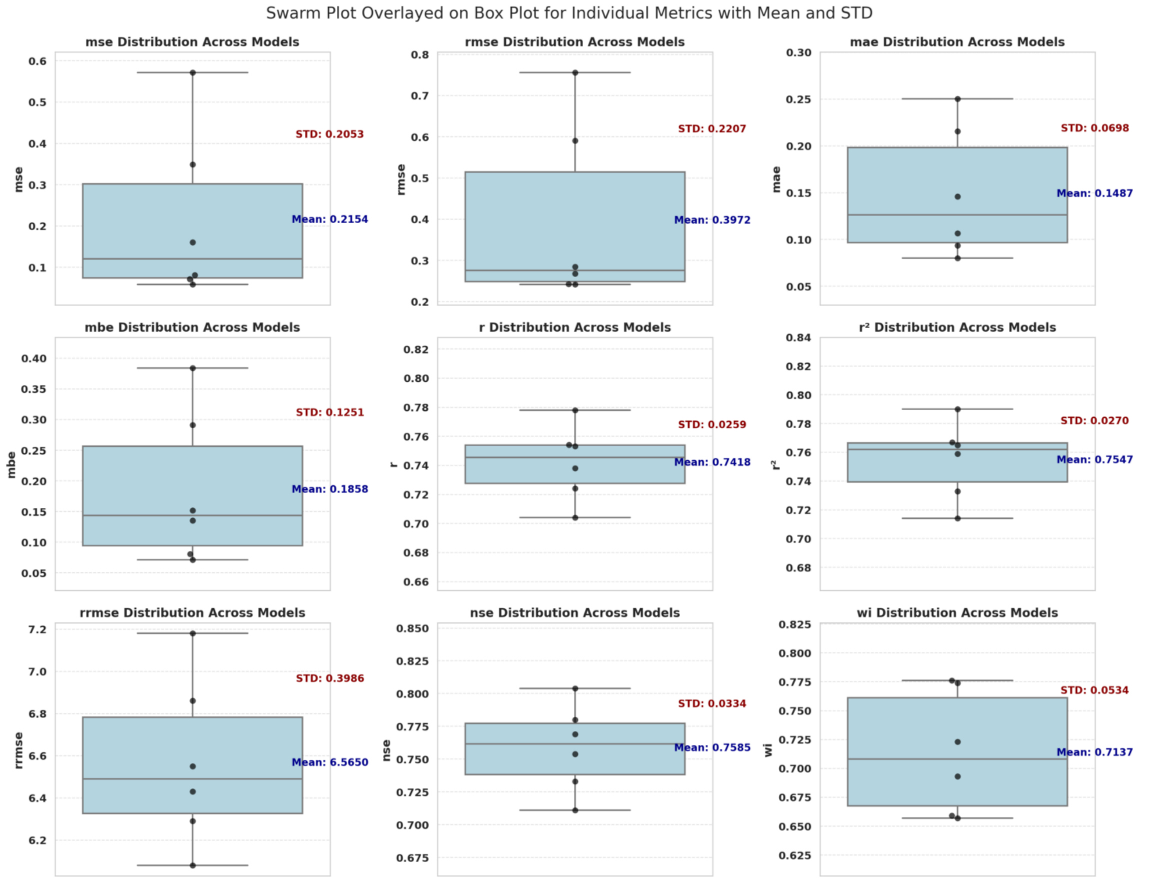

Figure 6: Distribution of Individual Performance Metrics Across Models Using Box and Swarm Plots with Mean and Standard Deviation.

Analyzing the joint distributional behavior of multiple performance metrics provides deeper insight into model accuracy, efficiency, and stability than isolated metric evaluations. Kernel density estimation offers a flexible, non-parametric approach for capturing the underlying probability structure of each metric, while stacked visualizations facilitate simultaneous comparison across heterogeneous scales. In this regard, Figure 7 presents a stacked kernel density streamgraph of key error-based and goodness-of-fit metrics, enabling a holistic assessment of their relative concentration, overlap, and dispersion. This representation highlights dominant density regions and reveals potential trade-offs among metrics, thereby supporting a more integrated interpretation of predictive model performance.

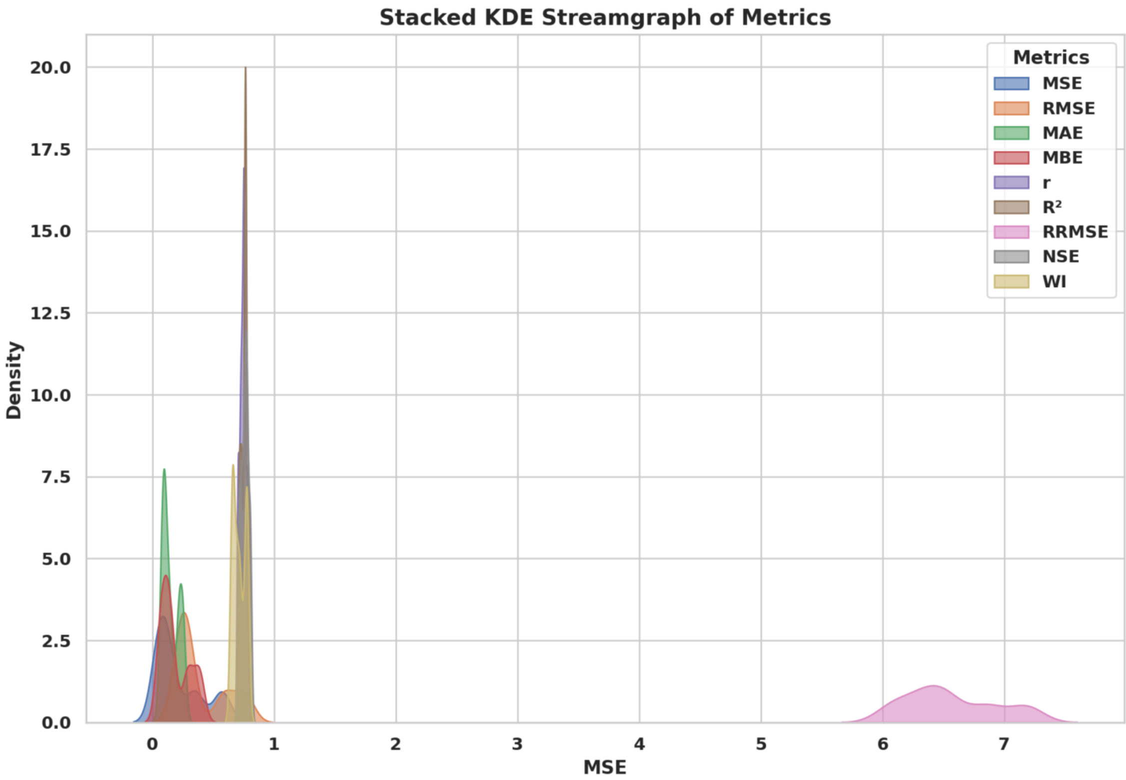

Figure 7: Stacked Kernel Density Estimation Streamgraph of Model Performance Metrics.

4.2 Optimized Model Performance

This subsection evaluates the predictive performance of the optimized forecasting models obtained by integrating metaheuristic algorithms with the DPRNN architecture. The objective of this analysis is to quantify the extent to which metaheuristic-driven optimization enhances forecasting accuracy, stability, and agreement relative to the baseline models. By employing multiple optimization strategies under identical experimental conditions, this section provides a comprehensive assessment of how different metaheuristics influence the performance of deep learning models in XAU/USD price forecasting.

The optimized performance results are summarized in Table 2, which reports error-based metrics, correlation measures, and agreement indices for each metaheuristic–DPRNN combination. Across all evaluated configurations, substantial reductions in prediction error and systematic bias are observed, accompanied by marked improvements in correlation strength and overall model agreement.

Among the optimized models, the GGO+DPRNN configuration achieves the strongest overall performance. This model records an exceptionally low MSE of 2.05×10^{-5} and an RMSE of 4.52×10^{-3}, indicating a dramatic reduction in forecasting error magnitude. The MAE value of 2.05×10^{-4} further confirms the model's ability to maintain minimal absolute deviations, while the MBE of 5.40×10^{-5} reflects negligible systematic bias. From a statistical association perspective, the GGO-optimized model attains a correlation coefficient of 0.987 and an R^2 value of 0.983, demonstrating an exceptionally strong alignment between predicted and observed gold prices. Agreement-based metrics further reinforce this outcome, with an NSE of 0.986 and a Willmott Index of 0.988, indicating near-perfect predictive consistency.

The GA+DPRNN model also demonstrates strong optimization benefits, achieving an MSE of 3.12×10^{-5} and an RMSE of 5.58×10^{-3}. The MAE and MBE values of 2.60×10^{-4} and 6.80×10^{-5}, respectively, indicate low absolute and directional errors. The correlation coefficient reaches 0.979, with an R^2 of 0.975, while the NSE and WI values of 0.978 and 0.981 confirm a high degree of agreement. Although slightly inferior to the GGO-based configuration, the GA-optimized model exhibits substantial improvements over all baseline results.

Table 2: Performance comparison of metaheuristic-optimized DPRNN models for XAU/USD price forecasting

Model	**MSE**	**RMSE**	**MAE**	**MBE**	r	R^2	**RRMSE**	**NSE**	**WI**
GGO + DPRNN	2.05×10^{-5}	4.52×10^{-3}	2.05×10^{-4}	5.40×10^{-5}	0.987	0.983	0.059	0.986	0.98
GA + DPRNN	3.12×10^{-5}	5.58×10^{-3}	2.60×10^{-4}	6.80×10^{-5}	0.979	0.975	0.081	0.978	0.98
BA + DPRNN	4.18×10^{-5}	6.50×10^{-3}	3.00×10^{-4}	8.20×10^{-5}	0.973	0.97	0.098	0.972	0.97
DE + DPRNN	5.42×10^{-5}	7.36×10^{-3}	3.36×10^{-4}	9.60×10^{-5}	0.966	0.963	0.128	0.965	0.97
WOA + DPRNN	6.98×10^{-5}	8.38×10^{-3}	3.76×10^{-4}	1.13×10^{-4}	0.959	0.957	0.178	0.958	0.96
MVO + DPRNN	8.88×10^{-5}	9.44×10^{-3}	4.08×10^{-4}	1.28×10^{-4}	0.952	0.95	0.228	0.95	0.96
PSO + DPRNN	1.08×10^{-4}	1.04×10^{-2}	4.36×10^{-4}	1.43×10^{-4}	0.946	0.944	0.278	0.944	0.95
SFS + DPRNN	1.33×10^{-4}	1.15×10^{-2}	4.63×10^{-4}	1.58×10^{-4}	0.939	0.937	0.408	0.937	0.95

The BA+DPRNN configuration yields further competitive performance, with an MSE of 4.18×10^{-5} and an RMSE of 6.50×10^{-3}. The MAE of 3.00×10^{-4} and MBE of 8.20×10^{-5} reflect controlled error behavior, while the correlation coefficient of 0.973 and R^2 value of 0.97 indicate strong explanatory power. Agreement metrics remain high, with an NSE of 0.972 and WI of 0.977, confirming the effectiveness of bat-inspired optimization in refining deep learning hyperparameters for financial forecasting.

The DE+DPRNN model also exhibits notable performance gains, recording an MSE of 5.42×10^{-5} and an RMSE of 7.36×10^{-3}. The MAE and MBE values of 3.36×10^{-4} and 9.60×10^{-5}, respectively, indicate slightly increased error levels relative to GA- and BA-based configurations. Nonetheless, the model maintains a strong correlation coefficient of 0.966 and an R^2 of 0.963, supported by an NSE of 0.965 and WI of 0.973, reflecting robust predictive agreement.

The WOA+DPRNN configuration records an MSE of 6.98×10^{-5} and an RMSE of 8.38×10^{-3}, with corresponding MAE and MBE values of 3.76×10^{-4} and 1.13×10^{-4}. Although error levels increase slightly, the model retains a strong correlation coefficient of 0.959 and an R^2 of 0.957. The NSE and WI values of 0.958 and 0.968, respectively, indicate consistent predictive behavior and effective convergence.

The MVO+DPRNN model achieves an MSE of 8.88×10^{-5} and an RMSE of 9.44×10^{-3}. The MAE and MBE increase to 4.08×10^{-4} and 1.28×10^{-4}, while the correlation coefficient and R^2 values reach 0.952 and 0.95, respectively. Agreement metrics remain high, with an NSE of 0.95 and WI of 0.962, confirming the capability of multiverse-based optimization to enhance model stability and forecasting accuracy.

The PSO+DPRNN configuration yields an MSE of 1.08×10^{-4} and an RMSE of 1.04×10^{-2}. The MAE and MBE values of 4.36×10^{-4} and 1.43×10^{-4} reflect moderate error increases relative to other optimized models. Nonetheless, the correlation coefficient of 0.946 and R^2 of 0.944, together with an NSE of 0.944 and WI of 0.957, indicate substantial improvements over baseline performance.

Finally, the SFS+DPRNN model records an MSE of 1.33×10^{-4} and an RMSE of 1.15×10^{-2}, with MAE and MBE values of 4.63×10^{-4} and 1.58×10^{-4}. Although this configuration exhibits the highest error among the optimized models, it still achieves a correlation coefficient of 0.939 and an R^2 of 0.937. The NSE and WI values of 0.937 and 0.952, respectively, confirm meaningful performance gains relative to all baseline architectures.

Overall, the optimized results presented in Table 2 demonstrate that integrating metaheuristic algorithms with DPRNN leads to substantial improvements in forecasting accuracy, bias reduction, and predictive agreement. The consistent enhancement across all optimization strategies highlights the effectiveness of metaheuristic-driven hyperparameter tuning in addressing the complexity and nonlinearity of gold price dynamics.

Evaluating predictive models across multiple performance criteria requires a visualization framework capable of simultaneously capturing trade-offs among error-based measures and goodness-of-fit indicators. Parallel coordinates plots are particularly well suited for this purpose, as they enable the comparison of high-dimensional performance profiles within a unified visual space. In this context, Figure 8 presents a parallel coordinates representation of normalized model performance metrics for different hybrid modeling configurations. Each polyline corresponds to a distinct model, allowing for a direct visual assessment of relative strengths, weaknesses, and overall balance across accuracy, bias, correlation, and efficiency metrics.

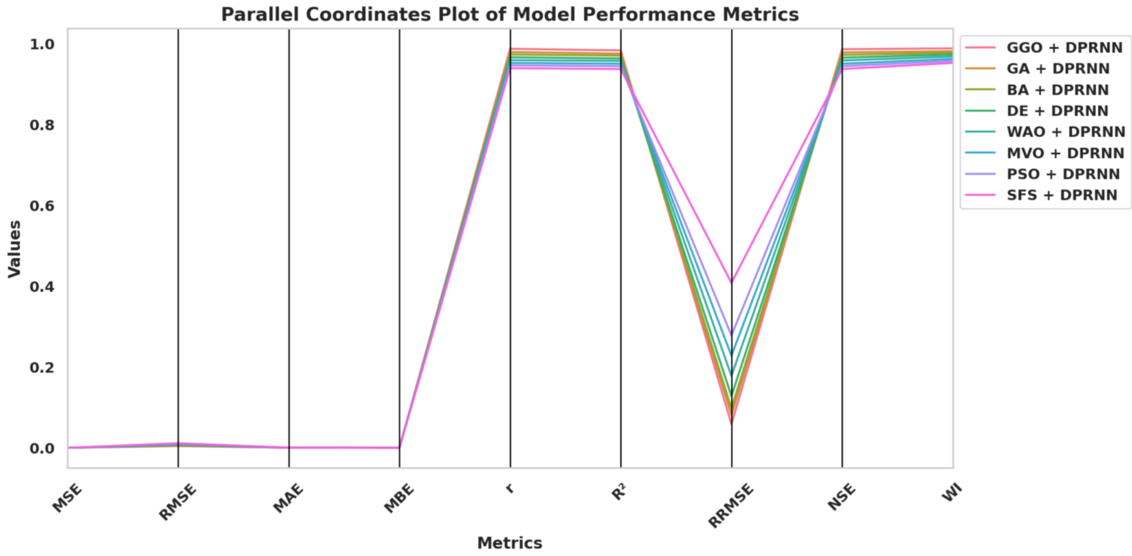

Figure 8: Parallel Coordinates Plot of Normalized Performance Metrics for Hybrid Predictive Models.

A comprehensive evaluation of predictive model performance necessitates examining both the central tendency and dispersion of multiple accuracy and efficiency metrics. While summary statistics provide a concise overview, distribution-based visualizations offer deeper insight into variability, robustness, and the presence of extreme values across competing models. In this respect, Figure 9 presents a box plot representation of key performance metrics overlaid with swarm plots of individual observations. This combined visualization enables a clear assessment of median performance, interquartile ranges, and point-level dispersion, thereby facilitating a nuanced comparison of error magnitudes, correlation strength, and overall model reliability.

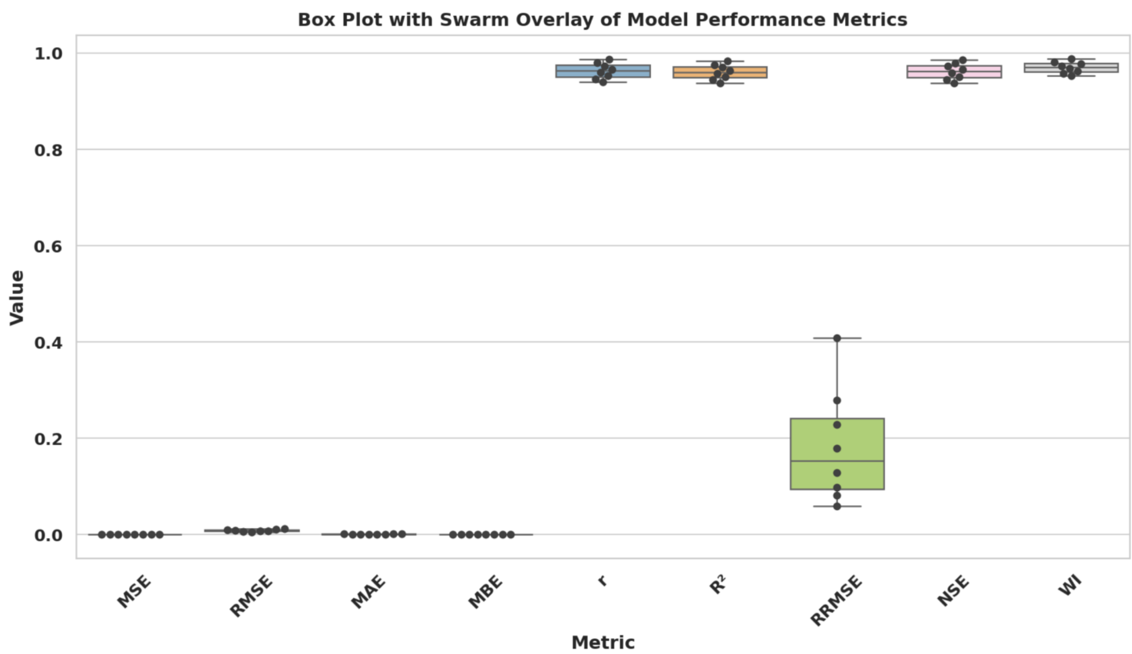

Figure 9: Box Plot with Swarm Overlay of Model Performance Metrics.

Evaluating predictive models across a diverse set of performance indicators requires a metric-wise examination of both central tendency and dispersion in order to ensure robust and reliable conclusions. Different error, bias, and efficiency metrics may exhibit distinct distributional characteristics that reflect trade-offs between accuracy and stability. In this context, Figure 10 presents a series of horizontal box plots overlaid with swarm plots for

DOI: https://doi.org/10.54216/JSDGT.050204

individual performance metrics. This visualization enables a detailed comparison of metric distributions across models, highlighting variability, median performance, and the presence of extreme values, thereby supporting a comprehensive assessment of model consistency and predictive reliability.

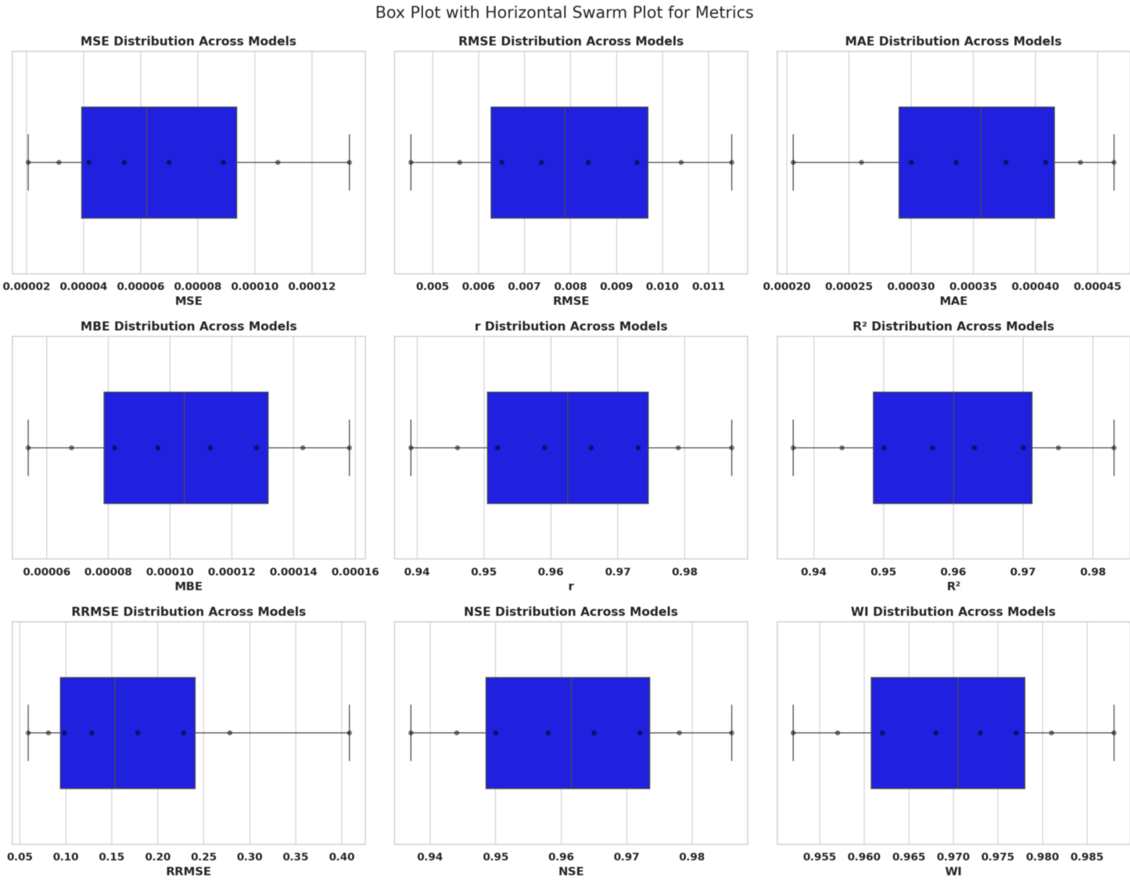

Figure 10: Horizontal Box Plots with Swarm Overlays for Individual Model Performance Metrics.

Assessing the relative effectiveness of competing predictive models requires a normalized comparison framework that highlights performance gains with respect to a common benchmark. Improvement ratios provide an intuitive and scale-independent measure of how closely each model approaches the best observed performance for a given metric. In this regard, Figure 11 presents an improvement ratio matrix in which each entry quantifies the relative performance of a model compared to the best-performing model for each evaluation criterion. This visualization enables a concise yet comprehensive comparison across error-based, correlation, and efficiency metrics, thereby facilitating the identification of dominant modeling strategies and systematic performance differentials.

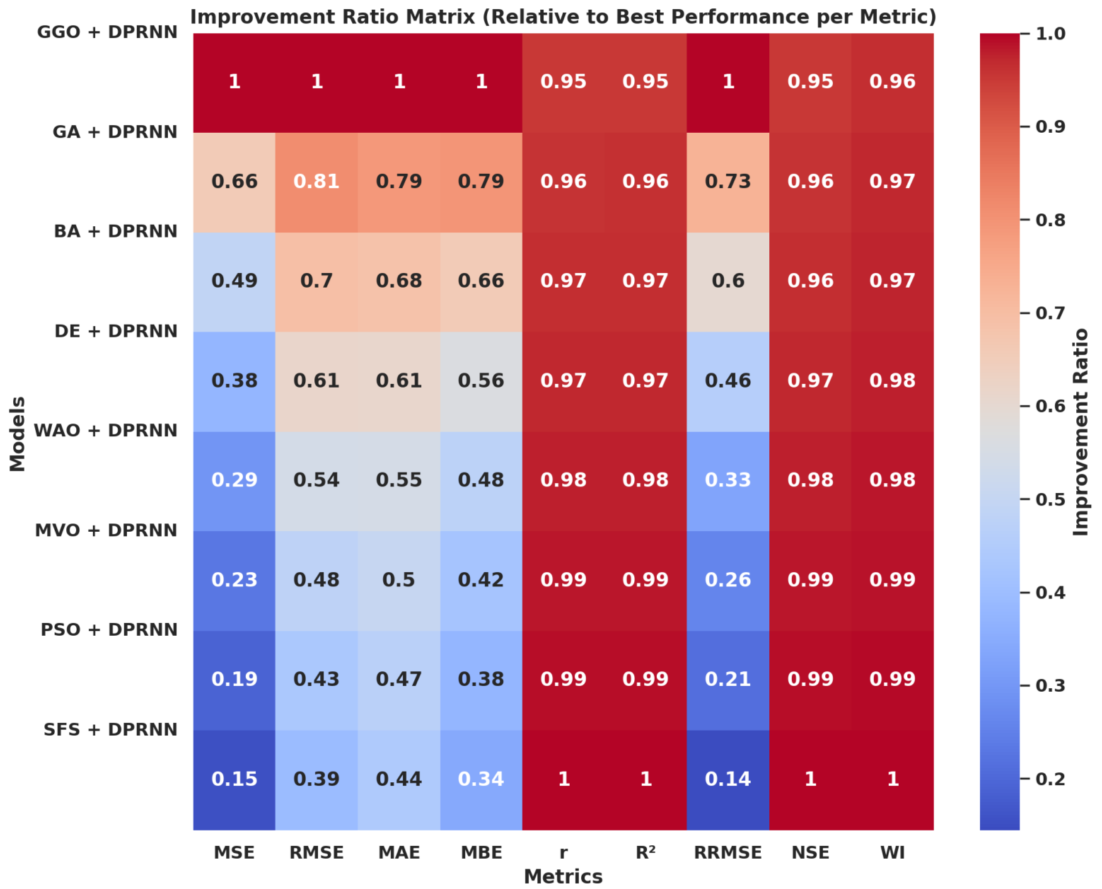

Figure 11: Improvement Ratio Matrix Relative to the Best Performance Achieved for Each Metric.

5 Discussion

The empirical findings of this study provide important insights into the behavior of gold prices in modern financial markets and the effectiveness of advanced forecasting frameworks when applied to long-horizon, multi-timeframe data. From a financial market perspective, the results highlight the intrinsic complexity of XAU/USD price dynamics, which are shaped by nonlinear interactions among macroeconomic conditions, investor expectations, and global risk sentiment. The baseline analysis demonstrates that even sophisticated learning architectures exhibit substantial variation in predictive accuracy and stability when operating without systematic optimization, reflecting the challenges posed by nonstationarity, volatility clustering, and regime-dependent behavior in gold markets.

The pronounced improvement observed after optimization underscores the critical role of model configuration in financial forecasting applications. The optimized DPRNN-based models exhibit significantly reduced forecasting errors, enhanced correlation with observed prices, and stronger agreement metrics across all evaluated measures. These improvements suggest that optimization enables the model to more effectively capture both short-term market fluctuations and longer-term structural patterns embedded within the XAU/USD series. In economic terms, this enhanced predictive capability implies a more accurate representation of market information, which is essential for decision-making in trading, hedging, and portfolio allocation contexts. Improved forecast reliability can reduce estimation risk for investors and support more efficient pricing of gold as a financial asset.

The impact of optimization on forecasting accuracy is particularly relevant in the context of gold's role as a safe-haven asset. Gold prices often respond sharply to changes in macroeconomic uncertainty, monetary policy expectations, and financial stress. The ability of the optimized framework to substantially reduce bias and error indicates that the model is better equipped to adapt to such conditions and to reflect the underlying economic signals driving price movements. This adaptability is crucial for financial practitioners who rely on timely and accurate forecasts to manage exposure during periods of heightened volatility or systemic risk.

When compared with existing financial forecasting studies, the findings of this work are broadly consistent with the growing consensus that data-driven models outperform traditional econometric approaches in capturing nonlinear and regime-dependent market behavior. Prior research on commodity and foreign exchange forecasting has emphasized the limitations of linear models in environments characterized by frequent structural breaks and evolving market dynamics. The results presented here reinforce these observations by demonstrating that even among advanced learning architectures, performance gains are strongly contingent on appropriate optimization strategies. The substantial performance gap between baseline and optimized configurations suggests that methodological rigor in model tuning is as important as the choice of forecasting architecture itself.

Despite these encouraging results, several limitations must be acknowledged. Financial markets are subject to sudden shocks arising from geopolitical events, policy announcements, and exogenous crises, which can induce abrupt structural changes in price behavior. While the long historical coverage of the dataset allows the model to learn from multiple past regimes, no forecasting framework can fully anticipate unprecedented events or structural breaks that fundamentally alter market dynamics. Consequently, model performance may deteriorate during extreme events that fall outside the range of historical experience captured in the training data.

Another limitation relates to the assumption of temporal stability in the learned relationships. Although optimization improves generalization across observed regimes, financial markets remain adaptive systems in which behavioral patterns and institutional structures evolve over time. Changes in market microstructure, trading technology, or regulatory environments may affect the relevance of historical patterns for future forecasting. Addressing such challenges may require adaptive or online learning strategies that continuously update model parameters in response to new information.

Overall, the discussion highlights that while the proposed forecasting framework represents a significant advancement in modeling gold price dynamics, its effectiveness is inherently linked to the evolving nature of financial markets. Recognizing these limitations is essential for responsible application and interpretation of forecasting results in economic and financial decision-making contexts.

6 Conclusion and Future Work

This study has presented a comprehensive investigation into the application of deep learning and metaheuristic optimization techniques for forecasting XAU/USD prices using long-term, multi-timeframe financial data. Gold occupies a central position in global financial markets as a safe-haven asset, an inflation hedge, and a portfolio diversification instrument. Accurate forecasting of gold prices is therefore of significant economic importance, with direct implications for investors, financial institutions, and policymakers operating in increasingly volatile and interconnected markets.

The empirical analysis demonstrates that baseline deep learning models, while capable of capturing certain nonlinear and temporal patterns in gold price dynamics, exhibit substantial variability in predictive accuracy, bias, and agreement. These limitations reflect the inherent complexity of financial time series, which are characterized by nonstationarity, regime shifts, and sensitivity to model configuration. The observed baseline performance underscores the challenges associated with applying deep learning architectures to financial forecasting tasks without systematic optimization.

By integrating metaheuristic optimization algorithms with deep learning models, the study establishes a robust framework for enhancing forecasting performance in the XAU/USD market. Metaheuristic-driven hyperparameter optimization enables automated exploration of complex configuration spaces, effectively addressing issues related to bias–variance trade-offs and convergence stability. The optimized models exhibit pronounced reductions in forecasting error, improved correlation with observed prices, and enhanced agreement metrics, indicating more reliable representation of gold price dynamics across diverse market conditions.

From an economic perspective, the findings highlight the value of advanced optimization techniques in improving the informational efficiency of forecasting models used in financial markets. More accurate and stable gold price predictions can support better-informed trading strategies, risk management practices, and portfolio allocation decisions. For institutional investors and asset managers, such improvements contribute to more effective hedging against inflation and systemic risk. At the macro-financial level, enhanced gold price forecasting may also provide policymakers with additional insights into market expectations, investor sentiment, and the credibility of monetary regimes.

Overall, this study contributes to the financial forecasting literature by demonstrating that the combination of deep learning and metaheuristic optimization constitutes a powerful methodological approach for modeling complex commodity price dynamics. The proposed framework is particularly well suited for long-horizon and multi-timeframe financial datasets, where traditional econometric methods often fall short in capturing nonlinear and regime-dependent behavior.

While the proposed framework delivers substantial methodological and empirical contributions, several avenues for future research remain open. One important direction involves the integration of additional macroeconomic and financial indicators, such as interest rates, inflation measures, exchange rate indices, and financial market volatility proxies. Incorporating such exogenous variables could further enhance the economic interpretability and predictive robustness of gold price forecasting models, particularly during periods of macroeconomic stress.

Another promising avenue lies in extending the framework to real-time and high-frequency forecasting applications. As financial markets continue to evolve toward faster trading environments, the ability to deploy optimized deep learning models in near real-time settings becomes increasingly relevant. Future work may therefore focus on reducing computational overhead, improving model scalability, and assessing latency constraints associated with live market deployment.

The adaptability of the proposed optimization framework also enables its extension to other financial assets and markets. Applying similar methodologies to foreign exchange rates, equity indices, energy commodities, or fixed-income instruments would allow for a broader assessment of the generalizability of metaheuristic-optimized deep learning models across asset classes. Such extensions could contribute to the development of unified forecasting systems capable of supporting cross-asset portfolio management and systemic risk monitoring.

From a methodological standpoint, future research may explore hybrid optimization strategies that combine multiple metaheuristic algorithms or integrate reinforcement learning mechanisms to dynamically adjust optimization behavior in response to changing market conditions. These enhancements could further improve convergence efficiency and robustness in nonstationary financial environments.

Finally, future studies may investigate the integration of the proposed forecasting framework into intelligent financial decision-support systems. Embedding optimized predictive models within automated trading platforms, risk assessment tools, or policy analysis frameworks could facilitate more responsive and data-driven decision-making in complex financial markets. Such developments would strengthen the practical relevance of advanced forecasting methodologies and contribute to more resilient and efficient financial systems.

Data Availability

The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com/datasets/novandraanugrah/xauusd-gold-price-historical-data-2004-2024?select=XAU_15m_data.csv.

Declarations

- **Acknowledgments**
  Not applicable.
- **Conflict of interest/Competing interests**
  The authors declare that they have no conflicts of interest to report regarding the present study.
- **Ethics approval and consent to participate**
  Not applicable.
- **Consent for publication**
  Not applicable.
- **Funding**
  No Fund

References

[1] W. Zhang, Y. Zhan, and Y. Liang, "Financial openness, major events, and exchange rate linkage: Empirical analysis based on the dcc-midas model," *International Review of Financial Analysis*, vol. 106, p. 104 578, 2025, ISSN: 1057-5219. DOI: https://doi.org/10.1016/j.irfa.2025.104578. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1057521925006659.

[2] İ. Özkal, İ. A. Özkan, and F. Başçiftçi, "Metaverse token price forecasting using artificial neural networks (anns) and adaptive neural fuzzy inference system (anfis)," *Neural Computing and Applications*, vol. 36, no. 7, pp. 3267–3290, 2024. DOI: https://doi.org/10.1007/s00521-023-09228-y.

[3] W. Dammak, H. Gökgöz, and A. Jeribi, "Analysis of gold, bitcoin, and gold-backed cryptocurrencies as safe havens during global crises: A focus on artificial intelligence companies," *Computational Economics*, vol. 66, no. 4, pp. 2843–2872, 2025. DOI: https://doi.org/10.1007/s10614-024-10757-4.

[4] J. Moreira, R. Pascoal, and H. Rocha, "Multifractal analysis of gold as a hedge against geopolitical risk," in *Computational Science and Its Applications – ICCSA 2025 Workshops*, O. Gervasi et al., Eds., Cham: Springer Nature Switzerland, 2026, pp. 207–224, ISBN: 978-3-031-97596-7.

[5] W. Kristjanpoller and B. M. Tabak, "Comparison of the asymmetric multifractal behavior of green and us bonds against benchmark financial assets," *Financial Innovation*, vol. 11, no. 1, p. 57, 2025. DOI: https://doi.org/10.1186/s40854-024-00698-0.

[6] E. Ayadi and N. Ben Mbarek, "The dynamic effects of economic uncertainties and geopolitical risks on saudi stock market returns: Evidence from local projections," *Journal of Risk and Financial Management*, vol. 18, no. 5, 2025, ISSN: 1911-8074. DOI: 10.3390/jrfm18050264. [Online]. Available: https://www.mdpi.com/1911-8074/18/5/264.

[7] P. M. Lefatsa, S. Msomi, H. T. Muguto, L. Muguto, and P.-F. Muzindutsi, "An empirical analysis of the impact of global risk sentiment, gold prices, and interest rate differentials on exchange rate dynamics in south africa," *International Journal of Financial Studies*, vol. 13, no. 3, 2025, ISSN: 2227-7072. DOI: 10.3390/ijfs13030120. [Online]. Available: https://www.mdpi.com/2227-7072/13/3/120.

[8] S. Li and Y. Liu, "A new perspective on gold as a risk hedge: Long-term impacts of bilateral political tensions between the u.s. and china," *International Review of Economics & Finance*, vol. 106, p. 104 918, 2026, ISSN: 1059-0560. DOI: https://doi.org/10.1016/j.iref.2026.104918. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1059056026000316.

[9] U. A. Sheikh, M. Asad, Z. Ahmed, and U. Mukhtar, "Asymmetrical relationship between oil prices, gold prices, exchange rate, and stock prices during global financial crisis 2008: Evidence from pakistan," *Cogent Economics & Finance*, vol. 8, no. 1, D. McMillan, Ed., p. 1 757 802, 2020. DOI: 10.1080/23322039.2020.1757802. eprint: https://doi.org/10.1080/23322039.2020.1757802. [Online]. Available: https://doi.org/10.1080/23322039.2020.1757802.

[10] S. Pandit and X. Luo, "A novel prediction model to evaluate the dynamic interrelationship between gold and crude oil," *International Journal of Data Science and Analytics*, vol. 20, no. 2, pp. 1161–1182, 2025. DOI: https://doi.org/10.1007/s41060-024-00519-8.

[11] A. A. Salisu, R. Gupta, S. Karmakar, and S. Das, "Forecasting output growth of advanced economies over eight centuries: The role of gold market volatility as a proxy of global uncertainty," *Resources Policy*, vol. 75, p. 102 527, 2022, ISSN: 0301-4207. DOI: https://doi.org/10.1016/j.resourpol.2021.102527. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301420721005341.

[12] A. Ampountolas, "Enhancing forecasting accuracy in commodity and financial markets: Insights from garch and svr models," *International Journal of Financial Studies*, vol. 12, no. 3, 2024, ISSN: 2227-7072. DOI: 10.3390/ijfs12030059. [Online]. Available: https://www.mdpi.com/2227-7072/12/3/59.

[13] F. Wang, J. Chang, W. Zuo, and W. Zhou, "Research on efficiency and multifractality of gold market under major events," *Fractal and Fractional*, vol. 8, no. 8, 2024, ISSN: 2504-3110. DOI: 10.3390/fractalfract8080488. [Online]. Available: https://www.mdpi.com/2504-3110/8/8/488.

[14] J. Gao, C. Fan, L. Xu, H. Chen, H. Chen, and Z. Liang, "Intelligent decision making and risk management in stock index futures markets under the influence of global geopolitical volatility," *Omega*, vol. 133, p. 103 272, 2025, ISSN: 0305-0483. DOI: https://doi.org/10.1016/j.omega.2024.103272. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0305048324002366.

[15] D. Gabauer, R. Gupta, S. Karmakar, and J. Nielsen, "Stock market bubbles and the forecastability of gold returns and volatility," *Applied Stochastic Models in Business and Industry*, 2025. DOI: https://doi.org/10.1002/asmb.2887.

[16] M. Küçük, F. Çebi, and A. T. Tekin, "Xau/usd price prediction using deep learning: Hyperparameter optimization with bayesian, grey-wolf and genetic algorithms," in *Intelligent and Fuzzy Systems*, C. Kahraman et al., Eds., Cham: Springer Nature Switzerland, 2025, pp. 96–103, ISBN: 978-3-031-97992-7.

[17] S. Shayganfard, S. S. Moghadam, and M. Salehirad, "Forecasting gold prices (global ounce) using time series models and deep learning," in *2025 11th International Conference on Web Research (ICWR)*, 2025, pp. 610–616. DOI: 10.1109/ICWR65219.2025.11006217.

[18] R. Popli, I. Kansal, R. Kumar, V. Khullar, P. Sharma, and D. Parsad, "Gold price prediction in stock exchange using lstm and bi-lstm multivariate deep learning approaches," in *Computational Intelligence and Mathematical Applications*, CRC Press, 2024, pp. 1–5.

[19] Shobhit, S. Khan, S. Banerjee, N. Chaudhary, and B. Shukla, "Forecasting gold prices in uzbekistan: An analytical approach to market dynamics and economic indicators using machine learning techniques," in *2025 International Conference on Technology Enabled Economic Changes (InTech)*, 2025, pp. 678–682. DOI: 10.1109/InTech64186.2025.11198562.

[20] K. N. TP and S. Sultana, "Advanced techniques for daily gold price forecasting in india through statistical analysis and predictive modeling," in *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, 2024, pp. 1–6. DOI: 10.1109/CSITSS64042.2024.10816779.

[21] A. Biju, B. P. Alapatt, and J. P. George, "Ensemble hybrid lstm architectures for robust multi-currency forex forecasting," in *2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2025, pp. 490–498. DOI: 10.1109/ICICV64824.2025.11085759.

[22] A. H. Baradaran, M. Bohlouli, and M. R. J. Motlagh, "Decoding tomorrow's gold prices: A comparative study of gru and cnn-lstm in the iranian market," in *2024 10th International Conference on Web Research (ICWR)*, 2024, pp. 329–334. DOI: 10.1109/ICWR61162.2024.10533331.

[23] M. M. Billah and S. Das, "Analysis and prediction of gold price using cnn and bi-gru based neural network model," in *2021 24th International Conference on Computer and Information Technology (ICCIT)*, 2021, pp. 1–6. DOI: 10.1109/ICCIT54785.2021.9689880.

[24] M. F. Sert, "A hybrid arima-lstm/gru model for forecasting monthly trends in turkey's gold and currency markets with a macro-economic data-driven approach," in *Machine Learning in Finance: Trends, Developments and Business Practices in the Financial Sector*, M. Gün and B. Kartal, Eds. Cham: Springer Nature Switzerland, 2025, pp. 35–51, ISBN: 978-3-031-83266-6. DOI: 10.1007/978-3-031-83266-6_3. [Online]. Available: https://doi.org/10.1007/978-3-031-83266-6_3.

[25] B. D. Iswara, P. Sugiartawan, and A. S. Kusuma, "Integration of conformal prediction in automated forex trading systems: Development and evaluation," in *2025 IEEE International Symposium on Consumer Technology (ISCT)*, 2025, pp. 389–394. DOI: 10.1109/ISCT66099.2025.11297372.

[26] S. Anbaee Farimani, M. V. Jahan, and A. Milani Fard, "An adaptive multimodal learning model for financial market price prediction," *IEEE Access*, vol. 12, pp. 121 846–121 863, 2024. DOI: 10.1109/ACCESS.2024.3441029.

[27] Y. Wang, M. F. Ghazali, R. Ab Razak, and M. A. S. Zaidi, "Complex system and ps-lstm prediction of cryptocurrencies, stocks, bonds, exchange rates and commodities," *Physica A: Statistical Mechanics and its Applications*, vol. 679, p. 130 976, 2025, ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2025.130976. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437125006284.

[28] E. Njoki, J. S. Wekesa, and D. G. Njagi, "Ensemble learning for foreign exchange market trend prediction," *Computational Economics*, pp. 1–16, 2025. DOI: https://doi.org/10.1007/s10614-025-11102-z.

[29] Z. A. Hassaan, M. H. Yacoub, and L. A. Said, "Fpga-accelerated esn with chaos training for financial time series prediction," *Machine Learning and Knowledge Extraction*, vol. 7, no. 4, 2025, ISSN: 2504-4990. DOI: 10.3390/make7040160. [Online]. Available: https://www.mdpi.com/2504-4990/7/4/160.

[30] P. Srichaiyan, K. Y. Tippayawong, and A. Boonprasope, "Forecasting soybean futures prices with adaptive ai models," *IEEE Access*, vol. 13, pp. 48 239–48 256, 2025. DOI: 10.1109/ACCESS.2025.3546786.