FORECASTING GOLD PRICE USING ENHANCED LSTM DEEP LEARNING MODEL

R.K. Keerthana[1], Uma Maheswari N[2], S. Suresh Kumar[3]

Abstract

The use of historical data and advanced algorithms has made machine learning an essential tool for predicting gold prices. The principal objective of this field of study is to elucidate patterns, trends, and interrelations among the myriad factors that impact gold prices, encompassing economic indicators, geopolitical occurrences, and the complicated dynamics of supply and demand. Through the utilization of machine learning algorithms, scholars are able to construct predictive models that provide significant insights into fluctuations in gold prices. These analytical insights empower traders, investors, and various stakeholders to engage in well-informed decision-making processes pertaining to gold investment strategies. In the present investigation, we investigate the area of data science and the methodologies of machine learning to predict the fluctuations in gold prices. We perform an exhaustive examination of historical gold price datasets, formulate advanced forecasting models, and meticulously assess their efficiency. A fundamental aspect of our scholarly inquiry involves the assessment of the dependability and precision of diverse machine learning frameworks employed in the forecasting of gold prices. We analyze various algorithms and techniques, evaluating their efficiency in capturing the fundamental trends in gold price variations. This analysis provides valuable findings and insights, aiding us in identifying the most effective models for producing precise gold price forecasts.

Keywords :

Machine learning, Gold price, Algorithms.

I. INTRODUCTION

Forecasting financial market, especially gold prices, presents a significant challenge due to the market's ever-changing and unpredictable nature[1]. Gold is often seen as a reliable place to invest, especially during uncertain times. Its price is influenced by big economic factors like inflation, global conflicts, and shifts in currency values[2]. Because of this, being able to predict gold prices accurately is incredibly important for people working in finance and investing. To make these predictions more reliable, a variety of machine-learning techniques have been developed[3]. These tools have changed the way experts approach financial forecasting by helping them sort through huge amounts of data and uncover patterns that aren't always easy to spot [4].

II. LITERATURE SURVEY

The US Dollar Index, the US Federal Funds Rate, the Retail Price Index (RPI), the US dollar's exchange rate with the Chinese yuan, the price of oil, and the S&P 500 Index were the six variables used in the experiment [5]. At a 5% significance level, the test findings show that, with the exception of the CPI and oil price, all of the aforementioned factors have a negative impact on precious metal prices.

Singhal et al. analyze the long-term relationship in Mexico between West Texas Intermediate (WTI) crude oil prices, the worldwide spot price of gold, the USD to Mexican peso (USD/MXN) exchange rate, and the Mexican stock market index (IPC)[6]. The study employs an autoregressive distributed lag (ARDL) bounds testing technique to investigate how these variables change over time [7]. The results show that when WTI crude oil prices go down, gold prices tend to rise, and that higher gold prices are linked to stronger share price valuations in Mexico.

S. Nallamothu et al. examine the factors influencing gold prices, focusing on elements linked with uncertainty. Their analysis includes four important indicators: the Global Economic Policy Uncertainty Index (GEPU), the Political Conflict Index (PCI), the Skewness Index (SKEW), and the Volatility Index (VIX). The last two indices indicate the risks and probable volatility in financial markets[9].

The tests took place considering 13 separate festive events noticed globally, like Religious and cultural celebrations. A structure using regression within a GARCH system was created, as well as the study by Schmidbauer and Rosch reveals that different festivities are linked to significant shifts in the cost of gold.

Also, countless research projects have made use of predictive modeling strategies to anticipate what gold prices

Department of AI & DS,[1]
Karpagam Academy of Higher Education[1]
Coimbatore, India[1]
keerthana.raju@kahedu.edu.in[1]

Department of AI & DS,[2]
Karpagam Academy of Higher Education[2]
Coimbatore, India[2]
umamaheswari.narayanan@kahedu.edu.in[2]

Department of Information Technology[3]
Sri Ramakrishna Mission Vidyalaya College of Arts and Science[3]
Coimbatore, India[3]
kumarsoftware@gmail.com[3]

* Corresponding Author

will be in the future or look at how price changes move. Al-Dhuraibi and Ali worked on predicting shifts in the price of gold using machine learning methods created to rank data [10]. To make an educated guess about gold price trends in the time to come, they made use of regression models, decision Tree-based models, Support Vector classifiers (SVC), and k-NN classifier (KNN). Their investigation centered on changes in weekly gold prices as a single element. Classification algorithms were implemented to sort the weekly price movement into a pair of groups: “+1” standing for a favorable shift (an increase in price) and “–1” standing for an unfavorable shift (a decrease in price)[11]. In the end, their research came to the conclusion that KNN was the single algorithm able to get to a good level of accuracy, which was 60.26%, with K=5.

Kemal, on the other hand, used a Multilayer Perceptron Neural Network (MLPNN) to guess at gold prices, which is different from strategies based on classification[8].

III. METHODOLOGY

The research is geared towards delivering a complete mobile platform that empowers individuals to observe and gain up-to-the-minute alerts spanning vital areas, specifically the monitoring of gold prices.

A. LSTM Model

A specific type of deep learning model called Long Short-Term Memory (LSTM) networks is designed to handle sequential input and discover long-term dependencies. Unlike conventional neural networks, Unlike conventional neural networks, LSTMs employ a more complex architecture consisting of memory cells and three crucial gates: the input gate, forget gate, and output gate. By controlling the network's information flow, these gates allow it to retain crucial data while gradually eliminating less pertinent information.

Vanishing and exploding gradients, which are prevalent in conventional recurrent neural networks, are successfully addressed by this method. As a result, LSTMs are ideal for examining time-dependent data and finding patterns in long sequences. They are extremely useful in fields like time-series forecasting and data mining, where identifying temporal correlations is crucial, because of their capacity to use past data.

B. LSTM for Gold Price Forecasting

Long Short-Term Memory (LSTM) networks provide an efficient deep learning solution for managing sequential financial data in order to forecast gold prices. Movements in the price of gold are not random; rather, they are impacted by past patterns, trends, and general market activity. Therefore, using a model that can learn from past data is essential, and LSTMs are particularly well-suited for this purpose.

A unique kind of neural network that functions somewhat differently from conventional ones is called Long Short-Term Memory (LSTM). Its three primary gates-input, forget, and output-as well as integrated memory cells aid in determining what data should be retained, discarded, and transmitted. As a result, LSTM is able to retain pertinent historical data while disregarding extraneous details. Because it is better able to comprehend both short-term price fluctuations and long-term trends, it is especially helpful for forecasting gold prices.

One of the biggest strengths of LSTM is that it can handle common problems faced by basic recurrent models, like vanishing and exploding gradients. This means it can learn from long sequences of data without forgetting important details along the way. When applied to historical gold price data, LSTM can spot hidden patterns and use them to make more accurate predictions about future price movements.

Overall, LSTM is a reliable and practical method for predicting gold prices, making it a helpful tool for financial analysis and better decision-making.

C. Enhanced LSTM Model

By adding new features to better handle the complexity of gold price prediction, the enhanced LSTM model expands the capabilities of the traditional LSTM. Although the standard LSTM is good at learning temporal relationships, it might not always be able to identify long-term trends, abrupt market swings, or subtle hidden patterns that affect gold prices.

The suggested model incorporates a number of cutting-edge elements to enhance performance. In order to better understand sudden price fluctuations, convolutional layers are used to extract local patterns and short-term characteristics from the data. Additionally, bidirectional LSTM layers are used to process the sequence both forward and backward, providing a more comprehensive understanding of the data.

To keep the model from overfitting and to improve its performance on fresh, unseen data, dropout regularization is employed. Combining these methods makes the enhanced LSTM model more reliable and produces predictions that are more accurate.

Overall, this enhanced design is able to capture both small details and long-term patterns in the data, leading to more reliable gold price forecasts while still being efficient enough for real-world financial applications.

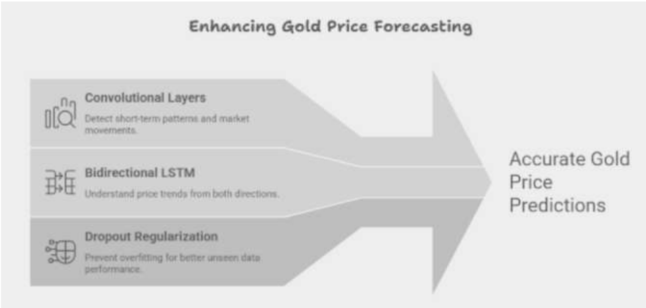

Figure:1 Gold Price Prediction using Enhanced LSTM

A. Enhanced LSTM for Gold Price Forecasting

To make gold price predictions more accurate, the enhanced LSTM model improves on the traditional version by adding extra features that can better handle the complexity of financial data. While a basic LSTM is good at learning patterns over time, it doesn't always do well with sudden price jumps, long-term trends, or hidden factors that influence the market.

To deal with these challenges, the proposed model uses a few advanced techniques. Convolutional layers help it pick up short-term patterns and spot quick changes in gold prices, so it can better react to sudden market movements. On top of that, bidirectional LSTM layers let the model look at information in both directions, which helps it understand price trends in a more complete and balanced way over time.

Dropout regularization is also included to help the model avoid overfitting and perform better on data it hasn't seen before. When all these techniques are combined, the enhanced LSTM model becomes more stable, dependable, and accurate in its predictions.

Overall, this improved approach strikes a good balance between short-term price changes and long-term trends, making it more practical and effective for forecasting gold prices.

VI. EVALUATION METRICS:

For assessing how well the models performed, we relied on three key measures. The first metric we employed, called Absolute Mean Error (MAE), provides a clear picture of how accurate the predictions were by calculating the average size of the errors in what the models predicted. Root Mean Squared Error (RMSE), our second measure, gives more weight to the bigger mistakes, making it very helpful when looking at time-based information that changes a lot. Lastly, we employed the Adjusted R-squared (R^2) value to judge how closely the expected gold prices matched the actual prices, which tells us about the model's total correctness.

A. Results and Discussion

We examined the results of the models we proposed using three important metrics: Adjusted R-squared (R^2), Square root of mean squared error (RMSE), and Absolute Mean Error (MAE). The discoveries from this evaluation are shown in Table 1. Furthermore, Figure 1 gives a visual depiction of these measures for both the LSTM [12] and Enhanced LSTM models, making it simple to contrast their predictive powers.

B. Model Performance in

The enhanced LSTM (ELSTM) model was created to address the shortcomings of the conventional design and boost performance by making specific changes to the core LSTM architecture [13]. Three key metrics were used to assess the efficacy of the Enhanced LSTM model and the Long Short-Term Memory (LSTM) [14]: Adjusted R-squared (R^2), Root Mean Squared Error (RMSE), and Absolute Mean Error (MAE). These metrics assess the models' ability to predict the price of gold by examining historical data; the results are shown in Table 1.

Table 1. Model Performance Metrics

Model	MAE	RMSE	R^2
LSTM	4.73	5.71	0.83
Enhanced LSTM	3.92	5.14	0.84

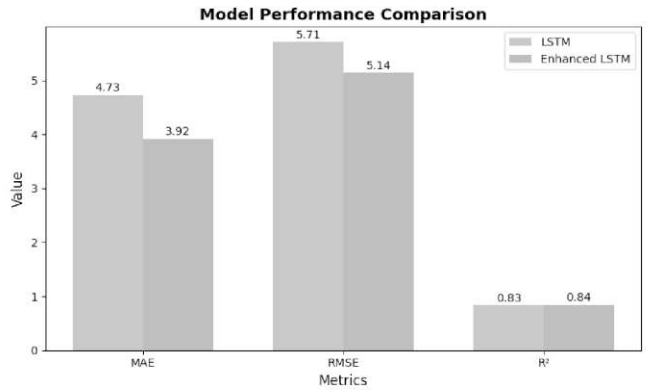

Figure:2 Simulation Comparison of LSTM and ELSTM

The performance comparison between the LSTM and Enhanced LSTM models reveals a clear improvement in the enhanced version. The Enhanced LSTM model outperforms the standard LSTM across all key evaluation metrics:

Absolute mean error (MAE): The Enhanced LSTM records a lower MAE. The Absolute Mean Error (MAE) decreases from 4.73 to 3.92, indicating that the Enhanced LSTM produces more accurate predictions on average, with an approximate 17% reduction in error.

Root Mean Squared Error (RMSE): The Enhanced LSTM's superior accuracy in lowering prediction errors is demonstrated by its significantly lower RMSE. Similarly, the Root Mean Squared Error (RMSE) drops from 5.71 to 5.14, showing that the Enhanced LSTM is more effective at minimizing larger prediction errors, which are penalized more heavily in this metric.

R^2 Value: The R-squared (R^2) value increases slightly from 0.83 to 0.84, suggesting a modest improvement in the model's ability to explain the variability in the data. Overall, these results confirm that the Enhanced LSTM provides better predictive performance, although the gain in explanatory power is relatively small. This suggests that the enhanced model accounts for a larger share of the variance in the target variable, offering a closer and more accurate fit to the data.

Overall, the Enhanced LSTM model shows a marked improvement in prediction accuracy and model performance, making it a more effective approach compared to the traditional LSTM in this scenario.

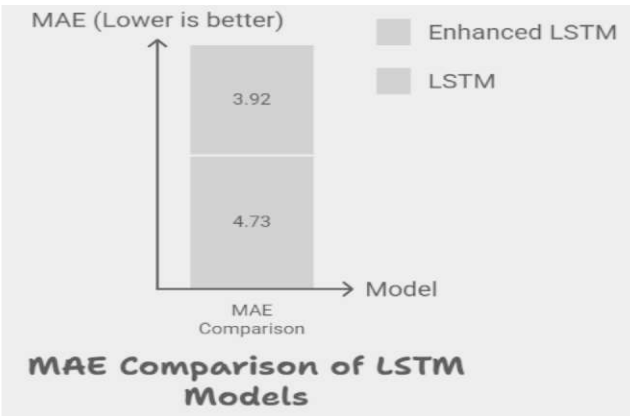

Figure : 3 MAE Comparison

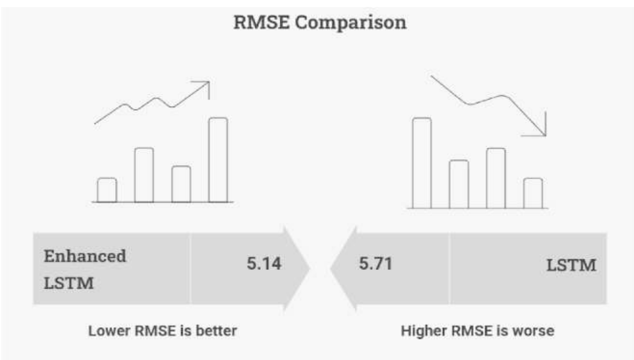

Figure:4 RMSE Comparison

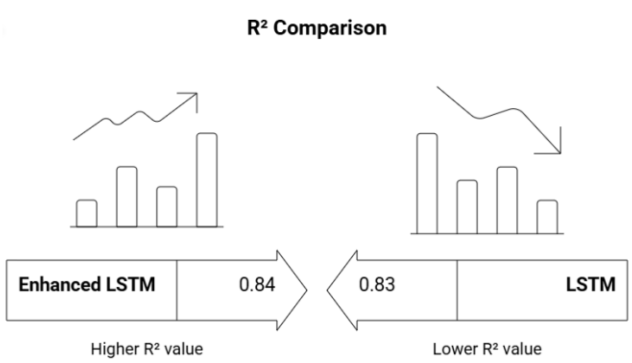

Figure : 5 R^2 Comparison

V. PREDICTED VS. ACTUAL GOLD PRICES:

To improve how we judge how well the model works, we looked at how the gold prices we predicted stacked up against what actually happened during the testing phase. This comparison gives us useful information about how correct and trustworthy the LSTM and Improved LSTM models are when used in the real world. The graph in Figure 3 shows a timeline that visually compares the real gold prices with the prices predicted by both models, letting us easily see how well they predict prices over a period of time. Table 2 below shows a comparison between the actual and predicted gold prices over a selected set of dates.

The results indicate that both LSTM and ELSTM models are effective for predicting gold prices using time series data. But because it can process data both forward and backward, the ELSTM model has an advantage [16]. This ability allows it to capture more context, making it particularly valuable for financial predictions where both past and future trends influence the results.

Table 2 Comparison Between The Actual And Predicted Gold Prices

Date	Gold Actual Price	Predicted Price (LSTM)	Predicted Price (ELSTM)
Oct 21, 2025	93,800	92878	93782
Oct 20, 2025	93,600	93458	93557
Oct 19, 2025	93,500	93257	93459
Oct 18, 2025	93,400	93274	93389
Oct 17,2025	93,450	93379	93421
Oct 16,2025	93,425	93392	93397
Oct 15,2025	93,400	93379	93394
Oct 14,2025	93,375	93365	93371
Oct 13,2025	93,350	93341	93347
Oct 12,2025	93,325	93318	93327

Figure : 6 Comparison Between The Actual and Predicted Gold Prices

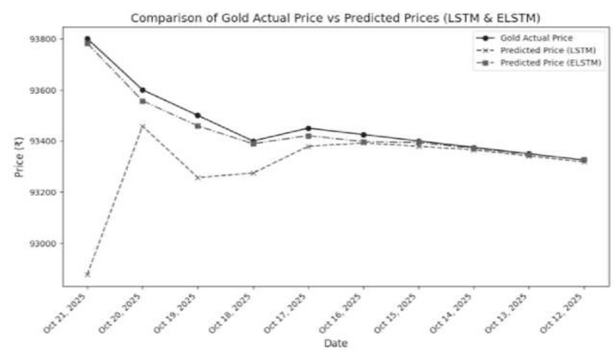

Figure:7 The actual gold prices against the predicted values generated by LSTM and ELSTM

In figure 7, the comparison between actual gold prices and predictions from both the LSTM and Enhanced LSTM (ELSTM) models over the selected dates in October 2025 shows that both models follow the overall trend of gold prices quite closely, but the Enhanced LSTM consistently provides predictions that are slightly closer to the actual values. While the standard LSTM tends to either underpredict or overpredict by a larger margin-such as on October 21, 2025, where it predicts 92,878 against an actual price of 93,800-the ELSTM prediction of 93,782 is significantly closer to the real value. This pattern is observed across most of the dates, where ELSTM reduces the deviation between predicted and actual prices.

VI. COMPARATIVE ANALYSIS OF LSTM AND ENHANCED LSTM MODELS

The performance of the Long Short-Term Memory (LSTM) and Enhanced Long Short-Term Memory (ELSTM) models was assessed using gold price data collected between October 12, 2025, and October 21, 2025. To evaluate the effectiveness of both models, commonly used regression metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2) were considered. In addition, the predicted prices were compared with the actual gold prices to understand how accurately each model captured market movements.

The evaluation results indicate that the ELSTM model performed slightly better than the traditional LSTM model across all performance measures. ELSTM achieved an MAE of 3.92, compared to 4.73 for LSTM, suggesting that its predictions were generally closer to the actual gold prices. Likewise, the RMSE value of ELSTM was lower (5.14) than that of LSTM (5.71), indicating that ELSTM produced fewer significant prediction errors and offered more stable forecasting performance.

The R^2 score also favored ELSTM, which recorded a value of 0.84 compared to 0.83 for LSTM. Although the difference is small, it shows that ELSTM was able to explain a slightly larger portion of the variation in gold prices and fit the observed data more effectively.

A closer look at the daily prediction results further highlights the advantages of the ELSTM model. Both models successfully identified the overall upward trend in gold prices during the testing period. However, ELSTM consistently generated values that were nearer to the actual market prices. This was especially evident on October 21, 2025, where the ELSTM prediction of 93,782 was very close to the actual price of 93,800, while the LSTM prediction of 92,878 showed a noticeably larger difference. Similar patterns were observed on the other dates, where ELSTM maintained greater accuracy and consistency.

These findings suggest that the modifications introduced in the ELSTM model enhanced its ability to capture complex time-dependent patterns and subtle fluctuations in gold price movements. As a result, ELSTM was better equipped to model market behavior and generate more reliable forecasts.

Overall, both models demonstrated their capability in predicting gold prices, but ELSTM consistently delivered more accurate and stable results. With lower error values, a slightly higher R^2 score, and predictions that closely matched actual prices, ELSTM can be considered a more effective and dependable approach for short-term gold price forecasting.

VII. GOLD PRICE PREDICTION COMPARISON

Figure: 8 Gold Price Forecasting Performance of LSTM vs Enhanced LSTM Models

Interpretation

- The Actual Price line represents the real gold market price.
- The ELSTM line closely follows the actual price trend throughout the period.
- The LSTM line deviates more, particularly on Oct 21, where it significantly underestimates the actual price.
- The visual trend confirms the evaluation metrics (lower MAE and RMSE, higher R^2) that ELSTM provides more accurate and stable forecasts than the conventional LSTM model.

Additionally, during periods of relatively stable prices (from October 12 to October 18), both models perform well, but ELSTM maintains a tighter alignment with actual values, indicating better consistency. The LSTM model shows slightly larger fluctuations in its predictions, whereas ELSTM outputs smoother and more accurate estimates. Overall, the Enhanced LSTM demonstrates improved precision and reliability in capturing both the level and trend of gold prices, reinforcing its superiority over the basic LSTM model for this dataset.

The performance of the Long Short-Term Memory (LSTM) and Enhanced Long Short-Term Memory (ELSTM) models was analyzed using gold price data collected between October 12, 2025, and October 21, 2025. The predicted values generated by both models were compared with the actual gold prices to evaluate their forecasting effectiveness.

The analysis shows that both LSTM and ELSTM were able to identify and follow the general trend of gold price movements during the study period. However, ELSTM demonstrated a higher level of accuracy by producing predictions that were consistently closer to the actual market prices. Although the LSTM model performed reasonably well, it exhibited larger deviations on certain dates, especially towards the end of the testing period. In contrast, ELSTM maintained relatively small prediction errors throughout the dataset.

A comparison of the prediction results reveals that ELSTM achieved better overall forecasting performance. Its ability to capture complex temporal patterns and subtle fluctuations in gold prices contributed to more precise and stable predictions. The standard LSTM model, while effective in recognizing trends, showed greater variation in its forecast values and was less accurate when market prices changed.

Another important observation is the consistency of the ELSTM model across all test samples. The model closely followed the gradual increase in gold prices and generated forecasts that aligned well with the actual values. This consistency highlights the robustness of ELSTM and its capability to generalize effectively when handling time-series financial data.

The results we observed align with the assertions, which highlighted that, generally, ELSTM models outperform their unidirectional counterparts when it comes to predicting future values in time series data.

Even though the advantage in effectiveness between the two types of models is not particularly significant, the enhanced capacity of the ELSTM to identify intricate trends and relationships within the data results in predictions that are more precise. This aspect holds substantial importance within the realm of financial markets, where projections spanning both shorter and longer durations are critical in the decision-making processes.

In summary, the results indicate that ELSTM provides a clear improvement over the traditional LSTM model for gold price prediction. Its higher accuracy, lower prediction error, and better consistency make it a reliable approach for short-term forecasting. These findings suggest that ELSTM can be a useful tool for supporting decision-making in financial markets, particularly for investors, traders, and analysts who rely on accurate price forecasts.

VIII. CONCLUSION

This study proposed and tested two different frameworks-namely LSTM and an improved version called Enhanced LSTM (ELSTM)-to anticipate the movement of gold prices. The results showed that the ELSTM framework was better than the basic LSTM framework when it came to predicting gold prices would be in the future. This enhancement is evident because it produced smaller Absolute mean error (MAE) and Root Mean Squared Error (RMSE) numbers, in addition to a bigger Adjusted R-squared number, which means the predicted values were much closer to actual values.

This study investigated and compared two forecasting approaches—Long Short-Term Memory (LSTM) and its improved variant, Enhanced LSTM (ELSTM)-for predicting gold prices. The performance evaluation was carried out using both statistical error metrics (MAE, RMSE, and R^2) and visual analysis of predicted versus actual gold prices from October 12 to October 21, 2025.

From the quantitative results, it is clearly observed that the ELSTM model outperformed the traditional LSTM model. ELSTM achieved a lower MAE (3.92 vs 4.73) and RMSE (5.14 vs 5.71), indicating that its predictions were consistently closer to the actual gold prices with fewer large deviations. In addition, the R^2 value of ELSTM (0.84) was slightly higher than that of LSTM (0.83), showing that ELSTM explained the variation in gold prices more effectively.

The graphical analysis further strengthens these findings. As shown in the comparison plot, the ELSTM curve closely follows the actual gold price trend, maintaining minimal deviation across all time points. In particular, on October 21, 2025, ELSTM almost exactly matches the actual value (93,782 vs 93,800), whereas the LSTM model shows a significant underestimation. Across the entire time period, the ELSTM predictions remain more stable and aligned with actual market behavior, while the LSTM model exhibits comparatively larger fluctuations and errors.

Overall, both models are capable of capturing the general upward trend in gold prices, but ELSTM demonstrates superior accuracy, stability, and reliability. The combination of numerical evaluation and graphical evidence confirms that the enhancements introduced in ELSTM significantly improve forecasting performance over the standard LSTM model.

Although the results are promising, further improvements are still possible. The study suggests that performance can be enhanced by fine-tuning hyperparameters, incorporating hybrid preprocessing techniques, and integrating additional external factors such as economic indicators, currency fluctuations, and global market conditions. Future research can also explore larger datasets and more advanced deep learning architectures to further improve prediction accuracy.

Even though the outcomes are promising, there remains room for further optimization. The performance of the proposed ELSTM model shows that even small adjustments to hyperparameters, combined with a mix of machine learning and deep learning preprocessing techniques, can significantly enhance the forecasting accuracy of LSTM models[15]. Subsequent studies may look into using larger collections of data, adding more outside factors that have an effect on the prices of gold, and researching more cutting-edge model designs in order to make forecasting even more accurate.

This research offers a solid starting point for using machine learning techniques to forecast financial time series. The ELSTM model, specifically, shows a strong potential for future gold prices prediction.

In conclusion, this work establishes a strong foundation for applying deep learning models in financial time-series forecasting. The ELSTM model, in particular, demonstrates strong potential for real-world gold price prediction and can serve as a reliable decision-support tool for investors, traders, and financial analysts.

REFERENCES

[1] Singh, S. K., Gupta, N., Baliyan, S., & Mishra, P. K. (2022). Gold price prediction using machine learning algorithm. NeuroQuantology, 20(20), 2998.

[2] Bala, R., & Singh, R. P. (2022). A dual-stage advanced deep learning algorithm for long-term and long-sequence prediction for multivariate financial time series. Applied Soft Computing, 126, 109317. https://doi.org/10.1016/j.asoc.2022.109317

[3] Mohtasham Khani, M., Vahidnia, S., & Abbasi, A. (2021). A deep learning-based method for forecasting gold price with respect to pandemics. SN Computer Science, 2(4), 335. https://doi.org/10.1007/s42979-021-00743-x

[4] Baradaran, A. H., Bohlouli, M., & Motlagh, M. R. J. (2024). Decoding tomorrow's gold prices: A comparative study of GRU and CNN-LSTM in the Iranian market. In Proceedings of the 2024 10th International Conference on Web Research (ICWR) (pp. 329–334). IEEE.

[5] Livieris, I. E., Stavroyiannis, S., Pintelas, E., & Pintelas, P. (2020). A novel validation framework to enhance deep learning models in time-series forecasting. Neural Computing and Applications, 32(23), 17149–17167. https://doi.org/10.1007/s00521-020-04948-1

[6] Hajek, P., & Novotny, J. (2022). Fuzzy rule-based prediction of gold prices using news affect. Expert Systems with Applications, 193, 116487. https://doi.org/10.1016/j.eswa.2021.116487

[7] Das, S., Sahu, T. P., & Janghel, R. R. (2022). Oil and gold price prediction using optimized fuzzy inference system–based extreme learning machine. Resources Policy, 79, 103109. https://doi.org/10.1016/j.resourpol.2022.103109

[8] Wang, J. (2022). A novel metal futures forecasting system based on wavelet packet decomposition and stochastic deep learning model. Applied Intelligence, 52(8), 9334–9352.

[9] Nallamothu, S., Rajyalakshmi, K., & Arumugam, P. (2024). Gold price prediction using skewness and kurtosis based generalized auto-regressive conditional heteroskedasticity approach with long short-term memory network. Journal of The Institution of Engineers (India): Series B, 1–13.

[10] Zhang, P., & Ci, B. (2020). Deep belief network for gold price forecasting. Resources Policy, 69, 101806. https://doi.org/10.1016/j.resourpol.2020.101806

[11] Srivastava, T., Mullick, I., & Bedi, J. (2024). Association mining–based deep learning approach for financial time-series forecasting. Applied Soft Computing, 155, 111469. https://doi.org/10.1016/j.asoc.2024.111469

[12] Salim, M., & Djunaidy, A. (2024). Development of a CNN-LSTM approach with images as time series data representation for predicting gold prices. Procedia Computer Science, 234, 333–340.

[13] Senthil Kumar, S., Krithika, N., & Kanakaraj, R. (2025). Deep learning-driven gesture and speech recognition for human–machine interaction: Enhancing virtual reality through AI integration. Karpagam Journal of Computer Science.

[14] Soumya, M., & Kanimozhi, J. (2025). A review on early detection and prediction of autism spectrum disorders using deep learning techniques. Karpagam Journal of Computer Science, 20(4).

[15] Ambiha, H. A., Kanna, M. K., & Inbavel, M. (2025). Intelligent sepsis detection in ICU: A real-time deep learning framework with hybrid optimization. Karpagam Journal of Computer Science, 20(4).

[16] Unnikrishnan, N. V., & Mythili, S. (2025). A survey on predicting and controlling air pollution using machine learning and deep learning techniques. Karpagam Journal of Computer Science, 20(1).

HYBRID EXPLAINABLE RETRIEVAL-AUGMENTED FRAMEWORK FOR HALLUCINATION AND NOISE-AWARE SUMMARIZATION IN LARGE LANGUAGE MODELS

G. Santhanakrishnan

Abstract

Large Language Models (LLMs) have achieved substantial performance in automatic summarization tasks; however, they remain prone to hallucinations, generating factually inconsistent or fabricated content - especially when presented with noisy or out-of-distribution (OOD) inputs. This paper presents the Hybrid Explainable Retrieval-Augmented Framework (HERAF), a unified five-module architecture that jointly addresses noise detection, OOD filtering, knowledge-grounded generation via Retrieval-Augmented Generation (RAG), semantic attention-flow explainability, and post-hoc hallucination verification. Experiments on CNN/DailyMail, XSum, and WikiHow demonstrate that HERAF achieves ROUGE-1 of 48.6, BERTScore F1 of 91.2, and a 38% hallucination rate reduction relative to BART and PEGASUS baselines. OOD detection accuracy reaches 94.3%, underscoring robust pipeline performance for safety-critical NLP applications.

Keywords : Large Language Models; Hallucination Detection; Explainable AI; Retrieval-Augmented Generation; Out-of-Distribution Detection; Transformer Models; NLP Robustness; Hallucination Mitigation; Factual Consistency; Trustworthy AI

1. INTRODUCTION

The proliferation of transformer-based language models - including BERT, GPT-4, T5, BART, and PEGASUS - has revolutionized automatic text summarization. These models generate fluent, coherent abstracts from lengthy documents; however, a persistent challenge is their propensity to hallucinate: producing statements plausible in form yet factually unsupported by the source text [1]. Such errors are especially consequential in high-stakes domains such as medical, legal, and financial summarization, where factual precision is non-negotiable.

Compounding this issue is the sensitivity of LLMs to noisy inputs - documents containing typographical errors, grammatical inconsistencies, or adversarially perturbed tokens - which significantly degrade summarization quality and exacerbate hallucination rates [2]. Furthermore, when LLMs encounter out-of-distribution (OOD) inputs, they tend to generate confidently incorrect outputs rather than abstaining or flagging uncertainty [3]. Despite growing awareness, existing solutions address these problems in isolation: dedicated noise-handling methods, separate OOD detectors, and independent hallucination mitigation techniques have not been integrated into a coherent framework [4].

Retrieval-Augmented Generation (RAG) has emerged as a promising direction for grounding LLM outputs in verified external knowledge [5] while attention-flow analysis provides transparency into model decision-making [6]. No prior work has unified these complementary approaches within a single pipeline. To address this gap, we propose HERAF - a five-module framework integrating noise detection, OOD filtering, RAG-grounded generation, attention-flow explainability, and hallucination verification for noise-aware, interpretable, and factually consistent summarization.

II. LITERATURE SURVEY

Evidence-grounded summarization frameworks generate concise outputs while maintaining traceability to original source documents [7]. Transformer-based language models achieve improved summarization performance when supported by retrieval and verification modules [8]. Re-ranking mechanisms refine retrieved results by prioritizing highly relevant and informative content for summary generation [9]. Semantic similarity assessment methods assist in selecting contextually appropriate information from large document collections [10]. Knowledge-guided generation techniques reduce unsupported content by integrating structured and unstructured information sources [11].

Multi-stage verification frameworks detect factual

Department of Computer Technology,
Karpagam Academy of Higher Education
santhanakrishnan.gurusaminathan@kahedu.edu.in

* Corresponding Author