Resources Policy 92 (2024) 105040

Contents lists available at ScienceDirect

Resources Policy

journal homepage: www.elsevier.com/locate/resourpol

How good are different machine and deep learning models in forecasting the future price of metals? Full sample versus sub-sample

Anu Varshini [a], Parthajit Kayal [a], Moinak Maiti [b,1,*]

[a] Madras School of Economics (MSE), Gandhi Mandapam Road, Behind Government Data Centre, Kottur, Chennai, 600025, India
[b] Department of Finance, School of Economics and Finance, University of the Witwatersrand, Johannesburg, South Africa

ARTICLE INFO

Keywords:
Machine learning
Deep learning
Forecasting
Commodity
Futures

ABSTRACT

This study aims to forecast metal futures in commodity markets, including gold, silver, copper, platinum, palladium, and aluminium, using different machine and deep learning models. Prevalent models such as Stacked Long-Short Term Memory, Convolutional LSTM, Bidirectional LSTM, Support Vector Regressor, Extreme Gradient Boosting, and Gated Recurrent Unit are utilized. The model performance is assessed by multiple factors such as Root Mean Squared Error, Mean Absolute Error, and Mean Absolute Percentage Error. The study stands out by considering multiple metal commodity futures simultaneously, incorporating both Machine Learning and Deep Learning models, and conducting two sets of experiments with a full sample and subsample analysis. In addition, it uses different inputs of 30- and 60-days periods for robustness checks. Mean Absolute Percentage Error values suggest that different machine and deep learning models are efficient on prediction the future metal prices. However, the model performance varies significantly with the influence of metal choice, sample period, and inputs on prediction performance. Therefore, it suggests that constructing a theory based on machine and deep learning models is challenging.

1. Introduction

Commodity markets play a significant role in the global economy, offering benefits like as price discovery, hedging opportunities, and ensuring liquidity. They are particularly instrumental in propelling economic growth, especially in resource-rich nations (Chang and Fang, 2022). Extensive research has underscored the importance of accurate price predictions in commodity markets, offering valuable insights for traders, investors, and policymakers. By assessing the price volatility, traders and investors can hedge their positions and mitigate potential losses. In a recent study, Kumar et al. (2022) emphasize that price prediction can help manage risks, optimize profits, and enhance market efficiency. Moreover, reliable price forecasts are important for improving portfolio performance and guiding investment decisions (Horák and Jannová, 2023). Investing in commodity futures helps investors achieve excess returns. Further, changes in commodity prices reflects direction of future inflation. This relationship is particularly noticeable for future prices compared to spot prices (Fama, 1984). Unlike spot prices, future prices provide more accurate forecasts that consider inflation and effectively reflect the impact of supply-demand dynamics on commodity prices (Fama, 1984; Verheyen, 2010). Commodity futures markets have played a vital role in reducing price volatility, enhancing market efficiency, and providing benefits to producers and consumers by mitigating the risk of possible price shocks (Jacks et al., 2011).

Recent studies underscore the significance of employing artificial intelligence (AI) techniques, such as machine learning (ML) and deep learning (DL) algorithms, to forecast commodity prices. These techniques have demonstrated substantial improvements, achieving accuracy and robustness improvements of over 60% (Li et al., 2017). Previous work primarily relied on traditional time series econometric techniques like ARIMA, ARCH, and GARCH models (Arowolo, 2013; Bhardwaj et al., 2014). However, it has been demonstrated that these models struggle to capture non-linear patterns, prompting the adoption of ML and Artificial Neural Networks (ANN) as alternatives, particularly in the presence of non-linearity (Shahwan and Odening, 2007). ML methodologies proficiently identify stochastic variables and capture latent non-linear characteristics that conventional econometric models are incapable of detecting (Herrera et al., 2019). Artificial Recurrent Neural Networks (ARNN) equipped with memory, such as Long-Short

* Corresponding author.
E-mail addresses: maitisoft@gmail.com, moinak.maiti@wits.ac.za (M. Maiti).
[1] Moinak Maiti was Independent Researcher, India, when this paper was submitted to the journal.

https://doi.org/10.1016/j.resourpol.2024.105040
Received 24 August 2023; Received in revised form 3 January 2024; Accepted 25 April 2024
Available online 30 April 2024

Term Memory (LSTM), have demonstrated considerable superiority over ARIMA models in terms of performance (Siami-Namini et al., 2019). Furthermore, hybrid models have gained popularity due to their improved forecasting accuracy, though conflicting results persist regarding their superiority (Herrera et al., 2019).

Metals represent a substantial share of commodity markets and assume a pivotal role in the global economy, being influenced by diverse economic and geopolitical factors (Gong and Xu, 2022). Given the growing interconnectedness of factors influencing metal prices, accurate predictions of future metal prices present challenges for investors and speculators seeking risk management. The primary aim of this research is to employ various machine learning (ML), and deep learning (DL) techniques to forecast metal futures within commodity markets. The metals under consideration encompass gold, silver, copper, platinum, palladium, and aluminium. The analysis focuses on well-established models in time series analysis, namely Stacked-LSTM (Li et al., 2020; Ojo et al., 2019), Convolutional LSTM (Essien and Giannetti, 2019), Bidirectional LSTM (Siami-Namini et al., 2019), Support Vector Regressor (SVR) Model (Makala and Li, 2021), Extreme Gradient Boosting (XGBoost) model (Jabeur et al., 2021), and Gated Recurrent Unit (GRU) model (Torres and Qiu, 2018). The performance of these models is evaluated by Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics. Although some researchers favour MAE over RMSE, it is advantageous to consider a combination of these indicators (Chai and Draxler, 2014).

This study differentiates itself from previous literature in several aspects. First, it considers multiple metal commodity futures simultaneously, which is unique compared to previous work focused on predicting the prices of individual metal commodities (Astudillo et al., 2020; Hussein et al., 2011; Livieris et al., 2020). Second, it incorporates both ML and DL models to derive the study conclusions. Third, the study comprises two sets of experiments: one involving the construction of prediction models using a full 20-year sample from 2003 to 2023, and the other conducting sub-sample analysis to identify potential models that consistently outperform the rest. Fourth, different robustness measures are used to validate the study findings such as different input periods. The results of this study do not yield conclusive evidence regarding the superiority of any specific model for accurate commodity price forecasting. The choice of commodities, sample periods, and inputs significantly influences the predictive performance of the models, suggesting the absence of a consistently superior model. The study findings regarding the variations in model performance across metals are highly relevant in the context of the Efficient Market Hypothesis, Portfolio Theory, Behavioural finance, and hedging strategies.

section 2 offers an extensive literature review, drawing comparisons between our research and previous studies. Section 3 describes the data source and the models employed in this study to predict future commodity prices. The analysis, results, and subsequent discussion are offered in Section 4. Finally, the study concludes in the last section, providing recommendations for future research.

2. Review of literature

For more than five decades, autoregression (AR) and moving averages (MA), which are statistical techniques, have been considered the cutting edge in time series modelling and prediction (Parmezan et al., 2019). Prior studies on forecasting prices in commodity markets have mainly focused on linear statistical methods, particularly ARIMA models. However, these techniques possess limitations concerning their modelling capabilities and computational demands (Zulauf et al., 1999). Additionally, statistical models often hinge on unrealistic assumptions like variable independence and normal distribution, which do not align with real-world scenarios (Onour and Sergi, 2011). In recent times, there has been a surge in interest towards non-parametric modelling using Machine Learning (ML) algorithms for time series prediction. This approach offers enhanced flexibility as it avoids assuming a specific data distribution (Parmezan et al., 2019).

The emergence of ML techniques has prompted a shift in the focus of commodity price forecasting from statistical methods to ML-based approaches. ML algorithms bring the advantage of adeptly managing extensive and complex datasets prevalent in commodity markets, which are characterized by substantial volatility and uncertainty stemming from geopolitical events, macroeconomic trends, and supply chain dynamics (Verheyen, 2010). These algorithms are aptly equipped to capture the dynamic nature of commodity markets, where prices experience frequent fluctuations (Schnepf et al., 2005). Time series forecasting, with the availability of abundant data and increased computing power, has become an area of extensive research, and ML models have become an integral part of it (Lim and Zohren, 2021). Unlike traditional statistical methods, ML algorithms excel at identifying patterns in sequential data that exhibit time-dependent patterns, which can be challenging for traditional methods to analyze (Herrera et al., 2019). As an example, Artificial Neural Networks (ANN) have demonstrated superior performance compared to classical statistical methods such as linear regression and Box-Jenkins approaches when forecasting excess returns in large stocks (Desai and Bharati, 1998). Similar findings are reported by Krollner et al. (2010) in their study on modelling non-linear time series using artificial neural networks. Other popular ML models used in time series forecasting include Support Vector Regressor (SVR) (Cao and Tay, 2003), Extreme Gradient Boosting (XGBoost) (Zhang et al., 2021b), and Recurrent Neural Network (RNN) such as Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU), which capture temporal dependencies and patterns in data (Yamak et al., 2019). Hybrid models, including CNN-LSTM, LSTM-CNN, ConvLSTM, and Encoder-Decoder, have exhibited encouraging performance when compared to classical models (Kilinc and Yurtsever, 2022).

ML models are increasingly applied in predicting various financial assets, including stocks, bonds, commodities, and cryptocurrencies, to assist investors in making informed decisions. In stock markets, ML models demonstrate superior performance compared to statistical and econometric models by capturing the non-linear nature of equity markets (Ampomah et al., 2020). In bond markets, ML algorithms are used to predict changes in bond prices and forecast future interest rate movements (Bianchi et al., 2021). In commodity markets and cryptocurrencies, ML models help identify hidden patterns and forecast price movements. Notable models used in these asset classes include.

- Stocks - SVM (Reddy, 2018), Random Forest (RF), LSTM (Nikou et al., 2019); ANN (Maiti et al., 2022; Vukovic et al., 2020)
- Bonds – Linear Regression, Boosted Regression Trees, RF, and Deep Neural Networks (Bianchi et al., 2021)
- Commodity Markets – ARIMA, SVR, Prophet, XGBoost and LSTM (Chen et al., 2021)
- Cryptocurrencies – Neural Networks, SVR, RF (Valencia et al., 2019); ANN, LSTM (Maiti et al., 2020), XGBoost (Ranjan et al., 2023)

The prices of metals in commodity markets exhibit extreme volatility and are challenging to predict, necessitating appropriate modelling techniques (Kayal and Maheswaran, 2021). Deep learning (DL) models have exhibited substantial advancements in the literature of financial time series forecasting, surpassing traditional ML models (Jabeur et al., 2021). DL approaches excel in extracting key features from complex data, and recent advancements in attention mechanisms have further enhanced prediction performance beyond baseline models (Zhou et al., 2019). The study by Zhang et al. (2021a) employ a range of algorithms, including Multi-Layer Perceptron Neural Network, SVM, Gradient Boosting Tree, and RF, to forecast monthly copper prices. Similarly, the study by Baser et al. (2023) employ robust tree-based prediction models such as Decision Tree, Adaptive Boosting, Random Forest (RF), Gradient Boosting, and XGBoost to predict the price of Gold, considering the presence of outliers commonly observed in commodity markets. RNN, particularly variations like LSTM and Convolutional Neural Network

(CNN), are commonly used DL models for commodity price forecasting due to their ability to handle complex sequential data in commodity markets (Yamak et al., 2019). Vidya and Hari (2020) used variants of RNN such as CNN and LSTM to forecast gold price series. It is further extended to syndicate models such as CNN-LSTM, Bi-directional-LSTM, and Bi-LSTM-CNN models by Peng (2021). However, the utilization of hybrid approaches in forecasting precious metal commodities is relatively novel and has received limited exploration.

The main objective of this paper is to enhance the existing literature by establishing a connection between metal commodity futures price point forecasting and the utilization of various ML and DL techniques. Specifically, the study incorporates Stacked-LSTM, ConvLSTM, Bi-LSTM, SVR, XGBoost, and GRU models in this endeavour. These models are effective in analyzing time series data with long-term dependencies, allowing for the extraction of sequential patterns in the data. Additionally, this study considers all available metal futures for trading over a 20-year period from 2003 to 2023. Furthermore, we conduct sub-sample analysis to identify the best model for commodity price prediction by detecting structural breaks in the data.

3. Data and methodology

3.1. Data

The daily future prices of metals (gold, silver, copper, platinum, palladium, and aluminium[2]) from February 2003 to January 2023 were collected from the Yahoo Finance website. The data was directly imported into a Jupyter notebook using the yfinance package and stored as a data frame. Pre-processing steps were applied to handle missing values, and the data were scaled using Min-Max scaling. The applied scaling technique encompassed subtracting the minimum value of each feature and subsequently dividing it by its range. Scaling plays a pivotal role in the construction of machine learning (ML) models due to their sensitivity to the scale of input features, and it facilitates the more efficient convergence of the gradient descent algorithm. Subsequently, the data was transformed into a series of input-output pairs suitable for feeding into the deep learning (DL) models. The data is then split into training and testing datasets and reshaped to a compatible format for input into the DL models. The primary analysis includes a window of 60 days' time periods is chosen as inputs (See Tables 1–3). For robustness checks the study also runs all the models with 30 days' time periods to compare the results (See Tables 4–6).

ML and DL methods have demonstrated effectiveness in addressing time series forecasting problems with complex non-linear relationships among multiple input variables (Shahwan and Odening, 2007). Parameters settings of all the selected models are presented in Appendix-I. Fig. 1 illustrates the normalized closing prices of gold, silver, copper, platinum, and palladium. By examining the price series of different metals in Fig. 1, we observe that palladium exhibits the highest volatility, while platinum has the least. Gold, silver, and copper show similar levels of volatility, with copper slightly higher than the other two It is noteworthy that the volatility levels can vary across different periods, even for the same metal. In financial time series predictions, the level of volatility can significantly impact prediction accuracy. High volatility leads to larger price swings and increased uncertainty, making it challenging for models to identify underlying patterns. This distinct volatility patterns among the metals can also be linked to the patterns related to storage costs or supply disruptions (Raza et al., 2023).

3.2. Methodologies

3.2.1. LSTM model (long -short term memory)

The LSTM model, which is a variation of Recurrent Neural Network (RNNs) proposed by Hochreiter and Schmidhuber (1997), enables the capture of long-term dependencies within time series data. This model excels at modelling non-linear relationships between input and output sequences, rendering it suitable for tackling intricate time-series forecasting challenges. The primary purpose behind developing LSTM was to tackle the vanishing gradient problem encountered by RNNs, which hinders their ability to handle long-term dependencies (Sherstinsky, 2020).

The LSTM architecture, depicted in Fig. 2, incorporates a cell state in addition to hidden states, thereby providing the model with an extended memory of past events. The cell state can be modified by employing gates, with three types of gates existing in an LSTM model: the forgetting gate, input gate, and output gate. The initial step involves determining whether to retain or discard information from the previous timestamp. This process considers the input of h_{t-1} and x_t which subsequently passes through the sigmoid layer, generating values ranging between 0 and 1. If the output is close to 0, the input is considered irrelevant and is not propagated further. Conversely, if the output is close to 1, the input is deemed relevant and continues to propagate through the network.

$$F_t = \sigma\left(W_f * (h_{t-1}, x_t) + b_f\right) \tag{1}$$

The subsequent stage involves determining the relevant information to be incorporated into the cell state, considering the previous hidden state and input data. This process occurs in two parts (See equations (1)–(6)). Firstly, the input gate, comprising a sigmoid layer, determines which values should be updated. The resulting output is then multiplied by the output from the tanh layer, generating a vector of new candidate values represented as $\widetilde{C}_t$. While the sigmoid function ensures that the model remains differentiable, the tanh activation function distributes the gradients as it has a zero cantered range of −1 to 1 thereby mitigating the exploding or vanishing gradient problem and allowing the cell information to flow longer.

$$I_t = \sigma(W_i * (h_{t-1}, x_t) + b_i) \tag{2}$$

$$\widetilde{C}_t = \tanh\left(W_c * (h_{t-1}, x_t) + b_c\right) \tag{3}$$

Subsequently, the old state is multiplied by F_t and add $I_t * \widetilde{C}_t$. This computation yields the new candidate value, which is appropriately scaled based on the decision to update each state value.

$$C_t = F_t * C_{t-1} + I_t * \widetilde{C}_t \tag{4}$$

In the final output layer, the first step involves applying a sigmoid function to determine the inclusion of specific components of the cell state in the output. Following this, the tanh function is applied, resulting in the generation of a new hidden state.

$$O_t = \sigma(W_o * (h_{t-1}, x_t) + b_o) \tag{5}$$

$$H_t = O_t * \tanh C_t \tag{6}$$

Furthermore, LSTM models find applications in various domains, such as machine translation (Zhou et al., 2016), speech recognition (Han et al., 2017), video tagging (Gao et al., 2017), and the forecasting of co-movements in cryptocurrencies (Maiti et al., 2020; Ranjan et al., 2023).

3.2.2. Stacked-LSTM (Stacked Long-Short Term Memory)

The Stacked-LSTM model, depicted in Fig. 3, differs from the original LSTM model in that it has multiple hidden LSTM layers, each containing multiple memory cells. Due to its increased complexity and depth compared to the original model, this advanced model enables us to achieve even more accurate forecasts. In this architecture, the output of one LSTM layer serves as the input for the subsequent LSTM layer. Each LSTM layer possesses its own distinct set of parameters, encompassing the three gates: the input gate, forget gate, and output gate. These gates play a crucial role in regulating the information flow within the layer.

[2] Aluminium data is only available from 2013 onwards.

Table 1
RMSE, MAE, and MAPE Scores for different models across metals for full sample data for 60-days' input.

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		Stacked LTSM	Convolutional-LSTM	SVR	Bidirectional LSTM	XG Boost	GRU
Gold	RMSE	17.85	30.28	21.54	23.85	19.86	18.08
	MAE	13.86	25.91	16.79	19.65	15.52	14.16
	MAPE	0.74%	0.81%	0.93%	1.01%	0.86%	0.79%
Silver	RMSE	0.90	0.49	0.48	0.44	0.51	0.54
	MAE	0.77	0.37	0.38	0.34	0.39	0.42
	MAPE	3.49%	2.11%	1.76%	1.52%	1.82%	1.56%
Copper	RMSE	0.09	0.10	0.10	0.07	0.09	0.14
	MAE	0.07	0.08	0.08	0.05	0.07	0.12
	MAPE	1.47%	1.41%	2.02%	1.56%	1.89%	1.44%
Platinum	RMSE	19.30	21.95	20.70	21.11	20.17	18.85
	MAE	15.22	17.78	16.41	16.47	16.18	14.92
	MAPE	1.69%	1.73%	1.73%	1.57%	1.71%	3.33%
Palladium	RMSE	89.06	76.33	108.45	70.62	81.21	77.57
	MAE	69.72	57.91	79.11	53.54	61.79	59.81
	MAPE	4.09%	3.00%	5.20%	2.58%	3.01%	4.08%
Aluminium	RMSE	58.54	52.53	56.92	67.75	51.03	49.10
	MAE	44.89	41.37	45.92	56.92	40.63	37.55
	MAPE	1.59%	1.76%	1.94%	1.70%	1.72%	1.69%

Table 2
RMSE, MAE and MAPE Scores for different models across metals for sub sample data (2003–13) for 60-days' input.

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		Stacked LTSM	Convolutional-LSTM	SVR	Bidirectional LSTM	XG Boost	GRU
Gold	RMSE	14.00	53.81	56.92	70.62	17.91	29.00
	MAE	10.47	47.92	45.92	53.54	14.27	26.00
	MAPE	3.00%	5.50%	0.97%	0.73%	0.84%	1.84%
Silver	RMSE	0.59	1.02	0.62	0.64	0.80	0.57
	MAE	0.46	0.87	0.48	0.50	0.61	0.45
	MAPE	1.74%	1.43%	1.52%	2.58%	1.89%	1.62%
Copper	RMSE	0.11	0.04	0.05	0.05	0.04	0.05
	MAE	0.10	0.03	0.04	0.04	0.03	0.04
	MAPE	1.08%	0.84%	1.15%	3.08%	1.02%	2.12%
Platinum	RMSE	19.28	23.73	26.06	19.75	20.79	36.75
	MAE	15.40	19.06	21.06	16.61	16.90	31.47
	MAPE	1.89%	5.35%	1.30%	0.94%	1.05%	5.28%
Palladium	RMSE	32.68	13.87	28.21	15.50	13.35	22.46
	MAE	30.70	10.81	21.81	13.03	10.41	19.94
	MAPE	2.26%	1.49%	3.34%	1.36%	1.58%	1.96%

Table 3
RMSE, MAE, and MAPE Scores for different models across metals for sub sample data (2013–23) for 60-days' input.

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		Stacked LTSM	Convolutional-LSTM	SVR	Bidirectional LSTM	XG Boost	GRU
Gold	RMSE	26.55	29.82	31.61	47.54	24.36	25.92
	MAE	22.94	26.40	26.54	44.09	18.19	22.56
	MAPE	0.88%	1.12%	1.52%	0.73%	1.05%	0.77%
Silver	RMSE	0.45	0.45	0.68	0.46	0.44	0.54
	MAE	0.35	0.35	0.56	0.35	0.34	0.42
	MAPE	1.92%	1.79%	2.66%	3.18%	1.64%	1.67%
Copper	RMSE	0.14	0.15	0.14	0.06	0.07	0.06
	MAE	0.12	0.13	0.11	0.04	0.06	0.03
	MAPE	1.30%	1.40%	3.10%	1.86%	1.64%	1.63%
Platinum	RMSE	23.70	27.20	21.78	19.06	20.74	19.27
	MAE	18.66	21.58	17.32	15.49	16.67	15.05
	MAPE	1.66%	1.73%	1.80%	1.59%	1.74%	2.38%
Palladium	RMSE	73.66	136.08	89.90	90.33	69.54	59.05
	MAE	59.33	120.52	70.36	75.00	56.27	47.98
	MAPE	4.47%	2.80%	3.69%	2.80%	2.95%	2.52%

Through this, the model attempts to learn the hierarchical representation of data where the lower layers of the model capture the simple patterns and the higher layers of the model capture the complex temporal patterns by building on the simple patterns.

Stacked-LSTMs can learn more complex representations of the data as each layers captures different aspect of the input data and therefore are more prone to overfitting. To avoid over fitting and to improve generalization, dropout layers can be added in between the LSTM layers. Studies show that deep-LSTM models with one LSTM layer built on top of another can significantly augment the learning capability of neural networks (Cui et al., 2018). Such models have emerged as a prevalent approach for accurately predicting stock market prices (Cao et al., 2019)

Table 4
RMSE, MAE and MAPE Scores for different models across metals for full sample data for 30-days' input.

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		Stacked LTSM	Convolutional-LSTM	SVR	Bidirectional LSTM	XG Boost	GRU
Gold	RMSE	17.67	33.16	18.84	18.17	18.47	22.24
	MAE	13.64	29.12	14.25	14.01	14.25	18.42
	MAPE	0.76%	1.62%	0.79%	0.78%	0.79%	1.03%
Silver	RMSE	0.59	0.69	0.43	0.42	0.51	0.47
	MAE	0.47	0.59	0.32	0.33	0.38	0.37
	MAPE	2.14%	2.78%	1.51%	1.52%	1.76%	1.73%
Copper	RMSE	0.07	0.07	0.07	0.06	0.10	0.07
	MAE	0.06	0.05	0.06	0.05	0.07	0.05
	MAPE	1.52%	1.45%	1.54%	1.33%	1.88%	1.46%
Platinum	RMSE	29.32	19.58	19.58	20.76	20.11	26.03
	MAE	23.85	15.90	15.68	16.39	16.34	20.88
	MAPE	2.51%	1.67%	1.65%	1.73%	1.73%	2.17%
Palladium	RMSE	69.52	105.53	92.42	91.29	80.00	68.22
	MAE	53.13	81.07	66.97	74.17	59.77	51.97
	MAPE	2.58%	4.01%	3.19%	3.60%	2.92%	2.55%
Aluminium	RMSE	58.40	54.02	52.10	50.41	49.82	58.31
	MAE	46.82	43.17	39.85	40.00	39.40	44.00
	MAPE	1.98%	1.84%	1.68%	1.70%	1.66%	1.84%

Table 5
RMSE, MAE and MAPE Scores for different models across metals for sub sample data (2003–13) for 30-days' input.

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		Stacked LTSM	Convolutional-LSTM	SVR	Bidirectional LSTM	XG Boost	GRU
Gold	RMSE	75.49	175.67	15.75	16.96	15.99	13.49
	MAE	74.28	174.82	12.32	13.70	12.38	10.44
	MAPE	4.38%	10.30%	0.72%	0.81%	0.73%	0.62%
Silver	RMSE	0.57	0.58	0.56	0.53	0.66	0.75
	MAE	0.32	0.48	0.43	0.41	0.52	0.61
	MAPE	1.43%	1.49%	1.35%	1.31%	1.65%	1.94%
Copper	RMSE	0.14	0.06	0.03	0.04	0.05	0.07
	MAE	0.13	0.05	0.03	0.03	0.04	0.06
	MAPE	3.73%	1.59%	0.84%	0.86%	1.13%	1.85%
Platinum	RMSE	22.53	21.59	20.73	23.06	14.71	24.07
	MAE	18.86	17.80	16.79	19.52	11.06	18.43
	MAPE	1.16%	1.10%	1.04%	1.21%	1.67%	1.14%
Palladium	RMSE	12.28	11.03	12.02	12.92	23.52	19.53
	MAE	10.23	9.04	9.45	10.89	18.60	16.67
	MAPE	1.52%	1.37%	1.43%	1.63%	1.15%	2.52%

Table 6
RMSE, MAE and MAPE Scores for different models across metals for sub sample data (2013–23) for 30-days' input.

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		Stacked LTSM	Convolutional-LSTM	SVR	Bidirectional LSTM	XG Boost	GRU
Gold	RMSE	24.44	19.12	18.54	17.72	21.79	21.20
	MAE	20.97	15.39	14.91	13.69	16.71	16.90
	MAPE	1.21%	0.88%	0.86%	0.78%	0.96%	0.96%
Silver	RMSE	0.47	0.48	0.46	0.44	0.51	0.54
	MAE	0.37	0.37	0.36	0.34	0.40	0.41
	MAPE	1.78%	1.81%	1.73%	1.63%	1.93%	1.97%
Copper	RMSE	0.06	0.06	0.08	0.06	0.08	0.06
	MAE	0.04	0.04	0.06	0.04	0.06	0.05
	MAPE	1.33%	1.34%	1.77%	1.29%	1.88%	1.42%
Platinum	RMSE	32.31	19.35	20.23	19.54	19.77	35.13
	MAE	26.71	15.44	15.82	16.02	15.36	31.10
	MAPE	2.77%	1.60%	1.65%	1.67%	1.61%	3.25%
Palladium	RMSE	63.03	59.34	66.77	58.76	63.60	64.65
	MAE	52.02	47.72	53.48	47.35	50.67	52.72
	MAPE	2.66%	2.45%	2.76%	2.43%	2.64%	2.72%

and have found applications in various domains, including index, forex, and commodity price prediction (Sezer et al., 2020).

3.2.3. GRU (Gated Recurrent Unit)

The GRU, proposed by Chung et al. (2014), is a variant of RNN. Like LSTM, GRU does not include an output gate. Nevertheless, GRU exhibits a lower parameter count compared to LSTM, leading to decreased memory consumption and improved computational speed. Empirical studies have demonstrated that GRUs exhibit superior performance compared to conventional RNNs (Chung et al., 2014).

Within the depicted architecture in Fig. 4, the GRU employs a solitary update gate for determining the pertinent information to retain,

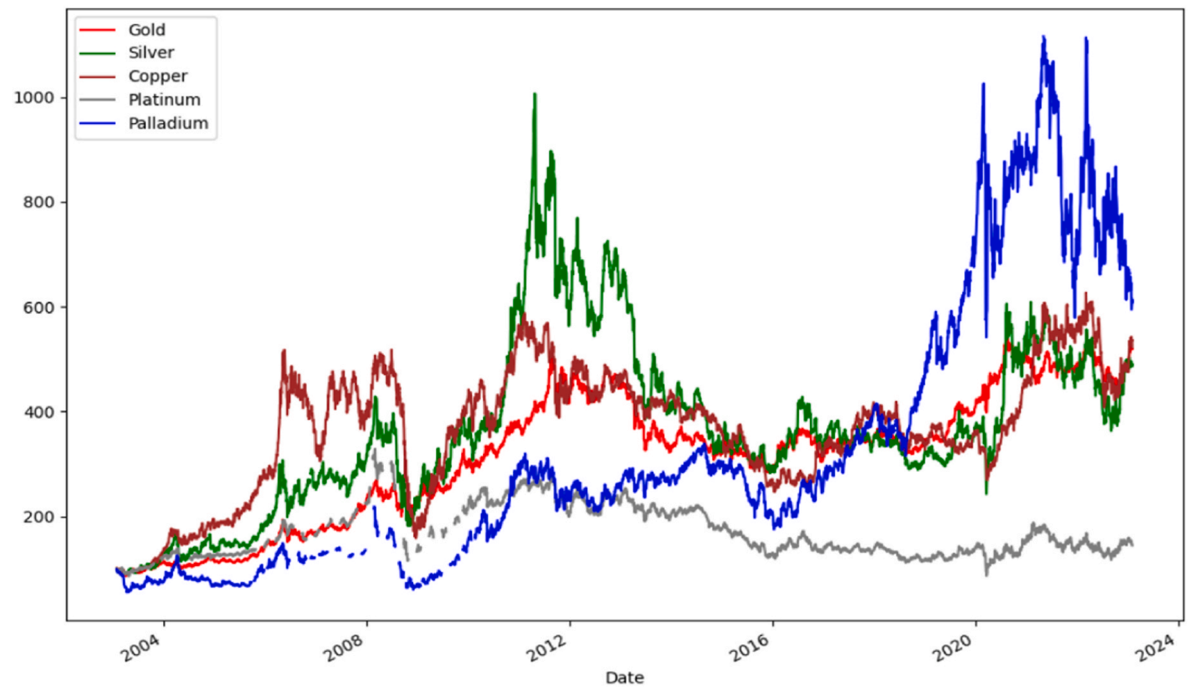

Fig. 1. Normalized prices for commodity futures [Note: Aluminium is not included as the data for Aluminium is only available from 2013 onwards.

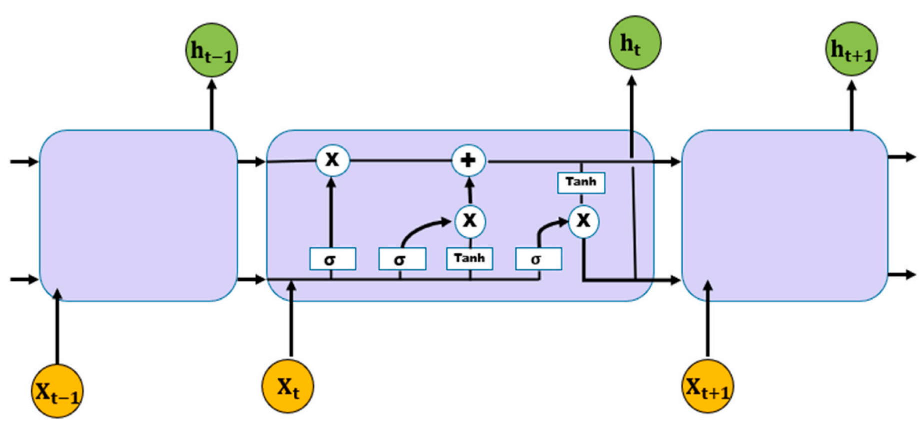

Fig. 2. The architecture of LSTM.

along with a reset gate to determine the information to discard (See equations (7)–(9)). Despite its simpler architecture compared to LSTM, the GRU model has demonstrated competitive performance and generally yields superior outcomes. This advantage arises from the presence of gates in GRU, particularly the update and reset gates, which facilitate the selective retention or omission of information. This selective processing of information proves particularly advantageous when addressing tasks characterized by longer dependencies.

$$U_t = \sigma(W_u * (h_{t-1}, x_t)) \tag{7}$$

$$R_t = \sigma(W_r * (h_{t-1}, x_t)) \tag{8}$$

$$H_t = (1 - U_t) * H_{t-1} + U_t * \tanh(W[R_t * H_{t-1}, U_t]) \tag{9}$$

GRU also addresses the issue of vanishing gradients encountered by vanilla recurrent neural networks. In such networks, the gradient diminishes over time during the backpropagation process, rendering the neural networks untrainable as it becomes too small to significantly impact the learning process. GRUs find applications in multivariate time series analysis, as well as in predicting financial indicator values such as price-to-sales ratio, return on tangible assets, and price-earnings ratio (Yang, 2021).

3.2.4. Bi-LSTM (bidirectional Long-Short Term Memory)

Bi-LSTM represents a variant of bidirectional RNN. In contrast, the LSTM cells, as illustrated in Fig. 5, can solely process the data in either the forward or backward direction.

Given its ability to process data simultaneously in both forward and backward directions, the Bi-LSTM model proves advantageous in capturing dependencies between past and future prices, thereby facilitating accurate predictions of future price movements. This model receives a sequence of past prices as input and predicts the subsequent price in the sequence. The input sequence is processed through separate LSTM layers for the forward and backward directions. The final prediction is obtained by concatenating the outputs of each LSTM layer and passing them through a dense layer equipped to deal with nonlinearity.

The Bi-LSTM model has demonstrated outstanding performance in various domains, including phoneme classification (Graves and Schmidhuber, 2005), speech recognition (Graves et al., 2013), and language modelling (Sundermeyer et al., 2012). Furthermore, Bi-LSTMs

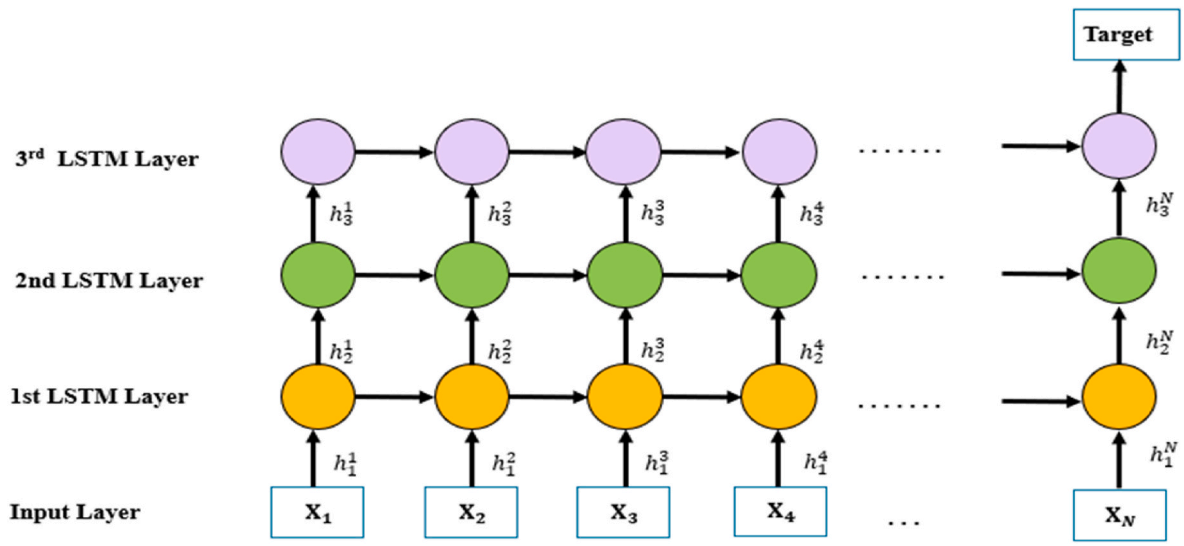

Fig. 3. Architecture of Stacked-LSTM.

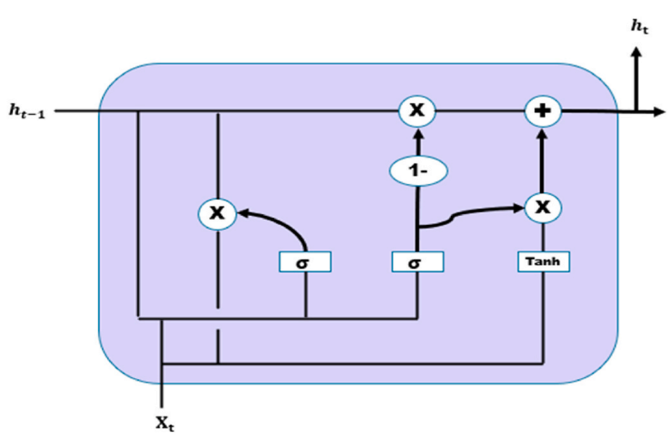

Fig. 4. Architecture of GRU

find application in financial time series modelling for predicting stock and index prices (Siami-Namini et al., 2019).

3.2.5. ConvLSTM (convolutional long short-term memory)

ConvLSTM represents a variant of RNN that was first proposed by Shi et al. (2015), designed specifically to handle spatiotemporal data. It achieves this by combining convolutional layers with long-short term layers. By incorporating this combination, the model can effectively capture the temporal dependencies as well as the spatial relationships inherent in the data. Consequently, ConvLSTM can identify the connections between different variables across various spatial areas, thus enhancing its ability to analyze complex spatiotemporal patterns.

The ConvLSTM model comprises two main components: Convolutional layers and LSTM layers (refer to Fig. 6). The convolutional layers function similarly to those in standard convolutional neural networks, extracting features from the input data using learnable filters. Nevertheless, traditional convolutional neural networks tend to focus on extracting temporal features, whereas ConvLSTM models integrate both convolutional and LSTM layers to extract spatial as well as temporal features. As a result, ConvLSTM models are widely employed in multivariate time series forecasting tasks to investigate the spatial-temporal correlations inherent in the data (Xiao et al., 2021).

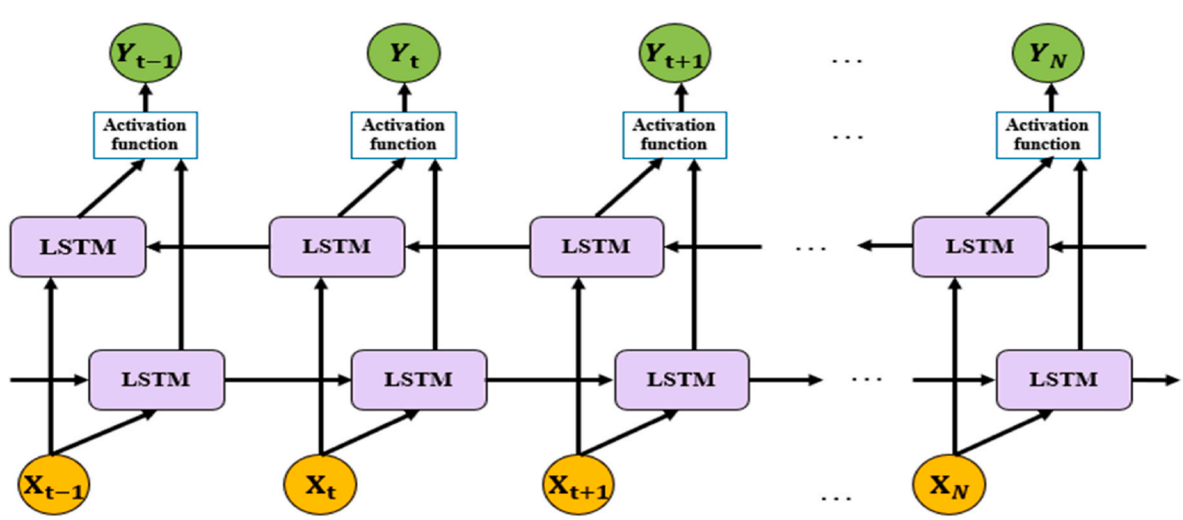

Fig. 5. The architecture of Bidirectional-LSTM.

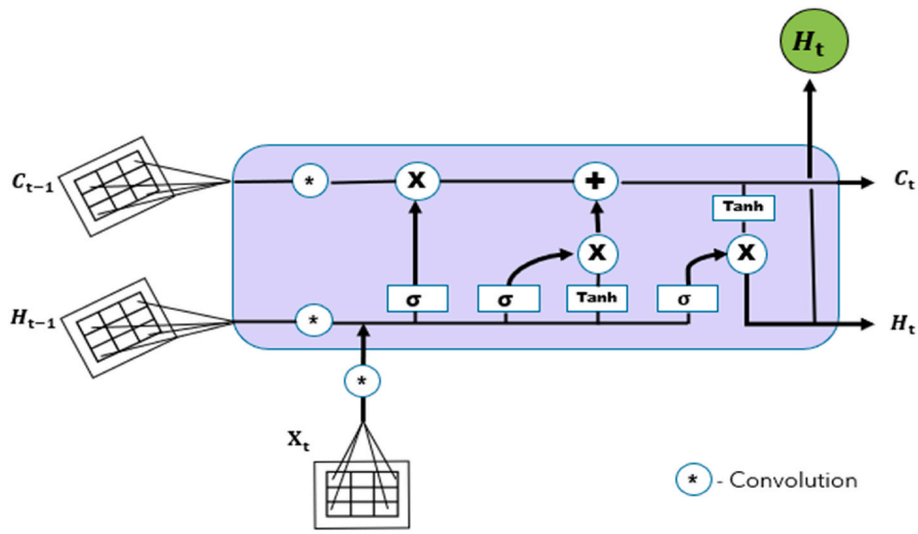

Fig. 6. The architecture of ConvLSTM.

3.2.6. SVM (Support Vector Regressor model)

The SVR is a variant of the SVM algorithm introduced by Drucker et al. (1996) and is commonly employed for regression analysis. Unlike standard regression techniques that attempt to fit a line to the data points, SVR aims to find a non-linear function for prediction. To accomplish this, SVR utilizes a mapping of variables to a higher-dimensional space, enabling their separation by means of a hyperplane. The primary goal of SVR is to determine a hyperplane that minimizes the errors, within a predetermined threshold (see Fig. 7).

SVR has broad applications in time series forecasting across diverse industries, encompassing the stock market, banking sector, and the prediction of volatility (Chen et al., 2023). It is also employed as a trend-based segmentation method. In SVR, data is provided as input-output pairs, and predictions can be made for future time steps. This offers a significant advantage over traditional regression techniques since SVR enables the capture of complex non-linear relationships between past and future time steps.

In the context of SVR modelling, the choice of a suitable kernel function is an important step. There are four kinds of kernel functions available: Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. The Linear kernel, being the simplest, is suitable for cases where a linear relationship exists or when the input data can be separated by a straight line. On the other hand, the Polynomial kernel is employed when there is a non-linear relationship in the data and the input cannot be separated by a single line. The RBF Kernel is the most commonly utilized kernel function as it allows for capturing intricate non-linear relationships in the data. Conversely, the Sigmoid kernel is less frequently used in SVR and finds its primary application in modelling neural networks.

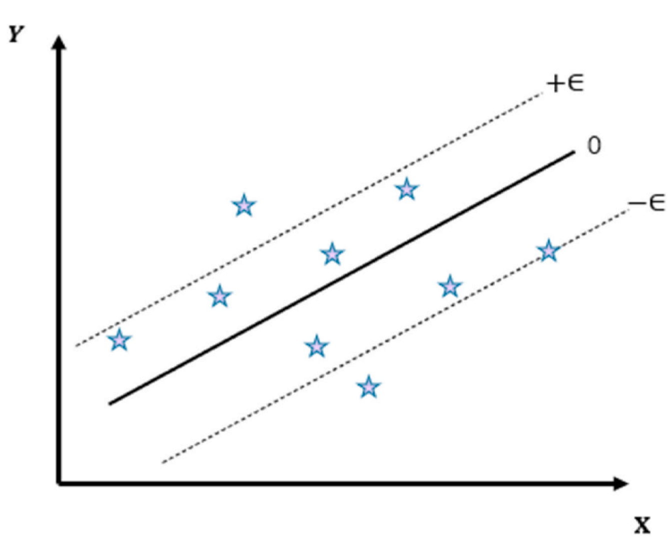

Fig. 7. Illustration of Support Vector Regressor function.

3.2.7. XGBoost

XGBoost is an ensemble learning method (see Fig. 8). Boosting involves combining multiple weak learners to create a robust learner by training a series of weak models on the errors made by previous models. XGBoost assigns higher weightage to the errors made by the previous model, allowing for subsequent correction and leading to improved performance and higher accuracy. XGBoost demonstrates a particular suitability for time series forecasting tasks as it possesses the capability to effectively handle temporal dependencies among variables while identifying non-linear relationships that exist between them. Compared to standard ML models, XGBoost offers easier handling of missing data. The model is an ensemble of decision trees built in a depth-wise manner, wherein these models are sequentially trained to address errors made by the previous weak learners. The algorithm incorporates various optimization and regularization techniques, which contribute to its power, flexibility, and popularity in the field of ML. XGBoost has found applications in the financial time series domain for forecasting stock market volatility, predicting commodity prices (Jabeur et al., 2021; Baser et al., 2023), and imputing time series data.

4. Results and discussion

4.1. Evaluation measures

To assess the performance of these machine learning (ML) and deep learning (DL) models, a comparative analysis is conducted among Stacked-LSTM, ConvLSTM, Support Vector Regressor (SVR), Bi-LSTM, Extreme Gradient Boosting (XGBoost), and Gated Recurrent Unit (GRU). We employ metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) to check for the forecasting performance of these models. These metrics facilitate model comparison to identifying the most suitable model. RMSE is one of the most adopted metrics to assess models' performance. It quantifies the squared deviation between the observed and predicted values and then calculates the square root of the mean of these squared deviations as shown in equation (10) (Astudillo et al., 2020;

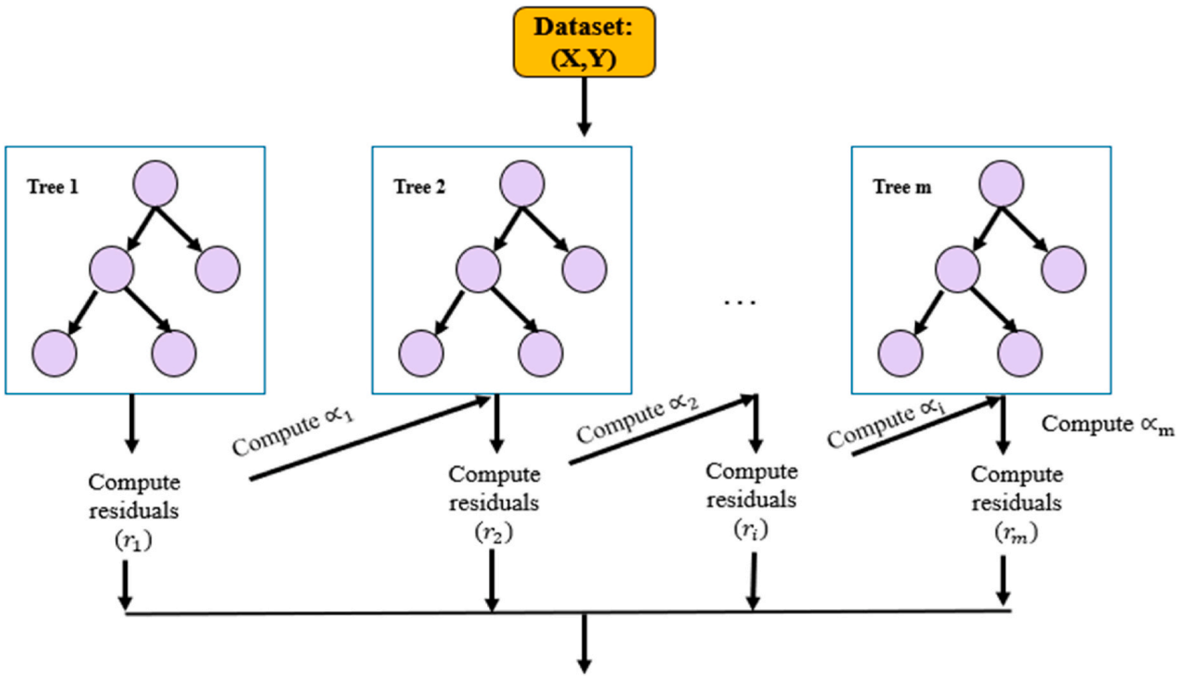

Fig. 8. Architecture of XGBoost.

Baser et al., 2023).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y - \dot{y}_i)}{n}} \tag{10}$$

One of the major reasons for choosing RMSE as the evaluation metric because it serves as a loss function for optimization in the neural networks that we have trained. Therefore, it allows us to directly evaluate the predictive quality across different models. Furthermore, RMSE applies a greater penalty to large errors, thereby emphasizing the importance of accurately capturing deviations from the actual values. For time series forecasting performance, a lower RMSE score is better.

MAE is an alternative metric that offers greater robustness to outliers compared to RMSE, as it does not involve squaring the errors. MAE computes the mean absolute discrepancy between the observed and predicted values, providing a straightforward measure of the overall discrepancy between them as shown in equation (11).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | y - y_i | \tag{11}$$

MAE is an easily interpretable metric that is less affected by outliers, as it does not penalize extreme values as severely as RMSE does (Chai and Draxler, 2014). A lower MAE indicates better accuracy, as it signifies that the forecasted values are, on average, closer to the actual values.

Along with MAE and RMSE, we also consider MAPE. It is a metric used to evaluate the accuracy of forecasts by measuring the average percentage difference between the forecasted values and the corresponding actual values. MAPE gives a sense of how well a forecasting model's predictions align with the actual data in terms of percentage deviations as shown in equation (12).

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{y - y_i}{y} * 100 \tag{12}$$

MAPE is expressed as a percentage, and a lower MAPE value indicates better accuracy. However, it is important to note that MAPE can encounter issues when dealing with small or zero actual values, as the percentage difference might become undefined or excessively large.

MAE, RMSE, and MAPE each possess their own set of advantages and disadvantages. Consequently, it is imperative to employ these metrics collectively to achieve a comprehensive assessment of the models' performances (Baser et al., 2023).

4.2. Findings and discussion

The study discovered that RMSE, MAE, and MAPE yield similar results regarding the performance of the models in most of the cases, as shown in Table 1.

The RMSE, and MAE values for the following metals (Gold, Platinum, Palladium, and Aluminium) are high. The lower MAPE values for gold (0.74%), and aluminium (1.59%) denote superiority of stacked LSTM over other models for forecasting their future prices. Similarly, Bidirectional LSTM yields better forecasting estimates for silver (1.52%), platinum (1.57%), and palladium (2.58%). Then, Convolution LSTM finds to be better forecasting model for forecasting the copper (1.41%) future prices. Notably, the GRU finds to be the second-best performing models for forecasting the future prices of following metals namely gold (0.79%), silver (1.56%), copper (1.44%), and aluminium (1.69%). The SVR model finds to be the least performing for copper (2.02%), palladium (5.20%), and aluminium (1.94%).

This implies that the best or least-performing models differ depending on the specific metal being analyzed. Given the varying results exhibited by different ML models for different metals within the same sample period, we proceed to conduct a sub-sample analysis to assess whether this variation persists within the same metal but with a different sample period. To accomplish this, we divide the full sample data into two equal periods: February 2003 to January 2013 and February 2013 to January 2023. This division is motivated by the presence of a structural break in the future prices of metals during early 2013. However, due to limited aluminium future price data, we exclude aluminium from the sub-sample analysis.

Table 2 displays the outcomes of the initial sub-sample analysis. Here some estimates differ from the full sample estimations as follows. The Bidirectional LSTM model MAPE value finds to be best for forecasting gold future prices. Whereas Convolutional LSTM MAPE value shows the worst performance for forecasting gold future prices among all the models. Similarly, for forecasting the silver future prices the estimates show vice-versa to gold estimates. Then for forecasting the copper future prices Bidirectional LSTM model finds to be the worst among all the models. These findings highlight the variation in model performance across different metals and different data samples. No single model consistently outperforms the others for any metal. Next, we aim to investigate whether similar observations persist in the second sub-sample.

Table 3 exhibits the second sub-sample analysis results. In the second sub-sample, we observe relatively more variation in relation to the performance of the model with the full sample as follows. Based on the MAPE values Bidirectional LSTM (0.73%), XG Boost (1.64%), Stacked LSTM (1.30%), and GRU (2.52%) find to be the best-performing models for forecasting future prices of the metal's gold, silver, copper, and palladium respectively. Overall, our observations indicate three important findings: (i) the best-performing model varies across metals, (ii) the choice of data period significantly impacts the models' performance, and (iii) inputs to the models.' These findings imply that there is no consensus on the best model for predicting metals' future prices. Assigning higher importance to recent data when utilizing these models for price prediction may lead to improved performance. However, none of the particular model able to maintain constant performance over time for predicting metals' future prices except in the case of forecasting the future prices for platinum (60 days inputs). Irrespective of the sample periods, Bidirectional LSTM, and GRU models are the best and worst performance models for forecasting the future prices for platinum with 60 days inputs. However, the model performance varies significantly with 30 days input even in case of platinum. This indicates that market efficiency is not a binary concept but exists along a spectrum. Thus, different assets may exhibit varying degrees of efficiency with time. This inherent challenge in making accurate predictions suggests that market participants would find it difficult to consistently outperform the market in predicting metals' future prices. Hence, our observations provide empirical support for the efficient market hypothesis (Fama, 1970). That suggests that it is hard to beat the market constantly with a single strategy. The forecasts for all the considered series using different techniques are visualized in Appendix II. The estimates with 30 days' inputs are shown in Tables 4–6. The obtained estimates show that the best and worst-performing models for forecasting metal future prices vary significantly from 60 days as inputs. It confirms that ML and DL models' performance can vary based on sample periods and inputs due to a combination of factors related to data, model complexity, and the underlying characteristics of the problem being solved.

The consistency of the best-performing model across different periods and choice of metals is not evident in our analysis as shown in Fig. 9. It shows that in most of the cases MAPE values are less than 5% that overall confirm the superiority of ML and DL models under consideration in forecasting the future metal prices. It also confirms that each model's performance under consideration on forecasting the future prices of the metals greatly varies with choice of metals, period, and inputs. While metals like gold and silver are considered safe-haven assets offering portfolio diversification, their speculative nature cannot be ignored (Janani Sri et al., 2022). Additionally, the price variations among different metals may not be directly linked, but they often move in the same direction, which can impact price prediction and model performance. Explaining the variations in model performance empirically is a challenge, as they can be influenced by multiple factors.

The key reasons for such variability or uncertainty in the model performances is due to the following. First, timeseries data distribution changes over a period. Second, variability in data distribution results in concept drift and structural breaks. Due to which there are deviations in characteristics of the underlying relationship between features and target variable over time. Third, dissimilar input features might be relevant during different time periods. Fourth, there could be demand and supply variations in different metals. Fifth, different models have their own strengths and weaknesses that affect its overall performance. Sixth, various external factors such as geopolitical, technological, legal, and other spheres affect the metal prices. Seventh, another important factor is "tail risk," which most ML models do not appropriately account for. As a result, they may underestimate the potential for extreme events, leading to inaccurate predictions during periods of heightened volatility or market stress. The presence of non-linearity, non-stationarity, and volatility in price series introduces complexity, posing challenges in consistently and accurately predicting metal prices, even when utilizing ML and DL models (Luo et al., 2022). Lastly, varying model performance across metals can also be linked to investor sentiment and biases specific to different metals. The study findings have high relevance for portfolio theory, hedging and risk Management. The variation in the model performance across metals can influence portfolio construction and asset allocation decisions. Consequently, it can have a direct influence on the effectiveness of hedging strategies and risk management techniques. Thus, to address these issues continuous innovation of models is essential to enhance the accuracy and robustness of forecasting.

5. Conclusion

Commodity markets play a crucial role in the global economy as they provide essential raw materials for various industries, with a significant portion of commodities traded on the futures market (Chang and Fang, 2022). Among the commodities, metals are particularly important and often considered safe-haven assets during economic crises (Peng, 2021). This study focuses on forecasting metal futures (e.g., gold, silver, copper, platinum, palladium, and aluminium) in commodity markets using different machine learning (ML) and deep learning (DL) techniques. The

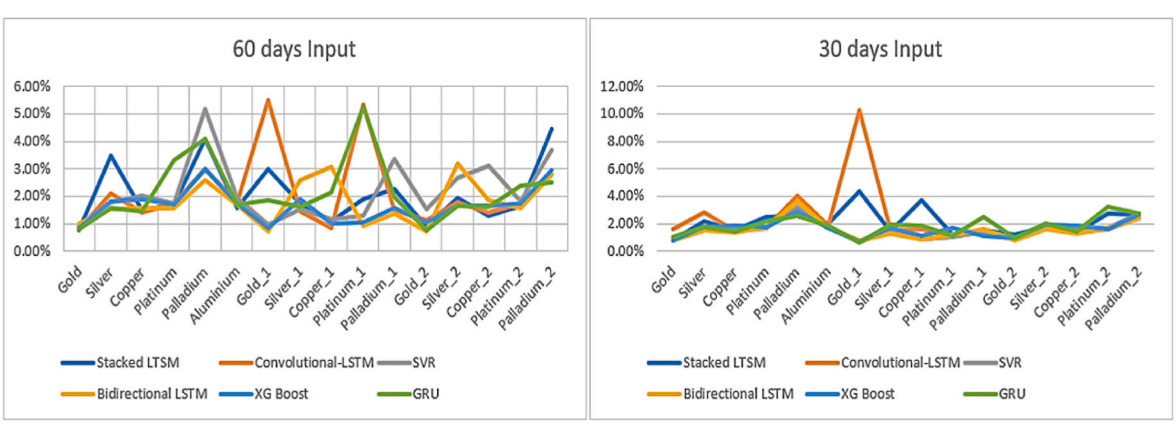

Fig. 9. MAPE values of the different metals across the various metals with different inputs
Note: Metal with no subscript represents full sample; subscript_1 represent subsample (2003–2013); and subscript_2 represent subsample (2013–2023).

models utilized in this study comprises of the Stacked-LSTM, Convolutional LSTM, Support Vector Regressor (SVR), Bidirectional LSTM, Extreme Gradient Boosting (XGBoost), and Gated Recurrent Unit (GRU). The analysis is conducted in two parts. First, the models are trained on 20 years of data, from February 2003 to January 2023. Second, the dataset is further bifurcated into two subsamples each of ten years for robust analysis. The models' performance is assessed by employing evaluation metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

Most of the research in this domain concentrates on forecasting the prices of individual metals (Vidya and Hari, 2020; Yamak et al., 2019), and only a few have performed sub-sample analysis due to limited data availability. This study stands out by considering a unique combination of metal assets and conducting sub-sample analysis, including data normalization and identification of structural breakpoints to determine the sub-sample periods. Additionally, the present study uses different inputs periods of 30 and 60 days respectively for robust analysis. The RMSE, MAE, and MAPE scores indicate that the performance of models varied across different metals, inputs, and periods indicating the absence of a consistently superior model. The results support the efficient market hypothesis and indicate that consistently outperforming the market in predicting metal prices may be challenging. Given the heterogeneity in predictive performance across different metals, and time periods, market participants, including traders and investors, should consider diversifying their investment strategies. The study findings show that conducting comprehensive sub-sample and rolling window analyses (Kayal and Balasubramanian, 2021; Som and Kayal, 2022) is essential before establishing the reliability of various ML and DL models for accurate price predictions.

Future studies in commodity price forecasting should explore several ways to enhance our understanding of this dynamic field. Combining technical and fundamental analysis can provide a more holistic view of commodity market behaviour, incorporating factors like supply and demand, geopolitical influences, technological, legal, and macroeconomic indicators. Therefore, building or validating any theory based on the ML and DL models estimations alone is not advisable. The traditional, ML, and DL models differ in purpose, focus, and methodologies. The future studies may consider hybrid models by the integration of different traditional, ML, and DL for forecasting future prices of the metals. The future studies may also investigate the application of deep reinforcement learning (DRL) for trading strategies in commodity markets holds promise for adaptive and responsive trading approaches.

CRediT authorship contribution statement

Anu Varshini: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Parthajit Kayal:** Conceptualization, Formal analysis, Project administration, Supervision, Validation, Writing – review & editing. **Moinak Maiti:** Conceptualization, Formal analysis, Supervision, Writing – review & editing.

Declaration of competing interest

There is no conflict of interest.

Acknowledgement

Authors are thankful to the University of Witwatersrand, Johannesburg, South Africa for supporting Open Access Publications.

Appendix-I

Table A
Parameters settings of Stacked LTSM, Convolutional-LSTM, Bidirectional LSTM, and GRU

Parameters/Hyperparameters	Stacked LTSM	Convolutional-LSTM	Bidirectional LSTM	GRU
filters		128		
kernel_size		(3, 3)		
strides		(1, 1)		
padding		same'		
activation		tanh'		
units	128	128	128	128
activation	tanh	tanh'	tanh	tanh
recurrent_activation	sigmoid	sigmoid'	sigmoid	sigmoid
use_bias	TRUE	TRUE	TRUE	TRUE
kernel_initializer	glorot_uniform	glorot_uniform	glorot_uniform	glorot_uniform
recurrent_initializer	orthogonal	orthogonal	orthogonal	orthogonal
bias_initializer	zeros	zeros	zeros	zeros
kernel_regularizer	None	None	None	None
recurrent_regularizer	None	None	None	None
bias_regularizer	None	None	None	None
kernel_constraint	None	None	None	None
recurrent_constraint	None	None	None	None
bias_constraint	None	None	None	None
dropout	0	0	0	
recurrent_dropout	0	0	0	

Table B
Parameters settings of SVR and XGBoost Models

XG Boost		SVR	
Parameter/Hyperparameter	Value	Parameter/Hyperparameter	Value
objective	reg: squarederror	kernel	rbf
booster	None	C (Penalty Parameter)	10
n_estimators	1000	epsilon	0.01
learning_rate	0.01	gamma	0.8

(*continued on next page*)

Table B (*continued*)

XG Boost		SVR
max_depth	3	
min_child_weight	1	
subsample	1	
colsample_bytree	1	

Appendix-II

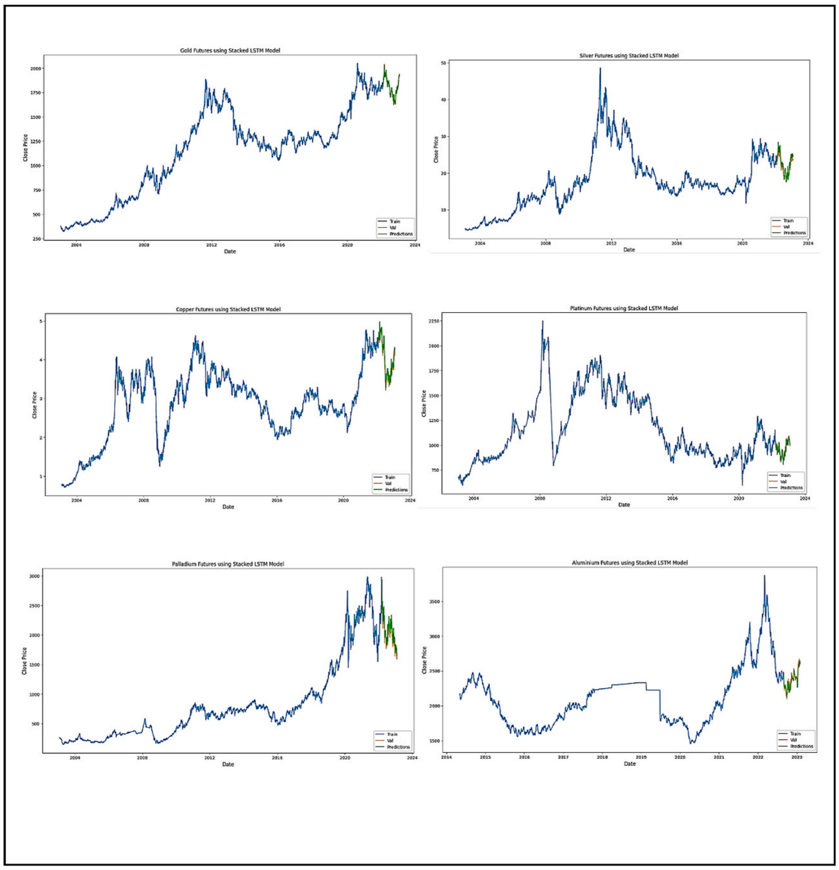

Fig. A. Predictions of prices of Gold, Silver, Copper, Platinum, Palladium and Aluminium through Stacked-LSTM Model.

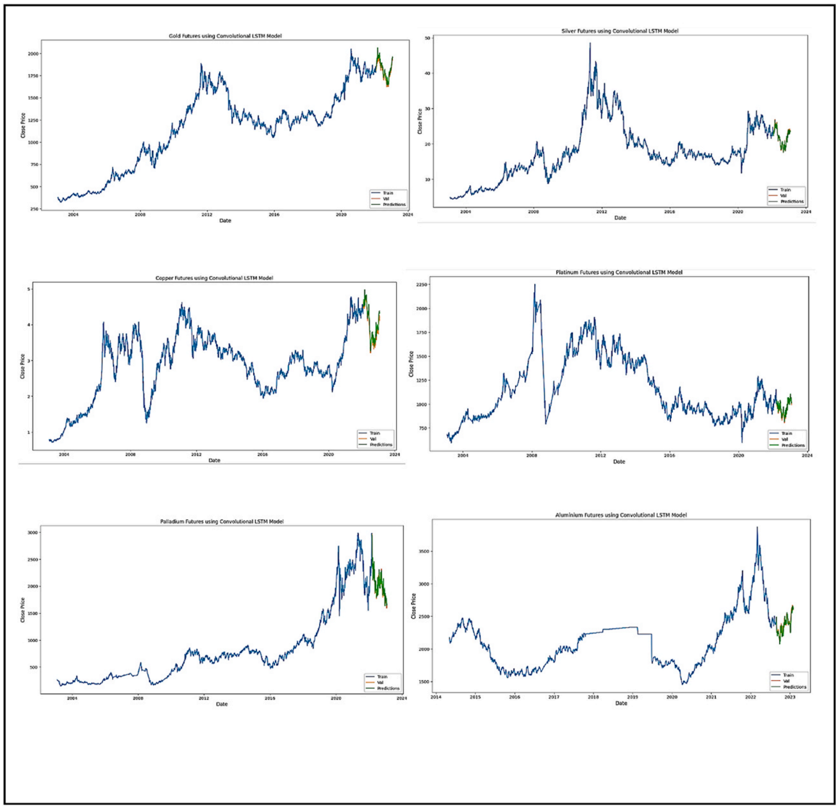

Fig. B. Predictions of prices of Gold, Silver, Copper, Platinum, Palladium and Aluminium through ConvLSTM Model.

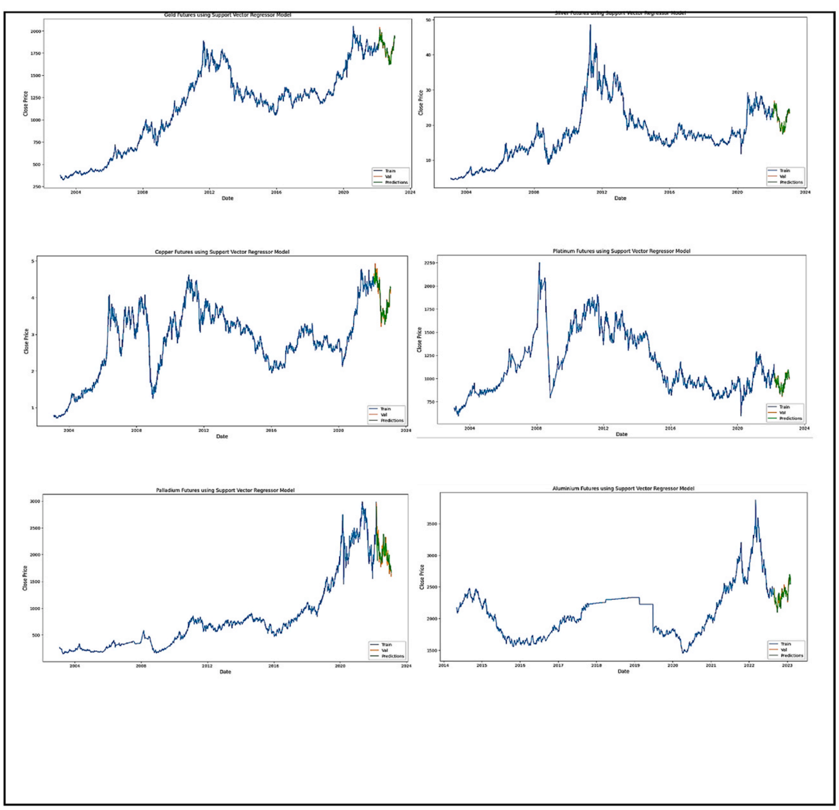

Fig. C. Predictions of prices of Gold, Silver, Copper, Platinum, Palladium and Aluminium through SVR Model.

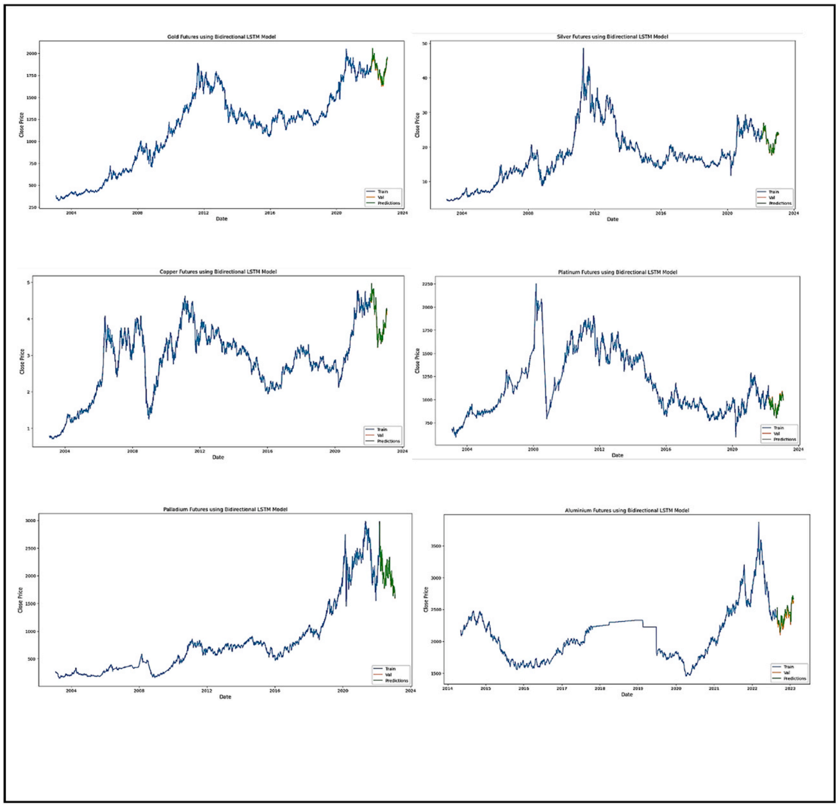

Fig. D. Predictions of prices of Gold, Silver, Copper, Platinum, Palladium and Aluminium through Bi-LSTM Model.

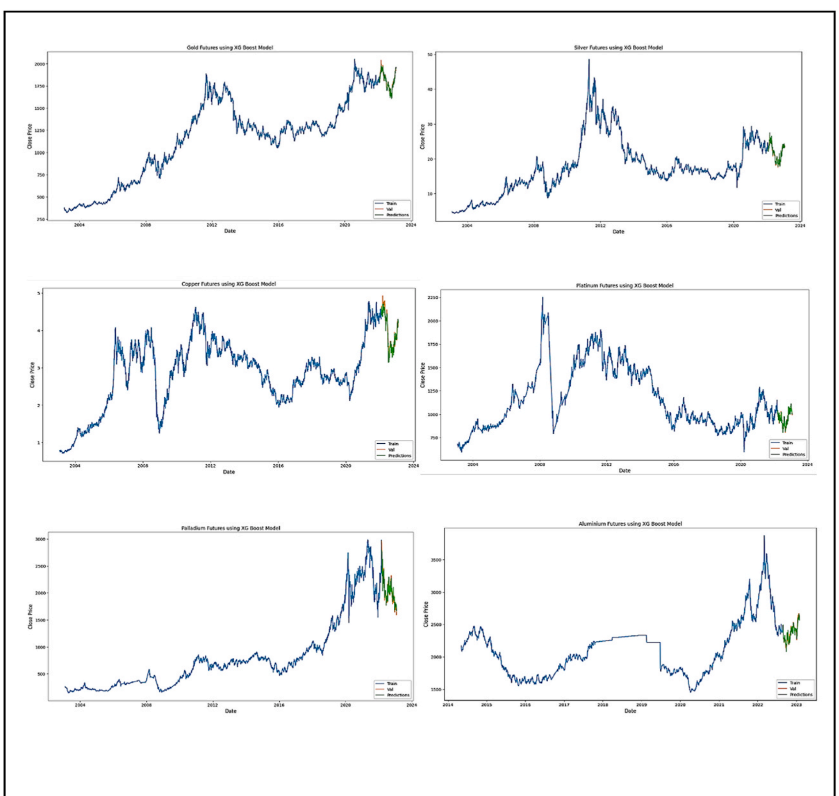

Fig. E. Predictions of prices of Gold, Silver, Copper, Platinum, Palladium and Aluminium through XGBoost Model.

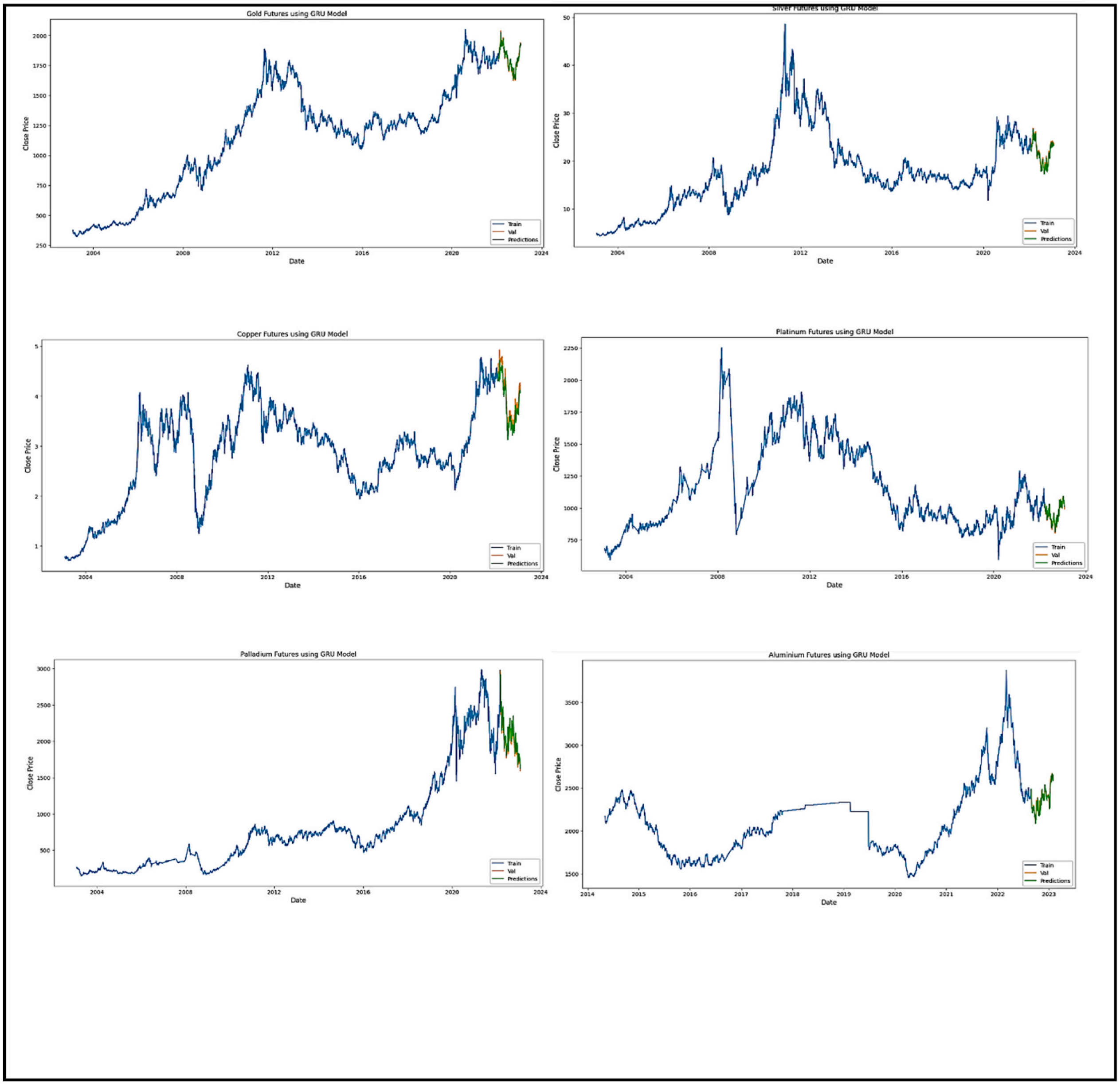

Fig. F. Predictions of prices of Gold, Silver, Copper, Platinum, Palladium and Aluminium through GRU Model.

References

Ampomah, E.K., Qin, Z., Nyame, G., 2020. Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement. Information 11 (6), 332.

Arowolo, W.B., 2013. Predicting stock prices returns using GARCH model. Int. J. Eng. Sci. 2 (5), 32–37.

Astudillo, G., Carrasco, R., Fernández-Campusano, C., et al., 2020. Copper price prediction using support vector regression technique. Appl. Sci. 10 (19), 6648.

Baser, P., Saini, J.R., Baser, N., 2023. Gold commodity price prediction using tree-based prediction models. Int. J. Intel. Syst. Appl. Eng. 11 (1s), 90–96.

Bhardwaj, S.P., Paul, R.K., Singh, D.R., et al., 2014. An empirical investigation of ARIMA and GARCH models in agricultural price forecasting. Econ. Aff. 59 (3), 415–428.

Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. Rev. Financ. Stud. 34 (2), 1046–1089.

Cao, J., Li, Z., Li, J., 2019. Financial time series forecasting model based on CEEMDAN and LSTM. Physica A 519, 127–139.

Cao, L.J., Tay, F.E.H., 2003. Support vector machine with adaptive parameters in financial time series forecasting. IEEE Trans. Neural Network. 14 (6), 1506–1518.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geoscien. model develop 7 (3), 1247–1250.

Chang, C.L., Fang, M., 2022. The connectedness between natural resource commodities and stock market indices: evidence from the Chinese economy. Res. Policy 78, 102841.

Chen, W.J., Yao, J.J., Shao, Y.H., 2023. Volatility forecasting using deep neural network with time-series feature embedding. Econ. Resch.-Ekonomska Istraživanja 36 (1), 1377–1401.

Chen, Z., Goh, H.S., Sin, K.L., et al., 2021. Automated agriculture commodity price prediction system with machine learning techniques. arXiv preprint arXiv: 2106.12747.

Chung, J., Gulcehre, C., Cho, K., et al., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv Preprint arXiv:1412.3555.

Cui, Z., Ke, R., Pu, Z., et al., 2018. Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. arXiv preprint arXiv: 1801.02143.

Desai, V.S., Bharati, R., 1998. A comparison of linear regression and neural network methods for predicting excess returns on large stocks. Anl. Oper. Resch. 78 (0), 127–163.

Drucker, H., Burges, C.J., Kaufman, L., et al., 1996. Support vector regression machines. Adv. Neural Inf. Process. Syst. 9.

Essien, A., Giannetti, C., 2019. A deep learning framework for univariate time series prediction using convolutional LSTM stacked autoencoders. In: 2019 IEEE Int. Symp. Innov. Intelligent SysTem. IEEE, pp. 1–6. *Appl. (INISTA)*.
Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. J. Finance 25 (2), 383–417.
Fama, E.F., 1984. Forward and spot exchange rates. J. Monetary Econ. 14 (3), 319–338.
Gao, L., Guo, Z., Zhang, H., et al., 2017. Video captioning with attention-based LSTM and semantic consistency. IEEE Trans. Multimed. 19 (9), 2045–2055.
Gong, X., Xu, J., 2022. Geopolitical risk and dynamic connectedness between commodity markets. Energy Econ. 110, 106028.
Graves, A., Mohamed, A.R., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 6645–6649.
Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Network. 18 (5–6), 602–610.
Han, S., Kang, J., Mao, H., Hu, Y., et al., 2017. Ese: efficient speech recognition engine with sparse lstm on fpga. In: Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, pp. 75–84.
Herrera, G.P., Constantino, M., Tabak, B.M., et al., 2019. Long-term forecast of energy commodities price using machine learning. Energy 179, 214–221.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neu. Comp. 9 (8), 1735–1780.
Horák, J., Jannová, M., 2023. Predicting the Oil price movement in commodity markets in global economic Meltdowns. Forecasting 5 (2), 374–389.
Hussein, S.F.M., Shah, M.B.N., Abd Jalal, M.R., et al., 2011. Gold price prediction using radial basis function neural network. In: 2011 Fourth International Conference on Modeling, Simulation and Applied Optimization. IEEE, pp. 1–11.
Jabeur, S.B., Mefteh-Wali, S., Viviani, J.L., 2021. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Anl. Oper. Resch 1–21.
Jacks, D.S., O'rourke, K.H., Williamson, J.G., 2011. Commodity price volatility and world market integration since 1700. Rev. Econ. Stat. 93 (3), 800–813.
Janani Sri, S., Kayal, P., Balasubramanian, G., 2022. Can equity be safe-haven for investment? J. Emerg. Mark. Finance 21 (1), 32–63.
Kayal, P., Balasubramanian, G., 2021. Excess volatility in bitcoin: extreme value volatility estimation. IIM Kozhikode Soc. Manag. Rev. 10 (2), 222–231.
Kayal, P., Maheswaran, S., 2021. A study of excess volatility of gold and silver. IIMB Manag. Rev. 33 (2), 133–145.
Kilinc, H.C., Yurtsever, A., 2022. Short-term streamflow forecasting using hybrid deep learning model based on grey wolf algorithm for hydrological time series. Sustainability 14 (6), 3352.
Krollner, B., Vanstone, B.J., Finnie, G.R., 2010. Financial time series forecasting with machine learning techniques: a survey. In: ESANN *(April*.
Kumar, K.A., Spulbar, C., Pinto, P., et al., 2022. Using econometric models to manage the price risk of Cocoa Beans: a case from India. Risks 10 (6), 115.
Li, L., Wu, Y., Ou, Y., et al., 2017. Research on machine learning algorithms and feature extraction for time series. In: 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). IEEE, pp. 1–5.
Li, Y., Bao, T., Gong, J., et al., 2020. The prediction of dam displacement time series using STL, extra-trees, and stacked LSTM neural network. IEEE Access 8, 94440–94452.
Lim, B., Zohren, S., 2021. Time-series forecasting with deep learning: a survey. Philosop. Transac. Royal Societ. A 379 (2194), 20200209.
Livieris, I.E., Pintelas, E., Pintelas, P., 2020. A CNN–LSTM model for gold price time-series forecasting. Neural Comput. Appl. 32, 17351–17360.
Luo, H., Wang, D., Cheng, J., et al., 2022. Multi-step-ahead copper price forecasting using a two-phase architecture based on an improved LSTM with novel input strategy and error correction. Res. Policy 79, 102962.
Maiti, M., Vukovic, D., Vyklyuk, Y., et al., 2022. BRICS capital markets Co-movement analysis and forecasting. Risks 10 (5), 88.
Maiti, M., Vyklyuk, Y., Vuković, D., 2020. Cryptocurrencies chaotic co-movement forecasting with neural networks. Intern. Tech. Letters 3 (3), e157.
Makala, D., Li, Z., 2021. Prediction of gold price with ARIMA and SVM. In Journal of Physics: Conference Series. 1767 (1), 012022.
Nikou, M., Mansourfar, G., Bagherzadeh, J., 2019. Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. Intell. Syst. Account. Finance Manag. 26 (4), 164–174.
Ojo, S.O., Owolawi, P.A., Mphahlele, M., et al., 2019. Stock market behaviour prediction using stacked LSTM networks. In: 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC). IEEE, pp. 1–5.
Onour, I., Sergi, B.S., 2011. Modeling and forecasting volatility in the global Food commodity prices. Agri. Econ. -Czech 57 (3), 132–139.
Parmezan, A.R.S., Souza, V.M., Batista, G.E., 2019. Evaluation of statistical and machine learning models for time series prediction: identifying the state-of-the-art and the best conditions for the use of each model. Inf. Sci. 484, 302–337.
Peng, Z., 2021. Precious Metal Price Prediction Using Deep Neural Networks.
Ranjan, S., Kayal, P., Saraf, M., 2023. Bitcoin price prediction: a machine learning sample dimension approach. Comput. Econ. 61, 1617–1636.
Raza, S.A., Masood, A., Benkraiem, R., Urom, C., 2023. Forecasting the volatility of precious metals prices with global economic policy uncertainty in pre and during the COVID-19 period: novel evidence from the GARCH-MIDAS approach. Energy Econ. 120, 106591.
Reddy, V.K.S., 2018. Stock market prediction using machine learning. Int. Res. J. Eng. Tech. 5 (10), 1033–1035.
Schnepf, R.D., Resources, Science, Division, Industry, 2005. Price determination in agricultural commodity markets: a primer. Cong. Res. Serv. Library of Congress. (December).
Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M., 2020. Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. App. Soft Comput. 90, 106181.
Shahwan, T., Odening, M., 2007. Forecasting agricultural commodity prices using hybrid neural networks. Comp. Intel. Econ. Fin. 2, 63–74.
Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys. Nonlinear Phenom. 404, 132306.
Shi, X., Chen, Z., Wang, H., et al., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst. 28.
Siami-Namini, S., Tavakoli, N., Namin, A.S., 2019. A comparative analysis of forecasting financial time series using arima, lstm, and bilstm. arXiv preprint arXiv:1911, 09512.
Som, A., Kayal, P., 2022. A multicountry comparison of cryptocurrency vs gold: portfolio optimization through generalized simulated annealing. BLOCK: Res. Appl. 3 (3), 100075.
Sundermeyer, M., Schlüter, R., Ney, H., 2012. LSTM neural networks for language modeling. In: Thirteenth Annual Conference of the International Speech Communication Association.
Torres, D.G., Qiu, H., 2018. Applying Recurrent Neural Networks for Multivariate Time Series Forecasting of Volatile Financial Data. KTH Royal Institute of Technology, Stockholm, Sweden.
Valencia, F., Gómez-Espinosa, A., Valdés-Aguirre, B., 2019. Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. Entropy 21 (6), 589.
Verheyen, F., 2010. Monetary Policy, Commodity Prices and Inflation-Empirical Evidence from the US. Commodity Prices and Inflation-Empirical Evidence from the US *(November, 2010)*.
Vidya, G.S., Hari, V.S., 2020. Gold price prediction and modelling using deep learning techniques. In: 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS). IEEE, pp. 28–31.
Vukovic, D., Vyklyuk, Y., Matsiuk, N., et al., 2020. Neural network forecasting in prediction Sharpe ratio: evidence from EU debt market. Physica A 542, 123331.
Xiao, Y., Yin, H., Zhang, Y., et al., 2021. A dual-stage attention-based Conv-LSTM network for spatio-temporal correlation and multivariate time series prediction. Int. J. Intell. Syst. 36 (5), 2036–2057.
Yamak, P.T., Yujian, L., Gadosey, P.K., 2019. A comparison between arima, lstm, and gru for time series forecasting. In: Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence, pp. 49–55.
Yang, S., 2021. A novel study on deep learning framework to predict and analyze the financial time series information. Future Generat. Comput. Syst. 125, 812–819.
Zhang, H., Nguyen, H., Vu, D.A., et al., 2021a. Forecasting monthly copper price: a comparative study of various machine learning-based methods. Res. Pol. 73, 102189.
Zhang, L., Bian, W., Qu, W., et al., 2021b. Time series forecast of sales volume based on XGBoost. In Journal of Physics: Conference Series 1873 (No. 1), 012067. IOP Publishing.
Zhou, J., Cao, Y., Wang, X., et al., 2016. Deep recurrent models with fast-forward connections for neural machine translation. Trans. Asso. Comp. Ling. 4, 371–383.
Zhou, J., He, Z., Song, Y.N., et al., 2019. Precious metal price prediction based on deep regularization self-attention regression. IEEE Access 8, 2178–2187.
Zulauf, C.R., Irwin, S.H., Ropp, J.E., et al., 1999. A reappraisal of the forecasting performance of corn and soybean new crop futures. J. Futures Mark. 19 (5), 603–618.