Prediction of Metal Prices Using Learning Machine Methods: Case of Indian Economy

Said Jaouadi & Mohammed Arif Riaz

Assistant Professor, Accounting and finance department, Jazan University, Gizan, Saudi Arabia.

DOI: https://doi.org/10.47772/IJRISS.2026.10200254

Received: 18 February 2026; Accepted: 24 February 2026; Published: 05 March 2026

ABSTRACT

This study evaluates machine learning and deep learning methodologies for forecasting metal prices (Titanium PPI, Bauxite & Aluminum, Coal Australia, and Iron Ore) within the context of the Indian economy. To enhance predictive precision, the analysis integrates key macroeconomic indicators, including the Indian CPI, industrial production, and the U.S. Federal Funds Rate. Empirical results demonstrate that tree-based ensembles, specifically the Random Forest model, consistently outperform other tested algorithms by achieving the lowest Root Mean Squared Error (RMSE) across all four commodities. Despite the theoretical advantages of deep learning for temporal sequence modeling, our findings indicate that tree-based models provide superior generalization and robustness against market volatility in this specific context. Furthermore, SHAP value analysis reveals that autoregressive target lags and specific macroeconomic variables are the primary drivers of price forecasts. These insights offer actionable guidance for policymakers and industrial stakeholders engaged in strategic planning and risk management.

Keywords: Metal Price Forecasting, Machine Learning, Deep Learning, Random Forest, Indian Economy, Macroeconomic Indicators.

INTRODUCTION

A considerable body of research is devoted to the stability of global supply chains and the macroeconomic impact of raw material costs. The volatility of metal prices represents a significant source of uncertainty for industries ranging from construction and automotive manufacturing to electronics and renewable energy. Sudden price shocks can disrupt production schedules, inflate infrastructure costs, and destabilize the fiscal balances of resource-dependent nations. Therefore, understanding and anticipating these fluctuations is essential for strategic industrial planning and national economic security. Price volatility refers to the rate at which the price of a security increases or decreases for a given set of returns, often indicating the level of risk associated with the asset. To address this, researchers are increasingly turning to Machine Learning (ML) methods. Unlike traditional linear statistical tools, ML algorithms—such as Neural Networks and Random Forests—are defined by their ability to model complex, non-linear relationships between independent variables (like geopolitical risk or currency rates) and dependent variables (metal prices), thereby providing more robust predictive capabilities.

The academic discourse on metal price forecasting has traditionally focused on the comparative performance of classical econometric models versus emerging machine learning techniques. Early research frequently benchmarked standard models like ARIMA and GARCH against algorithms such as Random Forest and Support Vector Regression (SVR). For instance, Varshini et al. (2024) and Sieradzki and Kwiatek (2025) demonstrate that while traditional statistical tools remain useful for baseline volatility estimates, they often fail to capture structural breaks and non-linear patterns. Consequently, studies by Ergin and Eren (2024) and Tripurana et al. (2022) have shown that non-linear ML models, specifically Artificial Neural Networks (ANN) and Random Forests, yield significantly lower error rates for precious metals like gold and silver by effectively mapping complex relationships between prices and economic indicators.

Building on these foundations, the literature has shifted toward more sophisticated deep learning and hybrid architectures to further enhance predictive precision. Recognizing the limitations of standalone models, researchers have proposed hybrid frameworks that combine feature extraction with temporal memory. Notable contributions by Waleed et al. (2025) and Li et al. (2023) highlight the superiority of CNN-LSTM architectures, which utilize Convolutional Neural Networks to filter spatial features from financial data before processing them through Long Short-Term Memory networks. This evolution is further exemplified by the introduction of neurosymbolic ensembles by Lee et al. (2024), which integrate symbolic logic with neural networks to improve the detection of rare price spikes, marking a transition from simple regression to interpretability-focused AI systems.

Although the literature is replete with isolated applications of machine learning, there remains a lack of consensus regarding the comparative efficacy of ensemble, deep learning, and hybrid paradigms within a unified framework. Most previous research focuses on a narrow selection of algorithms, often failing to rigorously benchmark traditional non-linear models against complex architecture combinations on the same dataset. This study addresses these inconsistencies by performing a systematic evaluation of six distinct methodologies: Random Forest, Support Vector Regression (SVR), XGBoost, LSTM, a Stacking ensemble (RF+XGB+SVR), and a Hybrid architecture (LSTM+SVRres). The novelty of this work lies in this comprehensive cross-paradigm comparison, specifically isolating whether the added computational complexity of stacking and hybrid models yields statistically significant improvements over robust standalone algorithms like XGBoost and LSTM, thereby identifying the optimal algorithmic strategy for metal price forecasting.

The paper is structured as follows: Section 2 summarizes the existing literature and theoretical background. Section 3 describes the data sources and the mathematical formulations of the applied methods, including the stacking and hybrid configurations. Section 4 analyzes the results and discusses the predictive power of each model. The paper concludes in Section 5 with a summary of findings and policy implications.

LITERATURE REVIEW

The global market for precious and industrial metals is characterized by extreme volatility, non-linearity, and susceptibility to exogenous shocks, ranging from geopolitical conflicts to shifts in monetary policy. Accurate price forecasting in this sector is critical for investors, policymakers, and industrial stakeholders who must mitigate risk and optimize procurement strategies. While traditional econometric models such as ARIMA and GARCH have historically served as the analytical standard, their inability to fully capture complex, non-linear dependencies has necessitated a paradigm shift toward artificial intelligence. This literature review synthesizes recent research comparing machine learning (ML) and deep learning (DL) methodologies, highlighting the evolution from standard regression techniques to complex hybrid architectures designed to navigate market uncertainty.

Comparative Analysis of ML and Econometric Benchmarks

To determine the baseline efficacy of modern algorithms, several studies have directly compared machine learning models against traditional statistical benchmarks. Varshini et al. (2024) conducted a comprehensive evaluation of various models, including Stacked-LSTM and XGBoost, across multiple metal futures. Their findings support the Efficient Market Hypothesis, suggesting that no single model consistently outperforms others across all timeframes and that commodity-specific volatility is a major determinant of forecasting success.

Despite this variability, specific machine learning interventions have demonstrated clear advantages over linear models. In the silver market, Ergin and Eren (2024) found that non-linear algorithms, specifically Random Forest and k-Nearest Neighbors (k-NN), significantly minimized prediction errors compared to Linear Regression. Similarly, Tripurana et al. (2022) identified Artificial Neural Networks (ANN) as superior to Support Vector Regressors (SVR) and linear methods for gold price forecasting, utilizing economic indicators like the S&P 500 and crude oil prices to map multi-dimensional relationships. Sieradzki and Kwiatek (2025) further advanced this comparison by evaluating the transformer-based Amazon Chronos model against econometric GARCH models. They concluded that while GARCH remains competitive for minimizing squared errors, AI foundation models offer superior accuracy under risk-sensitive metrics, though all models struggle to anticipate sudden volatility spikes caused by structural breaks.

The Rise of Deep Learning and Hybrid Architectures

Building on the limitations of standalone models, recent scholarship has increasingly favored hybrid architectures to improve predictive precision. A consensus has emerged regarding the efficacy of combining feature extraction with temporal memory. Waleed et al. (2025) and Li et al. (2023) both demonstrated that hybrid models integrating Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks outperform standalone architectures. Waleed et al. (2025) found this combination optimal for precious metals, while Li et al. (2023) reported a 34% reduction in RMSE for copper forecasting by using CNNs to filter spatial features from financial attributes before LSTM processing.

The superiority of deep learning extends to various industrial contexts. Aggarwal et al. (2023) and Derakhshani et al. (2024) utilized LSTM and 1D-CNN architectures, respectively, to surpass traditional ARIMA and institutional benchmarks in forecasting metal futures. However, architecture selection remains nuanced; Foroutan and Lahmiri (2024) noted that Temporal Convolutional Networks (TCN) were superior for energy and silver markets, whereas Bidirectional GRU models performed best for gold. Addressing the "black box" nature of these complex models, Lee et al. (2024) introduced a neurosymbolic ensemble framework. By integrating neural models with symbolic logic rules, they significantly improved the detection of rare price spikes, offering a layer of interpretability often missing in pure deep learning approaches.

Macroeconomic Drivers and Market Dynamics

Beyond algorithmic architecture, the integration of macroeconomic variables remains a critical area of research. Kangalli Uyar et al. (2023) combined machine learning with the GSADF test to identify bubble predictors, finding that the US Consumer Confidence Index drives gold bubbles, while palladium is more sensitive to financial variables like the S&P 500. Wang et al. (2025) expanded on this by incorporating Geopolitical Risk (GPR) indices into a mixed-frequency model, revealing that geopolitical shocks significantly enhance predictive power for safe-haven metals like gold but have less impact on industrial metals.

Innovative data sourcing has also proven effective. Oikonomou et al. (2025) leveraged "cross-learning" by pooling mining company stock prices as auxiliary variables, effectively using equity markets as leading indicators for commodity prices. In the context of crisis modeling, Ghali (2023) confirmed that univariate LSTM models often outperform multivariate ones during extreme events like the Russia-Ukraine war, as traditional correlations between variables like oil prices and metal values can break down during systemic shocks.

In our model, we used the following macroeconomic variables as explanatory:

- India_CPI (India's consumer price index), appearing with multiple lags (e.g., t−12, t−11, t−5). These CPI lags show up among the higher-ranked non-target predictors, indicating that inflation conditions in India contribute meaningfully to the model's forecasts.

- India_Prod (India industrial production/output proxy), also included with multiple lags (e.g., t−12, t−11, t−9). Its presence across several lagged terms suggests that India's real-activity cycle is an important driver in the model's predicted target dynamics.

- Fed_Funds (U.S. Federal Funds Rate), included with several lags (e.g., t−12, t−10, t−9, t−1). The repeated appearance of lagged policy-rate terms implies monetary/financial conditions (interest-rate environment) are relevant explanatory inputs for forecasting the target series.

Data source and Sample Splitting: All time-series variables used in this study—including the dependent commodity/metal price series and the full set of explanatory (exogenous) predictors—were retrieved from the Federal Reserve Bank of St. Louis' Federal Reserve Economic Data (FRED) database. The dataset comprises monthly observations spanning an eight-year period from January 2018 to December 2025, yielding a total of 96 observations per series. To preserve the temporal structure of the data and prevent look-ahead bias, we employed a strict sequential train–validation–test split:

- Training Set (75%): January 2018 to December 2023 (72 months) used for initial model fitting.
- Validation Set (12.5%): January 2024 to December 2024 (12 months) utilized for hyperparameter tuning and model selection.
- Test Set (12.5%): January 2025 to December 2025 (12 months) reserved strictly for out-of-sample forecast evaluation.

Material and Methods:

This study conducts a systematic out-of-sample comparison of six supervised learning approaches for prediction: Random Forest (RF), Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory networks (LSTM), a Stacking ensemble (RF+XGB+SVR), and a Hybrid residual architecture (LSTM + SVR). The objective is to evaluate whether nonlinear, high-capacity learners and their ensembles improve predictive accuracy relative to single-model baselines, while preserving a transparent, reproducible evaluation protocol across all candidates. The use of tree ensembles, kernel methods, and recurrent neural networks is well established in time-series and financial/commodity forecasting contexts, where nonlinearity, regime shifts, and complex interactions are common.

Theoretical foundations

RF is an ensemble of randomized regression trees trained on bootstrap resamples and aggregated by averaging, designed to reduce variance and improve generalization; Breiman (2001) provides the seminal theoretical and empirical treatment, including regression settings.

XGBoost is a regularized gradient-boosted decision tree system that optimizes an additive objective with explicit complexity control and scalable split-finding; Chen and Guestrin (2016) formalize the method and its system-level innovations.

Stacking (stacked generalization) combines multiple base learners by training a meta-learner on their out-of-fold predictions to reduce generalization error; Wolpert (1992) introduced the core scheme and its rationale.

Let $\{(x_t, y_t)\}_{t=1}^{T}$ denote the supervised dataset, where $x_t \in \mathbb{R}^p$ is a feature vector available at forecast origin t and $y_t \in \mathbb{R}$ is the scalar outcome to be predicted (e.g., next-period value or return). For a horizon h, the forecasting task is to learn $f(\cdot)$ such that:

$$\hat{y}_{t+h} = f(x_t),$$

and evaluate forecasts on a held-out test segment that occurs strictly after the training segment (to prevent look-ahead bias).

2 Random Forest (RF)

RF constructs B regression trees $\{T_b\}_{b=1}^{B}$ using bootstrap samples; at each split, a random subset of features of size m_{try} is considered. The RF predictor is:

$$\hat{f}_{RF}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x).$$

Key hyperparameters include B, m_{try}, maximum depth, and minimum leaf size (or analogous regularization controls). Breiman (2001) discusses RF generalization behavior and regression performance.

3 Support Vector Regression (SVR)

SVR estimates a function $f(x) = w^{\top}\phi(x) + b$ by minimizing model complexity and an ε-insensitive loss. A common primal form is:

subject to

The kernel trick replaces inner products $\phi(x_i)^\top \phi(x)$ with $K(x_i, x)$ (e.g., RBF). In applied commodity-price contexts, SVR is often tuned via grid search over C, ε, and kernel parameters, and evaluated using time-consistent splits.

4 XGBoost

XGBoost fits an additive model of K regression trees:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F},$$

where $\mathcal{F}$ is the space of CART-like trees. The learning objective combines a loss term and a regularization penalty on tree complexity (e.g., number of leaves and leaf weights), enabling bias–variance control and robustness to overfitting in flexible nonlinear settings. Chen and Guestrin (2016) provide the canonical formulation and describe scalable split finding and sparsity-aware learning.

5 LSTM

LSTM is a recurrent neural network architecture designed to model sequential dependencies through gated memory. Given an input sequence $\{x_t\}$, the LSTM updates hidden state h_t and cell state c_t via input, forget, and output gates; in forecasting, the network maps a window of past observations/features into a prediction $\hat{y}_{t+h}$. In empirical forecasting applications, LSTM models are typically trained by minimizing a squared-error loss with stochastic gradient methods, using normalization and time-ordered train/validation/test splits to preserve temporal structure.

6 Stacking ensemble (RF + XGB + SVR)

Let base learners be $\hat{f}_1, \hat{f}_2, \hat{f}_3$ (RF, XGB, SVR). Stacking trains a meta-learner $g(\cdot)$ using out-of-fold predictions $\tilde{z}_i = (\hat{f}_1^{(-k)}(x_i), \hat{f}_2^{(-k)}(x_i), \hat{f}_3^{(-k)}(x_i))$ as inputs:

$$\hat{y}_i = g(\tilde{z}_i).$$

This procedure reduces leakage by ensuring the meta-learner sees only predictions generated from models not trained on observation i. Wolpert (1992) provides the foundational framework for stacked generalization and its motivation in minimizing generalization error.

7 Hybrid residual model (LSTM + SVR)

The hybrid is constructed as a two-stage error-correction system:

1. Fit an LSTM to obtain $\hat{y}_{t+h}^{(LSTM)}$.
2. Compute residuals on training data: $e_{t+h} = y_{t+h} - \hat{y}_{t+h}^{(LSTM)}$.
3. Train an SVR on the same feature set (or an augmented set including LSTM states/predictions) to predict residuals: $\hat{\mathrm{e}}_{\mathrm{t+h}} = \hat{\mathrm{f}}_{\mathrm{SVR}}(\mathrm{x_t})$.
4. Final hybrid forecast:

$$\hat{\mathrm{y}}_{\mathrm{t+h}}^{(\mathrm{Hybrid})} = \hat{\mathrm{y}}_{\mathrm{t+h}}^{(\mathrm{LSTM})} + \hat{\mathrm{e}}_{\mathrm{t+h}}.$$

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue II February 2026

This design targets remaining nonlinear structure not captured by the LSTM, while retaining a decomposition that can be diagnosed via residual analysis (e.g., whether residuals retain autocorrelation or regime dependence).

Evaluation metrics

Forecast accuracy is evaluated using standard loss-based metrics such as RMSE and MAE, consistent with prior ML forecasting studies in commodity and financial markets. Where appropriate, additional diagnostics (e.g., correlation between predicted and realized series) may be reported, but primary model ranking relies on the same core metrics across all approaches to maintain comparability.

Advantages and limitations

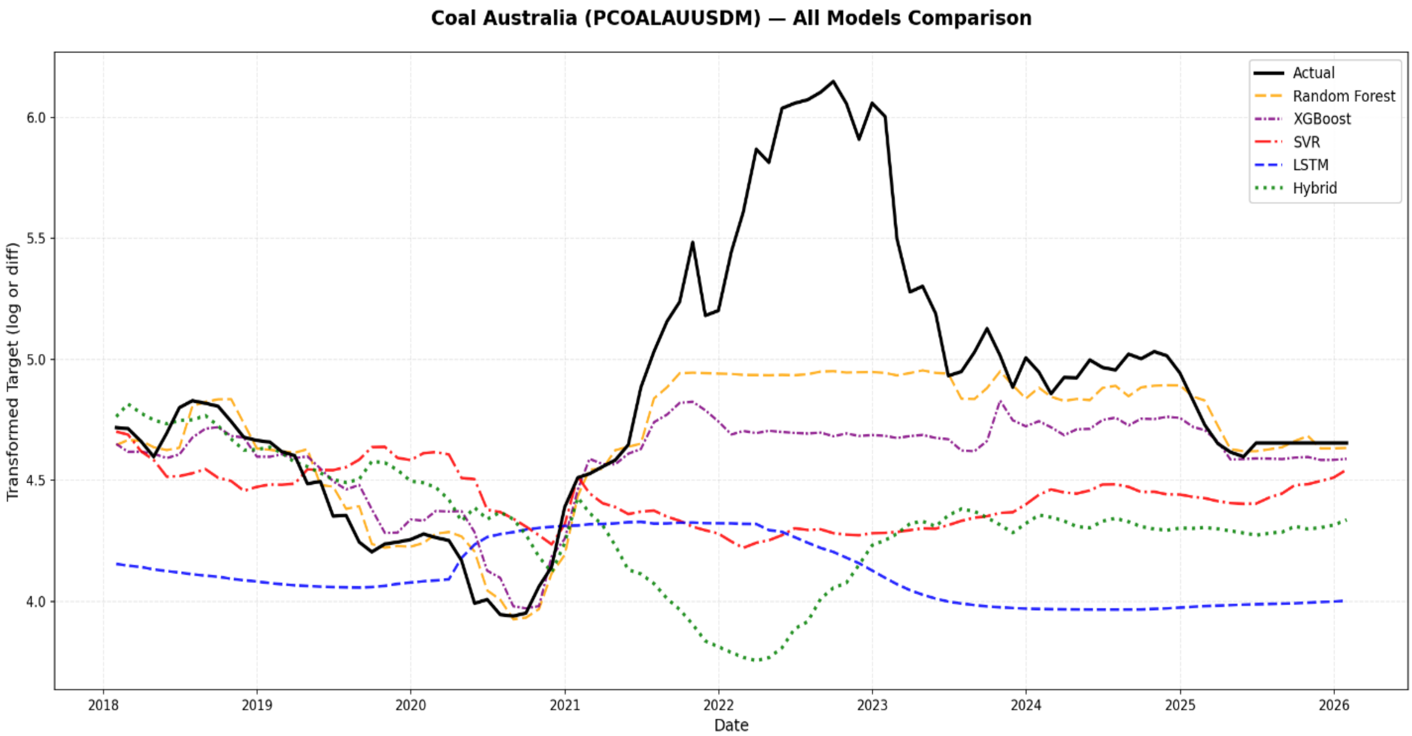
- Random Forest: Strong nonlinear function approximation and robustness via averaging; limitations include limited extrapolation ability and lack of explicit temporal-dependence modeling unless lagged/engineered features are provided.
- SVR: Effective in high-dimensional nonlinear regression with kernel flexibility; limitations include sensitivity to hyperparameters and computational cost for large samples.
- XGBoost: High predictive power with regularization and efficient training; can overfit without careful tuning and still requires time-aware validation to avoid leakage in sequential data.
- LSTM: Designed to capture temporal dependencies and nonlinear dynamics; limitations include training instability, sensitivity to scaling and architecture choices, and reduced interpretability relative to simpler models.
- Stacking: Often improves generalization by combining complementary inductive biases; limitation is added complexity and the need for strict out-of-fold construction to prevent optimistic bias.
- Hybrid (LSTM + SVR): Can correct systematic residual structure left by LSTM; limitation is that performance gains depend on residual predictability, and the two-stage design increases tuning burden and risk of overfitting if not validated properly.

Tree ensembles (RF, XGBoost) provide strong tabular nonlinear baselines, SVR offers a kernel-based alternative with different bias–variance tradeoffs, LSTM provides a sequence model for temporal dependence, and stacking/hybridization explicitly tests whether combining complementary learners improves robustness and accuracy.

RESULTS AND DISCUSSION

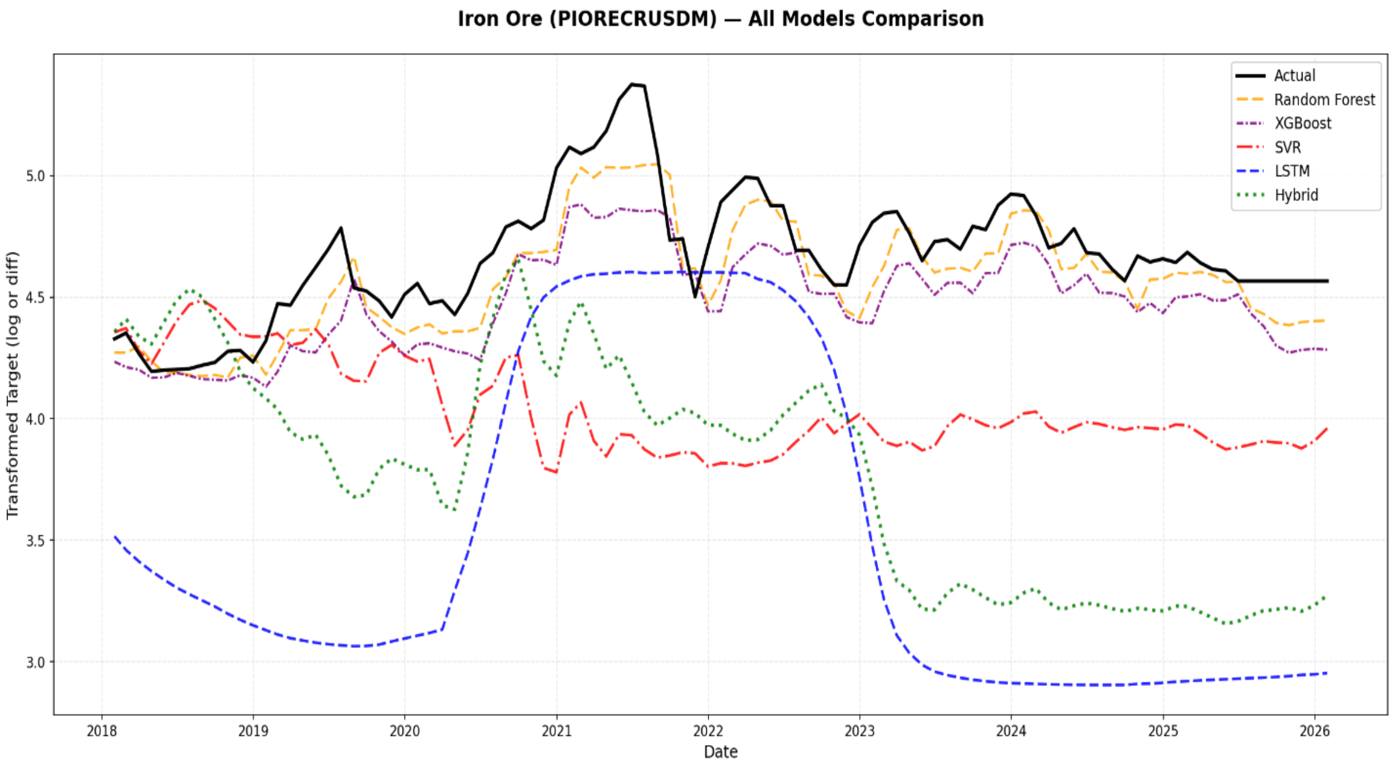
Across all four commodities, Random Forest achieved the lowest error (MAE and RMSE) compared with XGBoost, SVR, LSTM, and the Hybrid model, indicating the most accurate out-of-sample forecasts in this study. Visual inspection of the forecast plots further shows that the Random Forest predictions track the observed series closely across products, consistent with its superior error metrics.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue II February 2026



Model	MAE	RMSE
LSTM	0.7843	0.9226
SVR	0.5604	0.7564
Hybrid	0.6431	0.8960
XGBoost	0.3166	0.5125
Random Forest	0.2174	0.3961

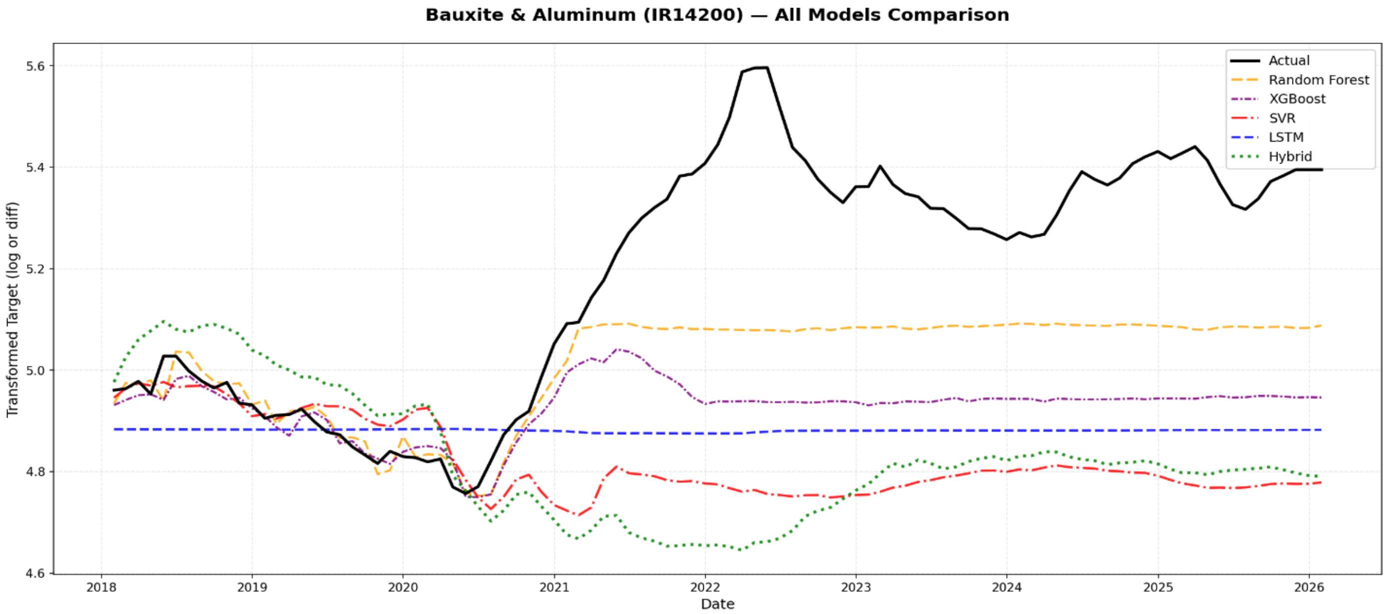
Figure 1: Prediction of Price of Coal.



Model	MAE	RMSE
LSTM	1.1711	1.3034
SVR	0.6575	0.7456
Hybrid	0.9220	1.0420
XGBoost	0.1966	0.2270
Random Forest	0.1196	0.1468

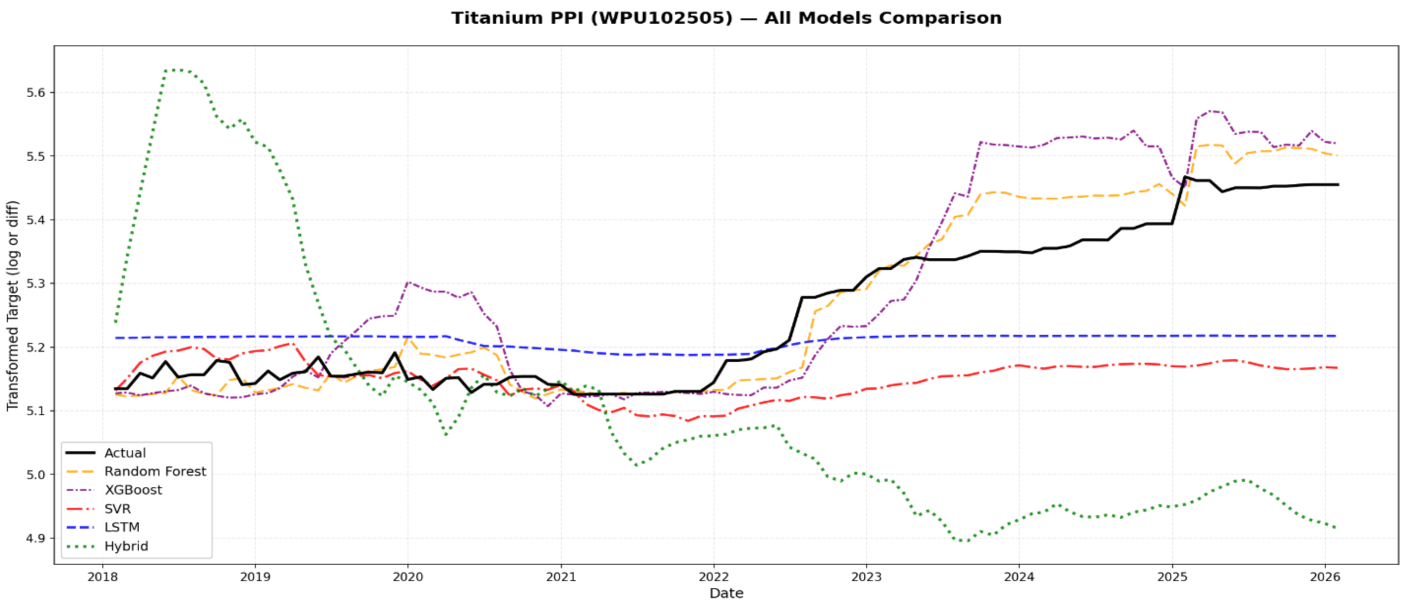
Figure 2: Prediction of Price of Iron.

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue II February 2026



Model	MAE	RMSE
LSTM	0.3244	0.3884
SVR	0.3852	0.4682
Hybrid	0.4119	0.4876
XGBoost	0.2639	0.3333
Random Forest	0.1806	0.2302

Figure 3: Prediction of Price of Bauxite and Aluminium.



Model	MAE	RMSE
LSTM	0.1005	0.1208
SVR	0.1126	0.1506
Hybrid	0.2667	0.3247
XGBoost	0.0702	0.0885
Random Forest	0.0360	0.0453

Figure 4: Prediction of Price of Titanium.

Product	**MAE (RF)**	**RMSE (RF)**	**MAE (XGB)**	**RMSE (XGB)**	**MAE (SVR)**	**RMSE (SVR)**	**MAE (LSTM)**	**RMSE (LSTM)**	**MAE (Hybrid)**	**RMSE (Hybrid)**
Titanium PPI	0.035975	0.045313	0.070187	0.088476	0.112631	0.150640	0.100452	0.120770	0.266693	0.324666
Bauxite &	0.180571	0.230171	0.263883	0.333308	0.385192	0.468238	0.324425	0.388387	0.411918	0.487628

Aluminum										
Coal Australia	0.217449	0.396095	0.316559	0.512465	0.560436	0.756415	0.784261	0.922605	0.643067	0.895986
Iron Ore	0.119629	0.146798	0.196648	0.226978	0.657489	0.745609	1.171145	1.303419	0.922022	1.041951

Table: comparative results of predicts of metals.

Random Forest was the best-performing model (lowest RMSE) for Titanium PPI (RMSE 0.0453), Bauxite & Aluminum (RMSE 0.2302), Coal Australia (RMSE 0.3961), and Iron Ore (RMSE 0.1468). This consistency suggests that tree-based ensembles handled the lagged target features and exogenous covariates more robustly than the other model classes in this setup.

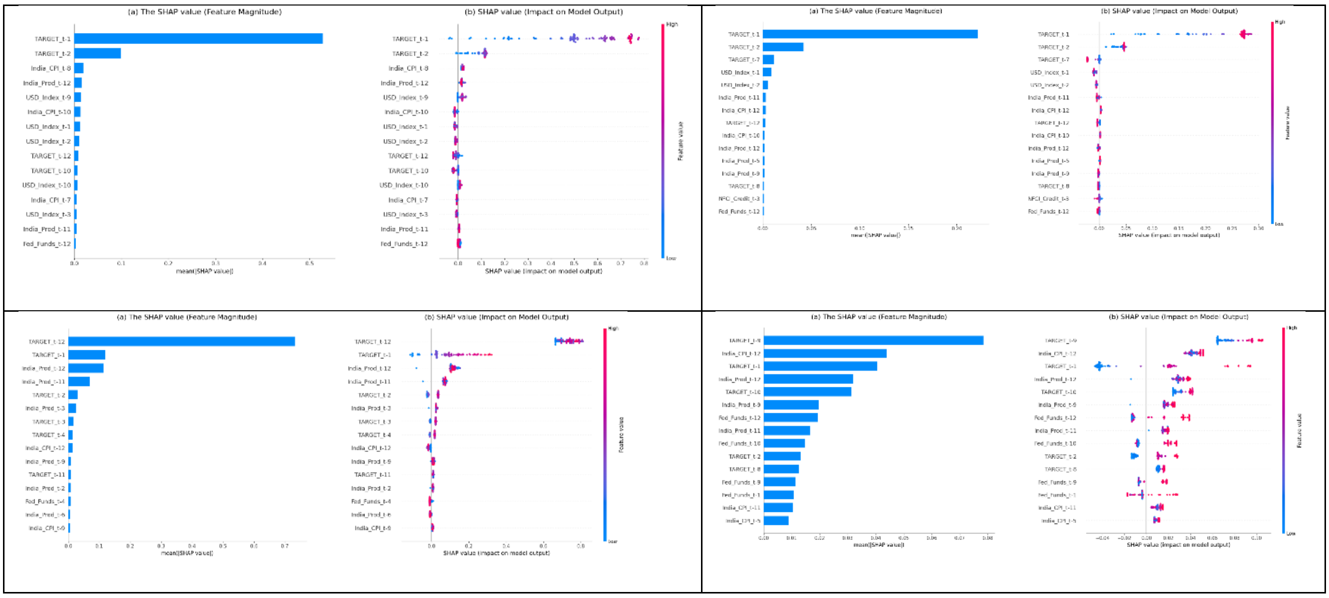


Figure 1: Shap values of predictive variables.

SHAP results indicate that forecasts are primarily driven by autoregressive dynamics (especially the first lag of the target), while exogenous macro-financial indicators (USD index, policy rate, and credit conditions) provide incremental predictive content that is smaller and potentially regime-dependent.

DISCUSSION

These findings are consistent with commodity-price theory emphasizing a mixture of persistent components (slow-moving fundamentals and expectations) and regime-dependent shocks (financial stress, dollar movements, and global activity). In such environments, tree-based ensembles can outperform parametric or more rigid nonlinear methods because they approximate piecewise nonlinearities and interactions that arise from inventory/stock-out logic, convenience yields, and financing constraints without requiring a correct structural specification. Moreover, commodity prices often respond differently to “demand-driven” versus “supply-driven” disturbances, making average linear responses fragile and motivating flexible predictive structures that can condition on observable states. Your multivariate design with exogenous variables aligns with this view: the best model is the one that can adapt most effectively to changing combinations of macro and financial conditions, which is precisely the comparative advantage of Random Forest in finite samples.

A notable theoretical tension is that, under frictionless storage and rational expectations, forward-looking information and arbitrage relationships should constrain near-term spot-price dynamics, potentially favoring models that explicitly encode temporal dependence or economic structure. The empirical dominance of Random

Forest over LSTM/Hybrid models therefore suggests either (i) the information set is adequately summarized by the chosen lag structure and covariates, (ii) the available sample size and signal-to-noise ratio are insufficient for high-capacity sequence models to generalize, or (iii) the price process is characterized by threshold effects and interaction terms that are easier to learn via tree partitions than via smooth function approximation. In addition, if the evaluation is conducted on a relatively short holdout (97 observations), model selection is more sensitive to finite-sample variance and hyperparameter instability—conditions under which robust, bagged learners often have an edge.

Several nuanced mechanisms can explain counterintuitive outcomes—most prominently, why models that are "more sophisticated" in principle (e.g., LSTM and Hybrid) underperform. First, deep recurrent models require large effective sample sizes to identify stable dynamics; with limited test windows and potentially nonstationary regimes, they may overfit training segments and degrade out-of-sample. Second, if exogenous drivers (e.g., global financial conditions, dollar index proxies, or international activity measures) enter nonlinearly and intermittently, LSTM training may underweight rare but influential episodes (crisis-like states), whereas Random Forest can isolate those states through splits and preserve predictive performance. Third, specification choices such as the forecast horizon, normalization, and lag depth can implicitly favor models that exploit local persistence; your error profile—especially the relatively low RMSE for Titanium PPI—suggests a strong predictable component that a tree ensemble can capture efficiently.

From a methodological perspective, these results also raise the importance of formal forecast comparison and robustness checks. Because forecast errors are often serially correlated, and because model comparisons can be sensitive to nestedness and parameter estimation error, the literature recommends tests such as Diebold–Mariano for equal predictive accuracy (with small-sample adjustments discussed in subsequent work) and, for nested-model comparisons, MSPE-adjusted procedures such as Clark–West. Embedding such tests would strengthen the inferential claim that the Random Forest improvements are not only economically meaningful but also statistically credible across horizons and commodities. In addition, the cross-commodity heterogeneity in RMSE magnitudes (e.g., Coal versus Titanium) highlights that volatility regimes differ substantially across markets, which can be connected to well-known macro-financial episodes where commodity prices reflect shifts in global demand, liquidity, and risk premia rather than purely physical supply disruptions.

Finally, your findings contribute to the broader interpretation that commodity prices are jointly shaped by fundamentals and financial channels, with the relative importance varying over time. Work on oil and commodity shocks emphasizes that the macroeconomic implications of price movements depend on the underlying source of the shock (aggregate demand, commodity-specific demand, or supply), implying that reduced-form predictive relationships may be nonlinear and state-dependent. The empirical success of Random Forest in your setting is therefore not merely a "black-box win," but an indication that the metal-price forecasting problem is characterized by interaction-heavy, regime-sensitive dynamics that can be captured with flexible empirical learners when the covariate set is appropriately chosen and evaluation is disciplined.

CONCLUSION

The empirical results of this study demonstrate that multivariate ML specifications augmented with exogenous macroeconomic variables significantly improve metal-price forecasting. Specifically, tree-based ensembles—notably the Random Forest model—provide the strongest overall accuracy and most robust generalization among the evaluated algorithms. Furthermore, SHAP analysis confirms that while autoregressive target lags dominate predictive power, India-centric macroeconomic factors provide crucial incremental information.

The results indicate that multivariate ML specifications with exogenous macro variables improve forecasting performance, with tree-based methods (notably Random Forest) providing the strongest overall accuracy among the evaluated algorithms.

Policy implications follow directly from this evidence. First, Indian policymakers and industry regulators should institutionalize a "macro-to-metals" monitoring dashboard that tracks industrial production and CPI (plus key financial conditions such as policy rates) as early-warning indicators for price pressures and volatility in strategic metals. Second, ministries and state agencies can use these macro-augmented ML forecasts to time buffer-stock rules, import duty adjustments, and procurement schedules to reduce budget exposure and procyclical purchasing. Third, firms in mining, refining, and metal-intensive manufacturing should integrate the model outputs into hedging and inventory policies (e.g., dynamic hedge ratios, procurement windows) and stress-test decisions under macro scenarios rather than only under price-based scenarios.

Several limitations should be acknowledged. The explanatory set is macroeconomic and lagged, which can miss high-frequency market information, supply disruptions, and structural breaks that also drive metals prices. Model comparisons are also conditional on the sample window, feature engineering choices (lag structure), and the specific ML algorithms evaluated, so performance rankings may shift with different horizons, markets, or periods of crisis.

Future research can extend the agenda in concrete ways. One path is to enrich India-specific fundamentals (e.g., mining output by metal, energy input costs, trade and inventory proxies) and explicitly test alternative lag structures and mixed-frequency designs to better match the timing of macro transmission. Another is to evaluate robustness under regime changes with rolling/recursive re-estimation, crisis subsamples, and structural-break-aware methods, and to benchmark against newer sequence models while preserving interpretability through SHAP/PDP-style diagnostics. A third is to translate forecast gains into economic value by backtesting policy/firm rules (procurement timing, hedging, buffer stocks) under realistic transaction costs and constraints.

Overall, the evidence motivates a forward-looking research agenda in which India-relevant macro fundamentals are treated not as background context but as first-class predictive signals within transparent ML forecasting systems for metals—systems designed to be both operationally usable and continuously validated as macro conditions and market structure evolve.

REFERENCES

1. Aggarwal, V., Ranjan, A., Shaurya, S., & Garg, S. K. (2023). To Determine the Futures pricing of Metal Commodities using Deep Learning. Evergreen, 10(2), 1027-1033. https://doi.org/10.5109/6793658
2. Derakhshani, R., GhasemiNejad, A., Amani Zarin, N., Amani Zarin, M. M., & Jalaee, M. S. (2024). Forecasting copper prices using deep learning: Implications for energy sector economies. Mathematics, 12(15), 2316. https://doi.org/10.3390/math12152316
3. Ergin, E., & Eren, B. S. (2024). Assessing the Effectiveness of Machine Learning Techniques for Silver Price Prediction: A Comparative Study. Bitlis Eren University Journal of Science, 13(4), 1293-1303. https://doi.org/10.17798/bitlisfen.1556171
4. Foroutan, P., & Lahmiri, S. (2024). Deep learning systems for forecasting the prices of crude oil and precious metals. Financial Innovation, 10(111). https://doi.org/10.1186/s40854-024-00637-z
5. Ghali, H. (2023). Price Prediction Models of Metals Considering International Crises [Master's thesis, the American University in Cairo]. AUC Knowledge Fountain. https://fount.aucegypt.edu/etds/2041
6. Kangalli Uyar, S. G., Uyar, U., & Balkan, E. (2023). Fundamental predictors of price bubbles in precious metals: a machine learning analysis. Mineral Economics. https://doi.org/10.1007/s13563-023-00404-z
7. Lee, N., Ngu, N., Sahdev, H. S., Motaganahall, P., Chowdhury, A. M. S., Xi, B., & Shakarian, P. (2024). Metal Price Spike Prediction via a Neurosymbolic Ensemble Approach. arXiv:2410.12785v1 [cs.LG]. https://arxiv.org/abs/2410.12785v1
8. Li, F., Zhou, H., Liu, M., & Ding, L. (2023). A Medium to Long-Term Multi-Influencing Factor Copper Price Prediction Method Based on CNN-LSTM. IEEE Access, 11, 69458-69473. https://doi.org/10.1109/ACCESS.2023.3288486
9. Oikonomou, K., Damigos, D., & Dimitriou, D. (2025). Globality in the metal markets: Leveraging cross-learning to forecast aluminum and copper prices. Resources Policy, 103, 105558.
10. Sieradzki, R., & Kwiatek, S. (2025). Forecasting Price Volatility of Non-Ferrous Metals: A Comparison of Econometric, Machine Learning, and AI Models. Working Paper. Cracow University of Economics / NYU Stern School of Business.
11. Tripurana, N., Kar, B., Chakravarty, S., Paikaray, B. K., & Satpathy, S. (2022). Gold Price Prediction Using Machine Learning Techniques. ACI'22: Workshop on Advances in Computation Intelligence, CEUR Workshop Proceedings (CEUR-WS.org), 274-281.

12. Varshini, A., Kayal, P., & Maiti, M. (2024). How good are different machine and deep learning models in forecasting the future price of metals? Full sample versus sub-sample. Resources Policy, 92, 105040.
13. Waleed, J., Hasan, T. M., Abdullah, A. J., & Alkhayyat, A. (2025). Precious Metal Prices Forecasting Using Optimally Configured Hybrid Deep Learning Approach. Journal of Machine and Computing. https://doi.org/10.53759/7669/jmc202505143
14. Wang, X., Wu, J., Pan, Y., Zhang, N., Wang, T., & Liu, J. (2025). What drives precious metals pricing? An explainable Mixed-frequency machine learning approach. Mineral Economics. https://doi.org/10.1007/s13563-025-00586-8