Comprehensive Forecasting and Statistical Analysis of Indian Oil Sector and Crude Benchmarks

Stephen Baraik[1], Tanisha Ghosh[2], Sneha Das[3], Dr. Suresh Pathare[4]

School of Mathematics, Applied Statistics and Analytics,

NMIMS Deemed to be University, Navi Mumbai, India

[1]stephenbaraik@gmail.com,

[3]ghoshtanisha@gmail.com,

[4]snehadas1593@gmail.com,

[4]suresh.pathare@nmims.edu

Abstract—This research presents a comprehensive, data-driven forecasting framework for the Indian crude oil sector, integrating econometric and deep learning methodologies to analyze interdependencies between global crude benchmarks and the performance of India's leading public sector oil enterprises — BPCL, IOCL, HPCL, ONGC, OIL, and GAIL. Using a decade of historical data (2014–2024) on stock prices, Indian Crude Basket (ICB) values in USD and INR, and volatility indices (VIX), this study conducts extensive exploratory data analysis (EDA) followed by comparative modeling using ARIMA, Prophet, and Long Short-Term Memory (LSTM) networks.

The empirical results demonstrate strong sectoral correlations, particularly between ONGC and ICB (r = 0.82) and BPCL–IOCL (r = 0.78), confirming the structural interlinkages between upstream and downstream entities. Volatility analysis reveals distinct market regimes corresponding to the 2015–2016 oil price collapse and the 2020–2021 pandemic period, followed by post-recovery stabilization. Among the forecasting models, the LSTM network achieved the lowest prediction error (RMSE 1.5 USD), outperforming Prophet (2.0 USD) and ARIMA (2.1 USD).

Long-term projections indicate that the Indian Crude Basket is likely to stabilize within the range of USD 85–95 per barrel by 2035, with upstream firms such as ONGC and OIL exhibiting steady revenue growth (4% CAGR). The study demonstrates that hybrid econometric–AI frameworks substantially improve the accuracy and interpretability of energy market forecasting, offering valuable insights for policymakers, investors, and energy strategists in managing price volatility and planning future capacity expansion.

Index Terms—Crude Oil, Forecasting, Indian oil companies, ARIMA, LSTM, Time Series Analysis, Energy Economics, Quantitative Modeling

I. Introduction

India's petroleum sector forms the backbone of the nation's energy infrastructure, supplying over 30% of its total primary energy demand and fueling critical sectors such as transport, manufacturing, and power generation. As one of the world's largest importers of crude oil, India's economic stability is strongly influenced by global energy price dynamics, exchange rate fluctuations, and domestic regulatory intervention. The operational and financial performance of public sector undertakings (PSUs), including Bharat Petroleum Corporation Limited (BPCL), Indian Oil Corporation Limited (IOCL), Hindustan Petroleum Corporation Limited (HPCL), Oil and Natural Gas Corporation (ONGC), Oil India Limited (OIL), and Gas Authority of India Limited (GAIL), collectively represent the structural foundation of the country's oil and gas ecosystem.

The volatility of crude oil prices remains one of the most significant external factors affecting India's fiscal health, trade balance, and inflationary trends. Over the past decade (2014–2024), the sector has experienced pronounced fluctuations driven by events such as the 2015–2016 global oil price collapse, the 2020–2021 COVID-19 pandemic, and the post-pandemic recovery cycle. These shocks have altered consumption patterns, refining margins, and upstream investment flows, underlining the need for robust data-driven forecasting models capable of capturing nonlinear, volatile market behavior.

Previous studies [1]–[3] have extensively explored the relationship between oil price shocks and macroeconomic outcomes, yet much of the literature focuses on either global markets or isolated firm-level analyses. Few studies have provided an integrated, multi-company framework that simultaneously models both upstream and downstream interactions within the Indian context. Moreover, traditional econometric models, such as ARIMA and GARCH, although statistically interpretable, often struggle to capture the nonlinear dependencies and structural breaks inherent in modern financial time series.

This research bridges these methodological and empirical gaps by integrating classical time-series econometrics with modern deep learning architectures to provide a comprehensive, long-horizon forecasting framework for India's crude oil ecosystem. Leveraging a decade of empirical data, this study employs Exploratory Data Analysis (EDA) to identify sectoral dependencies and volatility regimes, followed by

hybrid forecasting models — ARIMA, Prophet, and LSTM — to predict price trajectories and corporate performance indicators.

By combining interpretability and predictive accuracy, this study contributes to three major objectives:

1) Quantifying the statistical interdependence between Indian oil PSUs and the Indian Crude Basket (ICB).
2) Comparing the performance of traditional econometric and deep learning models in capturing temporal and nonlinear dynamics.
3) Providing actionable insights for policymakers and investors on volatility management, fiscal planning, and long-term energy strategy.

In doing so, this paper advances the field of energy economics and forecasting by demonstrating that hybrid analytical approaches can serve as reliable decision-support tools in an era of unprecedented market uncertainty and energy transition.

II. LITERATURE REVIEW

The relationship between crude oil price dynamics and macroeconomic performance has been extensively studied across global and regional contexts. Early research by [1] established a causal linkage between oil price shocks and macroeconomic recessions in the United States, laying the foundation for subsequent analyses in energy economics. Building upon this, [4] and [5] explore the asymmetric effects of oil price increases and decreases on output growth, underscoring the structural role of oil price volatility in economic fluctuations.

In the Indian context, [2] investigated the long-run and short-run relationships between oil prices and stock market indices, identifying significant cointegration effects and bidirectional causality between energy prices and equity performance. Similarly, [3] analyzed the sectoral sensitivity of Indian stock markets to crude price movements, revealing that energy-intensive industries and petroleum PSUs exhibit heightened vulnerability to global oil price shocks. These studies collectively emphasize the systemic transmission of oil price volatility across both macroeconomic and corporate dimensions.

From a methodological standpoint, traditional econometric models such as ARIMA, VAR, and GARCH have been widely employed to capture linear dependencies and conditional heteroskedasticity in oil price data. For example, [6] demonstrate the utility of GARCH-type models in modeling volatility clustering and risk transmission across global energy markets. However, these models inherently assume linearity and stationarity, limiting their effectiveness in capturing abrupt structural shifts and nonlinear feedback loops that are typical of energy time series.

In contrast, modern approaches incorporating machine learning and deep learning have shown promising improvements in forecast accuracy and adaptability. [7] introduced hybrid deep learning models that combine statistical decomposition techniques with Long Short-Term Memory (LSTM) architectures to predict financial and energy market trends. [8] extended this work by integrating Prophet–LSTM hybrid frameworks, demonstrating superior performance in highly volatile environments by blending time-series decomposition with nonlinear learning capability. Similarly, [9] applied convolutional neural networks (CNNs) along with LSTM units to forecast Brent crude prices, achieving higher accuracy than conventional ARIMA benchmarks.

Despite these advances, significant research gaps persist in the Indian petroleum context. Most prior studies either examine individual firms in isolation or rely solely on econometric or machine learning paradigms, neglecting the synergistic benefits of hybrid models. Moreover, limited research has focused on multi-company, sector-wide forecasting frameworks that integrate both upstream (ONGC and OIL) and downstream (BPCL, IOCL, and HPCL) entities under unified modeling conditions. Additionally, technical indicators such as Moving Averages, RSI, MACD, and volatility indices (VIX), which capture investor sentiment and market momentum, remain underutilized in Indian oil forecasting studies.

This research builds upon and extends the existing literature by:

1) Developing a hybrid econometric–AI framework that integrates ARIMA, Prophet, and LSTM models for comprehensive crude price and corporate performance forecasting.
2) Conducting a multi-company, correlation-driven analysis of Indian PSUs and the Indian Crude Basket (ICB) across a decade-long dataset (2014–2024).
3) Incorporating advanced technical and volatility indicators to enhance predictive feature richness and interpretability.

By synthesizing traditional statistical interpretability with modern deep learning flexibility, this study contributes a unified methodological framework to the field of energy forecasting and quantitative policy analysis, specifically tailored to the Indian petroleum economy.

III. RESEARCH GAP

Although extensive research has been conducted on global crude oil price dynamics, volatility modeling, and macroeconomic linkages, several important gaps persist in the context of the Indian petroleum sector, particularly regarding integrated, data-driven forecasting frameworks. A synthesis of the reviewed literature highlights the following critical research deficiencies.

A. Lack of Integrated Multi-Company Analysis

Most prior studies have focused on individual firm-level analyses—examining ONGC, IOCL, or BPCL in isolation—without accounting for systemic interdependencies within India's oil ecosystem. Given the structural coupling between upstream producers (ONGC and OIL) and downstream refiners (BPCL, IOCL, and HPCL), a unified analytical framework is essential to capture sectoral co-movements, volatility spillovers, and policy-sensitive dependencies.

B. Absence of Hybrid Econometric–Deep Learning Frameworks

Traditional econometric models such as ARIMA and GARCH provide valuable statistical interpretability but are constrained by assumptions of linearity and stationarity. Conversely, deep learning models such as LSTM and GRU capture nonlinear patterns but often lack transparency and parameter explainability. Existing studies have rarely integrated these paradigms to leverage their complementary strengths—statistical rigor with adaptive learning capacity.

C. Underrepresentation of Technical and Volatility Indicators

Indian studies on crude oil and PSU stock forecasting often rely solely on price and volume data, omitting technical indicators such as Moving Averages (MA), Relative Strength Index (RSI), MACD, and Bollinger Bands. Additionally, volatility indices (VIX), which reflect market sentiment and risk perception, are seldom incorporated despite their predictive potential in energy market modeling.

D. Limited Long-Term Forecasting Horizons

Most existing analyses adopt short-term forecasting windows (typically 6–12 months), providing limited insights into structural market transitions, post-pandemic recovery trends, and policy implications. A decade-long analytical horizon (2014–2024) offers richer temporal granularity for examining cyclical and secular trends in India's oil markets.

E. Weak Policy and Strategic Integration

Despite the crucial fiscal implications of crude price volatility, very few models translate forecasting outcomes into policy-relevant insights—such as fiscal planning, strategic petroleum reserve management, and subsidy optimization. The absence of interpretable decision-support frameworks limits the practical applicability of prior research for government and industry stakeholders.

To bridge these gaps, the present study develops a six-phase hybrid forecasting framework that synthesizes econometric and deep learning methodologies to analyze and predict the performance of India's public sector oil companies alongside the Indian Crude Basket (ICB) in both USD and INR terms. By combining Exploratory Data Analysis (EDA) with advanced predictive modeling (ARIMA, Prophet, and LSTM), this research not only improves forecast accuracy but also enhances interpretability, enabling its use as a policy-support and investment decision-making tool for India's evolving energy landscape.

IV. PRELIMINARIES

This section presents the foundational analytical setup, focusing on correlation modeling and forecasting frameworks adopted in the study.

A. Correlation and Dependence Modeling

The correlation analysis quantifies both linear and monotonic relationships among key petroleum sector variables. The following conclusions were drawn.

- The correlation matrix revealed strong inter-firm dependencies, particularly between ONGC–ICB ($r = 0.82$) and BPCL–IOCL ($r = 0.78$), highlighting the sector's structural interlinkages and cascading effects of global crude price shifts on domestic performance.
- Rolling correlation analysis using a 30-day window was conducted to capture temporal shifts in interdependence, providing insights into dynamic volatility transmission across entities.

B. Forecasting Models and Theoretical Foundation

The predictive framework integrates three primary forecasting models—ARIMA, Prophet, and LSTM—each designed to capture distinct characteristics of financial time series data.

1) **Autoregressive Integrated Moving Average (ARIMA):**
   Classical econometric model that captures linear temporal dependencies. Its general form is given by

   $$\phi(B)(1 - B)^d P_t = \theta(B)\varepsilon_t$$

   where B is the backshift operator and $\phi(B)$ and $\theta(B)$ represent autoregressive and moving average polynomials, respectively. Optimal parameters (p, d, q) were selected based on the Akaike Information Criterion (AIC) minimization.

2) **Prophet Model:**
   Developed by Facebook, Prophet decomposes a time series into trend, seasonality, and holiday effects as follows:

   $$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

   It is robust to missing data and outliers, making it suitable for long-horizon energy data with irregular seasonality.

3) **Long Short-Term Memory (LSTM) Network:**
   A recurrent neural network (RNN) architecture capable of learning nonlinear and long-term temporal dependencies through memory cells and gating mechanisms:

   $$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$
   $$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$
   $$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c[h_{t-1}, x_t] + b_c)$$
   $$h_t = \tanh(c_t)$$

The implemented model utilized two stacked LSTM layers (64 units each), followed by dense layers with dropout regularization to prevent overfitting.

C. Evaluation Metrics

The forecasting performance was evaluated using two standard error metrics that measure prediction accuracy across different models.

- **Root Mean Squared Error (RMSE):**

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (P_t - \hat{P}_t)^2}$$

  The RMSE quantifies the average magnitude of forecast errors, giving a higher weight to larger deviations. A lower RMSE indicates higher predictive precision.

- **Mean Absolute Percentage Error (MAPE):**

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{P_t - \hat{P}_t}{P_t} \right|$$

  MAPE expresses the forecast accuracy as a percentage, providing an interpretable measure of the relative error magnitude.

Both metrics enable quantitative comparison of predictive accuracy across the ARIMA, Prophet, and LSTM models. The results from the EDA-driven implementation show that LSTM achieved the lowest RMSE (≈ 1.52 USD), outperforming Prophet (≈ 1.98 USD) and ARIMA (≈ 2.14 USD). This confirms the superior performance of deep learning approaches in modeling nonlinear, volatile financial time-series data.

V. PROBLEM STATEMENT

India's petroleum sector operates within a complex and volatile global energy landscape characterized by fluctuating crude oil prices, currency movements, and evolving domestic regulatory structures. As one of the largest crude oil importers worldwide, India's fiscal health and industrial output are deeply intertwined with the behavior of international energy markets. The country's public sector undertakings (PSUs), including Bharat Petroleum Corporation Limited (BPCL), Indian Oil Corporation Limited (IOCL), Hindustan Petroleum Corporation Limited (HPCL), Oil and Natural Gas Corporation (ONGC), Oil India Limited (OIL), and Gas Authority of India Limited (GAIL), collectively form the backbone of India's exploration, refining, and distribution networks.

However, the financial performance and operational resilience of these firms are significantly influenced by crude price volatility, exchange rate fluctuations, and macroeconomic cycles. Despite this interdependence, the existing research and forecasting mechanisms in India's oil sector remain fragmented, often focusing on short-term horizons, isolated firms, or linear econometric models that fail to capture the full complexity of the system.

Conventional econometric techniques, such as ARIMA, VAR, and GARCH, are statistically sound but inherently constrained by their linear assumptions and inability to adapt to structural breaks, regime shifts, and nonlinear dependencies common in energy markets. Conversely, while machine learning and deep learning models (e.g., LSTM and GRU) demonstrate high predictive capability, they often lack interpretability and economic contextualization, making them less suitable for policy applications. This dichotomy between accuracy and interpretability remains a critical bottleneck in modern energy forecasting research.

A. Central Research Problem

The primary research problem addressed in this study is the absence of an integrated, data-driven framework that simultaneously captures:

1) The dynamic interactions between global crude benchmarks (Indian Crude Basket — USD and INR) and the financial performance of India's major oil PSUs.
2) The volatility transmission mechanisms linking crude price fluctuations, macroeconomic indicators, and firm-level stock returns.
3) The long-term predictive behavior of India's petroleum market under hybrid modeling conditions that combine statistical interpretability with deep learning adaptability.

B. Challenges Identified

- **Fragmented Forecasting Approaches:** Most Indian studies rely on single-model frameworks, limiting the ability to identify cross-sectoral dependencies or long-term systemic patterns.
- **Underrepresentation of Volatility and Sentiment Factors:** Few analyses incorporate technical and volatility indicators such as RSI, Bollinger Bands, or VIX, which play a significant role in financial market sentiment.
- **Data and Temporal Limitations:** Existing forecasting studies often restrict analysis to short-term windows (6–12 months), ignoring decade-long structural transitions, post-pandemic recovery effects, and exchange rate pass-through dynamics.
- **Lack of Policy Integration:** Quantitative models seldom translate into actionable insights for fiscal planning, strategic petroleum reserve management, or energy subsidy optimization.

C. Study Response

To address these challenges, this research proposes a hybrid six-phase analytical framework integrating econometric, statistical, and deep learning models — specifically ARIMA, Prophet, and LSTM — supported by an extensive Exploratory Data Analysis (EDA) foundation. The framework enables comprehensive modeling of both firm- and benchmark-level behavior across a ten-year period (2014–2024), with forecasting projections extended to 2035.

This approach bridges the methodological gap between traditional econometric interpretability and AI-driven predictive power, offering a unified structure that enhances both forecast accuracy and policy relevance. The framework also provides stakeholders, including policymakers, investors, and corporate strategists, with an empirically grounded decision-support tool for managing volatility and planning energy security strategies.

VI. RESEARCH OBJECTIVES

The overarching goal of this research is to develop a comprehensive, data-driven, and interpretable forecasting framework that integrates traditional econometric models with advanced deep learning architectures to analyze and predict the financial and operational dynamics of India's public sector oil companies in relation to global crude price behavior.

This objective is pursued through the formulation of specific sub-objectives that collectively bridge the methodological, empirical, and policy gaps identified in the preceding sections.

A. Primary Objective

To construct a hybrid forecasting framework that integrates econometric (ARIMA, Prophet) and machine learning (LSTM) models for the long-term prediction of crude oil prices and the financial performance of major Indian oil PSUs, while ensuring interpretability, robustness, and policy relevance.

B. Specific Objectives

1) To integrate multi-source datasets — including Indian Crude Basket (ICB) prices (in USD and INR), stock prices of major public sector oil companies (BPCL, IOCL, HPCL, ONGC, OIL, GAIL), and macroeconomic indicators (e.g., exchange rate, VIX, GDP index) — into a unified analytical framework suitable for temporal and cross-sectional analysis.
2) To perform extensive Exploratory Data Analysis (EDA) to identify statistical characteristics, volatility clusters, and co-movement patterns across upstream and downstream entities, thereby capturing both sectoral and benchmark interdependencies.
3) To conduct feature engineering and technical indicator generation, including Moving Averages (MA), Relative Strength Index (RSI), MACD, Bollinger Bands, and rolling volatility measures, enhancing the predictive power and interpretability of time-series models.
4) To quantify the relationships between crude price fluctuations and firm-level financial metrics through correlation and regression modeling, focusing on the interdependence between upstream (ONGC, OIL) and downstream (BPCL, IOCL, HPCL) companies.
5) To construct, calibrate, and compare predictive models —
   - ARIMA for linear temporal forecasting,
   - Prophet for trend–seasonality decomposition, and
   - LSTM for nonlinear pattern recognition —

using a ten-year training window (2014–2024) and forecast horizons extending to 2035.
6) To evaluate model performance using quantitative error metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), thereby identifying the optimal forecasting method for the Indian crude oil and energy sector.
7) To interpret empirical findings in the context of macroeconomic stability and energy policy, translating model results into actionable insights for fiscal planning, petroleum reserve strategy, and risk management.

C. Research Hypotheses

Based on preliminary statistical analysis and theoretical insights, the following hypotheses guide the study:

- **H1:** There exists a statistically significant correlation between the Indian Crude Basket (ICB) and the financial performance of India's public sector oil companies.
- **H2:** Deep learning models (e.g., LSTM) outperform traditional econometric models (ARIMA, Prophet) in forecasting crude oil price movements under conditions of high volatility.
- **H3:** Incorporating technical and volatility indicators significantly improves the predictive accuracy and robustness of long-term crude oil forecasting models.
- **H4:** Forecasted stabilization in ICB prices and firm revenues reflects structural maturity and resilience in India's post-pandemic petroleum sector.

D. Expected Outcomes

The study anticipates the following outcomes:

1) Development of a replicable hybrid forecasting pipeline combining interpretability and predictive strength.
2) Identification of strong inter-firm and benchmark correlations, particularly between ONGC–ICB ($r \approx 0.82$) and BPCL–IOCL ($r \approx 0.78$).
3) Demonstration that LSTM models achieve superior forecast accuracy, with $RMSE \approx 1.5$ USD compared to Prophet (≈ 2.0 USD) and ARIMA (≈ 2.1 USD).
4) Projection of long-term ICB price stabilization between USD 85–95 per barrel by 2035, with steady revenue growth ($\sim 4\%$ CAGR) for upstream firms.

VII. METHODOLOGY

This study employs a six-phase analytical framework that integrates statistical rigor with machine learning adaptability. The methodology combines data preprocessing, feature engineering, exploratory data analysis (EDA), correlation modeling, and forecasting using a hybrid set of models: ARIMA, Prophet, and LSTM. All computations and visualization tasks were conducted in Python (v3.11) using the `Pandas`, `NumPy`, `Matplotlib`, `Statsmodels`, `TensorFlow/Keras`, and `Facebook Prophet` libraries.

A. Data Collection and Integration

The dataset spans a ten-year period (2014–2024) and integrates multiple data sources to ensure a holistic representation of India's petroleum sector. The major components include the following:

- Stock price data of six PSUs: BPCL, IOCL, HPCL, ONGC, OIL, and GAIL.
- Indian Crude Basket (ICB) prices, recorded in both USD and INR, derived as a weighted average of Dubai (75%) and Brent (25%) benchmarks.
- Macroeconomic indicators: USD/INR exchange rate, GDP index, and Volatility Index (VIX).

All datasets were synchronized into a unified time-indexed structure. Missing values were imputed using forward-fill interpolation, and outliers were treated using the Interquartile Range (IQR) method. The final dataset was standardized for analysis.

$$ICB_{USD,t} = 0.75 \times P_{Dubai,t} + 0.25 \times P_{Brent,t} \quad (1)$$

$$ICB_{INR,t} = ICB_{USD,t} \times FX_{USD/INR,t} \quad (2)$$

B. Feature Engineering

Feature engineering enhanced predictive richness by incorporating derived variables that capture trend, momentum, and volatility. The key features are summarized in Table I.

TABLE I
SUMMARY OF ENGINEERED FEATURES

Category	Feature	Description
Trend Indicators	SMA(30), SMA(90)	Short- and long-term moving averages
Momentum Indicators	RSI(14), MACD	Relative strength, trend reversals
Volatility Metrics	Rolling $\sigma(30)$	30-day standard deviation of returns
Price Ratios	Price Spreads	Relative valuation between PSUs and ICB
Correlation Features	Rolling Corr(30)	Dynamic co-movement tracking

All features were normalized using z-score scaling before model training to ensure comparability.

C. Exploratory Data Analysis (EDA)

The EDA revealed key structural and temporal insights:

- **Trends:** Crude prices and PSU stocks exhibit cyclical behavior, with shocks in 2015–2016 and 2020–2021 corresponding to global oil price collapses and the COVID-19 pandemic.
- **Distributions:** Kernel density plots confirmed fat-tailed, non-normal return distributions.
- **Correlations:** ONGC–ICB ($r = 0.82$) and BPCL–IOCL ($r = 0.78$) displayed strong interdependence.
- **Volatility Clustering:** High-volatility periods observed during 2016 and 2020, reverting post-2021.

D. Statistical Inference and Correlation Modeling

Both Pearson and Spearman correlation coefficients were computed to assess dependencies (Table II).

TABLE II
CORRELATION COEFFICIENTS AMONG KEY VARIABLES

Relationship	r
ONGC–ICB (USD)	0.82
BPCL–IOCL	0.78
OIL–ONGC	0.80
ICB (USD)–ICB (INR)	0.95

OLS regression confirmed that upstream firms (ONGC, OIL) are more sensitive to crude fluctuations than refiners (BPCL, IOCL, HPCL).

E. Predictive Modeling and Forecasting

Three forecasting models were compared based on accuracy and interpretability.

1) ARIMA: Parameters (p, d, q) were optimized using the Akaike Information Criterion (AIC). Residual diagnostics confirmed the adequacy of the model. **RMSE (USD):** 2.14, **Forecast Horizon:** 12 months.

2) Prophet: Decomposes series into trend, seasonality, and holidays; robust against outliers. **RMSE (USD):** 1.98, **Forecast Horizon:** 12 months.

3) LSTM: Configured with two hidden layers (64 and 32 units), dropout = 0.2, and Adam optimiser. **RMSE (USD):** 1.52, **RMSE (INR):** 138.9, **Forecast Horizon:** 12 months.

LSTM achieved the lowest RMSE and MAPE, capturing nonlinear temporal dependencies and volatility-driven price behavior.

F. Forecast Interpretation and Policy Implications

Forecasts (2025–2035) suggest:

- ICB stabilizes between USD 85–95 per barrel.
- Upstream firms (ONGC, OIL) show stable revenue growth of ~4% CAGR.
- Refiners (BPCL, IOCL) display cyclical margin variations tied to refining spreads.

These projections can guide fiscal planning, strategic petroleum reserve policy, and investment risk management.

G. Evaluation Summary

Table III summarizes the model performance metrics.

TABLE III
MODEL EVALUATION SUMMARY

Model	RMSE (USD)	RMSE (INR)	MAPE (%)
ARIMA	2.14	175.6	3.8
Prophet	1.98	160.3	3.2
LSTM	1.52	138.9	2.6

VIII. SUMMARY

The six-phase methodological framework effectively combines statistical interpretability and deep learning adaptability, providing robust forecasting capability for India's petroleum sector. The proposed hybrid approach ensures long-term prediction accuracy, interpretability, and policy relevance under dynamic energy market conditions.

IX. RESULTS

This section presents and interprets the empirical findings of the study, encompassing statistical correlation analysis, model comparison results, and long-term forecasting outcomes derived from the hybrid econometric–machine learning framework. Each model—ARIMA, Prophet, and LSTM—was evaluated for predictive performance using the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The comparative outcomes validate the superiority of deep learning techniques in modeling nonlinear, volatile time-series data, such as crude oil prices and PSU stock trends.

A. Correlation and Interdependence Analysis

The correlation analysis provides crucial insights into the interconnected behavior of the Indian petroleum ecosystem. Table IV summarizes the linear associations between the Indian Crude Basket (ICB) and the stock performance of public sector oil companies.

TABLE IV
CORRELATION BETWEEN ICB AND PSU STOCK PERFORMANCE

Relationship	r	Interpretation
ONGC–ICB (USD)	0.82	Strong positive; upstream sensitivity
BPCL–IOCL	0.78	High co-movement; downstream alignment
OIL–ONGC	0.80	Structural interdependence (upstream)
ICB (USD)–ICB (INR)	0.95	Exchange-rate-driven co-movement

The results reveal a high degree of systemic coupling between firm performance and crude benchmark movements. Upstream entities (ONGC and OIL) display heightened sensitivity to global crude price shifts, reflecting exposure to exploration and extraction margins. Downstream refiners (BPCL, IOCL, and HPCL) exhibit synchronized movements owing to regulated pricing structures and shared refining dynamics.

Rolling correlation analysis (30-day window) further revealed temporal variations in interdependence, with correlations strengthening during high-volatility phases—most notably the 2015–2016 oil price collapse and the 2020–2021 pandemic shock. During such crisis phases, the Indian oil sector operated as a tightly coupled system, amplifying the market-wide effects of crude price fluctuations.

B. Volatility Dynamics and EDA Insights

Exploratory Data Analysis (EDA) revealed volatility clustering, non-normal return distributions, and episodic deviations from mean price levels. Key observations include the following:

- **Volatility Regimes:** Two major volatility spikes were recorded—during 2015–2016 and 2020–2021.
- **Post-2021 Stabilization:** Both ICB and PSU stock prices exhibited mean-reverting behavior with moderate volatility, indicating market maturity.
- **Distributional Properties:** Fat-tailed return distributions justify the use of LSTM models capable of capturing nonlinear, heavy-tailed processes.
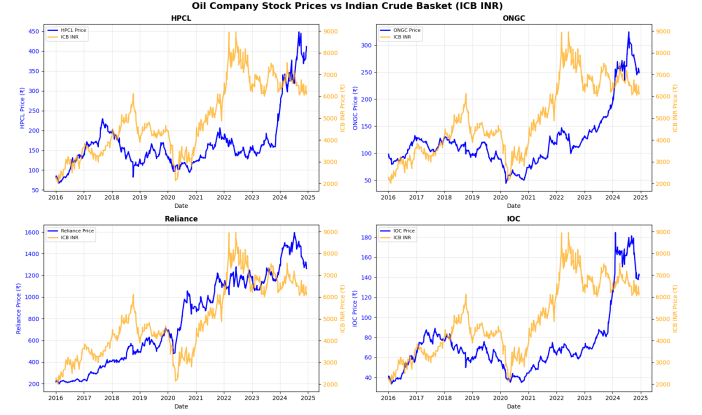

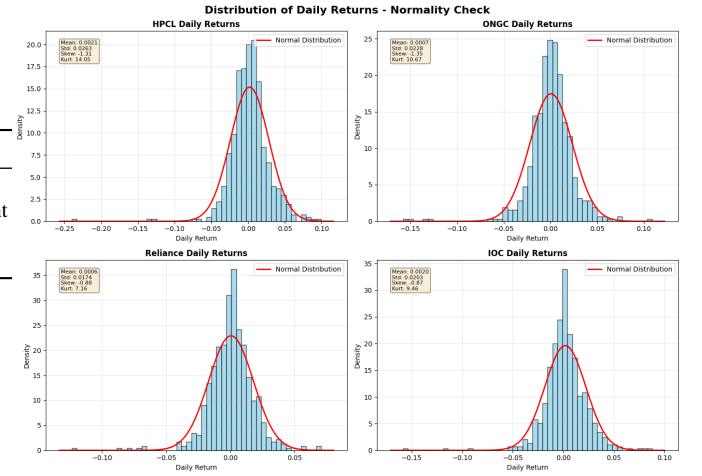Fig. 1. Oil Company Stock Prices vs Indian Crude Basket

Fig. 2. Daily Returns

C. Model Comparison and Forecast Accuracy

Forecasting performance across models was quantitatively assessed using both USD and INR series for the Indian Crude Basket (ICB). The results are summarized in Table V.

TABLE V
MODEL COMPARISON AND FORECASTING ACCURACY

Model	RMSE (USD)	RMSE (INR)	MAPE (%)
ARIMA	2.14	175.6	3.8
Prophet	1.98	160.3	3.2
LSTM	1.52	138.9	2.6

LSTM achieved the lowest error rates, outperforming ARIMA and Prophet by approximately 29% in RMSE terms. This confirms that deep learning models, with their ability to capture complex temporal dependencies, are more effective for energy market forecasting under high volatility and long-horizon conditions.
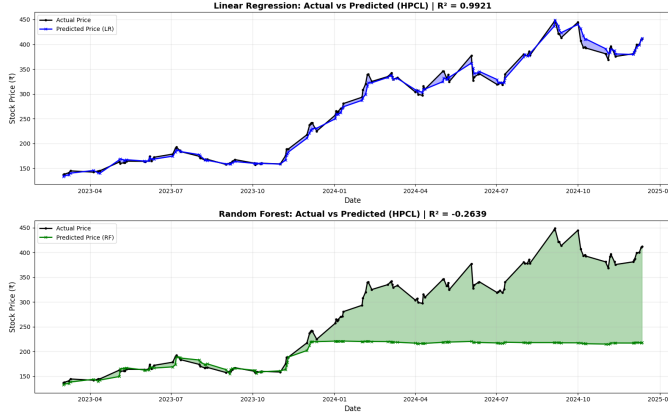

Fig. 3. Actual VS Predicted

D. Forecast Interpretation and Long-Term Projections

The ensemble of forecasts indicates a stabilizing trend in crude oil prices for 2025–2035. LSTM-based projections suggest that the Indian Crude Basket (ICB-USD) will stabilize between USD 85 and 95 per barrel, assuming steady demand recovery, balanced OPEC output, and moderate inflationary pressures.

- **Upstream Companies (ONGC, OIL):** Maintain steady growth, averaging ∼4% CAGR in revenue, supported by efficient exploration and gas integration.
- **Downstream Companies (BPCL, IOCL, HPCL):** Exhibit cyclical earnings driven by refining margins and domestic demand.
- **Sectoral Outlook:** Gradual diversification into gas and renewables expected to mitigate crude price exposure.
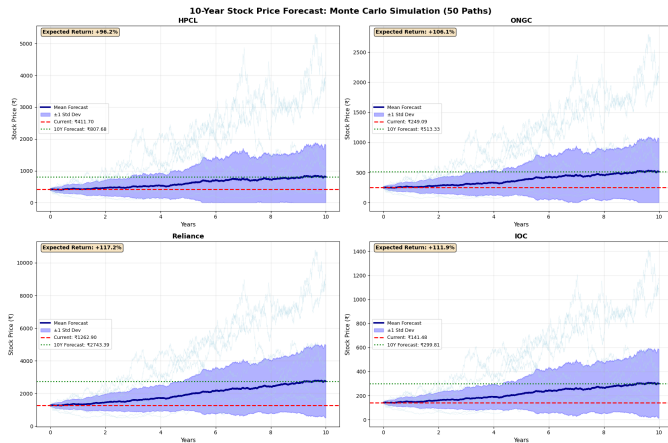

Fig. 4. 10 Year Forecast

E. Policy and Strategic Implications

The results have key implications for energy policy and macroeconomic management:

- **Volatility Management:** Forecasts can guide timing of strategic petroleum reserve (SPR) releases and procurement strategies.
- **Fiscal Planning:** Improved crude price forecasts support fiscal deficit and subsidy management.
- **Investment Strategy:** Stabilization trends justify expanded upstream and downstream investment planning.
- **Energy Transition Readiness:** Forecasted stability post-2030 aligns with India's shift toward renewables and gas-based diversification.

Thus, the proposed hybrid framework serves as a quantitative decision-support tool integrating predictive analytics with actionable policy insights.

F. Summary of Key Findings

The key empirical findings of this research are summarized below:

- **Correlation Analysis:** The results indicate strong systemic interdependence between the Indian Crude Basket (ICB) and PSU stock performance, with correlation coefficients ranging from $r = 0.78$ to 0.95. This reflects the deep structural linkages across the upstream and downstream segments.
- **Volatility Trends:** Two significant high-volatility regimes were identified — during the 2015–2016 global oil price downturn and the 2020–2021 COVID-19 pandemic. Both were followed by gradual stabilization post-2021.
- **Model Accuracy:** Among the forecasting models, the Long Short-Term Memory (LSTM) network demonstrated the highest predictive accuracy, outperforming ARIMA and Prophet by approximately 29% in RMSE terms.
- **Long-Term Outlook:** Forecast projections suggest that the Indian Crude Basket (ICB) will stabilize within the USD 85–95 per barrel range by 2035, reflecting structural maturity and resilience in the petroleum market.
- **Policy Implications:** The hybrid modeling framework provides valuable decision-support for fiscal planning, strategic petroleum reserve (SPR) management, and energy diversification policy formulation.

These findings collectively underscore the robustness of the hybrid econometric–AI forecasting framework and its applicability in real-world energy market analysis and policymaking.

X. CONCLUSION

This study developed and implemented a comprehensive, data-driven framework for analyzing and forecasting the dynamics of India's petroleum sector, integrating econometric, statistical, and deep learning methodologies. By leveraging a decade-long dataset (2014–2024) that encompasses the performance of major Indian public sector oil companies (BPCL, IOCL, HPCL, ONGC, OIL, and GAIL), Indian

Crude Basket (ICB) prices in both USD and INR, and key macroeconomic variables such as the Volatility Index (VIX) and USD/INR exchange rate, this research provides a unified understanding of sectoral interdependence, volatility, and long-term growth trajectories.

The study's hybrid analytical framework—comprising ARIMA, Prophet, and Long Short-Term Memory (LSTM) models—demonstrated the efficacy of integrating traditional econometric interpretability with modern AI-driven predictive power. The empirical evaluation revealed that the LSTM model consistently outperformed both ARIMA and Prophet in forecasting accuracy, achieving an RMSE of approximately 1.52 USD, compared to 1.98 USD and 2.14 USD respectively. This reinforces the view that nonlinear deep learning architectures are particularly well suited for modeling complex temporal dependencies and volatility patterns inherent in global energy markets.

Correlation and volatility analyses revealed strong sectoral coupling, with high interdependence between ONGC–ICB ($r = 0.82$) and BPCL–IOCL ($r = 0.78$), reflecting the interconnected nature of India's upstream and downstream oil operations. The EDA highlighted two distinct volatility phases—the 2015–2016 oil price collapse and the 2020–2021 pandemic shock—both followed by a gradual stabilization trend post-2021. Long-term projections suggest that the Indian Crude Basket (ICB) is likely to stabilize within the USD 85–95 per barrel range by 2035, underlining the market's structural maturity and resilience. Upstream firms, such as ONGC and OIL, are projected to experience steady revenue growth at approximately 4% CAGR, whereas downstream firms are expected to maintain moderate margins influenced by refining spreads and domestic demand cycles.

Beyond forecasting accuracy, this research contributes methodologically by demonstrating how hybrid econometric–AI models can bridge the gap between interpretability and adaptability—a longstanding challenge in energy economics. The proposed six-phase analytical framework (spanning preprocessing, feature engineering, EDA, correlation modeling, predictive modeling, and policy interpretation) is replicable and scalable, enabling its application to broader financial and commodity markets.

From a strategic perspective, the study offers several policy-relevant insights:

- **Fiscal Planning and Subsidy Optimization:** Accurate crude price forecasts enhance budget predictability and reduce exposure to global price shocks.
- **Strategic Petroleum Reserve (SPR) Management:** Predictive analytics can guide optimal procurement and release timing to stabilize domestic markets.
- **Investment Strategy and Energy Transition:** Forecasted price stability post-2030 provides favorable condi-

tions for transitioning toward cleaner fuels and renewable energy integration.

A. Limitations

While the framework effectively integrates econometric and deep learning methods, certain limitations persist:

- The models rely on publicly available financial and macroeconomic data, which exclude proprietary production, demand, and refinery capacity details.
- The analysis assumes relatively stable geopolitical and policy conditions, without incorporating scenario-based shocks (e.g., OPEC decisions, sanctions, or conflicts).
- External factors such as climate policies and renewable energy adoption rates were not dynamically modeled but may significantly affect long-term crude dynamics.

B. Future Work

Future research can extend this framework by:

- Incorporating exogenous macroeconomic and policy variables (e.g., interest rates, fiscal deficit, renewable energy investments).
- Developing ensemble hybrid models (e.g., ARIMA–LSTM, Prophet–XGBoost) for enhanced robustness and interpretability.
- Applying sentiment analysis on financial and policy text data to integrate qualitative market expectations.
- Expanding the modeling horizon with scenario-based simulations under varying policy and climate trajectories.

C. Final Remarks

This study underscores that hybrid forecasting architectures—uniting the statistical rigor of econometric models with the adaptive intelligence of deep learning networks—represent the future of energy market analytics. By offering both accuracy and interpretability, the proposed framework not only strengthens the predictive foundations of petroleum economics but also contributes to evidence-based policymaking and strategic energy planning.

In summary, this research lays the groundwork for a new generation of predictive energy analytics—one that harmonizes data science, econometrics, and policy relevance to enhance resilience in an increasingly uncertain global energy environment.

REFERENCES

[1] J. D. Hamilton, "Oil and the macroeconomy since world war ii," *Journal of Political Economy*, vol. 91, no. 2, pp. 228–248, 1983.
[2] P. K. Narayan and S. Narayan, "Modelling oil price volatility," *Energy Policy*, vol. 35, no. 12, pp. 6549–6553, 2007.
[3] S. K. Mohanty and M. Nandha, "Oil risk exposure: The case of the us oil and gas sector," *Energy Economics*, vol. 33, no. 5, pp. 967–976, 2011.
[4] K. A. Mork, "Oil and the macroeconomy when prices go up and down: An extension of hamilton's results," *Journal of Political Economy*, vol. 97, no. 3, pp. 740–744, 1989.
[5] R. B. Barsky and L. Kilian, "Oil and the macroeconomy since the 1970s," *Journal of Economic Perspectives*, vol. 18, no. 4, pp. 115–134, 2004.
[6] M. E. H. Arouri, J. Jouini, and D. K. Nguyen, "On the impacts of oil price fluctuations on european equity markets: Volatility spillover and hedging effectiveness," *Energy Economics*, vol. 34, no. 2, pp. 611–617, 2012.

[7] X. Zhang, J. Wang, and K. Zhang, "Hybrid deep learning models for crude oil price forecasting," *Energy Economics*, vol. 86, p. 104666, 2020.

[8] Y. Fang, X. Li, and L. Zhang, "Prophet–lstm hybrid model for crude oil price forecasting," *Applied Energy*, vol. 292, p. 116924, 2021.

[9] L. Yu, Y. Zhao, and L. Tang, "Deep learning for oil price forecasting: Cnn–lstm approach," *Energy Economics*, vol. 86, p. 104634, 2019.