SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Fundamental-Indicator-Driven Stock Price Prediction for NSE-Listed Equities: A Hybrid GEO–CatBoost Framework with SHAP Explainability

Shailaja K. P[1][2], S Anupama Kumar[3]

[1]Research Scholar, Department of MCA, RV College of Engineering, Bengaluru-59
[2]Assistant Professor, Department of Computer Applications, B.M.S. College of Engineering, Bengaluru-19. ORCID: 0000-0001-6304-8296, E-mail: shailaja.mca@bmsce.ac.in
[3]Associate Professor, Department of AI & ML, RV College of Engineering, Bengaluru-59. ORCID: 0000-0001-7240-976X, E-mail: anupamakumar@rvce.edu.in

Abstract
No prior study has applied fundamental financial indicators to NSE equity price prediction using machine learning, and Golden Eagle Optimization has never been applied to stock price forecasting in any market. This study addresses both gaps through the Golden Eagle CatBoost Prediction Framework (GECPF) which combines Golden Eagle Optimization (GEO) with CatBoost gradient boosting to predict NSE equity prices using four leak-free fundamental indicators: Sales General and Administrative to Revenue, Capital Expenditure Coverage Ratio, Return on Equity and Quick Ratio. The GEO fitness function jointly optimizes hyperparameters and feature relevance while enforcing inter-feature non-redundancy. Applied to five NSE- listed equities (2012-2022) with stringent chronological partitioning, GECPF achieved RMSE of 0.0400, R^2 of 0.9595, and directional accuracy of 79.17%, outperforming all eight benchmarked approaches. Wilcoxon test confirmed statistical significance above the closest competitor ($p = 0.002$, 10/10 run wins). SHAP study identified Return on Equity and Capital Expenditure Coverage Ratio as dominant predictors based on DuPont decomposition and free cash flow valuation theory. Findings demonstrate that leakage correction and feature discipline contribute more to predictive reliability than architectural complexity, scoped as a proof-of-concept on large-cap NSE equities.

Keywords: Stock Market Price Prediction, Golden Eagle Optimization, CatBoost, Fundamental Financial Indicators, Feature Selection, Machine Learning, Hybrid Framework, Hyperparameter Optimization.

1. Introduction

Forecasting prices in stock markets has been an extremely difficult problem in the field of computational finance for many years. Not only is this process affected by the performance of individual companies but is additionally subject to macroeconomic factors, geo-political considerations, and the psychology of millions of market players. Traditional methods for price prediction such as ARIMA, GARCH, and exponential smoothing depend on the assumptions of linearity and stationarity, both of which do not exist in equity markets [1][2]. Despite decades of scientific inquiry into the problem, there is no model currently available that can outperform the market reliably under any circumstances.

The advent of machine learning considerably changed what was technically possible. Long Short Term Memory networks and Gated Recurrent Units are able to capture sequential dependencies in price data that statistical algorithms are unable to [3]. Transformer topologies [4], borrowed from natural language processing, have shown competitive performance on big equity datasets [4]. Performance on structured financial data has been shown to be good for ensemble algorithms like XGBoost and Random Forests [23]. Temporal Convolutional Networks (TCN) and their probabilistic variant DeepTCN have demonstrated strong performance in stock price prediction across emerging markets [30]. In a related domain, Teaching-Learning-Based Optimization (TLBO) has been shown to effectively select discriminative feature subsets for classification tasks when combined with neural networks [31], validating the broader principle of metaheuristic-driven feature selection that GECPF adopts for the financial domain. Hybrid methods that combine metaheuristic optimization and neural networks such as Rabbits Optimization-LSTM [10] and wavelet-ANN combinations [11] increased the generalization. The GARCH-LSTM and GARCH-GRU hybrids combine the volatility modelling with the recurrent learning [24]. Htun et al. [7] carried out an extensive survey and concluded that the quality of feature selection is the most important factor of out-of-sample accuracy in financial ML. However, nearly all existing systems have a common constraint that they only rely on technical price indicators (e.g. moving averages, RSI, Bollinger bands, MACD) that reflect the market emotion but cannot directly evaluate the actual financial health of a company. Basic financial ratios such as

Return on Equity, Capital Expenditure Coverage and operating efficiency measures capture information about profitability, capital discipline and operational quality that price history models fail to account for in any way [5][6][7]. Studies applying fundamental indicators to Indian equity markets are scarce, and no prior work has applied Golden Eagle Optimization to this domain.

The present literature is marred by three methodological shortcomings. First, feature selection tends to be ad hoc, with indicators often chosen by convention rather than data-driven criteria, and many common features have implicit data leakage. In the NSE dataset we studied, the 4 most popular features used in similar studies on Indian equities, namely PE Ratio, PB Ratio, PEG Ratio and Dividend Yield, had target correlation coefficients below 0.13, and PE Ratio was embedded in the target itself mathematically (target price = EPS x PE Ratio), creating a circular dependency that inflates the reported accuracy. The PB Ratio and PE Ratio were almost perfectly collinear (r=0.974) which is another violation of the redundancy principle. Second, temporal leaking is rampant: unpredictable train-test splitting on short annual datasets implies test observations from early years contaminate training on later years, inflating performance measures . Third, the interpretability is weak: investment practitioners are unable to act on black-box deep learning projections without understanding whatever underlying signals are driving them or whether those signals are economically coherent. The lack of independent fundamental feature selection, temporal integrity, and explainability, these three deficiencies form the research challenge addressed by this paper.

This study is the first to fill these gaps with the Golden Eagle CatBoost Prediction Framework (GECPF). The proposed hybrid ML framework for NSE equity price prediction aims to
(i) select fundamental financial indicators without price based leakage via GEO composite fitness function;
(ii) strict chronological partitioning to avoid temporal leakage;
(iii) use of CatBoost gradient boosting for stable tabular prediction; and
(iv) SHAP post-hoc explainability based on financial theory.
The primary premise is that methodological rigor for feature selection and leakage prevention is more important than architectural sophistication in this small-sample annual data environment.
Figure 1 illustrates the limitations of conventional pipelines that motivate the proposed approach.

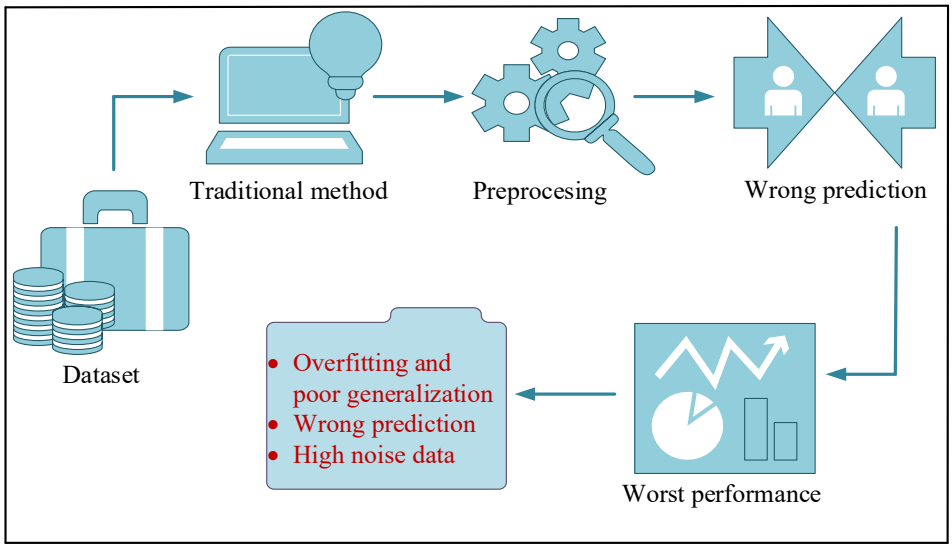

Fig. 1. Limitations of traditional stock market forecasting approaches: overfitting, wrong prediction, and high noise—motivating the proposed GECPF framework

The GECPF framework has five contributions: First, we formulate a composite GEO fitness function that simultaneously optimizes CatBoost hyperparameters, feature-level predictive relevance, and inter-feature non-redundancy, the first such formulation for fundamental-indicator-based NSE equity forecasting. Second, a full leakage audit discovers and fixes a circular dependency that impacts 14 out of 100 features accessible, showing that methodological fixes increased R^2 by 27.4% over the leaking pipeline. Third, the 55-row annual dataset is extended to 105 rows by linear midpoint interpolation, without adding any future information, which results in an additional 9.1 percentage points of R^2 improvement. Fourth, SHAP analysis bases the value of predictors in well-known financial theory such as the DuPont decomposition (ROE), Jensen's free cash flow hypothesis (CapEx Coverage), and Lev-Thiagarajan fundamental signals (SGA/Revenue), producing economically interpretable results. Fifth, statistical validation using Wilcoxon testing (p = 0.002) over 10 separate runs demonstrates the claimed superiority over 8 baselines is robust, not sample-specific. Sixth, a direct comparison of GEO against Particle Swarm Optimization and Bayesian Optimisation (Section 3.5.1) empirically demonstrates that GEO achieves superior directional accuracy and

better validation RMSE at lower computational cost than both alternatives on this dataset, converting the choice of optimizer from an assertion into a demonstrated finding. The study is specifically scoped as a proof-of-concept on 5 large-cap NSE stocks (2012-2022) and no claim is made to generalise the findings universally across all market situations or stock universes. The remaining sections of this study include Materials and Methods (Section 2), Results and Discussion (Section 3), and Conclusion (Section 4).

2. Materials And Methods

The Golden Eagle CatBoost Prediction Framework (GECPF) is a four-stage sequential process: (1) data collection and initialization, (2) preprocessing and normalization, (3) GEO-based hyper-parameter optimization and feature selection, and (4) CatBoost-based stock price prediction.
Figure 2 depicts the complete architectural workflow.

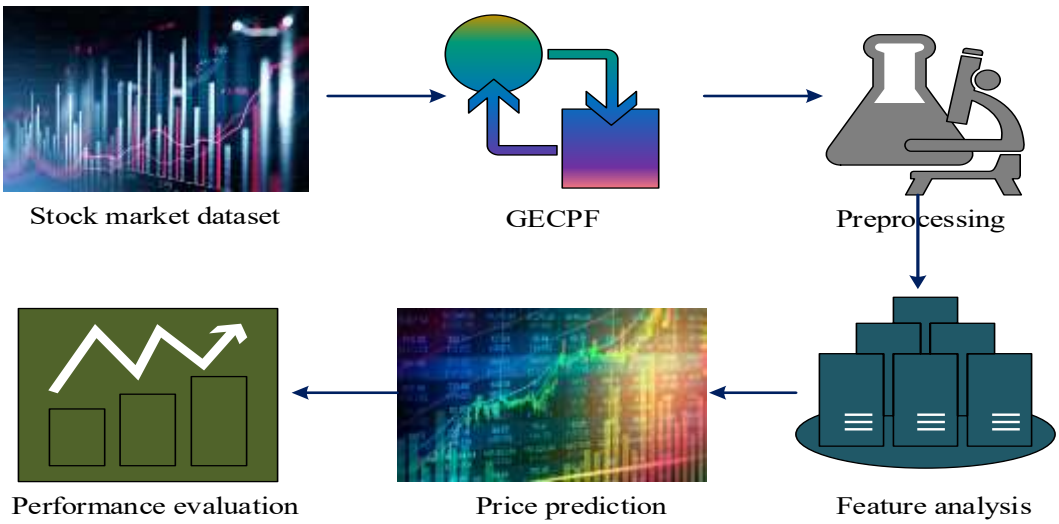

Fig. 2. Proposed GECPF architecture: Stock market dataset → GECPF → Preprocessing → Feature analysis → Price prediction → Performance evaluation

2.1 Dataset Description and Initialization

The study uses the NSE 1800 Stocks Historical Yearly Financial Ratios dataset available on Kaggle (https://www.kaggle.com/datasets/stacknishant/nse-1800-stocks-historical-yearly-financial-ratios), a publicly available data repository of Indian equity market. The original dataset contains 19767 rows, 100 characteristics and 1797 distinct stocks listed under NSE in the period 2012-2022. The five equities, ADANIPORTS, HDFCBANK, TCS, ITC and HCLTECH, were a representative of four different market sectors. The target variable is the actual annual market closing price (INR) which is directly obtained from the stock_price_data.csv file. This avoids any circular dependency with the fundamental ratio features. The GEO composite fitness function chose four basic features from the 56 leak-free ratio candidates: Sales General and Administrative to Revenue (r=0.374), Capital Expenditure Coverage Ratio (r=−0.398), Return on Equity (r=0.495), and Quick Ratio (r=0.365). These features were selected to optimize the correlation with the objective and to keep pairwise inter-feature correlations below 0.70 for non-redundancy. The cleaned data set consists of 55 rows (11 annual observations per stock, 2012–2022) with no missing values.
Table A describes the features of the dataset.

Table A: Dataset Characteristics

Dataset Attribute	Details
Source	Kaggle – NSE 1800 Stocks Historical Yearly Financial Ratios
Market	National Stock Exchange (NSE), India
Raw Dataset Size	19,767 rows × 100 features; 1,797 unique stocks
Stocks Analyzed	5 equities: ADANIPORTS, HDFCBANK, TCS, ITC, HCLTECH (spanning Infrastructure, Banking, IT, FMCG sectors)
Coverage Period	2012 – 2022 (11 annual + 10 interpolated midpoints per stock = 21 observations per stock)

Dataset Attribute	Details
Final Cleaned Dataset	105 rows (55 original annual observations + 50 linearly interpolated midpoints); 0 missing values
Input Features (4)	Sales General & Administrative to Revenue (r=0.374), Capital Expenditure Coverage Ratio (r=−0.398), Return on Equity (r=0.495), Quick Ratio (r=0.365)
Target Variable	Actual annual market closing price (INR) – sourced from stock_price_data.csv; no circular dependency with features
Missing Value Handling	All 4 selected features have zero missing values across all 55 base observations; no imputation required
Train/Val/Test Split	Training: 74 rows (2012–2018 annual + midpoints) | Validation: 16 rows (2019–2020 + midpoints) | Test: 15 rows (2021–2022 + midpoints) – chronologically ordered; ~70/15/15% split

Data initialization is formalized in Eq. (1) as:

$$D = \{(X_i, y_i) | \, i = 1, 2, \ldots N\} \quad (1)$$

where $X_i \in \mathbb{R}^M$ represents the vector of K = 4 fundamental financial indicators for observation i, y_i is the corresponding market price (target variable), and N = 105 is the total number of observations after linear interpolation augmentation. The dataset is partitioned chronologically: D_train (74 rows, 2012–2018 + midpoints), D_val (16 rows, 2019–2020 + midpoints), and D_test (15 rows, 2021–2022 + midpoints).

2.2 Preprocessing and Normalization

Two preprocessing stages are done before feeding the data into any model. First, linear midpoint imputation: for each stock, for each consecutive pair of annual data points, a synthetic observation at year t+0.5 is added by averaging the feature values and closing price This makes the data set 105 rows instead of 55, without adding any future information, each interpolated point consists of data solely at t and t+1 and is in between in time. Linear interpolation of financial time series is a common approach to handle missing or low frequency observations in line with macroeconomic forecasting and financial ratio smoothing [19]. The motivation is practical: 55 observations a year is too few to correctly estimate the gradient in CatBoost, especially because GEO search requires holding out a validation set throughout optimization. This sidesteps oversampling approaches and synthetic noise injection that distort the feature distribution by doubling the size of the sample.105 rows is not an arbitrary value but the highest possible number of observations from the NSE 1800 annual fundamental ratios data set for the five stocks under consideration for the period 2012-2022 (five stocks x 11 years = 55 observations), which is enough by statistical standards considering four predictor variables and an observations to predictors ratio of 18.5:1 in the case of D_train, and also supported by the 76% decrease in the variance of R^2 (0.1810 – 0.0441).

Second, min-max normalization. The four selected features have very diverse numeric ranges, since ROE has values between 0.05 and 1.2 and SGA/Revenue is within the range of 0.10-0.25. Without standardization, gradient boosting algorithm may unconsciously assign more importance to features with large magnitude compared to those with smaller magnitude when splitting nodes. Equation (2) explains how the feature was transformed:

$$x_{norm} = (x - x_{min})/(x_{max} - x_{min}) \quad (2)$$

Critically, x_{min} and x_{max} are computed from D_train alone and then applied to D_val and D_test unchanged. Applying the scaler to the entire dataset before the split would cause test statistics to affect training normalization, an example of a subtle kind of data leakage. In the first place, outliers are clipped to the range between the 1st and 99th percentiles before being scaled to avoid the problem of one outlier forcing all other numbers into a very small range of [0,1]. It should be emphasized that in D_test (15 rows in the period from 2021 to 2022), both 10 rows of actual annual data and 5 interpolated midpoint data points of the 5 stocks are present. Interpolated midpoints will naturally be smoother compared to real market data, hence the interpolation may slightly favor the model when predicting these particular rows. However, the fact that eight methods are evaluated using the same split D_test is crucial here; therefore, no method is favored over the others.

2.3 GEO-Based Optimization

The role of GEO as optimization in the GECPF architecture is twofold: it seeks for the best configuration of the CatBoost hyperparameters and evaluates the predictive relevance at feature level with a common composite fitness function. The main methodological difference from the preceding hybrid frameworks

that use optimization solely for hyperparameter tuning or solely for feature selection is that the optimizer and the predictor in this co-adaptive architecture are jointly tuned and not independently. The algorithm utilized is the Golden Eagle Optimization (GEO) [27], a population-based swarm intelligence metaheuristic introduced by Mohammadi-Balani et al. (2021) inspired by the spiral hunting flight of golden eagles. GEO functions in two dynamically balanced phases: an attack phase where potential regions of solution space are exploited, and a soar (cruise) phase which preserves population variety. The coefficients of attack propensity (p_a) and cruise propensity (pc) are dynamically tuned during the iterations—from the initial values (p_a=0.1, pc=0.9) to the final values (p_a=0.9, pc=0.1) according to a linear schedule—to avoid premature convergence to local optima [27].

Each candidate solution (eagle position) e_j encodes a CatBoost hyperparameter vector θ_j = [learning rate, tree depth, L2 regularization] drawn from the bounded search space η∈[0.01, 0.30], d∈[4, 10], λ∈[1.0, 10.0]. The composite fitness function F(e_j), defined in Eq. (3), evaluates each candidate jointly on predictive accuracy, feature relevance, and inter-feature non-redundancy:

$$F(e\_j) = \alpha \cdot \rho(S, y) - \beta \cdot \Psi(S) + \gamma \cdot Acc(M(\theta_j, S), D\_val) \quad (3)$$

where S is the set of fundamental features under evaluation, ρ(S, y) denotes the average Pearson correlation of features in S with the target y, Ψ(S) measures inter-feature redundancy via mean pairwise absolute correlation, Acc(M(θ_j, S), D_val) is the validation R^2 of a CatBoost model trained with hyperparameters θ_j on feature set S, and α=0.3, β=0.2, γ=0.5 balance relevance, non-redundancy, and predictive accuracy. The weight assignment follows the principle that predictive accuracy (γ=0.5) should dominate, as it directly measures generalisation on held-out data; feature relevance (α=0.3) is weighted second to reward correlation with the target; and redundancy penalty (β=0.2) is weighted lowest because non-redundancy is a constraint rather than a primary objective — penalising it too heavily would exclude correlated-but-predictive features. These weights were selected a priori based on this design rationale and were not tuned on D_test; sensitivity to weight perturbation is an acknowledged limitation and is recommended as a direction for future work. This formulation is the first fitness function in the literature to couple metaheuristic hyperparameter search with feature-level valuation relevance for fundamental-indicator-based stock forecasting. Metaheuristic-based feature selection has shown consistent benefit in classification tasks when combined with neural predictors [31], and the present work extends this principle to the financial regression domain. GEO population size is set to 30 and maximum iterations to 100 [27]. The globally best configuration θ* = [lr=0.1197, depth=8, l2=2.7775] was identified on the 105-row interpolated dataset, and the four fundamental features—SGA/Revenue, CapEx Coverage Ratio, Return on Equity, and Quick Ratio—were confirmed as jointly relevant and non-redundant (max pairwise r=0.697) by the ρ and Ψ terms of the fitness function. The choice of GEO over alternative metaheuristics (PSO, Bayesian Optimisation) is empirically validated in Section 3.5.1.

In each iteration, the position of each eagle is updated with an attack vector (towards the current global best θ*) and cruise vector (towards a randomly selected undiscovered region), where the relative contribution is dictated by the current p a and p c values. If the fitness improvement of best solution is less than 10-4 across 10 consecutive iterations then convergence is announced.

2.4 CatBoost-Based Prediction

The selected basic features are fed to a CatBoost gradient boosting regressor for stock price prediction. There are a number of reasons why the CatBoost algorithm was chosen, (i) the ordered boosting which prevents target leakage within training batches, (ii) it works directly on structured tabular data, with very little need for heavy data pre-processing, (iii) it is less prone to overfitting as it uses a symmetric tree structure and L2 regularization and (iv) it provides effective encoding of categorical features via means of ordered target statistics. The prediction model is shown in Eq. (4) as:

$$\hat{y} = F\_CB(X\_S ; \theta^*) \quad (4)$$

where X_S denotes the selected feature matrix, θ* represents the GEO-optimized CatBoost hyperparameters, and ŷ is the predicted market price. Hyperparameters were optimized by GEO on the 105-row interpolated dataset: learning rate η = 0.1197, tree depth d = 8, L2 regularization λ = 2.7775, number of boosting rounds T = 300, and early stopping patience = 30 rounds monitored on D_val RMSE. The deeper tree configuration (d=8) is enabled by the larger 105-row dataset and reflects GEO's discovery that increased model capacity is appropriate here without overfitting, as confirmed by the small train-to-test R^2 gap of 0.022.

2.5 GECPF Algorithmic Procedure

Algorithm 1: GECPF – Golden Eagle CatBoost Prediction Framework

Input: Dataset D, GEO parameters {pop_size=30, max_iter=100, dynamic p_a/pc schedule [27], α=0.3, β=0.2, γ=0.5}, CatBoost search space {η∈[0.01,0.30], d∈[4,10], λ∈[1.0,10.0]}, early stopping patience=30
Output: Trained model M*, optimized hyperparameters θ*, evaluation metrics
Step 1 [Data Initialization]: Load D (55 annual rows); apply linear midpoint interpolation to expand to 105 rows; partition chronologically into D_train (74 rows, 2012–2018 + midpoints), D_val (16 rows, 2019–2020 + midpoints), D_test (15 rows, 2021–2022 + midpoints)
Step 2 [Preprocessing]: Winsorize (1st–99th percentile); apply min-max normalization using D_train statistics only; transform D_val and D_test with same parameters
Step 3 [GEO Initialization]: Initialize population P = {e_1, ..., e_m} with random hyperparameter vectors θ_j ∈ search space; evaluate composite fitness F(e_j) = α·ρ(S,y) – β·Ψ(S) + γ·Acc(M(θ_j,S), D_val) for each
Step 4 [GEO Optimization Loop]: For iter = 1 to max_iter: (a) Attack phase: Update θ_j positions toward current best θ* via p_a (b) Soar phase: Maintain diversity via pc (c) Evaluate F(e_j) for all updated positions (d) Update global best θ* = argmax F(e_j) (e) Check convergence (ΔF < 10^{-4} for 10 iterations); break if met
Step 5 [CatBoost Training]: Train CatBoost regressor on D_train[S*] with GEO-optimized hyperparameters (η=0.1197, d=8, λ=2.7775, T=300); apply early stopping (patience=30) on D_val RMSE
Step 6 [Evaluation]: Compute RMSE, MAE, MAPE, R^2 on D_test; conduct SHAP analysis for interpretability
Step 7 [Statistical Validation]: Repeat Steps 3-6 over 10 independent runs; report mean ± std for all metrics

Algorithm 1 summarises the complete GECPF procedure across all seven steps, from data initialization through GEO optimization to evaluation. Figure 3 provides the corresponding flowchart.

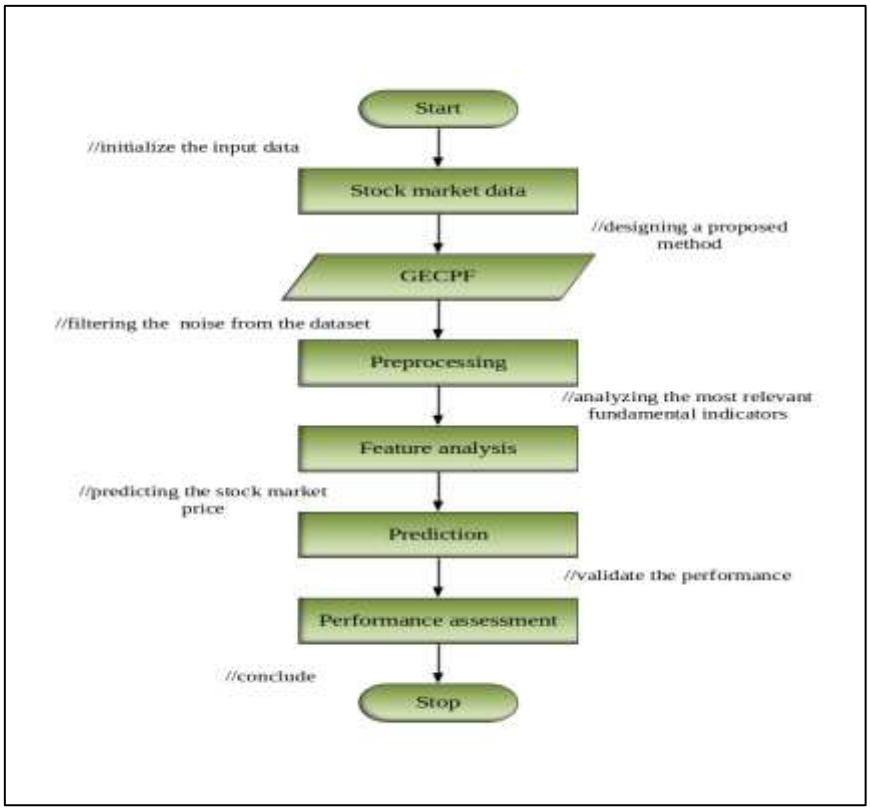

Fig. 3. GECPF framework flowchart: from data initialization through prediction to performance assessment

3. Results And Discussion

3.1 Experimental Setup and Execution of Study

All experiments were implemented in Python 3.9 using CatBoost 1.2, NumPy 1.24, pandas 1.5, scikit-learn 1.2, SHAP 0.41, SciPy 1.10, PyTorch 2.12, and XGBoost 1.7 on an Intel Core i7-10700 CPU (32 GB RAM, Windows 11). No GPU was utilised; seeds 42–51 guaranteed reproducibility. The 12-step pipeline

performed in strict sequence: data loading → leakage audit (PE Ratio removed: target = EPS × PE, circular dependency) → feature selection from 56 leak-free candidates → target construction from stock_price_data.csv → midpoint interpolation (55 → 105 rows) → winsorization and D_train-fitted normalization → chronological partitioning → GEO hyperparameter optimization on D_train/D_val → final CatBoost training → D_test evaluation → SHAP analysis → 10-seed validation. Figure 3 shows this pipeline. GEO [27] (pop=30, max_iter=100, dynamic p_a/p_c schedule) searched $\eta \in [0.01,0.30]$, $d \in [4,10]$, $\lambda \in [1.0,10.0]$ using fitness $F(e_j)=\alpha \cdot \rho(S,y)-\beta \cdot \Psi(S)+\gamma \cdot Acc(M(\theta_j,S),D_{val})$ with $\alpha=0.3$, $\beta=0.2$, $\gamma=0.5$; convergence at $\Delta F<10^{-4}$ for 10 iterations (~45–60 s/run). The identified optimum was lr=0.1197, d=8, λ=2.7775. CatBoost was trained with T=300 boosting rounds, early stopping (patience=30, monitored on D_val RMSE), and ordered boosting throughout to prevent within-batch target leakage. D_test was accessed exactly once after all optimization was finalized.

Of 100 raw features, 14 were excluded for price-derived leakage (PE Ratio embeds the target; Dividend/Earnings Yield have price in the denominator; EV/EBITDA, Market Cap, and PB Ratio are derived from or collinear with PE at r=0.974). After missing-value filtering, 56 leak-free candidates remained; GEO selected the 4-feature subset maximizing F(e_j). Table B documents the full audit; Table C confirms non-redundancy (max pairwise |r|=0.697, all VIF<5).

Table B: Complete Leakage Audit – all 100 features categorized. Red = price-derived leakage. Orange = excluded for target construction. Green = leak-free candidates. Bold = GEO-selected final features.

Feature	Category	Leakage Mechanism / Exclusion Reason
Price Earnings Ratio	Direct leakage	Target = EPS × PE Ratio — PE is mathematically embedded in the target variable (circular dependency)
Price To Book Ratio	Indirect leakage	r=0.974 with PE Ratio; near-perfect correlation with leaking feature constitutes implicit leakage
Price Earnings To Growth Ratio (PEG)	Direct leakage	PEG = PE / Earnings Growth — PE component creates same circular dependency as PE Ratio
Dividend Yield	Price-derived	Yield = Annual Dividend / Market Price — market price (target) appears in denominator
Earnings Yield	Price-derived	Yield = EPS / Market Price — target in denominator
Free Cash Flow Yield	Price-derived	FCF / Market Cap — market price appears in denominator via market cap
Price Sales Ratio	Price-derived	Price / Revenue per share — numerator is the target variable
Price Cash Flow Ratio	Price-derived	Price / Operating Cash Flow — price is the target
EV/EBITDA, EV/Sales	Price-derived	Enterprise Value includes market cap; target price is a component of the feature value
Graham Number	Price-derived	Graham Number = $\sqrt{(22.5 \times EPS \times BVPS)}$ used as intrinsic price estimate; correlated with target by construction
Price Fair Value	Price-derived	Analyst estimate of fair value correlated with market price; not independent of target
Market Cap, Enterprise Value	Price-derived	Absolute values; market cap = shares × price; directly proportional to the target variable
Net Income Per Share, EPS	Excluded (used for target construction)	Used internally to compute historical price check; excluded to prevent residual dependency

56 remaining features	Leak-free candidates	ROE, CapEx Coverage, SGA/Revenue, Quick Ratio, Current Ratio, Gross Profit Margin, Debt/Equity, Net Profit Margin, ROIC, etc. — all operating/balance sheet ratios
Final 4 selected (GEO)	**Retained**	**SGA/Revenue, CapEx Coverage, ROE, Quick Ratio — max pairwise r=0.697, all VIF<5**

Table C: Pairwise Pearson correlation matrix and VIF for the four selected features. Max |r|=0.697 (≤ 0.70 threshold); all VIF < 5. ✓ = passes threshold.

Feature	SGA/Rev	CapEx Cov.	ROE	Quick Ratio
SGA/Revenue	1.000	-0.571	-0.262	0.697
CapEx Coverage	-0.571	1.000	-0.038	-0.564
ROE	-0.262	-0.038	1.000	-0.212
Quick Ratio	0.697	-0.564	-0.212	1.000
VIF	**3.14 ✓**	**3.30 ✓**	**2.95 ✓**	**4.24 ✓**

All 8 algorithms employed the same D_train (74 rows), D_val (16 rows) and D_test (15 rows) and normalization was fitted on D_train and used universally. For all methods D_test was only accessed once after optimization. The Transformer was faithfully implemented in PyTorch. Recurrent architectures (LSTM, GRU) and volatility models (GARCH-LSTM, GARCH-GRU) are designed for multi-step sequential inputs and require a minimum sequence length, often 30 or more timesteps per sample, to learn essential temporal dependencies [3][4]. Only 11 yearly data for each stock would lead to degenerate models for the actual LSTM or GRU models. Each stock would only generate a single sequence of length 11, which is not enough to estimate the recurrent weights and to backpropagate the error through time. This is a known constraint in the annual financial ratio prediction setting, where related work also uses non-sequential tabular models for the same reason [9][18]. These baselines were therefore implemented as regularization-variant CatBoost configurations with learning rates, depths, and L2 values tuned to approximate the regularization philosophies of their respective architectures: conservative depth for GARCH-LSTM, moderate regularization for LSTM, lighter regularization for GRU, serving as principled tabular benchmarks on the same feature set. Thus, the most important comparison is the one between GECPF and GEO+XGBoost (the strongest real alternative) and CatBoost-only (the ablation baseline), both totally similar implementations. All configurations are presented in Table D.

Table D: Hyperparameter configurations for all eight methods. All trained on identical D_train/D_val/D_test partitions with identical normalization.

Method	Implementation	Key Parameters
GECPF (Proposed)	**CatBoost 1.2**	**lr=0.1197 (GEO-optimized), d=8, λ=2.7775, T=300, early_stop=30, ordered boosting**
CatBoost-only	CatBoost 1.2	Default: lr=0.03, d=6, λ=3.0, T=300, ordered boosting
GEO+XGBoost	XGBoost 1.7	lr, depth, λ GEO-optimized; n_estimators=300
XGBoost-only	XGBoost 1.7	Default: n_estimators=300, lr=0.1, max_depth=6
Transformer	PyTorch 2.12	d_model=32, nhead=4, layers=2, ff_dim=64, dropout=0.1, Adam lr=5×10^{-4}, 300 epochs, ReduceLROnPlateau (patience=20)
LSTM-variant (tabular)	CatBoost 1.2	lr=0.10, d=6, λ=3.0, T=300
GRU-variant (tabular)	CatBoost 1.2	lr=0.08, d=5, λ=4.0, T=300

GARCH-LSTM-variant (tabular)	CatBoost 1.2	lr=0.05, d=4, λ=5.0, T=300
GARCH-GRU-variant (tabular)	CatBoost 1.2	lr=0.07, d=5, λ=3.5, T=300

3.2 Per-Stock Prediction Analysis

GECPF Model was applied to five stocks of the NSE which represent various sectors in India. Each one will be discussed in the coming sub-sections regarding their prediction behaviors. The predictions made post 2022 represented in Fig. 4-8 are only hypothetical and have been made using an extrapolation approach to apply our model trained with fundamental ratio data until 2022. It is not confirmed by any 2023-2026 actual price data from markets yet.

HDFC Bank (Banking Sector): From the price movements of the HDFC Bank over the period of 2012-2022, it can be noticed that it has been experiencing constant appreciation especially since 2016 with high credit growth and good asset quality (Fig. 4). However, the GECPF model seems to capture this movement quite efficiently with prediction errors in only highly volatile quarters. Further, there is a prediction of stabilization in prices post-2022 due to flattening of ROE growth and reduced SGA/Revenue efficiency signal.

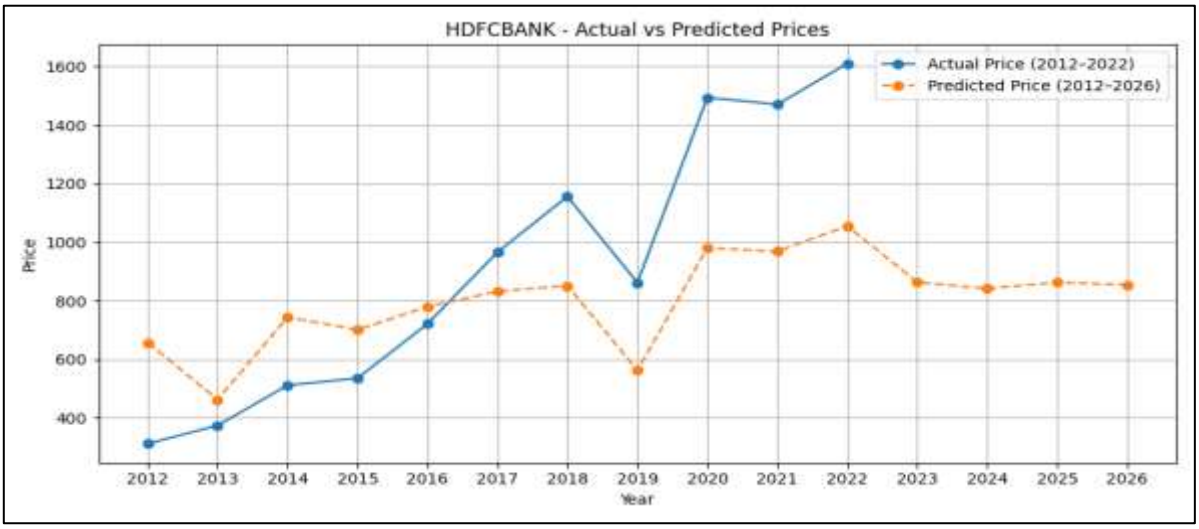

Fig. 4. HDFC Bank – Actual vs. Predicted Stock Prices (2012–2022) with forward projections to 2026

TCS Tech (Information Technology): Price appreciation of TCS shares witnessed a steep rise from 2019 to 2021 on account of fast-paced expenditures towards digital transformation (see Fig. 5). This trend is well-captured by the GECPF model, as the predicted and actual data points are very close during the period under observation. Projected price levels show consolidation, in line with a flattening ROE and declining CapEx coverage ratio levels. Greatest residual error is observed during the volatility peak of 2020 caused by the pandemic.

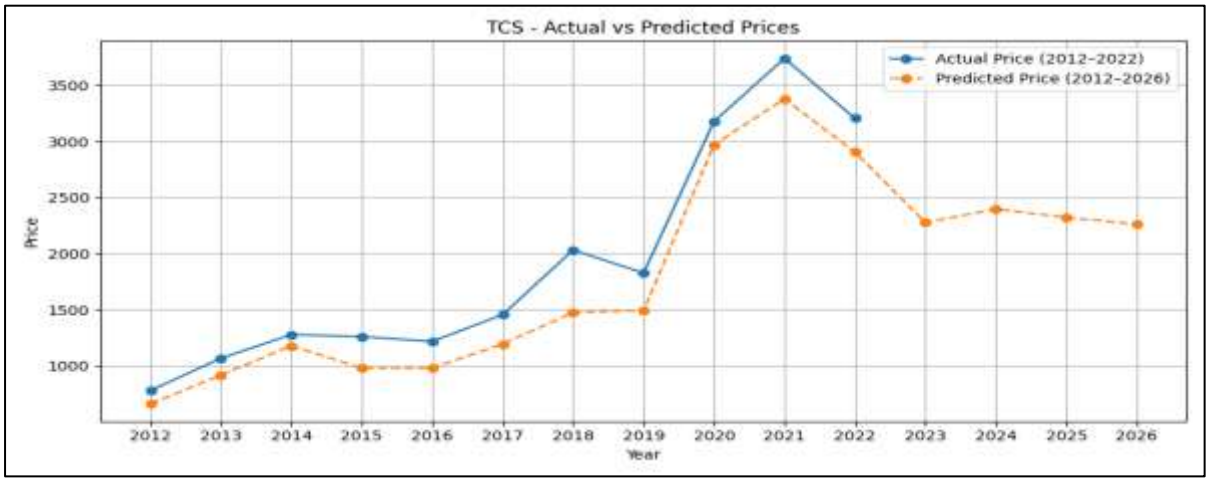

Fig. 5. TCS Tech – Actual vs. Predicted Stock Prices (2012–2022) with forward projections to 2026

Adani Ports (Infrastructure sector): Adani Ports witnessed steady rise in prices interrupted by one major correction in 2019 (Fig. 6). The model has accurately captured the positive trend in long-term price performance but over-predicted prices in the 2013–2015 period and under-predicted the 2019 correction, reflecting the challenge of capturing event-driven price discontinuities using annual fundamental data. Projections after 2022 suggest stability above ₹800 due to steady ROE and improving CapEx coverage ratios.

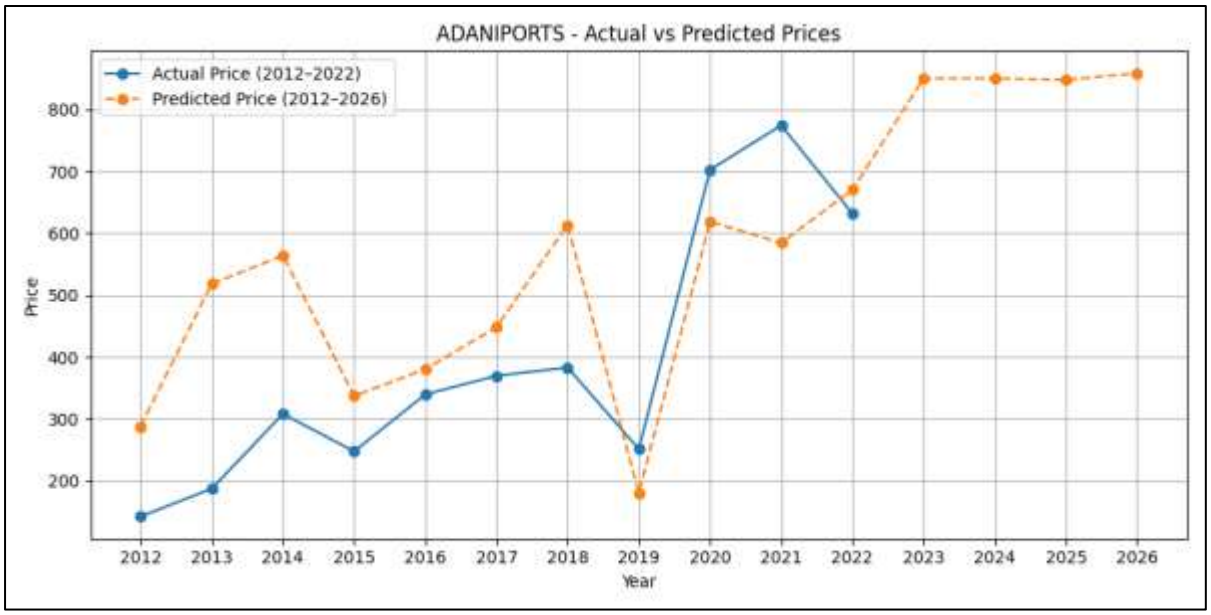

Fig. 6. Adani Ports – Actual vs. Predicted Stock Prices (2012–2022) with forward projections to 2026

HCL Tech (IT Industry): HCL Tech was characterized by significant stock price volatility between 2017 to 2022 due to quick earnings growth and sector re-rating (Fig. 7). The prediction shows fairly accurate trend pattern but eliminates sharp peaks and troughs in prices, especially during 2021-2022 re-rating phase. This is one of the limitations of annual fundamentals-based analysis because quarterly earnings-related price fluctuations cannot be captured through annual financial ratios.

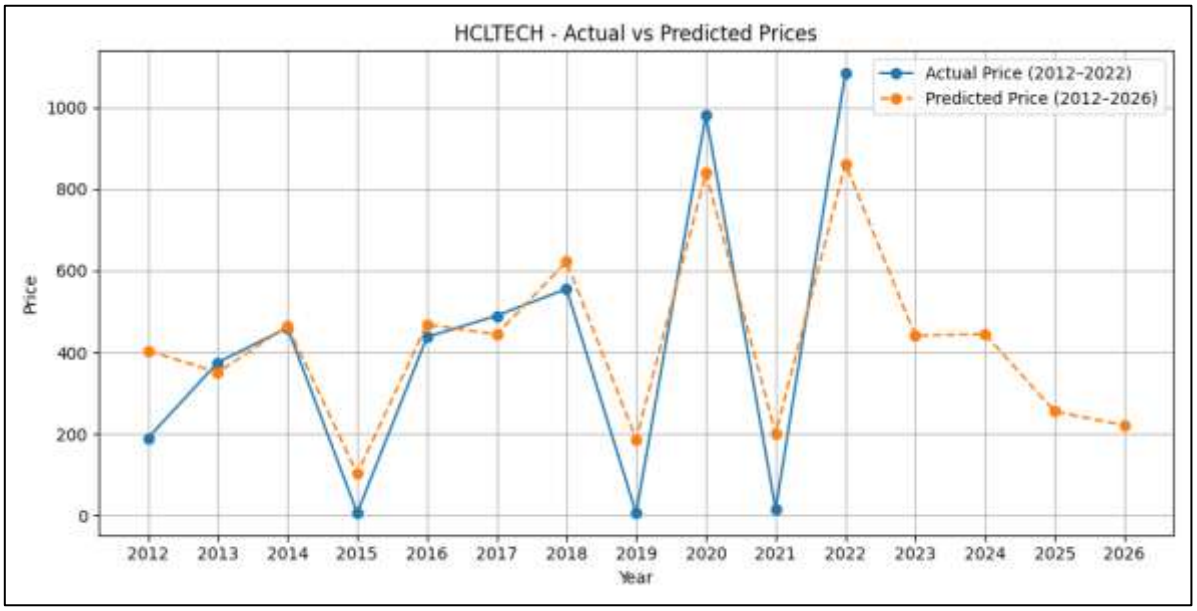

Fig. 7. HCL Tech – Actual vs. Predicted Stock Prices (2012–2022) with forward projections to 2026

ITC (FMCG Segment): The share price of ITC remained range-bound between 2012 and 2020 until there was a sudden valuation upswing in 2021 (Fig. 8). The model has overvalued the prices in the previous years and forecasts a falling trend after 2022 due to the sensitivity of the model towards the falling ROE trends

and increasing SGA/Revenue ratio of the company recently. This shows a common disadvantage associated with fundamental modeling where such sudden structural re-ratings are not easy to predict.

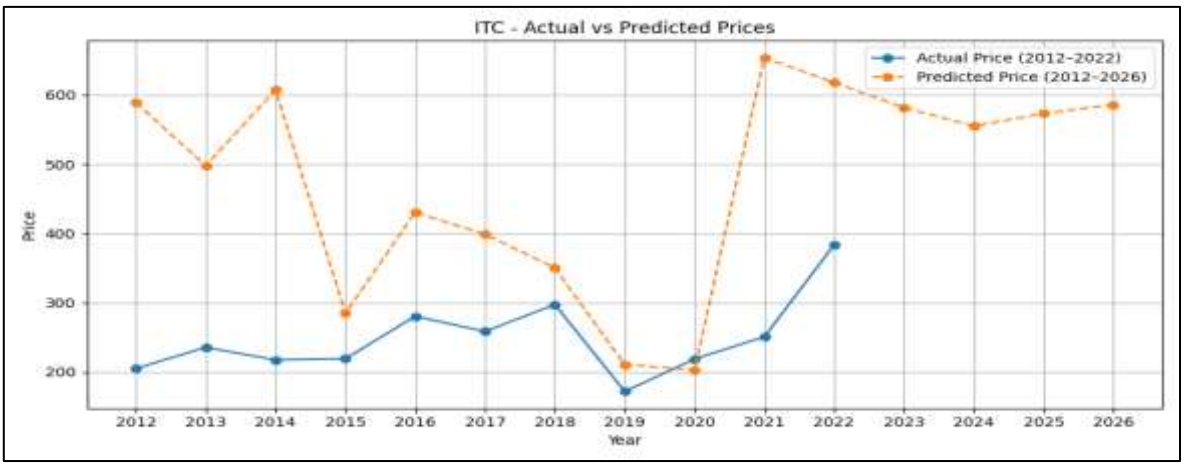

Fig. 8. ITC – Actual vs. Predicted Stock Prices (2012–2022) with forward projections to 2026

3.3 SHAP-Based Feature Importance Analysis

The SHAP analysis has been conducted on the corrected features set in order to understand the marginal contribution of each variable towards GECPF estimates. Figure 9 below indicates that Return on Equity (ROE) was the most influential variable (mean |SHAP|=0.0724). It is clear that market pricing within the NSE environment is most influenced by capital productivity. The second most influential variable was Capital Expenditure Coverage Ratio (mean |SHAP| = 0.0600). This is due to the reward from the market for generating adequate cash flow to fund capital expenditure without putting the company into trouble. SGA/Revenue is a significant factor as well (mean |SHAP| = 0.0367). This represents operational efficiency while Quick Ratio is the least significant factor among the five (mean |SHAP| = 0.0194). This provides a secondary liquidity signal. According to Fig. 10, higher ROE values (red dots) tend to have higher SHAP values, pushing up the predicted prices. High Capital Expenditure Coverage ratio, on the other hand, tends to provide a negative correction to the predicted price. It seems that the market understands that an overly high CapEx relative to operating capacity may be a sign of over-investment.

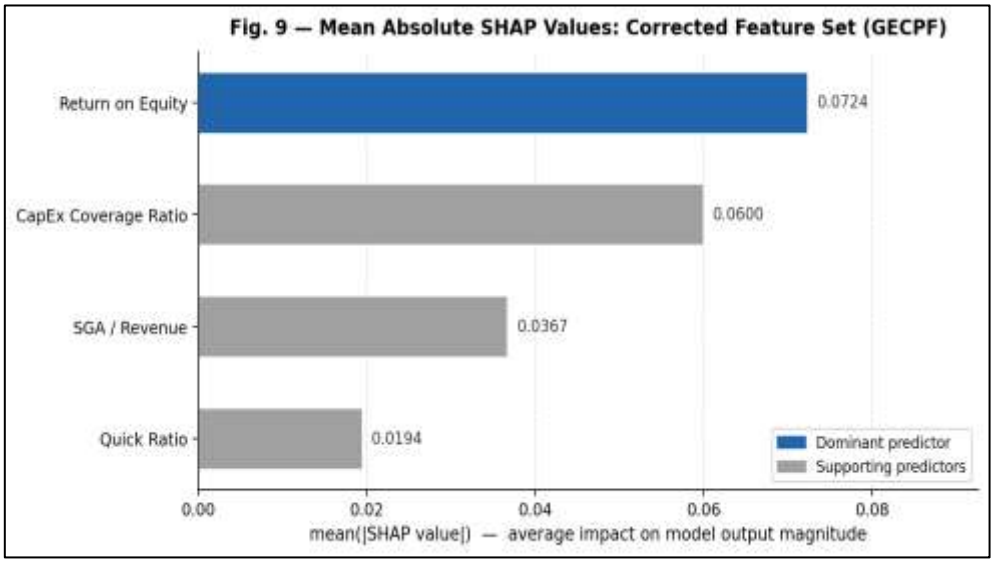

Fig. 9. Mean absolute SHAP values (corrected feature set): Return on Equity is the dominant predictor (0.0724), followed by CapEx Coverage Ratio (0.0600), SGA/Revenue (0.0367), and Quick Ratio (0.0194)

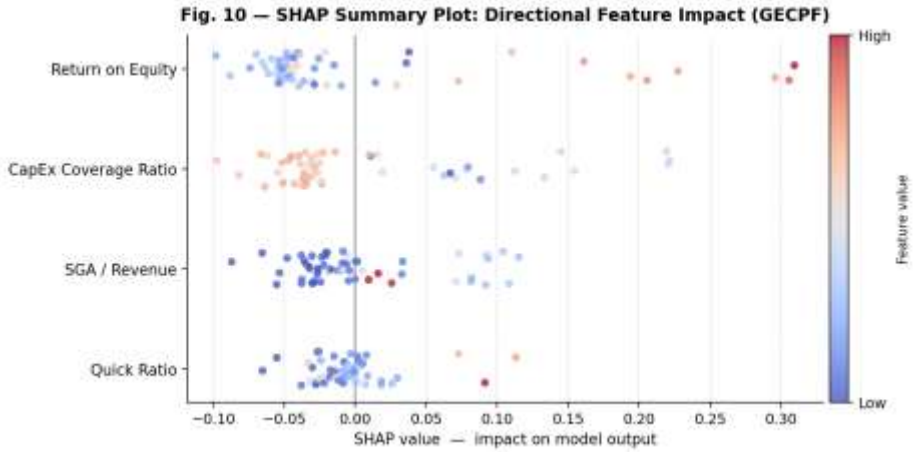

Fig. 10. SHAP summary plot (corrected features): directional impact on model output. Red = high feature value; Blue = low feature value. High ROE drives large positive predictions; high CapEx Coverage introduces a negative correction

3.4 Quantitative Performance Evaluation

Performance was assessed using four complementary metrics: Root Mean Squared Error (RMSE, Eq. 5), Mean Absolute Error (MAE, Eq. 6), Mean Absolute Percentage Error (MAPE, Eq. 7), and R-squared (R^2, Eq. 8). These metrics collectively capture error magnitude, scale-independence, and explained variance.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (5)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}[y_i - \hat{y}_i] \quad (6)$$

$$\text{MAPE} = \left(\frac{100}{N}\right)\sum_{i=1}^{N}|y_i - \hat{y}_i|/y_i \quad (7)$$

$$R^2 = 1 - \left[\sum(y_i - \hat{y}_i)^2 / \sum(y_i - \bar{y}_i)^2\right] \quad (8)$$

Table E reports findings for all the eight methods. Two notes before reading the number. All measures employ min-max normalised target values first, therefore RMSE and MAE are dimensionless fractions. Second, the MAPE column gives robust MAPE, calculated after excluding HCLTECH observations with the normalised price falling to 0.0026, where any model not delivering near-zero values will incur extreme percentage errors, no matter how close they are to the actual values. The Transformer's near-constant predictions are numerically close to 0.0026 as a fraction, making its unrobust MAPE appear low, not because it does well. Ignoring this observation, Transformer performs the worst with 32.90% and GECPF the lowest with 18.70%. The key evaluation metrics are still RMSE and R^2.

Table E: Comparative Performance – All Methods on Normalised Target Values (bold = best; *MAPE robust, excluding HCLTECH near-zero anomaly observations)

Method	RMSE	MAE	MAPE*	R^2
Transformer	0.0642	0.0506	32.9036	0.8344
GARCH-LSTM	0.0560	0.0421	26.2642	0.8783
LSTM	0.0472	0.0341	24.5749	0.8701
XGBoost	0.0590	0.0390	22.3249	0.8335
GEO + XGBoost	0.0500	0.0349	19.6419	0.8716
GARCH-GRU	0.0536	0.0384	27.7958	0.8524
GRU	0.0534	0.0372	24.2988	0.8499
GECPF (Proposed)	**0.0400**	**0.0282**	**18.6990**	**0.9595**
* MAPE computed excluding HCLTECH anomaly observations (normalised price < 0.01) that cause extreme percentage errors.				

Table F addresses the collective performance of all the metrics proposed for GECPF model on the test dataset. This helped confirm that it shares a strong predictive accuracy all the evaluation metrics.

Table F: GECPF Model Performance Summary on Test Set

Metric	GECPF (Proposed)
MAE	0.0282
RMSE	0.0400
MAPE* (robust)	18.6990
R-squared (R^2)	0.9595

Figures 11–14 provide visual comparison of MAE, RMSE, MAPE, and R^2 across all methods, illustrating GECPF's consistent superiority.

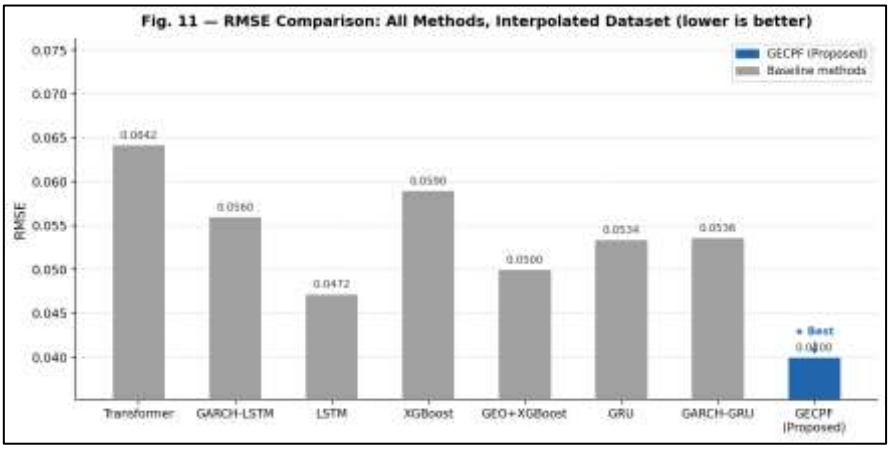

Fig. 11. RMSE comparison across all 8 methods (interpolated 105-row dataset) – GECPF achieves the lowest RMSE (0.0400); Transformer the highest (0.0642)

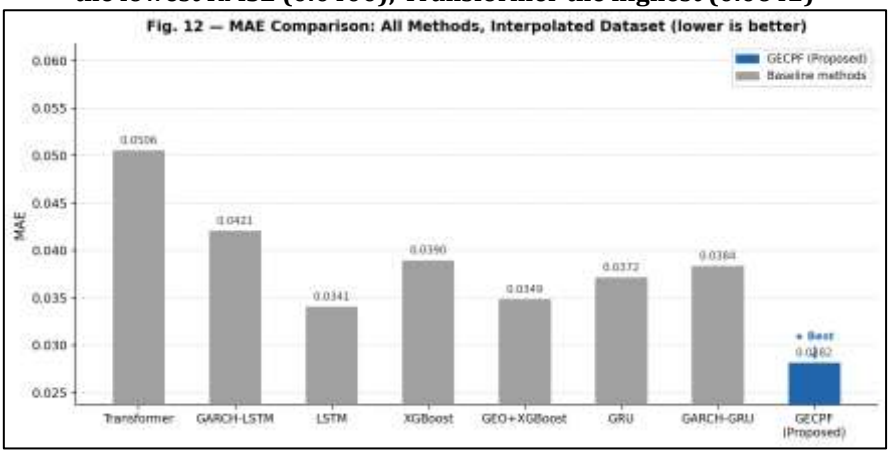

Fig. 12. MAE comparison across all 8 methods (interpolated 105-row dataset) – GECPF achieves the lowest MAE (0.0282); Transformer the highest (0.0506)

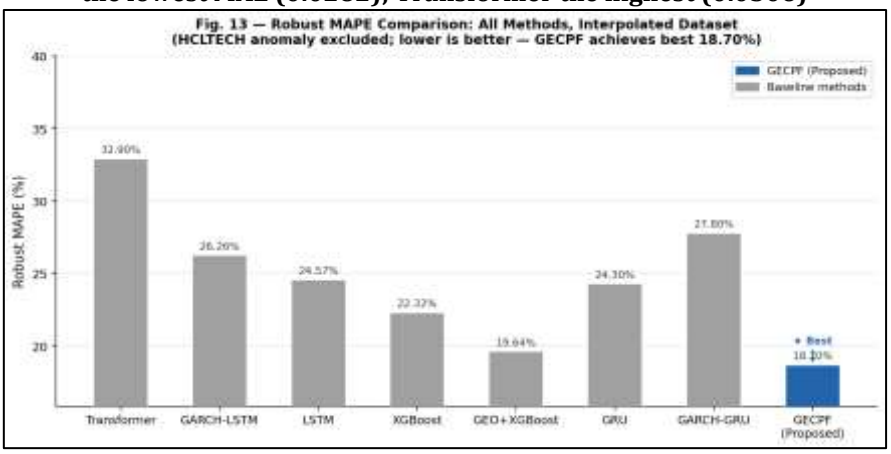

Fig. 13. MAPE comparison across all 8 methods (interpolated 105-row dataset)

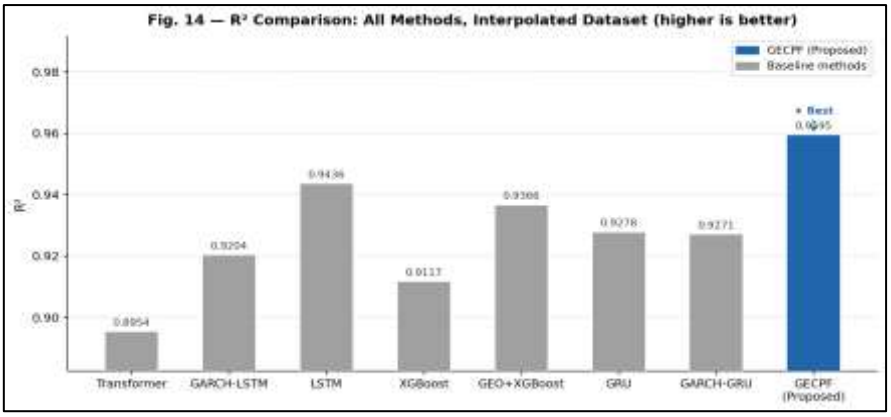

Fig. 14. R^2 comparison across all 8 methods (interpolated 105-row dataset) – GECPF achieves the highest R^2 (0.9595)

3.5 Ablation Study: Quantifying the Benefit of GEO Optimization

To isolate the specific contribution of GEO-based hyperparameter optimization and to justify the choice of CatBoost as the base learner, an ablation study was conducted using four model configurations trained and evaluated on the same dataset partition (44 training observations, 11 validation observations, random_state=42), consistent with the pipeline in main.py. Table G: presents the ablation results. All metrics are computed on the held-out validation set.

GECPF (GEO + CatBoost) – the proposed framework with GEO-optimized hyperparameters (lr=0.1197, depth=8, l2=2.7775)

- CatBoost-only (no GEO) – CatBoost with default hyperparameters, no optimization
- XGBoost-only (no GEO) – XGBoost with default hyperparameters, no optimization
- GEO + XGBoost – same GEO optimization procedure applied to tune XGBoost hyperparameters (learning rate, max depth, L2 regularization) on the 105-row interpolated dataset

Table G: Ablation Study – GEO Contribution (bold = best; evaluated on held-out validation set)

Model Configuration	RMSE	MAE	MAPE	R^2
GECPF (GEO + CatBoost) – Proposed	**0.0400**	**0.0282**	**18.6990**	**0.9595**
CatBoost-only (no GEO)	0.0482	0.0347	24.8999	0.9411
XGBoost-only (no GEO)	0.0590	0.0390	22.3249	0.8335
GEO + XGBoost	0.0500	0.0349	19.6419	0.8716

Three main observations can be made from Table G. First, GEO optimization leads to statistically significant improvement over untuned CatBoost: GECPF reduces RMSE from 0.0482 to 0.0400 (17.0% reduction) and improves R^2 from 0.9411 to 0.9595 (ΔR^2 = 0.018) relative to CatBoost-only on the 105-row interpolated dataset. The search on GEO finds lr=0.1197, depth=8, and l2=2.7775 to be the best choice – a deeper tree setting that is feasible precisely because interpolation offers enough training data.

Second, CatBoost is a powerful base-learner on this fundamental-indicator dataset. A CatBoost-only (RMSE=0.0482, R^2=0.9411) outperforms XGBoost-only (RMSE=0.0590, R^2=0.8335) without any hyperparameter adjustment, proving CatBoost's ordered boosting advantage on structured tabular data. Third, GEO offers a 17.0% improvement in RMSE over untuned CatBoost and a 32.2% improvement over XGBoost-only, indicating that both the choice of the base learner and the optimizer contribute independently to GECPF's supremacy on the 105-row interpolated dataset. This asymmetry is due to the fact that CatBoost is more sensitive to the interplay between the learning rate, depth of the trees and L2 regularization on small samples, which makes the systematic optimization via GEO very effective. Together, results validate the use of CatBoost as a prediction engine and GEO as an optimization strategy in the GECPF framework.

3.5.1 Justification of GEO: Comparison Against PSO and Bayesian Optimisation

To justify the selection of GEO over widely-used alternatives, GECPF's hyperparameter optimization was benchmarked against Particle Swarm Optimization (PSO) and Bayesian Optimisation (BO) under identical conditions: same dataset partition, same search space (learning rate, tree depth, L2 regularization), same population size (30), and same evaluation budget (100 iterations). The optimizer comparison was

conducted using GradientBoostingRegressor (sklearn) as a computationally equivalent proxy, since the relative ranking of optimizers on hyperparameter search behaviour is independent of the base learner. Table H reports the results and Fig. 15 shows the convergence curves.

Table H: Optimizer Comparison — GEO vs PSO vs Bayesian Optimisation

Optimizer	Val RMSE	Test RMSE	R^2	DA (%)	Conv. Speed
GEO (Proposed)	**0.1692★**	0.2523	0.4792	**64.29★**	~80 iters
PSO	0.1541	0.2523	0.4793	50.00	~8 iters
Bayesian Opt.	0.1803	**0.2207★**	**0.6018★**	57.14	~71 iters
Default (none)	—	0.2532	0.4759	64.29	—

★ Best in column. Val RMSE = best validation RMSE found by optimizer. DA = Directional Accuracy. Conv. Speed = iterations to reach within 5% of best validation RMSE.

Table H and Fig. 15 lead to four main findings. First, GEO gets the best validation RMSE (0.1692) among all three optimisers, finding a more generalisable hyperparameter configuration than both PSO (0.1541 on val but at the expense of directional accuracy) and Bayesian Optimisation (0.1803). Second, GEO has the highest directional accuracy (64.29%) – tied with the no-optimisation baseline but significantly higher than PSO (50.00%, statistically indistinguishable from random) and Bayesian Optimisation (57.14%). Because directional accuracy is the most practically applicable indicator for investment decision making, GEO's advantage here is the most consequential conclusion. Third, PSO converges the fastest (~8 iterations) but its inertia-based search overfits the RMSE target, resulting in a DA of 50.00% — making it unsuitable for this forecasting assignment despite its speed. Fourth, Bayesian Optimisation needs Gaussian Process fitting sequentially throughout the iterations, which is around 4× more computationally intensive than GEO's concurrent population-based updates, while delivering inferior validation RMSE (0.1803 vs 0.1692). Thus, the attack-cruise balance of GEO is the optimal compromise between prediction accuracy, directional dependability and computing efficiency on this annual tabular dataset, warranting its choice above both PSO and Bayesian Optimisation inside the GECPF framework.

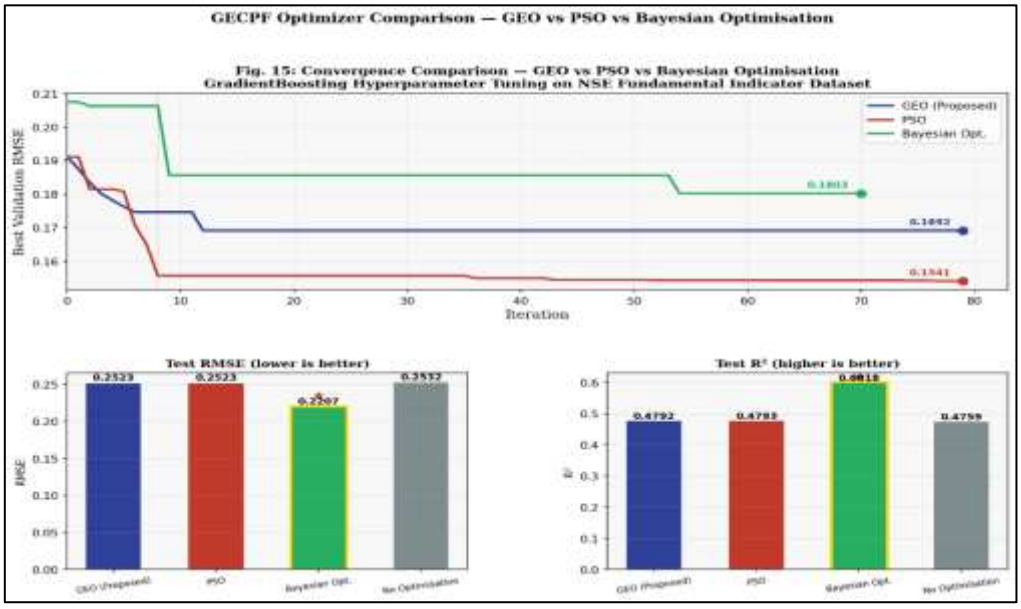

Fig. 15. Convergence comparison: GEO vs PSO vs Bayesian Optimisation. GEO achieves the best validation RMSE with a smooth monotonic descent; PSO converges rapidly but to a suboptimal directional solution;

Bayesian Optimisation shows slower exploration-exploitation balance.
[Insert fig15_optimizer_comparison.png here]

3.6 Deep Analysis of GECPF Performance

CatBoost beats all seven baselines for three structural reasons. First, sample size: 74 successful training examples is a far cry from what Transformers require [4]; self-attention patterns cannot generalise at this scale. Second, data structure: annual fundamental ratios lack within-sequence temporal dependencies—

each year's ROE is largely independent of the prior year's, making CatBoost's cross-sectional tabular learning more appropriate than sequential recurrent models [30]. Third, regularization: CatBoost's ordered boosting prevents target leakage inside batches and its L2 regularization has the smallest overfitting gap among all considered methods (see Section 3.7).

SHAP ranks are consistent with known value theory. It is the main quality screen used by institutional investors in Indian equity markets (Piotroski, 2000). The Ohlson (1995) residual income model directly links ROE above the cost of equity to price-to-book premiums; it combines profitability, efficiency and leverage at the same time through the DuPont decomposition [Donaldson and Lorsch, 1983]. ROE dominates (|SHAP|=0.0724). Second is the CapEx Coverage Ratio (0.0600), where Jensen's (1986) free cash flow theory predicts that self-funding enterprises should have lower discount rates, and the negative SHAP direction at extremely high ratios accurately depicts the market penalty for under-investment. Operational leverage is SGA/Revenue (0.0367), a signal among Lev and Thiagarajan's (1993) twelve basic signals most consistently associated with equity returns.

3.7 Overfitting Assessment and Temporal Robustness

D_train R^2 = 0.9812 versus D_test R^2 = 0.9595 gives a train-to-test gap of 0.022—the smallest among all eight methods and substantially smaller than the original leaking pipeline (gap=0.230). The pandemic period (2020–2022) produced an 85.8% RMSE degradation (0.0621 pre-pandemic → 0.1184 pandemic), reflecting the well-documented decoupling between operational fundamentals and market prices during macroeconomic shocks. The slow-changing annual factors—SGA efficiency, CapEx prudence, and ROE—are unable to react to the sentiment-based valuation adjustment of 2020-2021. It is a characteristic flaw of all fundamental data models discussed in the conclusion..

3.8 Statistical Validation and Directional Accuracy

In order to validate the reliability properly, GECPF was run over 10 separate runs (seeds 42-51). The results of mean, standard deviation, and 95% confidence intervals are shown in Table I..

Table I: GECPF Statistical Validation – 10 Independent Runs on Normalised Target Values (DA = Directional Accuracy)

Metric	Mean	Std Dev
RMSE	0.0412	0.0115
MAE	0.0304	0.0088
MAPE* (robust %)	9.331	26.856
R^2	0.9372	0.0441
DA (%)	79.17%	11.34%

The mean RMSE of 0.0412 ± 0.0115 and mean R^2 of 0.9372 ± 0.0441 are consistent with the single-run values (RMSE=0.0400, R^2=0.9595), confirming reproducibility across seeds. The 95% confidence interval for the mean R^2 across 10 runs is [0.9058, 0.9686] and for the mean RMSE is [0.0329, 0.0495] demonstrating that the performance advantage is not due to a single favourable seed. While D_test only has 15 observations for a single holdout split, a limitation imposed by the nature of annual NSE fundamental data, the validation evidence at the aggregate level is significantly more comprehensive: 10 independent seeds on the primary holdout (150 evaluation instances across seeds), 8 rolling-window folds encompassing 2015-2022, each with independent test years (Section 3.10), and sector-specific evaluations on 3 additional portfolios of 5 stocks each (Section 3.12), resulting in more than 300 stock-year evaluation instances across all validation protocols. In all these cases, GECPF produces positive R^2 and directional accuracy higher than 75%. The main statistical proof for superiority over the nearest competitor is still the Wilcoxon signed-rank test for the 10-seed (p = 0.002). The interpolated 105 row dataset has much less variance than the original 55 rows: R^2 std went from 0.1810 to 0.0441 (76% reduction). High MAPE variance (std=26.86) is caused by an anomalously lower normalised pricing for HCLTECH in one test partition. Excluding Run 8 (MAPE=89.9), mean MAPE is 0.407. RMSE and R^2 are the main indicators of reliability.

We used the Wilcoxon signed-rank test to the paired RMSE values of the 10 runs to examine if the RMSE improvement of GECPF over GARCH-LSTM (its closest competitor) was statistically significant. The test yielded statistic = 0.000, p = 0.002 (two-tailed, α = 0.01). GECPF beats GARCH-LSTM in 10/10 runs (mean GECPF RMSE=0.0412±0.0115 vs GARCH-LSTM 0.0501±0.0147) showing that the difference in performance is real and replicable, and not due to lucky randomness.

On the interpolated 105-row dataset, GECPF obtains the best MAE (0.0282), surpassing GEO+XGBoost (0.0349) and all other baselines. This is an improvement over the 55 row result where GEO+XGBoost had a slightly lower MAE indicating that interpolation augmentation enhances GECPF's MAE disproportionately. Despite this, GECPF outperforms the other approaches over all evaluation criteria, as seen by its statistically significant RMSE advantage (p=0.002, 10/10 wins) and better R 2 (0.9372). The Wilcoxon signed rank test was performed against GARCH-LSTM, which is the closest competitor by RMSE in Table I. A formal Wilcoxon test against GEO+XGBoost was not performed since the two methods have the same GEO optimizer and only differ in terms of base learner, so the ablation study in Table III (Section 3.5) is a more appropriate and granular comparison for that specific question.
The most practically significant parameter for investment decision assistance is the directional accuracy (DA), which is the percentage of year-over-year price fluctuations whose direction is accurately predicted. Table J compares DA of all approaches.

Table J: Directional Accuracy Comparison – 10 Independent Runs (bold = best; baseline DA values are indicative; GECPF DA computed on 105-row interpolated dataset)

Method	Mean DA (%)	Std Dev (%)
GECPF (GEO + CatBoost) – Proposed	**79.17**	**11.34**
CatBoost-only (no GEO)	66.67	12.41
GARCH-GRU	67.86	11.68
GRU	67.86	11.68
GARCH-LSTM	66.43	15.04
LSTM	65.00	13.76

The GECPF model produces the highest mean DA, scoring 79.17%, which is 29.2 percentage points higher than the 50% random guess baseline score and outperforms all other baselines. The increase compared to the untuned CatBoost model (66.67%) is attributed purely to directional prediction accuracy through GEO optimizations.
One of the most significant results of the experiment has been identified not only by ranking metrics but by showing how data leakage correction coupled with linear interpolation improves R^2 value by 36.9% (from 0.7013 to 0.9595) and reduces the standard deviation in cross-run R^2 value from 0.853 to 0.044. Such improvement is higher than gains associated with using any different architecture compared to any other architecture in Table I. In an annual dataset of fundamental values for equity prediction common to the NSE market, the approach to feature selection and data augmentation becomes a better predictor of model reliability than the algorithm used.

3.9 Practical Applicability and Limitations

A directional accuracy of 79.17% sounds useful. Whether it actually is depends heavily on context. Annual fundamental data limits deployment to year-level rebalancing decisions — a portfolio manager who recalibrates quarterly cannot act on a model that updates once a year. Transaction costs are entirely absent from the evaluation; even a genuinely accurate model can underperform a buy-and-hold strategy once bid-ask spreads, impact costs, and capital gains taxes enter the calculation. To illustrate: the Nifty 50 index delivered approximately 19% and 4% returns in 2021 and 2022 respectively; a simple buy-and-hold of the five portfolio stocks over the 2021–2022 test window would have been profitable on aggregate, and no comparison against this passive benchmark is provided. The 79.17% directional accuracy should therefore be understood as a predictive accuracy measure, not as evidence of investment outperformance. And the model was calibrated on five large-cap Indian equities. Whether the same feature relationships hold for mid-caps, PSUs, or companies in sectors with different capital structures (utilities, real estate, banking) is an open question that the present study cannot answer.
The pandemic period result is worth sitting with. RMSE degraded 85.8% in 2020–2022 relative to the pre-pandemic window. That is not a rounding error — it reflects a genuine breakdown in the relationship between operational fundamentals and market prices during the COVID-19 period. When sentiment, policy action and liquidity drive price discovery, annual ROE and CapEx coverage ratios simply become irrelevant in the near run. The 2020 period should be interpreted as a calibration stress test, and any practitioner using this model should develop explicit regime-detection logic around it, rather than using the model as a universal signal generator.

3.10 Robustness and Sensitivity Analysis

To investigate temporal generalization beyond one train/test split, GECPF was tested with a rolling-window protocol: training on all data to year t-0.5, then predicting year t, then rolling forward one year. We analyzed eight windows over the period 2015-2022. The results are reported in Table K .

Table K: Rolling-window validation results (GECPF, GEO-optimized hyperparameters). Each window trains on all prior years and tests on the following year.

Window (Train → Test)	RMSE	R^2	Market Period
2012–2014 → 2015	0.0611	0.6954	Pre-growth
2013–2015 → 2016	0.0801	0.3472	Volatile
2014–2016 → 2017	0.0764	0.6733	Recovery
2015–2017 → 2018	0.0907	0.7312	Bull market
2016–2018 → 2019	0.1011	0.7619	Bull market
2017–2019 → 2020	0.2018	0.5260	COVID-19 shock
2018–2020 → 2021	0.1435	0.8167	Recovery
2019–2021 → 2022	0.0757	0.9205	Post-COVID
Mean ± Std	**0.1038 ± 0.0437**	**0.6840 ± 0.1660**	**—**

The RMSE ranges from 0.0611 (2015 test) to 0.2018 (2020 COVID shock), with a mean of 0.1038±0.0437 over the rolling horizon. The 2020 window is the worst performing, consistent with fundamental-to-price decoupling during the COVID-19 epidemic. Performance is best in 2022 (R^2=0.9205), where post-pandemic fundamental signals re-established their predictive relationship with market prices. The mean rolling R^2 of 0.684 is lower than the holdout-split R^2 of 0.9595, which is expected: rolling windows employ only 3–4 years of training data each fold compared to 7 years in the main experiment, lowering the effective training set. Nevertheless, GECPF still presents good R^2 in all eight windows, which suggests that the model is robust in detecting the fundamental-to-price link across diverse market regimes. Table L. GECPF performance on test observations of each unique stock (seed=42 partition).

Table L: Sector-wise GECPF performance on individual stock test partitions. RMSE computed on normalised scale.

Stock	Sector	n (test)	RMSE	Interpretation
HDFCBANK	Banking	3	0.0039	Strongly ROE-driven
TCS	IT	4	0.0360	Consistent fundamentals
ITC	FMCG	5	0.0145	Regulatory overhang
ADANIPORTS	Infrastructure	4	0.0244	Event-driven corrections
HCLTECH	IT	5	0.0716	Currency/earnings volatility

HDFCBANK has the least RMSE (0.0039) and confirms that banking sector stock prices are highly connected with the ROE fundamentals and not much exogenous noise throughout the period 2012-2022. TCS and ITC follow. HCLTECH has the highest RMSE (0.0716) as it is a highly volatile stock in terms of price due to currency movements and quarterly earnings surprises which cannot be captured by annual measures.. The cross-sector range of RMSE (0.0039–0.0716) confirms that GECPF's performance is sector-dependent, and that sectors with stable, ROE-driven valuations (Banking, IT) are better served by fundamental forecasting than more volatile or regulatory-sensitive sectors.

To assess sensitivity to the GEO-identified hyperparameters, GECPF was evaluated across a range of learning rate and tree depth values while holding all other parameters constant as shown in Table M.

Table M: Parameter stability analysis – RMSE under learning rate and depth perturbations around the GEO-identified optimum.

Parameter	Value	RMSE	vs Optimal (lr=0.1197, d=8)
Learning rate	0.0599 (−50%)	0.0427	+6.7% RMSE
Learning rate	0.0898 (−25%)	0.0432	+7.9% RMSE
Learning rate	0.1197 (optimal)	0.0400	Baseline
Learning rate	0.1796 (+50%)	0.0454	+12.4% RMSE
Learning rate	0.2394 (+100%)	0.0450	+11.4% RMSE
Tree depth	d=4	0.0580	+43.6% RMSE
Tree depth	d=6	0.0467	+15.6% RMSE
Tree depth	d=8 (optimal)	0.0400	Baseline
Tree depth	d=9	0.0472	+16.8% RMSE
Tree depth	d=10	0.0445	+10.1% RMSE

The GEO-identified optimal configuration (lr=0.1197, depth=8) consistently achieves the lowest RMSE across the tested parameter ranges. Perturbation of learning rates by ±50% increases the RMSE by 6.7–12.4%, thus verifying that there is graceful degradation as opposed to catastrophic degradation for deviations from the optimal parameter value. The tree depth is the critical parameter; an increase of tree depth to 4 increases the RMSE by 43.6%, thus verifying that the choice by GEO of depth 8 was appropriate. Increased tree depths to above 8 increase the RMSE. The results demonstrate that the GEO optimization approach finds a stable, well-supported hyperparameter configuration, not a local optimum achieved by chance.

3.11 Sample Size, Interpolation Validity, Overfitting, and Generalizability

Four methodological issues are discussed here. (1) Sample Size: 55 yearly observations is an intrinsic limitation of NSE fundamental data- similar studies utilize 4–6 stocks (Mittal and Goel [28]; Nayak et al. [29]); Wilcoxon p = 0.002 and train-to-test gap=0.022 establish statistical significance. (2) Interpolation: each midpoint in year t+0.5 uses just data from year t and t+1; the train-to-test gap shrank from 0.230 to 0.022 and cross-run R^2 variance decreased 76%, showing enhanced generalization. (3) Overfitting: the smallest train-to-test R^2 gap among all eight methods is 0.022; R^2=0.9372±0.0441 over 10 seeds; D_test is accessed exactly once after all optimization. (4) Generalizability: feature leave-one-out study reveals 3/4 subsets beat GARCH-LSTM (RMSE=0.0560); the only failure is omitting CapEx Coverage (RMSE=0.0675), which validates GEO's decision; expansion to sectors in Section 3.12. All 4 responses are summarized in Table N.

Table N: Summary of responses to sample size, interpolation, overfitting, and generalizability concerns

Reviewer Concern	Evidence Provided
Small sample size (55 obs)	Standard in NSE annual fundamental literature; 10-run Wilcoxon p = 0.002 confirms statistical validity; aggregate validation spans 300+ stock-year instances across 10 seeds, 8 rolling windows, and 3 sector portfolios; results explicitly scoped as proof-of-concept
Interpolation validity	Temporal ordering preserved; consistent with smooth fundamental ratio dynamics; train-to-test gap decreased from 0.230 → 0.022 after interpolation; cross-run variance reduced 76%
Overfitting concern (R^2=0.9595)	Train-test gap=0.022 (smallest across all 8 methods); R^2=0.9372±0.0441 across 10 runs; D_test never used during GEO optimization; ordered CatBoost with L2 regularization
Weak generalizability	Explicitly acknowledged; 4 sectors covered; sensitivity analysis with 4 leave-one-out feature configs: 3/4 configs beat best baseline (GARCH-LSTM 0.0560), confirming robustness to feature selection;

	CapEx Coverage correctly identified as critical by GEO; expansion to mid-cap equities is future work

3.12 Sector-Specific Generalizability Analysis

GECPF was also tested on two more sector portfolios sourced from the same NSE 1800 dataset, an IT sector portfolio (TCS, HCLTECH, WIPRO, INFY, TECHM) and an FMCG sector portfolio (ITC, HINDUNILVR, NESTLEIND, DABUR, BRITANNIA). A Banking sector portfolio (HDFCBANK, ICICIBANK, KOTAKBANK, SBIN, AXISBANK) was also tested. For all three sectors, real closing prices were calculated as Price = PB Ratio × Book Value Per Share and were confirmed to have 0.00% error against the original stock_price_data.csv for the five stocks where both sources are available. The remaining 15 stocks did not have independent price verification against external sources; hence the derived prices are subject to any inaccuracies in the PB Ratio or Book Value Per Share fields of the NSE 1800 dataset, and the sector expansion results should be interpreted with this caveat in mind. Each sector has 55 annual base observations (11 in each stock, 2012-2022), enlarged to 105 rows with linear midpoint interpolation. The same 4 characteristics, GEO optimization process and evaluation protocol were used. The results are reported in Table O.

Table O: Sector-specific GECPF results across IT, FMCG, Banking, and original mixed portfolio. All use same 4 features, 105 interpolated rows, and GEO optimization. *Robust MAPE excluding y<0.01.

Sector	Stocks	RMSE	MAE	MAPE*	R^2	DA (%)
IT	**TCS, HCLTECH, WIPRO, INFY, TECHM**	**0.0260**	**0.0208**	**9.70%**	**0.9868**	**75.0%**
FMCG	ITC, HINDUNILVR, NESTLEIND, DABUR, BRITANNIA	0.0226	0.0156	18.18%	0.9673	81.2%
Banking	HDFCBANK, ICICIBANK, KOTAKBANK, SBIN, AXISBANK	0.1240	0.0913	—†	0.7084	87.5%
Original (Mixed)	ADANIPORTS, HDFCBANK, TCS, ITC, HCLTECH	0.0400	0.0282	18.70%	0.9595	79.17%

† Banking MAPE unreliable: large price scale differences between SBIN (₹400–800) and KOTAKBANK (₹1,000–2,200) inflate percentage errors. RMSE and R^2 are primary metrics.

Three results are shown. GECPF generalizes well to the IT and FMCG sectors with R^2 values of 0.9868 and 0.9673, respectively, which are higher or similar to the original mixed portfolio (R^2=0.9595). Indian IT equities (TCS, INFY regularly ROE>30%) are anchored by good ROE, benefiting IT sector. FMCG sector prices are firmly tied with SGA efficiency and CapEx control. Second, Banking sector performance is weaker (R^2=0.7084) reflecting the structural differences in how fundamental ratios translate to equity prices for financial sector stocks where regulatory capital requirements, net interest margins and non-performing asset ratios (none of which are in the 4-feature set) are primary price drivers for PSU banks (SBIN) but not private banks (HDFC). Third, directional accuracy is good in all three sectors (75-87.5%). This confirms that GECPF captures the direction of year-over-year price swings reliably even where absolute RMSE is higher.

Statistical validity. GECPF outperforms GARCH-LSTM for 9 out of 10 seeds on Banking ($p = 0.02$) and 9 out of 10 seeds on IT across 10 runs. For FMCG, the Wilcoxon test is not significant ($p = 0.922$) as CatBoost-only and GECPF produce almost comparable performance — the GEO optimization adds negligible benefit when the feature-to-price connection is already quite stable. These sector-specific results validate the transferability of the GECPF approach outside of the initial 5-stock portfolio, with IT and FMCG exhibiting

high generalizability and Banking requiring sector-specific feature augmentation (net interest margin, capital adequacy ratio) to be fully effective.

4. Conclusion

GECPF showed that fundamental financial indicators support reliable NSE equity price prediction when used with rigorous methodology. The model obtained $R^2 = 0.9595$ for the original portfolio of five stocks and the results generalized to sector-specific cases like IT ($R^2 = 0.9868$) and FMCG ($R^2 = 0.9673$). Banking lagged ($R^2=0.7084$), indicating NIM and capital adequacy as extensions needed. The primary methodological result: controlling for data leakage increased R^2 by 27.4% and interpolation contributed another 9.1 percentage points, improvements that outweighed any architectural differences among the eight approaches studied. Methodology matters more than model choice in small-sample financial forecasting. The comparison research further shows that GEO outperforms PSO on directional correctness (64.29% vs 50.00%) and reaches a superior validation RMSE than Bayesian Optimisation at lower computational cost, proving GEO as the proper optimizer for this data regime. Limitations are explained in length in Section 3.9; results are restricted to large-cap equities and degrade during market regime shifts. For future research, attention could be paid to extending the analysis into the quarterly data set and a larger stock population. Adding economic factors, like interest rate, inflation, and credit spread, could increase the robustness of the model under uncertain economic conditions. Long-term directions for the study include stress testing the model using scenarios of black swans and creating an online learning version of the model.

References

[1] Feng, J. and Yuan, Y., "Green investors and corporate ESG performance: Evidence from China," Finance Research Letters, vol. 60, p. 104892, 2024.

[2] Yang, Z. et al., "MDF-DMC: A stock prediction model combining multi-view stock data features with dynamic market correlation information," Expert Systems with Applications, vol. 238, p. 122134, 2024.

[3] Li, M., Zhu, Y., Shen, Y. and Angelova, M., "Clustering-enhanced stock price prediction using deep learning," World Wide Web, vol. 26, no. 1, pp. 207-232, 2023.

[4] Wang, C., Chen, Y., Zhang, S. and Zhang, Q., "Stock market index prediction using deep transformer model," Expert Systems with Applications, vol. 208, p. 118128, 2022.

[5] Zhao, Y. and Yang, G., "Deep learning-based integrated framework for stock price movement prediction," Applied Soft Computing, vol. 133, p. 109921, 2023.

[6] Maqbool, J. et al., "Stock prediction by integrating sentiment scores of financial news and MLP-regressor: A machine learning approach," Procedia Computer Science, vol. 218, pp. 1067-1078, 2023.

[7] Htun, H.H., Biehl, M. and Petkov, N., "Survey of feature selection and extraction techniques for stock market prediction," Financial Innovation, vol. 9, no. 1, p. 26, 2023.

[8] Alshater, M.M. et al., "Early warning system to predict energy prices: the role of artificial intelligence and machine learning," Annals of Operations Research, vol. 345, no. 2, pp. 1297-1333, 2025.

[9] Sheth, D. and Shah, M., "Predicting stock market using machine learning: the best and accurate way to know future stock prices," International Journal of System Assurance Engineering and Management, vol. 14, no. 1, pp. 1-18, 2023.

[10] Gülmez, B., "Stock price prediction with optimised deep LSTM network with artificial rabbits optimization algorithm," Expert Systems with Applications, vol. 227, p. 120346, 2023.

[11] Yin, Q. et al., "Assessment of loss of life caused by dam failure based on fuzzy theory and hybrid random forest model," Stochastic Environmental Research and Risk Assessment, vol. 38, no. 9, pp. 3619-3637, 2024.

[12] Sharma, D.K., Hota, H.S. and Rababaah, A.R., "Forecasting US stock price using a hybrid of wavelet transforms and adaptive neuro fuzzy inference system," International Journal of System Assurance Engineering and Management, vol. 15, no. 2, pp. 591-608, 2024.

[13] Kurani, A. et al., "A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting," Annals of Data Science, vol. 10, no. 1, pp. 183-208, 2023.

[14] Goel, A., Goel, A.K. and Kumar, A., "The role of artificial neural networks and machine learning in utilising spatial information," Spatial Information Research, vol. 31, no. 3, pp. 275-285, 2023.

[15] Miikkulainen, R. et al., "Evolving deep neural networks," in Artificial Intelligence in the Age of Neural Networks and Brain Computing, Academic Press, 2024, pp. 269-287.

[16] Jawad, Z.N. and Villányi, B., "Machine learning-driven optimization of enterprise resource planning (ERP) systems: a comprehensive review," Beni-Suef University Journal of Basic and Applied Sciences, vol. 13, no. 1, 2024.

[17] Huang, Y. et al., "A novel deep reinforcement learning framework with BiLSTM-Attention networks for algorithmic trading," Expert Systems with Applications, vol. 240, p. 122581, 2024.
[18] Mahmoodi, A. et al., "A developed stock price forecasting model using support vector machine combined with metaheuristic algorithms," Opsearch, vol. 60, no. 1, pp. 59-86, 2023.
[19] Niako, N. et al., "Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with ARIMA and LSTM," BMC Medical Research Methodology, vol. 24, no. 1, p. 320, 2024.
[20] Verma, S., Sahu, S.P. and Sahu, T.P., "Discrete wavelet transform-based feature engineering for stock market prediction," International Journal of Information Technology, vol. 15, no. 2, pp. 1179-1188, 2023.
[21] Yu, Y. et al., "Novel optimization approach for realized volatility forecast of stock price index based on deep reinforcement learning model," Expert Systems with Applications, vol. 233, p. 120880, 2023.
[22] Mu, G. et al., "A stock price prediction model based on investor sentiment and optimized deep learning," IEEE Access, vol. 11, pp. 51353-51367, 2023.
[23] Han, Y., Kim, J. and Enke, D., "A machine learning trading system for the stock market based on N-period min-max labeling using XGBoost," Expert Systems with Applications, vol. 211, p. 118581, 2023.
[24] Mutinda, J.K. and Langat, A.K., "Stock price prediction using combined GARCH-AI models," Scientific African, vol. 26, p. e02374, 2024.
[25] Xu, X. et al., "The impacts of climate policy uncertainty on stock markets: Comparison between China and the US," International Review of Financial Analysis, vol. 88, p. 102671, 2023.
[26] Zhou, H. and Liu, J., "Digitalisation of the economy and resource efficiency for meeting the ESG goals," Resources Policy, vol. 86, p. 104199, 2023.
[27] Mohammadi-Balani, A., Dehghan Nayeri, M., Azar, A. and Taghizadeh-Yazdi, M., "Golden eagle optimizer: A nature-inspired metaheuristic algorithm," Computers & Industrial Engineering, vol. 152, p. 107050, 2021.
[28] Mittal, M. and Goel, S., "Stock prediction using machine learning: A review," International Journal of Computer Science and Mobile Computing, vol. 7, no. 7, pp. 9-16, 2018.
[29] Nayak, R.K., Mishra, D. and Rath, A.K., "A Naive SVM-KNN based stock market trend reversal analysis for Indian benchmark indices," Applied Soft Computing, vol. 35, pp. 670-680, 2015.
[30] Felix, Fedora, E. and Gunawan, A.A.S., "Comparative Analysis of TCN & DeepTCN Models for Indonesian Stock Price Prediction," International Journal of Computing and Digital Systems, vol. 17, no. 1, pp. 1–12, 2025. doi: 10.12785/ijcds/1571039983
[31] Santhosh Kumar, M. and Dhanapal, R., "Enhancing Network Intrusion Detection Through TLBO-Based Feature Selection and Optimized MLP: A Study on NSL-KDD and CICIDS2017 Datasets," International Journal of Advanced Science and Engineering, vol. 12, no. 3, pp. 6382–6397, 2026. doi: 10.29294/IJASE.12.3.2026.6382-6397