JURNAL INFOTEL
VOL. 18, NO. 3, AUGUST 2026, PP. 455–471. DOI:10.20895/INFOTEL.V18I3.10792

RESEARCH ARTICLE

An Explainable Hybrid LSTM–XGBoost Framework for Portfolio Optimization with Walk-Forward Backtesting in the Indonesian Equity Market

Mohammad Hikam 'Abdul Karim [1] and Diandra Chika Fransisca[2,*],

[1,2] Informatics engineering study program, Telkom University, Purwokerto Campus, JL. DI Panjaitan No. 128, Purwokerto 53147, Central Java, Indonesia

Received: March 10, 2026; Revised: June 02, 2026; Accepted: June 13, 2026.

Abstract: The Indonesian equity market presents substantial challenges for portfolio construction due to its high volatility, sector concentration, and sensitivity to global macroeconomic shocks. Although recent studies employ machine learning for stock return prediction, most focus primarily on forecast accuracy without integrating predictive models into a complete portfolio decision-making framework. This study proposes an explainable hybrid LSTM–XGBoost framework for portfolio optimization with walk-forward backtesting in the Indonesian equity market, evaluated on LQ45 constituents of the IDX over 23 quarterly rebalancing periods from January 2020 to November 2025. The framework combines temporal sequence modeling (LSTM) with nonlinear ensemble learning (XGBoost), dynamic ensemble weighting, mean variance optimization with risk-parity fallback, sector diversification constraints, and stop-loss controls. An explainable feature attribution mechanism that combines LSTM gradient-based saliency and XGBoost gain-based importance enables regime-aware interpretation. The empirical results show a total return of 371.69% versus 35.41% for IHSG, a Sharpe ratio of 1.25 versus 0.03, and a maximum drawdown of -20.29% compared with -37.75%. The annualized alpha of 26.50% and the beta of 0.43 confirm that the performance arises from the ability to select stocks rather than the exposure of the market.

Keywords: hybrid LSTM–XGBoost, portfolio optimization, walk-forward backtesting, explainable AI, Indonesian stock market

© Jurnal Infotel ISSN: 2085-3688; e-ISSN: 2460-0997

1 Introduction

The Indonesian capital market has grown rapidly in recent years, mainly driven by rising retail participation. According to KSEI, by August 2025 the market recorded more than 18 million Single Investor Identification (SID) accounts, with individual investors comprising over 99 percent of total participants [1]. This expansion reflects increased engagement with equity instruments, particularly liquid benchmarks such as LQ45. Despite this growth, the Indonesian stock market remains highly volatile and sensitive to global macroeconomic shifts. Events such as the COVID-19 pandemic and heightened uncertainty during 2023–2024 caused significant disruptions in the Indonesia Composite Index (IHSG), resulting in sharp price swings that are difficult to anticipate [2].

Classical statistical models such as ARIMA are widely used for univariate time-series forecasting of stock prices and returns due to their simplicity and interpretability. However, their linear assumptions limit their ability to capture the nonlinear, volatile, and noisy dynamics of equity markets, particularly during periods of market regime shifts and financial turbulence [3,4]. Machine learning models, including LSTM and XGBoost, address these limitations by exploiting temporal dependencies and nonlinear feature interactions. However, each model has shortcomings when used independently: LSTM requires long training periods and large datasets, while XGBoost lacks sequential memory and cannot capture long-horizon dynamics effectively. Recent studies show that combining LSTM and XGBoost can improve prediction accuracy for financial time series [5]; nevertheless, most hybrid studies focus solely on forecasting and rarely integrate model outputs into a full decision-making pipeline for portfolio allocation [6]. Explainability also remains a key challenge: post-training feature importance methods can yield unstable explanations [7]; however, ensemble-based attribution provides greater reliability [8].

The research gap addressed by this study is threefold: first, the lack of an end-to-end pipeline linking hybrid machine learning forecasts to portfolio optimization in the Indonesian equity market; second, the absence of walk-forward backtesting protocols that prevent look-ahead bias; and third, insufficient model interpretability in existing hybrid frameworks. This study addresses these gaps by proposing an integrated framework that combines hybrid LSTM–XGBoost forecasting with mean-variance portfolio optimization and multi-year walk-forward backtesting on LQ45 stocks.

Accordingly, the main contributions of this study are twofold. First, this research presents a hybrid predictive model combining LSTM and XGBoost specifically designed to capture the nonlinear and highly volatile behavior of Indonesian equity market data. Second, this study integrates hybrid model outputs into a mean-variance portfolio allocation framework, enabling practical and actionable investment strategies. Furthermore, this research presents a structured walk-forward backtest protocol to systematically evaluate the robustness of the strategy across different market regimes, culminating in a reliable feature attribution mechanism that combines gradient-based saliency based on LSTM and gain-based importance based on XGBoost within a weighted ensemble framework to improve model interpretability and transparency in financial forecasting.

2 Research Method

2.1 Research Framework

The core problem addressed in this study is the absence of an end-to-end quantitative pipeline that bridges hybrid machine learning forecasting with portfolio optimization in the Indonesian equity market. Existing studies either focus on return prediction in isolation or apply naive allocation strategies, leaving a gap in practical, deployable frameworks that are simultaneously accurate, interpretable, and risk-aware.

To address this, the study proposes a five-stage pipeline. The first stage involves data collection and feature engineering from LQ45 daily historical data across multiple market regimes. The second stage trains a hybrid LSTM–XGBoost model using a dynamic weighted ensemble, where each sub-model is assigned a weight inversely proportional to its validation Mean Squared Error (MSE) at each rebalancing period. The third stage feeds the predicted returns into a mean-variance optimizer to construct a long-only, sector-constrained portfolio. The fourth stage runs a walk-forward backtest with stop-loss and quarterly rebalancing to simulate realistic investment conditions. The fifth and final stage applies a hybrid explainable AI module to decompose and interpret the model's predictions in terms of market regime and feature contribution. Each stage operates exclusively on historically available data at the time of each rebalancing point, preventing look-ahead bias throughout the entire pipeline.

2.2 Dataset and Feature Engineering

The dataset consists of daily Open, High, Low, Close, Volume (OHLCV) price data for LQ45 Index constituents on the Indonesia Stock Exchange (IDX), retrieved via the yfinance API [9], spanning January 2020 to November 2025. This period was selected to encompass multiple distinct market regimes: the COVID-19 pandemic shock (2020), the commodity supercycle and rate hike cycle (2022–2023), and the subsequent recovery (2024–2025). A static risk-free rate R_f of 4.75% is applied, based on the average BI-7 Day Reverse Repo Rate.

To ensure model generalization and prevent data leakage, the dataset is split into training and test sets prior to any scaling or transformation. This study employs Robust Scaler for feature normalization, which is specifically designed to handle financial data with significant outliers by using the median and the Interquartile Range (IQR) rather than mean and standard deviation [10]:

$$x_{\text{scaled}} = \frac{x - \text{median}(x)}{Q_3(x) - Q_1(x)} \tag{1}$$

where x is the original feature value, $Q_3(x)$ is the third quartile (75th percentile), $Q_1(x)$ is the first quartile (25th percentile), and $Q_3(x) - Q_1(x)$ represents the interquartile range used to reduce the impact of outliers in financial data.

Feature engineering transforms the raw OHLCV data into 18 predictive signals across six categories: momentum features (momentum 5D, momentum 10D, momentum 21D, momentum 63D), volatility features (volatility 21D, risk-adjusted momentum), technical indicators (RSI, RSI normalized, MA20/MA50 ratio, trend strength), price action features (price vs. 63D high, drawdown, daily range, candle body ratio), return statistics (log re-

turn mean, log return standard deviation), and volume features (volume ratio, log volume ratio).

2.3 Hybrid LSTM–XGBoost Prediction Model

The prediction framework of this study employs two complementary machine learning architectures that address different aspects of financial time-series modeling. LSTM is used to capture long-term temporal dependencies in sequential price and feature data [11, 12]. As a variant of Recurrent Neural Networks, LSTM overcomes the vanishing gradient problem through three gating mechanisms: a forget gate that selectively discards irrelevant historical information, an input gate that incorporates new information into the cell state, and an output gate that controls what is passed to the next time step. This architecture is particularly suited to financial data, where patterns spanning weeks or months, such as trend continuation, momentum buildup, and mean reversion, are critical predictive signals.

XGBoost, on the other hand, is applied to model nonlinear interactions among the engineered features at each time step [13, 14]. As a gradient-boosted ensemble of decision trees, XGBoost is highly effective at capturing complex, non-monotonic relationships between technical indicators, price action signals, and momentum features that LSTM's sequential structure cannot directly exploit. The two models thus address fundamentally different predictive challenges: LSTM handles the temporal structure of returns, while XGBoost handles cross-sectional feature complexity.

The non-standard contribution of this study lies in the dynamic ensemble weighting mechanism. Rather than assigning fixed weights, the contribution of each sub-model to the final prediction is adjusted at every quarterly rebalancing period based on its relative out-of-sample validation performance. Specifically, the weight of each sub-model is computed as the inverse of its validation MSE, normalized so that the total weight sums to one. The final integrated prediction of the hybrid model is then formulated as follows [15]:

$$\hat{y} = w_{\text{LSTM}} \cdot \hat{y}_{\text{LSTM}} + w_{\text{XGBoost}} \cdot \hat{y}_{\text{XGBoost}} \tag{2}$$

where $\hat{y}$ is the final ensemble prediction for stock return, $\hat{y}_{\text{LSTM}}$ is the prediction from the LSTM model, $\hat{y}_{\text{XGBoost}}$ is the prediction from the XGBoost model, w_{LSTM} and w_{XGBoost} are the corresponding ensemble weights. The weights satisfy $w_{\text{LSTM}} + w_{\text{XGBoost}} = 1$ and are recomputed at each quarterly rebalancing period based on the inverse MSE ratio from the most recent validation window.

2.4 Explainable AI With Feature Importance

A key limitation of many hybrid machine learning models in finance is their limited interpretability under changing market conditions [7]. To address this issue, this study proposes a hybrid Feature Importance framework that combines gradient-based saliency from the LSTM sub-model with gain-based importance from XGBoost into a unified importance score at each rebalancing period.

For the XGBoost sub-model, feature importance is derived from gain-based importance, which measures the contribution of each feature to reducing prediction error across the ensemble of trees. For the LSTM sub-model, feature importance is approximated using gradient-based saliency, where the magnitude of the prediction-loss gradient with respect to each input feature represents its influence on the model output.

The two importance vectors are combined using the same ensemble weights employed during the prediction stage:

$$H_i = w_{\text{LSTM}} \cdot \text{FI}_i^{\text{LSTM}} + w_{\text{XGBoost}} \cdot \text{FI}_i^{\text{XGBoost}} \tag{3}$$

where H_i denotes the final hybrid importance score for feature i, $\text{FI}_i^{\text{LSTM}}$ and $\text{FI}_i^{\text{XGBoost}}$ represent the normalized LSTM and XGBoost importance scores for feature i, respectively, and w_{LSTM} and w_{XGBoost} denote the ensemble weights (with $w_{\text{LSTM}} + w_{\text{XGBoost}} = 1$). Both importance vectors are normalized prior to combination to ensure scale comparability.

This mechanism enables regime-aware interpretation by identifying how feature relevance shifts across different market environments. This characteristic distinguishes the proposed framework from static post-hoc explanation methods such as SHAP or LIME.

2.5 Portfolio Optimization Strategy

The predicted returns from the hybrid model serve as inputs to the mean-variance optimization model [16, 17]. The objective is to construct a portfolio that minimizes variance given a target return, subject to full-investment and long-only constraints:

$$\text{Minimize } \sigma^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j C_{ij} \tag{4}$$

$$\text{subject to } \sum_{i=1}^{N} w_i E_i = \gamma \tag{5}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{6}$$

$$w_i \geq 0 \tag{7}$$

where σ^2 denotes the portfolio variance, N is the number of assets (stocks), w_i represents the allocation weight of asset i, C_{ij} denotes the covariance between the predicted returns of assets i and j, E_i represents the expected return of asset i (from the hybrid model prediction), and γ denotes the target portfolio return. Equation (6) ensures that the total portfolio allocation equals one (fully invested), while Equation (7) prevents short-selling by restricting portfolio weights to non-negative values (long-only constraint).

Several non-standard control mechanisms are incorporated into the allocation stage. When the number of candidate stocks falling above the 2.0% predicted return threshold is insufficient to solve the optimization, a risk-parity fallback mechanism is applied, allocating weights in inverse proportion to each asset's predicted volatility. A sector diversification constraint caps any single sector at 30% of total portfolio weight at each rebalancing, limiting concentration risk during periods when a single sector dominates model predictions. Stop-loss levels are computed individually per holding at each rebalancing date and triggered dynamically throughout the holding period, providing a continuous risk floor between rebalancing events.

2.6 Backtesting and Evaluation Metrics

The strategy is validated through a walk-forward backtesting protocol [18, 19] designed to simulate realistic portfolio management conditions. At each quarterly rebalancing date, the hybrid model is retrained exclusively on historical data available up to that point, preventing any information from future periods from influencing the model or portfolio construction. This strict sequencing ensures that all reported performance metrics reflect genuinely out-of-sample predictive capability. A 0.2% transaction cost is applied per trade to account for market friction.

The primary performance metric is the Sharpe ratio, which measures the excess return earned per unit of total risk [20]:

$$\text{SR} = \frac{E_p - R_f}{\sigma_p} \tag{8}$$

where E_p denotes the expected portfolio return, R_f represents the risk-free rate, and σ_p denotes the annualized portfolio volatility.

Additional evaluation metrics include: total return (cumulative portfolio value growth), annualized return (geometric mean annualized), annualized volatility (σ_p), maximum drawdown (the largest peak-to-trough decline over the evaluation period), alpha (α, the annualized excess return vs. IHSG after adjusting for market risk via Capital Asset Pricing Model (CAPM)), and beta (β, the sensitivity of portfolio returns to IHSG movements where $\beta = 1.0$ indicates equal market exposure) [20,21]. Prediction accuracy for individual stocks is measured using MSE. These metrics collectively capture return generation, risk management, and market exposure characteristics of the strategy.

3 Results

3.1 Experimental Setup

The backtest is conducted on LQ45 Index constituents sourced from the IDX over the period January 2, 2020, to November 30, 2025, comprising 23 quarterly rebalancing periods. At each rebalancing date, the model is retrained using all available historical data up to that point. A transaction cost of 0.2% is applied per trade, and a static risk-free rate of 4.75% (BI-7 Day Reverse Repo Rate) is used for Sharpe ratio computation. Evaluation metrics include total return, annualized return, volatility, Sharpe ratio, maximum drawdown, alpha, and beta relative to the IHSG benchmark.

3.2 Model Configuration and Hyperparameter Tuning

The LSTM model is configured with two stacked LSTM layers (64 and 32 units) followed by dropout regularization, a dense hidden layer, and a single regression output. The XGBoost regressor is configured with regularized parameters to balance model complexity and generalization. Final hyperparameters for both sub-models were selected based on validation MSE across a held-out validation window, and are detailed in Table 1 and Table 2 below.

Table 1: LSTM model configuration

Parameter	Value	Description
Lookback window	84 days	Number of past time steps used as input
LSTM layers	2 (64 $\rightarrow$ 32 units)	First layer: return_sequences=True
Dropout rate	0.3	Applied after LSTM layers to prevent overfitting
Hidden dense layer	16 units, ReLU	Intermediate fully connected layer
Output layer	Dense(1)	Single-unit regression output
Optimizer	Adam (lr = 0.001)	Adaptive learning rate optimizer
Loss function	Mean Squared Error	Average squared difference between prediction and actual
Max epochs	30	Maximum training epochs
Batch size	32	Samples processed before weight update
Early stopping	Patience = 10	Monitor val_loss; restore best weights

Table 2: XGBoost model configuration

Parameter	Value	Description
Objective	reg:squarederror	Regression task with squared error loss
n_estimators	300	Number of boosting trees
Learning rate	0.02	Step size shrinkage to prevent overfitting
Max depth	4	Maximum depth of each tree
Subsample	0.8	Fraction of training samples used per tree
Colsample bytree	0.8	Fraction of features used per tree
Random state	42	Fixed seed for reproducibility

3.3 Overall Backtest Performance

The hybrid LSTM–XGBoost portfolio optimizer was evaluated through a walk-forward backtest spanning January 2, 2020, to November 30, 2025, encompassing 23 quarterly rebalancing periods across IDX equities. The full backtest results compared to the IHSG benchmark are presented in Table 3.

Table 3: Backtest performance: proposed vs. IHSG (2020–2025)

Metric	Strategy	IHSG Benchmark	Difference
Total Return (%)	371.69	35.41	+336.28
Annualized Return (%)	30.02	5.26	+24.75
Volatility (Annual) (%)	20.19	17.02	+3.17
Sharpe Ratio	1.25	0.03	+1.22
Maximum Drawdown (%)	-20.29	-37.75	+17.46
Alpha (α) vs IHSG (%)	26.50	–	–
Beta (β) vs IHSG	0.43	–	–

The strategy achieved a total return of 371.69% over the evaluation period, substantially outperforming the IHSG benchmark return of 35.41%, yielding a differential of 336.28 percentage points. On an annualized basis, the strategy generated a return of 30.02% compared to 5.26% for the IHSG, representing an annualized outperformance of 24.75 percentage points.

Risk-adjusted performance was notably strong: the Sharpe ratio of 1.25 versus 0.03 for the IHSG indicates a substantial improvement in return per unit of risk. The alpha of 26.50% confirms that the strategy generates returns through stock-selection skill beyond what CAPM predicts from its market exposure. The beta of 0.43 indicates that for every 1% movement in the IHSG, the strategy moves approximately 0.43%, confirming low market sensitivity and strong idiosyncratic return generation. The maximum drawdown of -20.29% was considerably lower than the benchmark's -37.75%, confirming effective downside risk management, particularly during the COVID-19 market crash of March 2020.

3.4 Model Validation and Ensemble Weighting

Throughout the backtest, the hybrid model was retrained at each rebalancing date using only information available up to that point, adhering to a strict walk-forward validation discipline to prevent data leakage. Table 4 reports validation MSE for both LSTM and XGBoost sub-models across selected rebalancing periods, alongside the dynamically computed ensemble weights.

Table 4: Validation MSE and ensemble weights at selected rebalancing periods

Period	LSTM Val MSE	XGBoost Val MSE	LSTM Weight	XGBoost Weight
2020–01–02	0.019364	0.018809	0.49	0.51
2020–03–31	0.021996	0.022004	0.50	0.50
2020–12–21	0.022510	0.022440	0.50	0.50
2022–03–07	0.020373	0.020080	0.50	0.50
2024–05–08	0.016711	0.016654	0.50	0.50
2025–04–25	0.015977	0.015730	0.50	0.50
2025–10–20	0.015552	0.015468	0.50	0.50

The validation MSE values for both models exhibited a downward trend over time, decreasing from approximately 0.019–0.022 in early 2020 to 0.015 in 2025, reflecting progressive improvements in data quality and the growing availability of historical price information. The ensemble weights converged to approximately 0.50/0.50 in most rebalancing windows, confirming balanced and complementary predictive contributions from both sub-models. A slight asymmetry was observed at the initial rebalancing on 2020–01–02, where XGBoost held a marginally higher weight of 0.51 (LSTM: 0.49), reflecting a modest early advantage of the gradient-boosted model.

3.5 Explainable AI and Feature Importance Analysis

At each rebalancing period, a hybrid feature importance score was computed by combining gradient-based feature saliency from the LSTM sub-model with gain-based feature importance from XGBoost, weighted according to their ensemble contributions. Table 5 presents the average feature category importance across all 23 rebalancing periods.

Table 5: Average feature category importance across all rebalancing periods

Feature Category	Avg Importance (%)	Top Indicators	Role in Prediction
Technical Indicators	22.7	MA20/MA50 Ratio, RSI	Identifies trend direction and overbought/oversold conditions
Price Action	21.8	Price vs 63D High, Drawdown	Measures short-term price behavior and reversal pressure
Momentum Features	21.6	Momentum 63D, Momentum 21D	Captures multi-horizon trend strength and continuation
Return Statistics	12.8	Log Return Std, Log Return Mean	Captures expected return tendency and dispersion
Volatility Features	12.3	Risk-Adjusted Momentum, Log Return Std	Represents volatility-normalized performance
Volume Features	8.8	Volume Ratio, volume_ratio_log	Indicates liquidity pressure and abnormal trading activity

Technical indicators ranked highest on average at 22.7%, narrowly ahead of price action (21.8%) and momentum features (21.6%), reflecting the strong relevance of trend identification and multi-horizon price momentum in the Indonesian equity market. The relative importance of feature categories shifted dynamically across market regimes. Table 6 and Table 7 present the breakdown for two contrasting periods.

Table 6: Feature importance for March 2020 rebalancing (high-volatility regime)

Feature Category	Importance (%)	Key Features
Technical Indicators	23.1	MA20/MA50 Ratio, Trend Strength
Momentum Features	21.3	Momentum 63D, Momentum 21D
Price Action	20.3	Drawdown, Daily Range
Return Statistics	16.1	Log Return Mean, Log Return Std
Volatility Features	10.9	Volatility 21D, Risk-Adj Momentum
Volume Features	8.3	volume_ratio_log, Volume Ratio

This regime-adaptive redistribution of predictive power underscores a major strength of the walk-forward hybrid architecture. The model automatically re-evaluates feature relevance at every rebalancing period. The stock-level explainable AI explanations reinforce this: during March 2020, all eight selected stocks were classified under deep-value and mean-reversion regimes (RSI values predominantly below 50, negative risk-adjusted momentum); during December 2020, all seven selected stocks exhibited high-momentum trend-following characteristics (RSI values above 75, strongly positive trend strength), demonstrating that the explainable AI module accurately captures and communicates the prevailing market regime in each rebalancing.

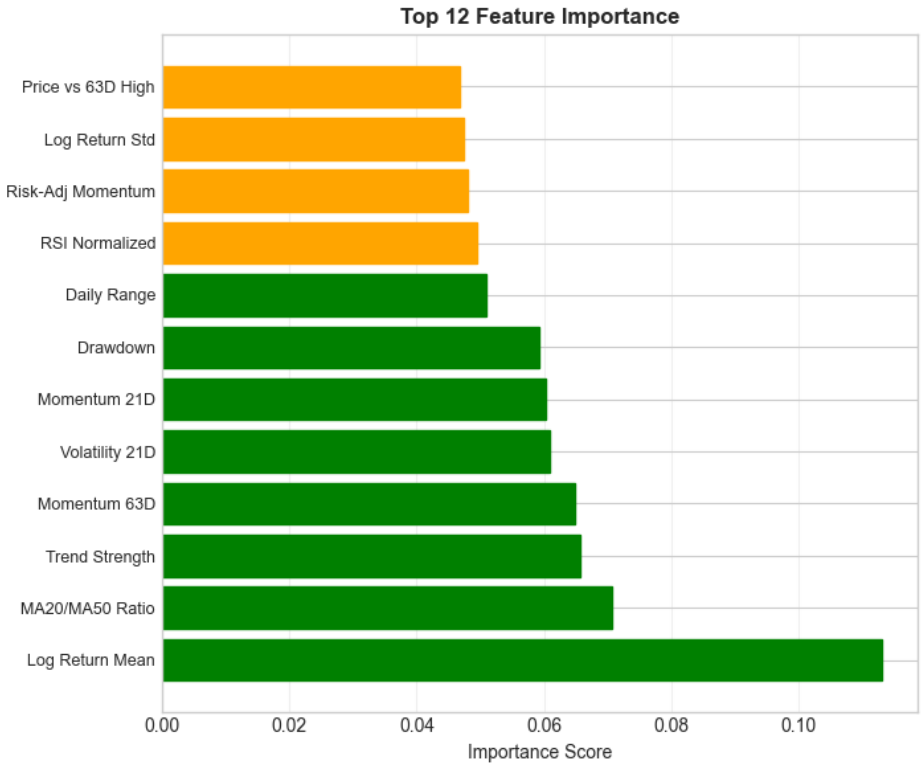

Figure 1: Top 12 feature importance scores for March 2020 (high-volatility regime).

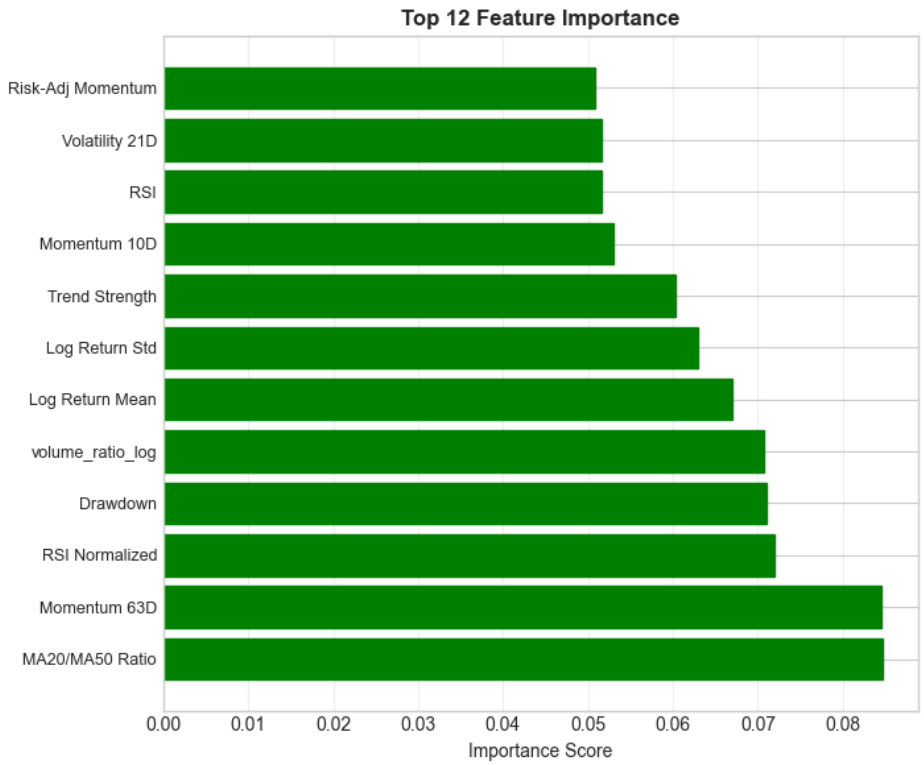

Figure 2: Top 12 feature importance scores for December 2020 (momentum-driven regime).

Table 7: Feature importance for December 2020 rebalancing (momentum-driven regime)

Feature Category	Importance (%)	Key Features
Technical Indicators	26.8	MA20/MA50 Ratio, RSI Normalized
Momentum Features	22.4	Momentum 63D, Momentum 10D
Price Action	17.6	Drawdown, Price vs 63D High
Return Statistics	13.0	Log Return Mean, Log Return Std
Volatility Features	10.2	Volatility 21D, Risk-Adj Momentum
Volume Features	9.9	volume_ratio_log, Volume Ratio

3.6 Portfolio Construction and Rebalancing

At each rebalancing date, the model generated predicted forward returns for all constituent stocks, applied a 2.0% threshold filter, enforced sector diversification constraints (maximum 30% per sector), and optimized portfolio weights using mean-variance optimization with a risk-parity fallback. Table 8 presents selected rebalancing outcomes across the backtest period.

Table 8: Selected rebalancing outcomes

Date	Top Stock	Return (%)	Market Context	n
2020–01–02	TOWR	2.38	Pre-COVID bull market	4
2020–03–31	SMRA	27.48	COVID crash (high volatility bounce)	8
2020–06–26	TOWR	7.40	Post-crash recovery	5
2020–09–23	BRPT	9.11	Second wave uncertainty	7
2020–12–21	ANTM	19.07	Vaccine rally, commodity surge	7
2022–03–07	ESSA	14.69	Energy price spike (Russia-Ukraine)	7
2023–02–22	TKIM	7.99	Post-rate-hike stabilization	8
2024–05–08	JPFA	1.28	Low signal environment	4
2025–04–25	INCO	6.82	Mid-2025 recovery	7
2025–10–20	BRPT	2.41	Late 2025 consolidation	3

The results of the portfolio construction reveal that the model was meaningfully adapted to different market regimes. During the COVID-19 crash (March 2020), the model identified high predicted returns of up to 27.48% for SMRA, driven by deep-value mean-reversion signals. During the commodity rally of Q1 2022, energy stocks like ESSA dominated. Periods of low market signal (May 2024, October 2025) resulted in smaller, more conservative portfolios of three to four stocks, triggering the risk-parity fallback mechanism. In general, the mean–variance optimization produced feasible allocations in 16 of 22 rebalancing periods, while the remaining periods relied on the risk-parity fallback allocation under limited candidate-stock conditions.

3.7 Equity Curve and Drawdown Analysis

The cumulative equity curve and drawdown comparison between the strategy and the IHSG benchmark are presented in Figure 3, Figure 4, and Figure 5. The rolling Sharpe ratio,

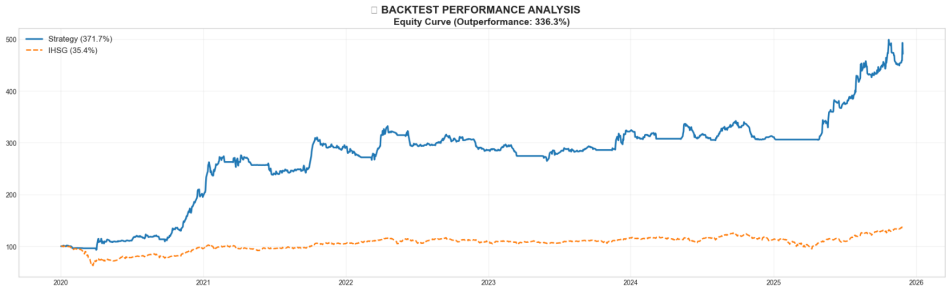

Figure 3: Cumulative equity curves: proposed strategy vs. IHSG benchmark (indexed to 100).

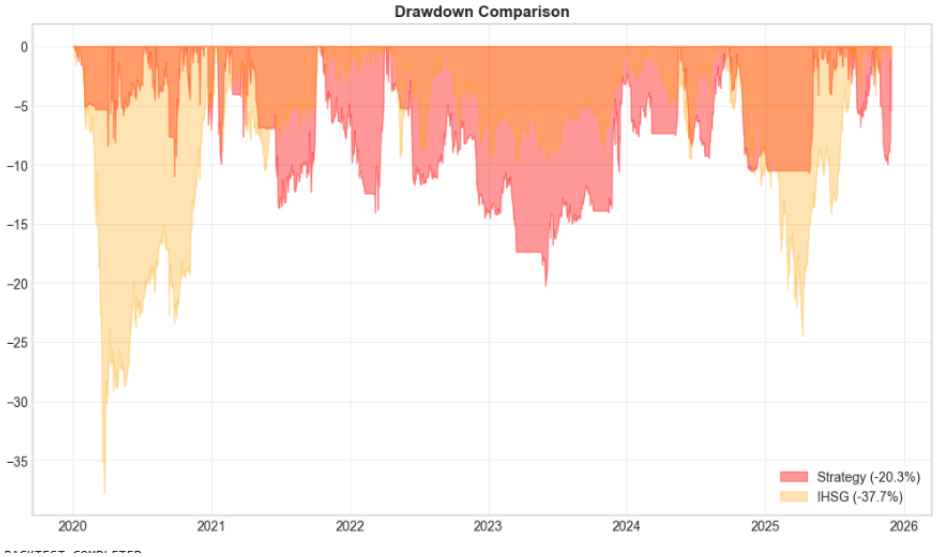

Figure 4: Drawdown comparison: proposed strategy vs. IHSG benchmark.

displayed in the third panel, captures the time-varying nature of risk-adjusted performance across the full evaluation period.

The equity curve demonstrates consistent outperformance across the evaluation horizon, with the largest divergence occurring during the 2020–2021 COVID-19 recovery period and the 2022 commodity rally. The strategy recorded a maximum drawdown of -20.29% during the initial COVID-19 shock, but recovered relatively quickly in subsequent rebalancing periods. In contrast, the IHSG benchmark experienced a deeper drawdown of -37.75% and a slower recovery trajectory. The rolling Sharpe ratio further confirms consistently positive risk-adjusted performance, with notable peaks during post-pandemic recovery and commodity-driven market expansion.

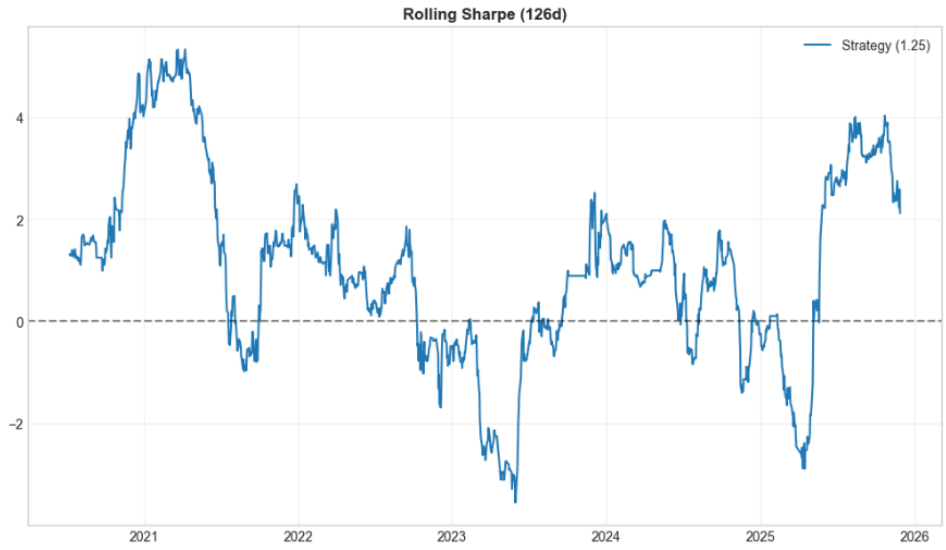

Figure 5: Rolling 126-day Sharpe ratio (average 1.25).

4 Discussion

This study evaluated a hybrid LSTM–XGBoost portfolio optimization framework through a walk-forward backtest on LQ45 constituents over January 2020 to November 2025, comprising 23 quarterly rebalancing periods. The simulation applies a 0.2% transaction cost per trade, a 4.75% risk-free rate, and enforces sector diversification constraints (max 30% per sector) with stop-loss controls. Performance is assessed using total return, annualized return, volatility, Sharpe ratio, maximum drawdown, alpha, and beta relative to the IHSG. This setup ensures all results reflect genuine out-of-sample predictive performance without look-ahead bias.

The strategy achieved a total return of 371.69% versus 35.41% for the IHSG, with a Sharpe ratio of 1.25 compared to 0.03 for the benchmark. These results confirm that combining temporal sequence modeling (LSTM) with nonlinear tree-based learning (XGBoost) captures complementary predictive signals that neither model can achieve alone. The near-equal ensemble weighting throughout the backtest (approximately 0.50/0.50) indicates balanced contributions from both sub-models, consistent with ensemble-learning theory. The low beta of 0.43 and annualized alpha of 26.50% confirm that excess returns arise from stock-selection skill rather than leveraged market exposure.

The explainable AI analysis demonstrates that feature importance shifts dynamically across market regimes. During stable trend-following conditions, technical indicators and momentum features dominate. During high-volatility periods such as the COVID-19 crash of March 2020, return statistics and price action features become more influential, as short-term behavioral price movements carry stronger predictive power when medium-term trends weaken. These shifts are driven by the walk-forward retraining process, which updates feature weights at every period using only historically available data, preventing overfitting to a single market regime. The complete inversion of stock classifications, from

deep-value oversold in March 2020 to overbought momentum-driven in December 2020, demonstrates that the explainable AI layer reflects the actual decision logic of the hybrid model, not merely post-hoc annotation.

From a risk management perspective, the sector diversification constraint proved effective in limiting concentration risk during periods when energy, mining, and industrial sectors dominated model predictions. The mean-variance optimization with risk-parity fallback successfully allocated weights in 16 of 22 rebalancing periods; the remaining 6 defaulted to equal risk contribution. The strategy's maximum drawdown of -20.29% compares favorably with the benchmark's -37.75%, reflecting the combined effect of sector diversification, stop-loss management, and the model's tendency to adopt smaller, more conservative portfolios during low-signal environments. These results contribute to the growing body of evidence that machine learning-based portfolio construction offers measurable advantages over passive benchmark strategies in emerging equity markets.

5 Conclusion

This study proposed an explainable hybrid LSTM–XGBoost framework for portfolio optimization in the Indonesian equity market, integrating walk-forward retraining, dynamic ensemble weighting, feature attribution, sector diversification, and stop-loss risk controls into a unified quantitative pipeline. Empirical evaluation over 23 quarterly rebalancing periods from 2020 to 2025 demonstrated substantial outperformance relative to the IHSG benchmark, achieving a total return of 371.69%, an annualized return of 30.02%, and a Sharpe ratio of 1.25 while maintaining a relatively low maximum drawdown and market beta. The consistently positive alpha further indicates that the framework generates excess returns primarily through predictive stock-selection capability rather than passive market exposure.

The explainability analysis further revealed that feature relevance shifts dynamically across market regimes, with momentum and technical indicators dominating during trending periods, while return statistics and price action features become more influential during high-volatility environments. These findings demonstrate that the proposed framework not only improves predictive and portfolio performance, but also provides interpretable insights into the evolving drivers of market behavior.

Overall, this study highlights the practical viability of explainable hybrid machine learning frameworks for risk-aware portfolio construction in emerging equity markets. Future work may extend the framework through macroeconomic integration, adaptive transaction-cost modeling, event-driven rebalancing, and cross-market validation across ASEAN exchanges.

Acknowledgments

The authors would like to express their sincere gratitude to Telkom University Purwokerto for the institutional support and resources that made this research possible.

References

[1] KSEI, "Statistik pasar modal indonesia – agustus 2025," tech. rep., KSEI, Jakarta, Indonesia, Aug. 2025. Accessed: March 7, 2026.

[2] I. S. Lesmana and S. Saadah, "Pandemic and indonesia stock market performance," *Ilomata International Journal of Management*, vol. 2, pp. 254–262, Oct. 2021.

[3] Z. Shi, Y. Hu, G. Mo, and J. Wu, "Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction," Apr. 2022. Accessed: March 7, 2026.

[4] A. Gifty and Y. Li, "A comparative analysis of LSTM, ARIMA, XGBoost algorithms in predicting stock price direction," *Engineering and Technology Journal*, vol. 9, pp. 4978–4986, Aug. 2024.

[5] H. Oukhouya, H. Kadiri, K. El Himdi, and R. Guerbaz, "Forecasting international stock market trends: XGBoost, LSTM, LSTM-XGBoost, and backtesting XGBoost models," *Statistics, Optimization & Information Computing*, vol. 12, no. 1, pp. 200–209, 2024.

[6] Y. Ma, Y. Wang, W. Wang, and C. Zhang, "Prediction-based mean–variance portfolios with risk budgeting based on neural networks," *Expert Systems with Applications*, vol. 230, p. 120638, Nov. 2023.

[7] D. Rengasamy *et al.*, "Feature importance in machine learning models: A fuzzy information fusion approach," *Neurocomputing*, vol. 511, pp. 163–174, Oct. 2022.

[8] W. Khan *et al.*, "Hybrid XGBoost–LSTM model for state-of-health prediction of lithium-ion batteries under different thermal conditions," *International Journal of Electrochemical Science*, p. 101218, Oct. 2025.

[9] Yahoo Finance, "Yahoo Finance – stock market live, quotes, business & finance news." Online. Accessed: March 7, 2026.

[10] J. M. H. Pinheiro *et al.*, "The impact of feature scaling in machine learning: Effects on regression and classification tasks," *IEEE Access*, vol. 13, pp. 199903–199931, Nov. 2025.

[11] X. Martínez-Barbero, R. Cervelló-Royo, and J. Ribal, "Portfolio optimization with prediction-based return using long short-term memory neural networks," *Computational Economics*, vol. 65, pp. 1479–1504, Mar. 2025.

[12] M. Ashrafzadeh, M. Sadrani, and S. H. Zolfani, "Deep learning and machine learning models for portfolio optimization: Enhancing return prediction with stock clustering," *Results in Engineering*, vol. 27, p. 106263, Sept. 2025.

[13] William, A. K. Wilham, and S. R. Manalu, "Stocks and models: A comparative benchmark of machine learning and deep learning approaches for stock market prediction," *Procedia Computer Science*, vol. 269, pp. 1532–1545, Jan. 2025.

[14] H. Kim, "Mean-variance portfolio optimization with stock return prediction using XGBoost," *Economic Computation and Economic Cybernetics Studies and Research*, vol. 55, no. 4, pp. 5–20, 2021.

[15] R. A. J. Alhatemi and S. Savaş, "A weighted ensemble approach with multiple pre-trained deep learning models for classification of stroke," *Medinformatics*, vol. 1, pp. 10–19, Dec. 2023.

[16] Y. Zhao, J. Wang, Y. Wang, and M. Lv, "How to optimize modern portfolio theory? a systematic review and research agenda," *Expert Systems with Applications*, vol. 263, p. 125780, Mar. 2025.

[17] A. Chaweewanchon and R. Chaysiri, "Markowitz mean-variance portfolio optimization with predictive stock selection using machine learning," *International Journal of Financial Studies*, vol. 10, p. 64, Aug. 2022.

[18] P. Liston, C. Gretton, and A. Lensky, "The role of stop-loss orders in market efficiency and stability: An agent-based study," in *Proceedings of the 16th International Conference on Agents and Artificial Intelligence (ICAART)*, vol. 1, pp. 280–288, 2024.

[19] M. Bağcı and P. Kaya Soylu, "Classification of the optimal rebalancing frequency for pairs trading using machine learning techniques," *Borsa Istanbul Review*, vol. 24, pp. 83–90, Dec. 2024.

[20] B. Benjlijel and H. Mansali, "The expected Sharpe ratio of efficient portfolios under estimation errors," *Cogent Economics & Finance*, vol. 9, Jan. 2021.

[21] Z. Yan, "Investment portfolio optimization with supervised learning and attention mechanism," *Egyptian Informatics Journal*, vol. 32, p. 100839, 2025.