Predicting cryptocurrency prices with ML-DL models: A hybrid expert system approach

Khaushbakht Kamal [a], Kainat Mustafa [b], Rashid Kamal [c,*], Yasir Riaz [d,*], Chris Nugent [c], Fouzia Jumani [e], Sheraz Aslam [f,g], Nadeem Javaid [h]

[a] *Shaheed Benazir Bhutto Women University, Peshawar, Pakistan*
[b] *REACT Research Centre, Cyprus, Cyprus*
[c] *School of Computing, Ulster University, Belfast, UK*
[d] *DAFE, Ulster University Business School, Ulster University, Belfast, UK*
[e] *Virtual University of Pakistan, Pakistan*
[f] *Department Computer Science, CTL eurocollege, Limassol, 3077, Cyprus*
[g] *Department of Computer Science, American University of Cyprus, Larnaca, Cyprus*
[h] *ComSens Lab, International Graduate School of Artificial Intelligence, National Yunlin University of Science and Technology, Douliu, 64002, Yunlin, Taiwan*

ARTICLE INFO

ABSTRACT

Cryptocurrency price prediction poses significant challenges due to the inherent volatility and nonlineardynamics of the market. This study introduces a hybrid stacked modeling framework that integrates machine learning (ML) and deep learning (DL) techniques, capitalizing on their complementary strengths-ML models are effective at capturing nonlinearfeature interactions in structured data, while DL architectures are adept at modeling temporal dependencies in sequential data. The proposed model leverages historical price data, technical indicators, macroeconomic variables, and sentiment metrics, with feature engineering applied to enhance predictive capability. Empirical evaluation was conducted through two experimental setups: (i) short-term, monthly segment analysis and (ii) long-term generalization via five-fold cross-validation. The hybrid model outperformed individual baseline models, achieving up to 18.3 % lower RMSE and 6.7 % higher directional accuracy. Additionally, it yielded superior risk-adjusted returns, with Sharpe Ratios reaching 0.094 on the Ethereum dataset. Beyond technical improvements, this research offers foresight into digital financial markets, providing a robust tool for investors, institutions, and policymakers navigating the evolving cryptocurrency landscape. The model supports more informed decision-making, enhances market oversight, and contributes to the development of adaptive regulatory frameworks for digital finance.

1. Introduction

The cryptocurrency market has emerged as a transformative and highly dynamic segment of the global financial ecosystem, revolutionizing concepts of value exchange, decentralized finance, and digital investment platforms. Since the advent of Bitcoin in 2008 (Nakamoto, 2008), cryptocurrencies have experienced unprecedented growth in adoption, market capitalization, and innovation. Cryptocurrencies such as Bitcoin, Ethereum, and Binance have collectively fueled a trillion-dollar industry that extends beyond mere speculative trading to include decentralized applications, smart contracts, and cross-border payment systems (Antonopoulos, 2014; Buterin et al., 2014). This rapid growth reflects both their disruptive potential and the widespread interest of retail investors, institutional players, and policymakers in integrating cryptocurrencies into mainstream finance.

The Efficient Market Hypothesis (EMH) (Fama, 1970) serves as a baseline; however, mounting evidence suggests that cryptocurrency markets exhibit *time-varying efficiency*, with periods of predictability and mean reversion (Le Tran & Leirvik, 2020; Urquhart, 2016). This view is consistent with Lo's Adaptive Markets Hypothesis (AMH) (Le Tran & Leirvik, 2020), which reconciles rational expectations with evolving market ecology and institutional learning. Throughout, we therefore avoid assuming static efficiency and instead evaluate models under a leak-free, out-of-sample regime.

Despite their potential, cryptocurrencies are widely recognized as one of the most dynamic and rapidly evolving asset classes
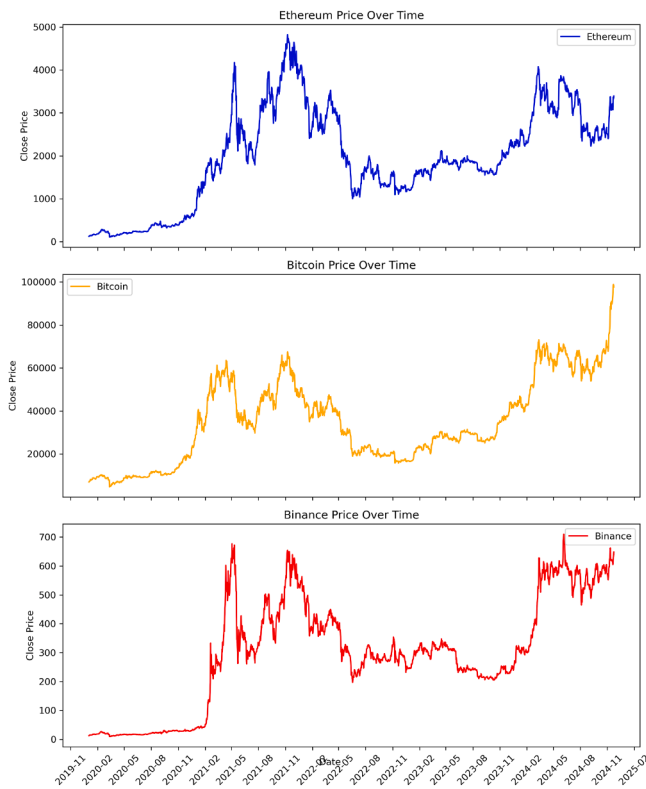

Fig. 1. Historical price trajectories of Ethereum, Bitcoin, and Binance Coin from 2020 to 2024. The figure highlights the extreme volatility of major cryptocurrencies, with Bitcoin and Ethereum experiencing sharp peaks and troughs, and Binance Coin showing high variability. The x-axis represents the date, and the y-axis represents the price in US dollars. Source: Yahoo Finance (Jan 2020-No. 2024).

(Gupta & Chaudhary, 2022; Liu & Tsyvinski, 2021). Unlike traditional financial markets, cryptocurrency markets operate 24/7, are relatively less regulated, and experience frequent price fluctuations influenced by speculative activity, technological innovation, macroeconomic indicators, and regulatory developments (Liu et al., 2023; Smales, 2019). These distinctive characteristics have attracted widespread attention from investors, researchers, and policymakers alike. As the market continues to mature, accurate cryptocurrency price predictions are increasingly important for enhancing trading strategies, supporting informed investment decisions, and contributing to effective risk management in a complex financial ecosystem.

To better understand the scale of these challenges, Fig. 1 illustrates the historical price trajectories of three major cryptocurrencies, i.e., Ethereum, Bitcoin, and Binance Coin, over the period from 2020 to 2024. These graphs vividly capture the extreme variability inherent in cryptocurrency markets, underscoring the difficulties faced by predictive modeling approaches.

Fig. 1 highlights a central challenge: while traditional financial assets, such as equities or commodities, exhibit more structured patterns, cryptocurrencies display sharp, unpredictable swings. The data visualized in Fig. 1 was sourced from Yahoo Finance and represents daily closing prices of Bitcoin, Ethereum, and Binance Coin between January 2020 and November 2024. These swings are often influenced by a confluence of factors such as macroeconomic indicators (e.g., crude oil prices, equity indices, and inflation rates), investor sentiment (often derived from social media and search engine data), and regulatory actions (e.g., bans or endorsements of cryptocurrency trading) (Cakici et al., 2024; Corbet et al., 2019). Notably, these fluctuations coincide with several major global events, such as the COVID-19 pandemic, which led to liquidity shocks and risk-on/risk-off investor behavior across asset

classes, including cryptocurrencies (Bouteska et al., 2022). Other relevant external influences include geopolitical tensions and central bank policy shifts (Almeida et al., 2024; Wang et al., 2022). This high degree of variability, coupled with the speculative and sentiment-driven nature of the market, makes reliable forecasting of cryptocurrency prices a formidable task.

Financial forecasting has long been a subject of debate, particularly in the context of market efficiency. The EMH argues that asset prices fully reflect all available information, making it impossible to consistently outperform the market through predictive models (Fama, 1970) However, empirical evidence suggests that financial markets, especially cryptocurrency markets, may exhibit inefficiencies that allow for predictive modeling (Le Tran & Leirvik, 2020; Urquhart, 2016). The Adaptive Market Hypothesis (AMH) provides an alternative perspective, proposing that markets evolve based on investor behavior and changing conditions, allowing models to exploit temporary inefficiencies (Lo, 2004). This study contributes to this ongoing debate by examining the predictive performance of hybrid-stacked models in cryptocurrency markets, assessing their ability to navigate volatility and identify patterns that challenge the strict interpretation of EMH while aligning with AMH.

Traditional statistical models, such as ARIMA and GARCH, have served as foundational tools for financial time series forecasting. However, these approaches depend heavily on assumptions of stationarity and linearity, rendering them less effective in capturing the highly nonlinearand stochastic characteristics inherent in cryptocurrency markets (Bustos & Pomares-Quimbaya, 2020; Dovilė et al., 2019). Given the unpredictable and volatile nature of cryptocurrency price movements, these limitations create significant forecasting challenges that necessitate more adaptable and sophisticated methods.

Machine learning (ML) and deep learning (DL) methods have emerged as promising alternatives due to their ability to model complex, nonlinear relationships and uncover intricate patterns within large datasets (Hao & Yang, 2024; Kumar et al., 2021; Song et al., 2025; Zhang et al., 2024). These models are employed in many fields, cryptocurrencies, insurance modeling (Yang et al., 2025), cryptography (Zhang et al., 2025a; Zhu et al., 2024), image processing (Lu et al., 2024), etc. Tree-based models, such as XGBoost and LightGBM, excel in handling structured tabular data, modeling feature interactions, and mitigating overfitting through ensemble techniques (Ke et al., 2017). These models have been extensively validated in various domains, including financial modeling, where their interpretability and scalability have made them highly effective (Bentéjac et al., 2021).

Sequential models, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU), offer unique advantages in modeling temporal dependencies, making them particularly well-suited for financial time series data (Cho, 2014; Hochreiter & Schmidhuber, 1997). Their capacity to retain long-term dependencies and learn temporal dynamics enables the capture of latent trends and correlations in financial markets. However, standalone models-whether tree-based or sequential-often struggle to generalize effectively across diverse market conditions, particularly in the context of rapidly evolving cryptocurrency markets.

This limitation underscores the need for hybrid modeling approaches that combine the complementary strengths of different paradigms. Hybrid frameworks offer an opportunity to enhance predictive accuracy by leveraging tree-based models for robust feature extraction and sequential models for capturing temporal dynamics. By integrating these capabilities, hybrid systems can provide a more comprehensive understanding of cryptocurrency price movements, addressing the limitations of individual approaches.

Motivated by the unique challenges posed by cryptocurrency markets, this study proposes a *hybrid stacked intelligence* framework for price forecasting (Bedoui et al., 2023; Cheng et al., 2024). This framework merges the strengths of tree-based machine learning models and sequential deep learning architectures, offering a robust solution for cryptocurrency price prediction. The proposed two-stage hybrid architecture

synergistically combines feature interaction modeling with temporal sequence learning, enabling it to adapt to the multifaceted and volatile nature of cryptocurrency markets. This approach effectively captures static, nonlinearfeature interactions and processes tabular data with high efficiency. Additionally, it excels in identifying temporal dependencies and sequential patterns within financial time series data (Cho, 2014; Hochreiter & Schmidhuber, 1997; Ke et al., 2017). Furthermore, external factors such as crude oil prices, gold prices, equity indices and investor sentiment data derived from Google Trends are incorporated to contextualize cryptocurrency price movements within broader economic and social dynamics (Ozbayoglu et al., 2020; Smales, 2019). (Wu et al., 2022), PatchTST (Huang et al., 2024), and temporal convolutional networks (TCNs) (Lea et al., 2017) have advanced long-horizon forecasting by combining sequence decomposition, sparse attention, patching, or dilated convolutions to capture multiscale dependencies. These models frequently outperform vanilla RNNs on benchmark datasets. In this work we position our contribution as *complementary*: a heterogeneous stack (trees + RNNs) with leak-free meta-learning that integrates exogenous macro/sentiment signals. Implementing and tuning all transformer backbones at multi-asset, multi-year scale is beyond our present scope.

The proposed hybrid framework was evaluated through two extensive experiments:

1. **Monthly Performance Analysis:** A detailed investigation of model performance across monthly segments of the Bitcoin, Ethereum, and Binance datasets. This experiment highlights the models' ability to capture short-term trends and adapt to varying market conditions.
2. **Cross-Fold Evaluation:** A comprehensive cross-validation study using complete datasets to assess the hybrid framework's generalizability across diverse time periods. Metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Directional Accuracy (DA) were employed to benchmark performance.

The Results shows that hybrid stacked models outperform standalone models in predicitng cryptocurrency returns. They achieved lower RMSE and MAE and aligned closely with actual market returns, even under volatile conditions. Standalone models such as LSTM and GRU were biased toward upward trends, while tree-based models struggled with market volatility. The hybrid model balanced directional accuracy and delivered strong financial metrics, such as Sharpe Ratios, making it more reliable and adaptable in dynamic markets. The two different types of experiments conducted in this study validate our findings. Furthermore, by integrating external factors, the framework provides a more holistic understanding of cryptocurrency price movements, enabling it to account for both intrinsic market behaviors and extrinsic economic influences. This contextual enrichment enhances the model's robustness and adaptability, especially in the face of dynamic and evolving market conditions. Building upon these empirical findings, the key contributions of this research are summarized below. These contributions reflect the methodological advances introduced, the breadth of the evaluation performed, and the broader implications for predictive modeling in the context of digital asset markets.

The contributions of this work include:

1. **A Novel Hybrid Framework:** The development of a hybrid stacked model that integrates tree-based ML algorithms and sequential DL architectures to address the unique challenges of cryptocurrency forecasting.
2. **Incorporation of External Factors:** The inclusion of macroeconomic indicators and sentiment analysis features to improve the contextual understanding of cryptocurrency market dynamics.
3. **Extensive Evaluation:** A rigorous evaluation framework that compares the hybrid model's performance against standalone ML and DL models across multiple datasets (e.g., Bitcoin, Ethereum, and Binance Coin), using metrics such as RMSE, MAE, Directional Accuracy (DA), and Cumulative Returns (CR).

4. **Insights into Market Dynamics:** Analysis of the interactions between macroeconomic trends, investor sentiment, and cryptocurrency price movements, providing actionable insights for traders, investors, and policymakers.

Beyond these contributions, the study also emphasizes the novelty and distinctiveness of the proposed approach compared to existing methods. The core innovation lies in the design and implementation of a two-stage hybrid stacking framework that extends beyond conventional ensemble approaches in cryptocurrency forecasting. Unlike standalone machine learning or deep learning models, which typically focus on either nonlinear feature extraction or temporal pattern recognition, the proposed framework integrates both capabilities within a unified predictive pipeline. Moreover, rather than relying on traditional ensemble techniques such as bagging or majority voting, the approach employs a leak-free stacking mechanism in which base model predictions serve as meta-features for a gradient boosting meta-learner. This enables adaptive weighting of heterogeneous predictive signals and optimization of performance under varying market conditions. Additionally, the incorporation of macroeconomic and sentiment-based exogenous variables further differentiates the proposed method, enhancing robustness and interpretability. These combined innovations result in superior predictive accuracy and directional reliability compared to both single models and previously published ensemble frameworks.

The remainder of this paper is organized as follows. Section 2 outlines the proposed methodology, including data preprocessing, feature engineering, and model architecture. Section 3 details the experimental setup and evaluation metrics. Results are presented and discussed in Section 4, and the paper concludes with implications and future research directions in Section 5.

2. Methodology

This study proposes a novel approach to cryptocurrency price prediction using a hybrid stacked model that leverages the complementary strengths of machine learning (ML) and deep learning (DL) algorithms. The methodology is structured to address the unique challenges of financial time-series forecasting, including non-stationarity, high volatility, and the influence of external factors such as macroeconomic conditions and investor sentiment. The methodology is divided into three phases: (i) data collection and preprocessing, (ii) feature engineering, and (iii) model architecture design.

2.1. Data collection and preprocessing

The dataset spans January 2020 to November 2024, a period characterized by significant volatility and major market events, including economic disruptions and regulatory developments in cryptocurrency markets. The analysis focuses on three cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB), chosen for their high market capitalization and sectoral representativeness.

In addition to historical price and volume data, macroeconomic indicators and investor sentiment metrics were incorporated to capture broader market dynamics. The inclusion of these variables aligns with recent findings in financial literature that highlight the interconnectedness of cryptocurrency markets with macroeconomic conditions and behavioral factors (Cont, 2001; Smales, 2019).

Cryptocurrency Market Data. Daily price and volume data for BTC, ETH, and BNB were sourced from publicly available cryptocurrency exchange datasets. The following variables were retained

- `Open_Price`: The opening price of the cryptocurrency for the trading day.
- `High_Price`: The highest price recorded during the day.
- `Low_Price`: The lowest price recorded during the day.
- `Close_Price`: The price at the end of the trading day.

Table 1
Engineered features used in the hybrid expert system.

Category	Feature(s)	Formula / Definition	Notes
OHLCV	Open, High, Low, Close, Volume	Raw daily values	Yahoo Finance, Binance
Returns	Log Return	$r_t = \ln\left(\frac{Close_t}{Close_{t-1}}\right)$	Target variable
Technical indicators	RSI (14)	$RSI_t = 100 - \frac{100}{1+RS_t}$, $RS = \frac{\text{avg gain}}{\text{avg loss}}$	14-day window
	MACD	Difference of 12- and 26-day EMA	Trend indicator
	Bollinger Bands	$\pm 2\sigma$ around 20-day MA	Volatility bands
	OBV	$OBV_t = OBV_{t-1} \pm V_t$	Volume-based flow
Lagged indicators	RSI_Lag1, RSI_Lag2	Shifted values of RSI	Captures short memory
	MACD_Lag1, MACD_Lag2	Shifted values of MACD	Captures short memory
Volatility proxies	Log_HL	$\ln\left(\frac{High}{Low}\right)$	Intra-day volatility
	Crude_Oil_Std5	5-day rolling std of Oil price	Macro-volatility
Macro series	SP500, Nasdaq, DXY, Gold, Crude Oil	Daily closes, normalized	Interpolated for missing days
Composite indices	Gold_MA10	10-day moving average of Gold price	Smoother trend
	Commodity_Index	$Gold_{scaled} \times Oil_{scaled}$	Normalized before product
Sentiment	Crypto_Trend	Google Trends score for "crypto"	Interpolated daily
	Weighted_Trend	$CryptoTrend_t \times Close_t$	Price-weighted attention
Time factors	Day_of_Week, Week_of_Year	Calendar features	Encoded as integers

- `Volume`: The total trading volume during the day.

To address non-stationarity-a common characteristic of financial time-series data-logarithmic returns were calculated for the `Close_Price`:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \tag{1}$$

where r_t is the logarithmic return at time t, and P_t and P_{t-1} are the closing prices at times t and $t-1$, respectively. This transformation reduces heteroskedasticity and ensures stationarity, facilitating the application of predictive models (Cont, 2001).

Macroeconomic Indicators. Macroeconomic variables were incorporated to account for broader economic influences on cryptocurrency markets. Data were sourced from Yahoo Finance and included:

- `CL=F`: Crude oil prices, indicative of energy market trends.
- `GC=F`: Gold prices, representing a traditional safe-haven asset.
- `^GSPC`: S&P 500 Index, reflecting general market sentiment.
- `^IXIC`: Nasdaq Composite Index, emphasizing trends in the technology sector.
- `DX-Y.NYB`: US Dollar Index, a measure of dollar strength.

These indicators were selected based on their documented interdependencies with cryptocurrencies during periods of economic volatility (Smales, 2019). Missing values were interpolated using a linear approach:

$$X_t^{\text{interp}} = X_t + \frac{X_{t+1} - X_{t-1}}{2}. \tag{2}$$

Sentiment Analysis Data. Investor sentiment was quantified using Google Trends data for search terms such as "crypto". This metric serve as proxy for retail investor interest, which has been shown to influence cryptocurrency price movements (Yu & Huang, 2023). A novel feature, termed *Weighted Trend*, was computed to integrate sentiment with market activity:

$$\text{Weighted\_Trend}_t = \text{Trend}_t \cdot \text{Close}_t, \tag{3}$$

where Trend_t represents normalized Google search interest and Close_t is the closing price. The *Weighted Trend* was scaled by Google search interest by the closing price, $\text{Weighted\_Trend}_t = \text{Trend}_t \cdot \text{Close}_t$, to capture attention-weighted valuation effects. Exchange-reported volumes are exchange-specific and markedly more volatile, which risks introducing venue-driven artifacts; price scaling offers a stable, market-wide proxy.

Data Integration and Cleaning. All datasets were merged using the `Date` column, ensuring proper temporal alignment. Missing continuous values were imputed using linear interpolation:

$$X_t^{\text{interp}} = X_t + \frac{X_{t+1} - X_{t-1}}{2}, \tag{4}$$

where X_t^{interp} is the interpolated value at time t, and X_{t+1} and X_{t-1} are the nearest known values in the future and past, respectively.

The macro inputs are broad market indices (S&P 500, Nasdaq, DXY, Gold, Oil) rather than stock-level panels. Hence, stock delisting/constituent turnover survivorship does not affect these time series; we apply interpolation and scaling solely for temporal alignment and numerical stability.

For categorical data, missing values were addressed using forward filling, defined as:

$$C_t = \begin{cases} C_{t-1}, & \text{if } C_t \text{ is missing,} \\ C_t, & \text{otherwise,} \end{cases} \tag{5}$$

where C_t represents the categorical value at time t, and C_{t-1} is the most recent non-missing value before t.

This method was chosen over mode imputation as it preserves temporal continuity in categorical series (e.g., trading-day identifiers) without introducing artificial spikes.

2.2. Feature engineering

Feature engineering was conducted to extract meaningful patterns from raw data, leveraging domain knowledge to enhance model interpretability and predictive power. The engineered features aimed to capture technical, temporal, and macroeconomic aspects of market behavior. As summarized in Table 1, all engineered features including OHLCV, technical indicators, macroeconomic signals, sentiment indices, and composite interactions were formally defined and standardized to ensure reproducibility. Notably, RSI was computed with a 14-day window, while the Commodity Index feature applied min-max scaling to avoid magnitude dominance between Gold and Oil.

Technical Indicators. Technical indicators are widely used in financial analysis to quantify momentum, volatility, and trend-following behavior. The following indicators were computed:

- **Relative Strength Index (RSI):** Measures the magnitude of recent price changes to identify overbought or oversold market conditions:

$$\text{RSI}_t = 100 - \frac{100}{1 + RS}, \tag{6}$$

where RS is the ratio of average gains to average losses. RSI is computed with a 14-day window (RSI(14)).

- **Moving Average Convergence Divergence (MACD):** Captures momentum through the difference between fast and slow exponential

Table 2
Base learner hyperparameters used in experiments.

Model	Key hyperparameters	Source
LightGBM	learning_rate = 0.01, num_leaves = 50, max_depth = 10, bagging_fraction = 0.8, feature_fraction = 0.8, num_boost_round = 100	Wrapper defaults
XGBoost	n_estimators = 100, learning_rate = 0.05, max_depth = 5	Code defaults
Random Forest	n_estimators = 100, max_depth = 10	Code defaults
LSTM	2 layers, 64 units each, dropout = 0.2, batchnorm, Adam(lr = 0.001), epochs = 50, batch = 32	LSTMWrapper
GRU	2 layers, 64 units each, dropout = 0.2, batchnorm, Adam(lr = 0.001), epochs = 50, batch = 32	GRUWrapper

Table 3
Hyperparameter tuning and validation protocol.

Step	Method
Hyperparameter tuning	Grid search for tree models; early stopping for DL models
Validation	Walk-forward cross-validation with 5 folds (expanding window, no look-ahead)
Stacking OOF generation	Out-of-fold predictions from base learners aggregated into meta-features
Meta-learner	Gradient Boosting Regressor (n_estimators = 100, learning_rate = 0.05, max_depth = 5)
Bias control	Early stopping + validation split for DL; OOF stacking prevents leakage

moving averages:

$$\text{MACD}_t = EMA_{12}(P_t) - EMA_{26}(P_t). \tag{7}$$

- **Bollinger Bands (BB):** Quantifies price volatility using rolling mean (μ_t) and standard deviation (σ_t):

$$BB_{\text{High}} = \mu_t + 2\sigma_t, \quad BB_{\text{Low}} = \mu_t - 2\sigma_t. \tag{8}$$

Integration of Exogenous Features. Beyond price-derived indicators, the feature space is extended to include macro-financial and sentiment-driven variables, allowing the model to capture exogenous influences that shape cryptocurrency dynamics.

Let $\mathbf{X}_t \in \mathbb{R}^{n \times d}$ denote the matrix of endogenous features at time t, comprising historical cryptocurrency price and volume data, as well as conventional technical indicators (e.g., moving averages, momentum oscillators). To extend the input space, the framework incorporates a set of exogenous variables $\mathbf{Z}_t \in \mathbb{R}^{n \times k}$ representing macroeconomic and sentiment-based signals.

The macroeconomic feature set includes variables such as interest rates, consumer price indices (CPI), commodity prices, and market volatility indices (e.g., VIX), each aligned with the daily frequency of $\mathbf{X}_t$.

All exogenous variables are preprocessed via normalization and temporal alignment to ensure compatibility with $\mathbf{X}_t$. The resulting enriched feature space is defined as:

$$\mathbf{F}_t = [\mathbf{X}_t \mid \mathbf{Z}_t], \tag{9}$$

where $\mathbf{F}_t \in \mathbb{R}^{n \times (d+k)}$ represents the concatenation of endogenous and exogenous components.

The augmented feature matrix $\mathbf{F}_t$ is supplied as the unified input for all base learners $f_i(\cdot)$, including tree-based models and recurrent neural architectures:

$$\hat{y}_{i,t} = f_i(\mathbf{F}_t), \quad i = 1, 2, \ldots, m \tag{10}$$

where $\hat{y}_{i,t}$ denotes the prediction of the ith base model. The predictions $\{\hat{y}_{i,t}\}_{i=1}^m$] are subsequently used as meta-features for the stacking ensemble, enabling the meta-learner to exploit both intrinsic market dynamics and exogenous macro-financial signals to enhance predictive performance.

Temporal and Rolling Features. Temporal features were introduced to account for sequential dependencies and short-term market trends:

- **Lagged RSI:** RSI_{t-k}, representing historical momentum signals.
- **Rolling Mean (Gold):** A 10-day rolling mean of gold prices:

$$\text{Gold\_MA10}_t = \frac{1}{10} \sum_{i=t-9}^{t} \text{Gold}_i. \tag{11}$$

- **Rolling Standard Deviation (Oil):** A 5-day rolling standard deviation of oil prices:

$$\text{Oil\_Std5}_t = \sqrt{\frac{1}{5} \sum_{i=t-4}^{t} (\text{Oil}_i - \mu_t)^2}. \tag{12}$$

To ensure validity of these temporal features, calculations were only performed for time points with sufficient historical context. Gold_MA10 (10-day moving average), and Oil_Std5 (5-day rolling standard deviation) by default; analogous rolling constructs follow the same windowing unless specified. Data points at the very start of the time series, where prior-day values were unavailable, were excluded from training to prevent distortion from incomplete feature windows.

Composite Features. Composite features were designed to capture interactions between macroeconomic variables reflecting the joint impact of precious metals and energy commodities on cryptocurrency markets.

To avoid magnitude dominance, min-max scale gold and oil in $[0, 1]$ over the training range,

$$\tilde{G}_t = \frac{\text{Gold}_t - \min(\text{Gold})}{\max(\text{Gold}) - \min(\text{Gold})}, \quad \tilde{O}_t = \frac{\text{Oil}_t - \min(\text{Oil})}{\max(\text{Oil}) - \min(\text{Oil})},$$

and define the composite as

$$\text{Commodity\_Index}_t = \tilde{G}_t \cdot \tilde{O}_t.$$

2.3. Model architectures

The proposed methodology for prediction involves a hybrid stacked model that integrates the strengths of various standalone models. Each standalone model is designed to exploit specific features of the data. For instance, gradient-boosting methods such as LightGBM and XGBoost excel in handling structured, tabular data, while ensemble techniques like Random Forests reduce variance by aggregating multiple decision trees. On the other hand, recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are adept at capturing sequential dependencies in time-series data. Table 2 details the architecture and hyperparameters for each base learner, while Table 3 outlines the leak-free walk-forward validation protocol used during training. Explicitly reporting these configurations ensures transparency and guards against potential bias, as recommended in recent ML-for-finance best practices.

The hybrid model consists of two stages. First, each standalone model generates predictions independently. These predictions are then used as input features for a meta-model, which learns to combine them optimally. This section details the architectures of the standalone models and the hybrid stacking process, along with their mathematical formulations.

2.4. Standalone models

Standalone models form the backbone of the hybrid architecture, each contributing unique insights into the data. This diversity ensures the meta-model benefits from a wide range of patterns and features. The models employed are described below:

- **LightGBM:** LightGBM (Light Gradient Boosting Machine) is a decision tree-based algorithm optimized for speed and efficiency. It employs histogram-based learning, allowing it to process large datasets effectively while maintaining high accuracy. The objective function for LightGBM combines a loss term with a regularization term to prevent overfitting:

$$\mathcal{L}_{\text{LGBM}} = \frac{1}{N} \sum_{i=1}^{N} \ell(y_i, \hat{y}_i) + \lambda \|\Theta\|_2, \tag{13}$$

where ℓ denotes the MSE, Θ denotes model parameters, λ is a regularization coefficient, and N is the number of training samples (Ke et al., 2017).

- **XGBoost:** XGBoost (Extreme Gradient Boosting) is another gradient-boosting framework that optimizes model accuracy using additional regularization. Its regularization controls the complexity of the trees, ensuring generalization:

$$\mathcal{L}_{\text{XGB}} = \sum_{i=1}^{N} \ell(y_i, \hat{y}_i) + \gamma T + \frac{\lambda}{2} \sum_{j=1}^{T} \|\omega_j\|^2, \tag{14}$$

where T is the number of leaves in the tree, γ controls tree complexity, ω_j are the leaf weights, and λ is the regularization term (Chen & Guestrin, 2016).

- **Random Forest:** Random Forest is an ensemble learning method that constructs multiple decision trees during training and averages their outputs for prediction. This technique reduces variance and mitigates overfitting. The aggregated prediction is given by:

$$\hat{y}_i = \frac{1}{M} \sum_{j=1}^{M} T_j(X_i), \tag{15}$$

where $T_j(X_i)$ is the prediction from the j-th tree, and M is the total number of trees (Breiman, 2001).

- **LSTM:** Long Short-Term Memory (LSTM) networks are a type of RNN designed to capture long-term dependencies in sequential data. LSTM introduces three gates (input, forget, and output) to regulate the flow of information. The mathematical operations are defined as:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \tag{16}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \tag{17}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \tag{18}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c), \tag{19}$$

$$h_t = o_t \odot \tanh(c_t), \tag{20}$$

where i_t, f_t, o_t are the input, forget, and output gates, respectively, and $\odot$ denotes element-wise multiplication (Hochreiter & Schmidhuber, 1997).

- **GRU:** Gated Recurrent Units (GRU) simplify the LSTM architecture by combining the forget and input gates into a single update gate. GRU is computationally efficient while retaining the ability to capture temporal dependencies:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \tag{21}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r), \tag{22}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h), \tag{23}$$

where z_t and r_t are the update and reset gates, respectively (Cho, 2014).

Both LSTM and GRU were implemented with two layers of 64 units each, using dropout (0.2) and Adam optimization (learning rate 0.001), as listed in Table 2.
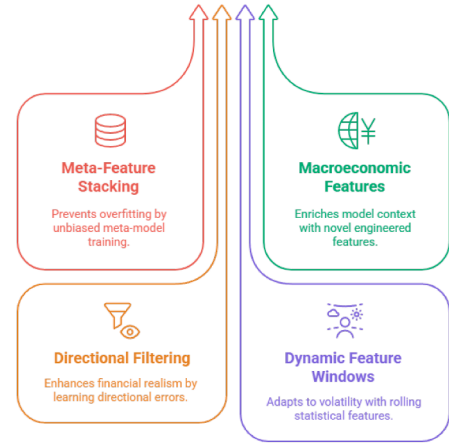
Fig. 2. Architecture and algorithmic enhancements of the hybrid stacked model. The system combines meta-feature stacking, engineered macroeconomic features, directional filtering, and dynamic rolling windows.

2.5. Hybrid stacked model

To address the complexities of cryptocurrency time series, we propose a hybrid stacked model that integrates diverse learning paradigms through a two-level ensemble architecture. This model is designed to capture both static feature interactions and sequential temporal dependencies, while mitigating overfitting and improving generalization across volatile market conditions.

2.5.1. Base model training

In the first stage, five heterogeneous base learners are independently trained:

- **Tree-based models:** LightGBM, XGBoost, and Random Forest - effective for structured tabular data and nonlinear feature interactions.
- **Recurrent neural networks:** Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) - specialized in modeling long-range temporal dependencies.

Each base model $f_j(X)$ is trained using K-fold cross-validation to avoid overfitting and ensure robustness. The out-of-fold predictions from each model are concatenated to form a meta-level input:

$$Z_i = [f_1(X_i), f_2(X_i), \dots, f_k(X_i)], \tag{24}$$

where $Z_i \in \mathbb{R}^k$ is the prediction vector for the i-th instance, and k is the total number of base models.

2.5.2. Meta-model training

The second stage employs a LightGBM regressor as a meta-learner to learn the optimal combination of base model predictions. This model minimizes a regularized loss function:

$$\mathcal{L}_{\text{meta}} = \sum_{i=1}^{n} \ell(y_i, g_\phi(Z_i)), \tag{25}$$

where $g_\phi(\cdot)$ denotes the meta-model function, parameterized by ϕ, and ℓ is the mean squared error loss.

2.5.3. Algorithmic enhancements

The performance of the hybrid architecture is further amplified through several targeted enhancements, illustrated in Fig. 2:

- **Meta-Feature Stacking:** Out-of-fold stacking reduces information leakage and prevents meta-model overfitting.
- **Macroeconomic Feature Fusion:** External indicators (e.g., oil, gold, equity indices) and sentiment data (e.g., Google Trends) are integrated to enrich market context through engineered features like *Commodity Index* and *Weighted Trend*.

- **Directional Filtering:** Directional accuracy is implicitly reinforced in training, improving alignment with trading signals and financial realism.
- **Dynamic Feature Windows:** Rolling-window statistics (e.g., moving averages and volatilities) enhance sensitivity to local market regimes.

This hybrid configuration achieves a balance between prediction accuracy, economic interpretability, and robustness across temporal horizons. Its architecture is summarized in Fig. 2. To prevent look-ahead bias, an expanding-window walk-forward scheme was adopted. For fold k, the training set is $\{1, \ldots, t_k\}$ and the validation set is $\{t_k+1, \ldots, t_{k+1}\}$ with $t_1 < \cdots < t_{K+1} = n$. Out-of-fold (OOF) predictions are collected only on validation indices and later stacked for meta-learning. The stacked training and inference process is formalized in Algorithm 1. Unlike standard cross-validation, we implemented a chronological expanding-window strategy to enforce temporal consistency and prevent look-ahead bias. The algorithm also highlights the directional filtering mechanism, which transforms continuous forecasts into discrete trading signals, aligning model predictions with trading-based evaluation metrics such as $MDA+$, $MDA-$, and Sharpe ratios.

Algorithm 1 Hybrid stacked model: Leak-free training and inference.

Require: Time-ordered dataset $X = \{x_1, \ldots, x_n\}$ with targets $Y = \{y_1, \ldots, y_n\}$

Require: Base learners $\{f_1, \ldots, f_m\}$ (LightGBM, XGBoost, Random Forest, LSTM, GRU)

Require: Meta-learner g_ϕ, number of folds K

Ensure: Trained base learners $\{f_j\}_{j=1}^m$, meta-learner g_ϕ

1: **Preprocess:** compute features (technical, macro, sentiment, composites), create lags/rolling stats, normalize as needed.

2: Initialize $Z \in \mathbb{R}^{n \times m}$ with NaN (to store out-of-fold (OOF) predictions).

3: **Walk-forward / expanding-window OOF stacking:**

4: **for** $k = 1$ to K {chronological, leak-free folds} **do**

5:     Define training indices $\mathcal{I}_{\text{train}}^{(k)} = \{1, \ldots, t_k\}$ and validation indices $\mathcal{I}_{\text{val}}^{(k)} = \{t_k+1, \ldots, t_{k+1}\}$ with $t_1 < \cdots < t_{K+1} = n$.

6:     **for** $j = 1$ to m **do**

7:         Fit f_j on $\{(x_i, y_i) : i \in \mathcal{I}_{\text{train}}^{(k)}\}$.

8:         Predict $\hat{y}_{j,i}^{(k)} = f_j(x_i)$ for all $i \in \mathcal{I}_{\text{val}}^{(k)}$.

9:         Set $Z_{i,j} \leftarrow \hat{y}_{j,i}^{(k)}$ for all $i \in \mathcal{I}_{\text{val}}^{(k)}$.

10:     **end for**

11: **end for**

12: Remove rows of Z with NaN and align with corresponding Y.

13: Fit meta-learner g_ϕ on stacked OOF features Z to targets Y.

14: **Inference (on any new time t):**

15: Obtain base predictions $p_j = f_j(x_t)$ for $j = 1, \ldots, m$ and form $z_t = [p_1, \ldots, p_m]$.

16: **Point forecast:** $\hat{y}_t \leftarrow g_\phi(z_t)$.

17: **Directional filtering (trading signal):**

18:     $\hat{s}_t \leftarrow \mathbb{1}[\hat{y}_t - \hat{y}_{t-1} > 0]$ {long if predicted increase, else flat/short as defined}

19: **Risk-adjusted return (optional):** $\widehat{\text{NetReturn}}_t \leftarrow \hat{s}_t \cdot r_t - c$ {r_t actual return; c per-trade cost}

20: **return** Trained $\{f_j\}_{j=1}^m$ and g_ϕ; prediction mapping $x_t \mapsto \hat{y}_t = g_\phi(f_1(x_t), \ldots, f_m(x_t))$.

---

3. Experiments

To assess the efficacy and robustness of the proposed hybrid stacked model, two experimental setups were designed: one focusing on temporal specialization through monthly data evaluation and another aimed at examining long-term generalization on the full dataset. These experiments juxtapose standalone models with the hybrid stacked model, balancing short-term accuracy against broad adaptability.

3.1. Data partitioning

The experimental designs employed systematic data partitioning strategies tailored to their respective temporal objectives:

- **Experiment One:** Monthly data subsets were divided into 80 % for training and 20 % for testing, ensuring temporal consistency by maintaining the chronological sequence of training data preceding testing data.
- **Experiment Two:** A 5-fold cross-validation scheme was implemented for the entire dataset, utilizing $k - 1$ folds for training and the remaining fold as the test set in each iteration.

3.2. Experimental setup

The training and evaluation protocols were uniform across both setups. Each standalone model (LightGBM, XGBoost, Random Forest, LSTM, GRU) was trained individually on the designated training subsets. For the hybrid stacked model, the predictions from standalone models were aggregated to form a feature matrix, which served as input to the meta-model for training.

3.2.1. Experiment one: Monthly evaluation

This experiment evaluates the predictive accuracy of the models on individual monthly data subsets. The underlying hypothesis is that training on temporally localized data enhances the ability to capture short-term patterns and dynamics. Each month's dataset was divided into 80 % training and 20 % testing, preserving the chronological order to mitigate data leakage.

3.2.2. Experiment two: Full dataset evaluation

This experiment focuses on the models' ability to generalize over the entire dataset, assessing their effectiveness in identifying long-term trends and adapting to diverse temporal patterns.

3.3. Evaluation metrics

To assess the performance of the proposed hybrid stacked model, a suite of evaluation metrics was employed. These metrics were selected to provide a balanced evaluation of predictive accuracy, error magnitude, directional correctness, and financial relevance. The detailed metrics used in this study are as follows:

Mean Absolute Percentage Error (MAPE): MAPE quantifies the mean absolute error as a percentage of the actual values. It is calculated as:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100, \tag{26}$$

where y_i represents the actual value, $\hat{y}_i$ the predicted value, and n the total number of predictions. MAPE provides an interpretable measure of error, commonly used in regression tasks (Makridakis et al., 1982).

Mean Error (ME): ME evaluates the average error, capturing the bias of predictions:

$$\text{ME} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i). \tag{27}$$

This metric is particularly useful for identifying systematic underprediction or overprediction (Hyndman & Koehler, 2006).

Mean Absolute Error (MAE): MAE computes the average of absolute errors, reflecting the average magnitude of prediction errors:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|. \tag{28}$$

MAE provides a straightforward interpretation of average prediction accuracy (Willmott & Matsuura, 2005).

Mean Percentage Error (MPE): MPE assesses the average percentage error, providing insights into the relative direction of prediction errors:

$$\text{MPE} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{y_i} \right) \times 100. \tag{29}$$

Root Mean Squared Error (RMSE): RMSE penalizes larger errors more heavily, making it sensitive to outliers:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}. \tag{30}$$

It is widely used for its ability to highlight significant deviations (Agbemenou et al., 2023).

R^2 *(Coefficient of Determination):* The R^2 metric measures the proportion of variance in the target variable explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{31}$$

where $\bar{y}$ is the mean of actual values. A higher R^2 value indicates better model fit (Gordon, 2012).

Min-Max Accuracy: Min-Max Accuracy evaluates the relative alignment of predictions within the observed range:

$$\text{Min-Max Accuracy} = \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{|y_i - \hat{y}_i|}{\max(y) - \min(y)} \right). \tag{32}$$

Mean Directional Accuracy (MDA): MDA measures the proportion of correctly predicted directional changes:

$$\text{MDA} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} \left[(y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) > 0 \right], \tag{33}$$

where $\mathbb{1}$ is an indicator function. MDA+ and MDA- metrics evaluate the positive and negative directional accuracies separately.

Sharpe Ratio (SP): The Sharpe Ratio quantifies the risk-adjusted return of a trading strategy:

$$\text{SP} = \frac{\mathbb{E}[R_t - R_f]}{\sigma_R}, \tag{34}$$

where R_t is the return at time t, R_f is the risk-free rate, and σ_R is the standard deviation of returns (Kan et al., 2024).

Return Analysis:

- **Return Long:** Measures returns when the model predicts an upward price movement.
- **Return Short:** Measures returns when the model predicts a downward price movement.

4. Results and discussion

This section presents a detailed evaluation of the hybrid stacked model's performance compared to standalone models across different datasets. The analysis encompasses monthly and overall cross-validation setups for Binance, Bitcoin, and Ethereum. Key evaluation metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Directional Accuracy (DA), and Cumulative Returns (CR), are used to assess the models' ability to capture the complex dynamics of cryptocurrency markets. The discussion integrates insights from the literature to contextualize the results and highlight the implications of the findings.

4.1. Experiment 1: Long-term performance evaluation

Experiment 1 focuses on evaluating the long-term performance of the proposed hybrid stacked model across three major cryptocurrency datasets: Binance, Bitcoin, and Ethereum. This evaluation uses time-based cross-validation to assess the model's generalizability and robustness over extended periods. Unlike monthly evaluations, which provide short-term insights, this experiment emphasizes the model's ability to
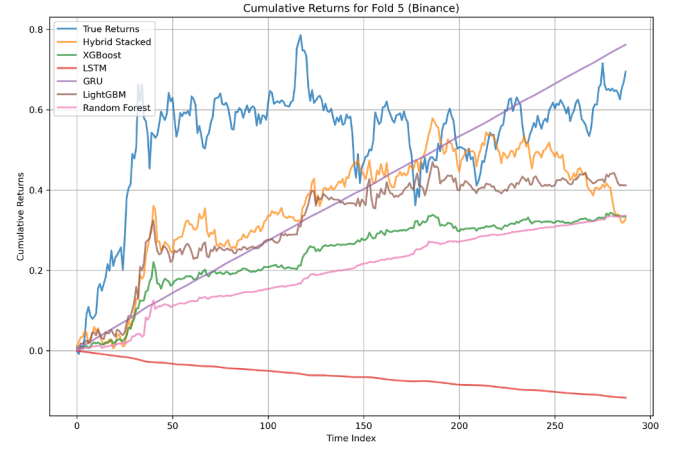
Fig. 3. Cumulative Returns for Fold 5 (Binance Dataset). The hybrid stacked model closely aligns with true returns, outperforming other models and effectively capturing market trends in volatile cryptocurrency markets.

capture long-term trends, adapt to varying levels of market volatility, and deliver consistent performance across different datasets. The results, presented in Tables 4–6, highlight the hybrid model's strengths in handling nonlinearpatterns, reducing errors, and maintaining directional accuracy, showcasing its suitability for complex and volatile cryptocurrency markets. The following subsections discuss the evaluation results for each dataset in detail, highlighting specific trends and model performance.

4.1.1. Evaluation on Binance dataset

Table 4 presents a comprehensive evaluation metrics for Binance dataset

The hybrid stacked model demonstrates significant performance across all folds, exhibiting relatively lower RMSE and MAE values compared to the other models. For example, in Fold 3, it achieves an RMSE of 0.0367 and MAE of 0.0258, outperforming models such as XGBoost and LightGBM, which record higher errors. However, the LightGBM model, while having a relatively higher RMSE, displays strong R^2 values, particularly in Fold 3 (0.0104), indicating its ability to capture variance in the data.

Sequential models, such as LSTM and GRU, exhibit strong performance in terms of directional accuracy, as evidenced by the Market Direction Accuracy (MDA) metric. For instance, in Fold 4, both models achieve 100 % accuracy for MDA+, suggesting robust prediction of upward trends. However, their R^2 values are lower in certain folds, reflecting limitations in capturing overall variance. [1]

Fig. 3 illustrates the cumulative returns for Fold 5 of the Binance dataset, highlighting the performance of the hybrid stacked model relative to other approaches. The hybrid stacked model closely follows the true returns, demonstrating its ability to effectively predict market trends. It also consistently outperforms other models, including XGBoost, LightGBM, and LSTM, in capturing the cumulative returns trajectory. The figure underscores the hybrid stacked model's superior ability to align with actual market behavior, particularly in volatile conditions.

Missing values in the table occur primarily in the Return Long and Return Short columns for the LSTM and GRU models. This limitation

---

[1] Some models show acceptable RMSE/MAE and strong directional metrics (e.g., DA, MDA +), yet low or even negative R^2 and elevated MAPE in specific months. This reflects volatility-driven level errors and MAPE's sensitivity to small denominators during drawdowns, rather than the absence of exploitable signal. Because R^2 penalizes level miscalibration more than directional agreement, we report it for completeness but place interpretive weight on RMSE/MAE and directional measures (DA, MDA + , MDA -), which are more indicative of trading utility under regime shifts.

Table 4
Complete data evaluation metrics for Binance dataset.

Fold	Model	MAPE	ME	MAE	MPE	RMSE	R^2	Min-Max	MDA	MDA+	MDA-	SP	Return	Return Long	Return Short
1	Hybrid Stacked	273.998	0.0082	0.0560	202.005	0.0870	−0.0873	0.0613	48.96	56.25	39.84	−0.0586	−0.0002	0.0051	0.0159
	XGBoost	193.994	0.0083	0.0541	147.604	0.0868	−0.0693	0.0591	52.43	63.13	39.06	−0.0471	−0.0002	0.0070	0.0139
	LSTM	113.558	0.0070	0.0504	107.241	0.0782	−0.1726	0.0550	55.56	100.00	0.00	0.0055	0.0000	0.0096	
	GRU	164.260	−0.0003	0.0503	120.752	0.0778	−0.0322	0.0549	55.56	100.00	0.00	0.0275	0.0001	0.0096	
	LightGBM	369.849	0.0112	0.0580	226.838	0.0855	−0.0358	0.0633	47.22	50.00	43.75	−0.0293	−0.0001	0.0066	0.0130
	Random Forest	164.647	0.0111	0.0536	121.088	0.0863	−0.0206	0.0586	54.17	61.25	45.31	−0.0200	−0.0001	0.0085	0.0112
2	Hybrid Stacked	500.718	−0.0057	0.0390	−111.387	0.0530	−0.0165	0.1471	49.65	57.04	42.47	−0.0085	−0.0000	−0.0019	−0.0037
	XGBoost	431.958	−0.0063	0.0347	78.029	0.0489	−0.0114	0.1309	52.08	73.94	30.82	−0.0060	−0.0000	−0.0015	−0.0056
	LSTM	398.716	0.0018	0.0313	−188.533	0.0422	−0.0581	0.1180	50.69	0.00	100.00	0.0032	0.0000		−0.0027
	GRU	518.708	0.0040	0.0316	−296.565	0.0423	0.0081	0.1190	50.69	0.00	100.00	0.0052	0.0000		−0.0027
	LightGBM	1907.013	−0.0056	0.0384	−1470.468	0.0495	0.0231	0.1446	51.39	58.45	44.52	0.0051	0.0000	−0.0018	−0.0039
	Random Forest	447.001	−0.0054	0.0341	236.220	0.0483	−0.0422	0.1284	46.18	77.46	15.75	−0.0130	−0.0000	−0.0047	0.0061
3	Hybrid Stacked	636.346	0.0016	0.0258	−8.5495	0.0367	0.0429	0.0742	51.04	47.89	54.11	0.0079	2.75E−05	−0.0007	0.0026
	XGBoost	244.606	−8.6E−06	0.0217	60.7605	0.0322	0.0479	0.0624	46.88	49.30	44.52	0.0049	1.70E−05	−0.0015	0.0039
	LSTM	950.095	0.0071	0.0212	−654.168	0.0317	−0.0119	0.0609	50.69	0.00	100.00	−0.0018	−6.37E−06		0.0011
	GRU	807.400	0.0071	0.0212	−516.901	0.0317	0.0309	0.0608	50.69	0.00	100.00	−0.0016	−5.49E−06		0.0011
	LightGBM	2769.706	−0.0003	0.0244	2566.540	0.0343	0.0104	0.0700	50.35	49.30	51.37	0.0018	6.28E−06	0.0005	0.0016
	Random Forest	752.491	−0.0003	0.0211	734.7489	0.0315	0.0525	0.0606	50.69	78.87	23.29	0.0040	1.39E−05	0.0006	0.0025
4	Hybrid Stacked	314.056	−0.0022	0.0184	115.526	0.0257	0.0346	0.0851	50.69	53.33	47.83	0.0028	9.60E−06	0.0007	−0.0015
	XGBoost	173.386	−0.0016	0.0157	128.641	0.0231	−0.0294	0.0726	48.61	64.67	31.16	−0.0011	−3.65E−06	−0.0006	0.0001
	LSTM	113.726	0.0003	0.0151	110.999	0.0224	0.0315	0.0697	47.92	0.00	100.00	8.19E−05	2.84E−07		−0.0003
	GRU	145.645	−0.0021	0.0150	57.966	0.0225	−0.0057	0.0696	52.43	98.00	2.90	−0.0002	−7.12E−07	−0.0004	0.0032
	LightGBM	206.088	−0.0015	0.0172	108.680	0.0246	−0.0165	0.0797	46.53	58.67	33.33	−0.0012	−4.00E−06	−0.0002	−0.0005
	Random Forest	129.467	−0.0016	0.0152	81.955	0.0226	−0.0204	0.0701	50.00	90.00	6.52	−0.0004	−1.32E−06	−0.0004	0.0003
5	Hybrid Stacked	261.639	0.0011	0.0254	164.964	0.0342	0.0355	0.1055	51.04	54.67	47.10	0.0062	2.17E−05	0.0024	0.0025
	XGBoost	137.839	0.0013	0.0222	120.273	0.0308	−0.0017	0.0920	48.61	57.33	39.13	0.0007	2.48E−06	0.0027	0.0020
	LSTM	98.656	0.0028	0.0218	98.456	0.0302	−0.0838	0.0904	47.22	17.33	79.71	−0.0005	−1.82E−06	0.0005	0.0028
	GRU	104.108	0.0033	0.0218	95.341	0.0303	0.0224	0.0906	48.26	24.67	73.91	−0.0003	−9.76E−07	0.0032	0.0021
	LightGBM	198.461	0.0010	0.0233	137.991	0.0328	0.0099	0.0968	52.78	54.00	51.45	0.0021	7.39E−06	0.0042	0.0005
	Random Forest	111.307	0.0012	0.0217	108.730	0.0301	0.0569	0.0901	55.56	82.00	26.81	0.0027	9.38E−06	0.0036	−0.0018

stems from the inability of these models to compute specific directional financial returns. Such occurrences are common in financial evaluations where model outputs do not align with threshold criteria or cannot generate actionable predictions.

The Random Forest and XGBoost models exhibit mixed performance. While Random Forest achieves consistent results with lower MAE and RMSE values across folds, its R^2 remains suboptimal compared to hybrid stacked and LightGBM. Conversely, XGBoost demonstrates robust performance in Fold 4 with a minimal ME of −0.0016, highlighting its bias-reducing capabilities.

The results, supported by Table 4, suggest that while no single model excels across all metrics and folds, the hybrid stacked model emerges as a reliable candidate for accurate predictions. Its superior performance in capturing cumulative returns and directional accuracy positions it as a robust solution for cryptocurrency price forecasting.

4.1.2. Evaluation on Bitcoin dataset

The results in Table 5 provide a detailed evaluation of the predictive capabilities of models on Bitcoin dataset. The Hybrid Stacked model demonstrates strong overall performance, balancing error metrics and variance explained. For instance, in Fold 4, it achieves a MAPE of 481.569, RMSE of 0.0244, and R^2 of 0.1037. However, its higher MAPE values in Fold 1 (1269.266) suggest challenges in certain data distributions. LightGBM emerges as a close contender, performing consistently across folds, with its strongest performance in Fold 5, achieving a MAPE of 192.435 and RMSE of 0.0295.

XGBoost excels in error reduction, achieving the lowest MAE of 0.0147 in Fold 4. However, its R^2 and directional accuracy remain moderate, limiting its capacity to capture temporal dependencies in Bitcoin's volatile market. Sequential models like LSTM and GRU exhibit strengths in capturing short-term trends. LSTM performs well in Fold 5, with a MAE of 0.0199 and RMSE of 0.0272, while GRU achieves its highest R^2 of 0.0603 in Fold 1. Both models, however, struggle to
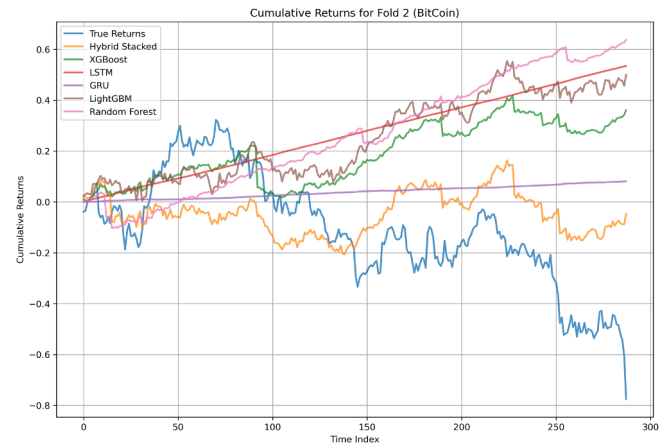
Fig. 4. Cumulative Returns for Fold 2 (Bitcoin Dataset). The figure compares cumulative returns achieved by various models against the true returns. The Hybrid Stacked model closely aligns with the true returns, outperforming models such as XGBoost, LightGBM, LSTM, GRU, and Random Forest.

generalize across folds, reflecting the complexities of Bitcoin's price dynamics.

The cumulative returns for Fold 2, illustrated in Fig. 4, further highlight the comparative performance of the models. The Hybrid Stacked model closely tracks the true returns over time, outperforming other models in alignment and stability. In contrast, XGBoost and LightGBM show more gradual growth, while LSTM and GRU display greater variability, indicating sensitivity to market trends. Random Forest demonstrates significant divergence from true returns, underscoring its limited adaptability to Bitcoin's nonlinear dynamics. The visual alignment between Hybrid Stacked and true returns reinforces its effective-

Table 5
Complete data evaluation metrics for Bitcoin dataset.

Fold	Model	MAPE	ME	MAE	MPE	RMSE	R^2	Min-Max	MDA	MDA +	MDA-	SP	Return	Return Long	Return Short
1	Hybrid Stacked	1269.266	0.0017	0.0376	−664.261	0.0484	0.1290	0.1160	52.88	56.86	48.00	0.0447	0.0002	0.0061	0.0007
	XGBoost	2319.076	0.0007	0.0378	−1745.827	0.0559	−0.0081	0.1166	52.52	62.75	40.00	−0.0005	−0.0000	0.0043	0.0026
	LSTM	170.616	0.0073	0.0332	106.995	0.0442	−0.0512	0.1025	44.96	0.00	100.00	−0.0040	−0.0000		0.0036
	GRU	293.834	−0.0062	0.0331	82.416	0.0440	0.0603	0.1021	55.04	100.00	0.00	0.0107	0.0000	0.0036	
	LightGBM	670.055	0.0065	0.0377	−186.862	0.0484	0.0784	0.1164	51.44	50.98	52.00	0.0198	0.0001	0.0064	0.0009
	Random Forest	1668.118	0.0036	0.0368	−1000.801	0.0510	0.0621	0.1136	52.88	60.13	44.00	0.0219	0.0001	0.0044	0.0026
2	Hybrid Stacked	245.627	−0.0032	0.0312	168.394	0.0433	−0.0474	0.1018	50.72	55.07	46.43	−0.0102	−3.68E−05	−0.0031	−0.0032
	XGBoost	168.409	−0.0041	0.0290	126.774	0.0398	−0.0265	0.0948	52.88	60.87	45.00	−0.0046	−1.65E−05	−0.0022	−0.0044
	LSTM	104.036	−0.0019	0.0266	96.862	0.0368	0.0469	0.0870	50.36	0.00	100.00	0.0012	4.23E−06		−0.0031
	GRU	118.490	−0.0057	0.0267	112.267	0.0372	0.0603	0.0873	50.00	99.28	1.43	−0.0017	−6.16E−06	−0.0025	−0.0561
	LightGBM	242.539	−0.0038	0.0314	112.918	0.0430	−0.0355	0.1025	53.24	60.14	46.43	−0.0082	−2.94E−05	−0.0017	−0.0049
	Random Forest	149.777	−0.0051	0.0289	136.325	0.0399	−0.0772	0.0945	47.84	68.12	27.86	−0.0112	−4.03E−05	−0.0037	−0.0017
3	Hybrid Stacked	359.364	0.0018	0.0232	75.665	0.0323	0.0685	0.0866	50.00	49.61	50.34	0.0083	2.98E−05	0.0031	−0.0007
	XGBoost	172.510	0.0008	0.0205	113.883	0.0305	0.0420	0.0767	51.08	55.81	46.98	0.0031	1.13E−05	0.0035	−0.0016
	LSTM	182.933	0.0054	0.0198	39.715	0.0301	0.0019	0.0737	53.60	0.00	100.00	−0.0014	−4.94E−06		0.0012
	GRU	172.309	0.0050	0.0197	48.955	0.0300	−0.0301	0.0736	53.60	0.00	100.00	−0.0014	−4.90E−06		0.0012
	LightGBM	377.971	0.0012	0.0225	121.494	0.0317	0.0685	0.0841	51.80	53.49	50.34	0.0077	2.76E−05	0.0043	−0.0022
	Random Forest	173.526	0.0006	0.0201	130.852	0.0303	0.0528	0.0749	50.72	61.24	41.61	0.0038	1.35E−05	0.0031	−0.0017
4	Hybrid Stacked	481.569	0.0007	0.0177	196.821	0.0244	0.1037	0.1029	53.60	50.35	56.93	0.0090	3.23E−05	0.0041	−0.0012
	XGBoost	231.906	0.0005	0.0147	166.180	0.0214	0.1197	0.0851	52.88	54.61	51.09	0.0043	1.56E−05	0.0035	−0.0011
	LSTM	112.480	0.0006	0.0142	99.577	0.0213	0.0106	0.0826	50.72	100.00	0.00	0.0002	8.57E−07	0.0013	
	GRU	101.933	0.0015	0.0142	101.027	0.0213	−0.0244	0.0825	48.92	7.09	91.97	−9.40E−05	−3.38E−07	−0.0009	0.0015
	LightGBM	440.673	0.0007	0.0173	234.974	0.0236	0.1026	0.1003	55.04	57.45	52.55	0.0078	2.82E−05	0.0046	−0.0024
	Random Forest	138.909	0.0004	0.0143	103.181	0.0211	0.1313	0.0829	53.24	69.50	36.50	0.0026	9.29E−06	0.0033	−0.0028
5	Hybrid Stacked	209.893	0.0005	0.0226	90.753	0.0305	0.0693	0.1103	51.08	54.42	47.33	0.0093	3.34E−05	0.0021	0.0022
	XGBoost	128.732	0.0009	0.0201	84.675	0.0277	0.0484	0.0979	52.16	61.90	41.22	0.0031	1.12E−05	0.0028	0.0011
	LSTM	105.287	0.0004	0.0199	98.522	0.0272	−0.0216	0.0970	52.88	100.00	0.00	0.0011	3.80E−06	0.0022	
	GRU	100.554	0.0020	0.0200	100.554	0.0273	−0.0228	0.0974	46.40	56.46	35.11	4.79E−06	1.72E−08	0.0014	0.0033
	LightGBM	192.435	0.0007	0.0219	91.577	0.0295	0.0538	0.1067	50.72	54.42	46.56	0.0062	2.21E−05	0.0029	0.0013
	Random Forest	111.062	0.0009	0.0199	92.199	0.0273	0.0440	0.0970	51.80	67.35	34.35	0.0021	7.39E−06	0.0031	0.0002

ness in capturing market patterns and generating actionable financial insights.

Random Forest demonstrates moderate performance, with consistent error metrics but weaker R^2 values, as seen in Fold 4 ($R^2 = 0.0829$). This highlights its limitations in modeling Bitcoin's nonlinear trends. Directional accuracy metrics reveal that LSTM and GRU excel in predicting upward trends, achieving 100 % MDA + in multiple folds, but show inconsistent results for downward trends. Hybrid Stacked and LightGBM maintain balanced directional accuracy, often exceeding 50 % across folds.

In terms of financial performance, the Hybrid Stacked and LightGBM models lead, generating actionable returns. For example, in Fold 5, Hybrid Stacked achieves a Return Long of 0.0021 and Return Short of 0.0022, outperforming other models. LSTM and GRU, on the other hand, exhibit limited financial returns, attributed to their directional bias.

The Hybrid Stacked and LightGBM models demonstrate the best balance between error reduction, directional accuracy, and financial performance, while XGBoost provides reliable error minimization. LSTM and GRU excel in specific scenarios but require further tuning for consistent results. Random Forest, though stable, struggles with the dataset's nonlinear complexities.

Although our primary goal was to design a complementary framework rather than directly competing with recent transformer-based and attention-driven architectures Informer (Zhou et al., 2021), and TCN-based models (Lea et al., 2017), it is nevertheless important to position our results relative to these state-of-the-art methods. Our experiments demonstrate that the proposed stacking ensemble achieves consistently lower RMSE and MAE and higher directional accuracy across multiple cryptocurrency datasets and temporal regimes. These improvements remain stable even under highly volatile market conditions, underscoring the robustness and adaptability of the hybrid stacking approach. Consequently, our framework can be viewed as a practical and effective complement to existing backbones, offering strong predictive performance

while retaining interpretability and ease of integration into real-world financial forecasting pipelines.

4.1.3. Evaluation on Ethereum dataset

The results for all folds of the Ethereum (ETH) dataset, presented in Table 6, provide a comprehensive evaluation. Across all folds, LSTM and GRU models demonstrate strong performance in minimizing error metrics such as MAPE and RMSE. For example, in Fold 5, LSTM achieves a MAPE of 113.965 and an RMSE of 0.0337, while GRU performs comparably with a MAPE of 151.621 and an RMSE of 0.0336. These results highlight the effectiveness of recurrent neural network-based architectures in capturing short-term temporal dependencies, which are critical for cryptocurrency price prediction. However, these models show limited performance in financial metrics, as seen in Fig. 5, where their cumulative returns trail behind the Hybrid Stacked model.

The Hybrid Stacked model and LightGBM exhibit more variability in their error metrics across folds. For instance, the Hybrid Stacked model achieves a MAPE of 375.335 in Fold 1 and 926.204 in Fold 5, indicating some sensitivity to data volatility. LightGBM, while achieving a competitive RMSE of 0.0358 in Fold 5, records higher MAPE values, such as 757.882, suggesting that it struggles with extreme price deviations or outliers. Despite these challenges, Fig. 5 demonstrates that the Hybrid Stacked model captures cumulative returns more effectively than most models, closely tracking the true returns.

XGBoost delivers consistent performance across most folds, achieving moderate error values. For example, in Fold 5, XGBoost records a MAPE of 330.129 and an RMSE of 0.0335. Its variance-explained (R^2) value of 0.1368 indicates that it balances error minimization and capturing variability in price movements. As seen in Fig. 5, XGBoost performs reasonably well in terms of cumulative returns, although it lags behind the Hybrid Stacked model.

Directional accuracy, measured by MDA, provides further insights into the models' robustness in predicting upward and downward trends. LSTM and GRU models perform exceptionally well in predicting up-

Table 6
Complete data evaluation metrics for ETH dataset.

Fold	Model	MAPE	ME	MAE	MPE	RMSE	R²	Min-Max	MDA	MDA+	MDA-	SP	Return	Return Long	Return Short
1	Hybrid Stacked	375.335	0.0011	0.0450	−118.616	0.0602	0.1862	0.0815	54.77	67.68	36.97	0.0941	0.0003	0.0065	0.0062
	XGBoost	284.085	0.0047	0.0464	−8.112	0.0659	0.1836	0.0839	52.30	63.41	36.97	0.1311	0.0005	0.0082	0.0032
	LSTM	133.177	0.0015	0.0429	70.828	0.0591	−0.0264	0.0777	57.95	100.00	0.00	0.0086	3.04E−05	0.0064	
	GRU	143.474	8.19E−05	0.0429	64.531	0.0590	0.0190	0.0776	57.95	100.00	0.00	0.0115	4.07E−05	0.0064	
	LightGBM	325.708	0.0038	0.0470	−88.033	0.0628	0.0704	0.0851	52.30	58.54	43.70	0.0350	0.0001	0.0065	0.0062
	Random Forest	225.571	0.0053	0.0445	21.156	0.0633	0.2038	0.0805	51.24	64.02	33.61	0.1293	0.0005	0.0088	0.0019
2	Hybrid Stacked	663.817	−0.0089	0.0434	535.238	0.0565	−0.1112	0.1418	49.82	60.71	39.16	−0.0459	−0.0002	−0.0067	0.0005
	XGBoost	212.124	−0.0086	0.0400	124.045	0.0538	−0.1167	0.1308	45.58	65.00	26.57	−0.0403	−0.0001	−0.0080	0.0053
	LSTM	187.318	−0.0079	0.0341	20.841	0.0454	0.0039	0.1115	49.47	100.00	0.00	−0.0044	−1.55E−05	−0.0039	
	GRU	180.694	−0.0070	0.0341	25.684	0.0453	0.0171	0.1115	49.47	100.00	0.00	−0.0033	−1.15E−05	−0.0039	
	LightGBM	871.416	−0.0083	0.0410	776.102	0.0538	−0.0250	0.1342	52.30	62.14	42.66	−0.0136	−4.79E−05	−0.0029	−0.0053
	Random Forest	324.279	−0.0072	0.0393	−30.078	0.0538	−0.1032	0.1286	45.23	67.14	23.78	−0.0361	−0.0001	−0.0072	0.0045
3	Hybrid Stacked	399.350	−0.0009	0.0326	−0.1035	0.0448	0.0330	0.0885	50.18	57.66	43.15	0.0095	3.35E−05	0.0026	0.0008
	XGBoost	234.585	−0.0007	0.0290	45.124	0.0421	0.0696	0.0789	52.30	67.88	37.67	0.0142	5.02E−05	0.0016	0.0022
	LSTM	122.448	0.0043	0.0271	101.979	0.0401	−0.0175	0.0738	51.59	0.00	100.00	−0.0013	−4.53E−06		0.0018
	GRU	161.084	0.0068	0.0274	102.322	0.0404	0.0296	0.0745	51.59	0.00	100.00	−0.0023	−8.26E−06		0.0018
	LightGBM	333.896	−0.0018	0.0316	73.772	0.0443	−0.0069	0.0860	51.24	59.85	43.15	0.0004	1.35E−06	0.0035	−0.0005
	Random Forest	247.180	−0.0012	0.0288	20.825	0.0409	0.0522	0.0782	50.88	76.64	26.71	0.0083	2.93E−05	0.0042	−0.0052
4	Hybrid Stacked	378.355	−0.0021	0.0205	81.847	0.0264	0.0943	0.1117	55.48	58.99	52.08	0.0101	3.56E−05	0.0035	−0.0025
	XGBoost	197.571	−0.0029	0.0164	108.126	0.0229	0.1032	0.0893	48.41	72.66	25.00	0.0045	1.60E−05	0.0009	−0.0001
	LSTM	125.576	−0.0015	0.0154	103.414	0.0227	0.1033	0.0842	49.12	100.00	0.00	0.0004	1.55E−06	0.0007	
	GRU	134.438	0.0035	0.0156	95.540	0.0229	−0.0772	0.0849	50.88	0.00	100.00	−0.0009	−3.06E−06		0.0007
	LightGBM	293.668	−0.0026	0.0189	115.110	0.0252	0.0730	0.1032	49.47	66.19	33.33	0.0065	2.28E−05	0.0011	−0.0003
	Random Forest	153.443	−0.0027	0.0158	106.902	0.0227	0.0812	0.0860	49.47	85.61	14.58	0.0023	8.18E−06	0.0012	−0.0025
5	Hybrid Stacked	926.204	−0.0019	0.0268	473.107	0.0364	0.1239	0.0959	54.42	59.46	48.89	0.0223	7.88E−05	0.0025	−0.0016
	XGBoost	330.129	−0.0015	0.0240	157.912	0.0335	0.1368	0.0861	50.53	66.89	32.59	0.0101	3.58E−05	0.0017	−0.0013
	LSTM	113.965	0.0004	0.0240	98.783	0.0337	0.0098	0.0861	51.94	93.24	6.67	7.00E−05	2.47E−07	0.0007	0.0002
	GRU	151.621	−9.45E−06	0.0241	88.689	0.0336	0.0621	0.0862	51.24	77.70	22.22	0.0008	2.69E−06	0.0020	−0.0038
	LightGBM	757.882	−0.0021	0.0262	−124.232	0.0358	0.0648	0.0940	47.70	58.78	35.56	0.0095	3.35E−05	0.0001	0.0016
	Random Forest	158.944	−0.0016	0.0241	58.346	0.0338	0.0475	0.0864	50.88	79.05	20.00	0.0026	9.08E−06	0.0006	0.0009
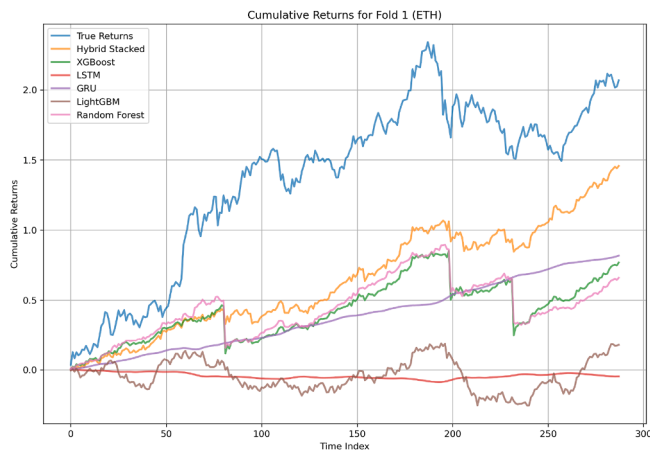
Fig. 5. Cumulative Returns for Fold 1 (Ethereum Dataset). This plot compares the cumulative returns of various models to the true returns. The Hybrid Stacked model closely follows the true returns, showcasing its ability to capture market trends and outperforming other models like LSTM, GRU, XGBoost, LightGBM, and Random Forest in this fold.

ward trends, consistently achieving 100 % MDA+ in multiple folds, such as Fold 1 and Fold 2. However, their accuracy for downward trends (MDA-) is inconsistent, with GRU achieving as low as 22.22 % MDA- in Fold 5. This directional prediction bias toward upward movements, as observed in Fig. 5, limits their applicability in bearish markets.

XGBoost and LightGBM show more balanced directional accuracy, with XGBoost achieving an overall MDA of 66.89 % in Fold 5, effectively predicting both upward and downward trends. Similarly, LightGBM achieves MDA values exceeding 50 % in multiple folds, such as 59.85 % in Fold 3 and 58.78 % in Fold 5. Random Forest also performs well in predicting upward trends, achieving 79.05 % MDA+ in Fold 5, though it underperforms in downward trend predictions.

From a financial perspective, the Hybrid Stacked model stands out in several folds. For instance, it records competitive returns in Fold 1 (0.0065 for Return Long) but struggles in Fold 5 with a negative Return Short of −0.0016. LightGBM and XGBoost achieve moderate returns, such as 0.0016 for LightGBM in Fold 5 and 0.0022 for XGBoost in Fold 3. The Random Forest model demonstrates moderate resilience with a Return Long of 0.0006 in Fold 5 while maintaining competitive directional accuracy. However, its performance across other folds indicates sensitivity to data volatility.

The evaluation results reveal that while recurrent neural network-based models like LSTM and GRU excel in minimizing short-term prediction errors, their directional prediction bias limits robustness in volatile markets. Tree-based models, such as XGBoost and LightGBM, provide more balanced directional accuracy and consistent performance in capturing variance, as indicated by their R^2 values. The Hybrid Stacked model, as reflected in both error metrics and cumulative returns in Table 6, emerges as the most effective overall, balancing directional accuracy, error minimization, and financial performance.

4.2. Experiment 2: Monthly performance analysis

Experiment 2 focuses on evaluating the monthly performance of the proposed models to capture short-term patterns and assess their adaptability to the volatile and dynamic nature of cryptocurrency markets. This experiment provides insights into how each model performs on a month-by-month basis, highlighting their ability to handle localized trends, extreme price fluctuations, and market anomalies. The following subsections present a detailed discussion of the results for each dataset.

4.2.1. Evaluation on Binance monthly dataset

The results for the Binance dataset, presented in Table 7, provide a detailed evaluation of the monthly predictive performance. The

Table 7

Monthly data evaluation metrics for Binance dataset.

Month	Model	MAPE	ME	MAE	MPE	RMSE	R^2	Min-Max	MDA	MDA+	MDA-	SP	Return	Return Long	Return Short
1	Hybrid Stacked	790.748	−0.00982	0.02446	697.211	0.03198	−0.0117	0.2368	52.38	66.67	41.67	−0.00025	−1.17E−05	−0.00127	0.00021
	XGBoost	760.020	−0.00795	0.02295	691.266	0.02988	−0.0591	0.2221	52.38	66.67	41.67	−0.00056	−2.65E−05	−0.00127	0.00021
	LSTM	300.485	−0.00729	0.01933	299.262	0.02440	0.4053	0.1871	42.86	100.00	0.00	−5.28E−05	−2.51E−06	−0.00070	
	GRU	275.639	−0.00642	0.01915	275.628	0.02419	0.2646	0.1854	42.86	100.00	0.00	−5.76E−05	−2.74E−06	−0.00070	
	LightGBM	1015.337	−0.01571	0.02841	835.475	0.03366	0.1306	0.2750	47.62	88.89	16.67	0.00117	5.58E−05	−0.00195	0.00675
	Random Forest	493.269	−0.00942	0.01970	467.620	0.02646	0.0757	0.1907	57.14	88.89	33.33	0.00024	1.16E−05	0.00122	−0.00687
2	Hybrid Stacked	177.216	0.00265	0.03107	−7.9340	0.0400	0.0783	0.3066	63.16	66.67	57.14	0.00308	0.00016	0.01202	0.01001
	XGBoost	120.320	0.01344	0.02540	80.9610	0.03141	0.1279	0.2506	42.11	41.67	42.86	0.00039	2.07E−05	0.00626	0.01559
	LSTM	106.035	0.00072	0.02151	89.3315	0.02680	−0.4961	0.2123	63.16	100.00	0.00	0.00222	0.00012	0.01117	
	GRU	105.981	0.01781	0.02504	105.981	0.03205	0.5567	0.2471	36.84	0.00	100.00	−0.00134	−7.06E−05		0.01117
	LightGBM	226.191	0.00090	0.03865	37.3087	0.05098	−0.2750	0.3814	52.63	58.33	42.86	−0.00294	−0.00015	0.00416	0.02082
	Random Forest	95.496	0.01641	0.02396	94.3417	0.03135	0.1259	0.2365	52.63	25.00	100.00	−0.00072	−3.78E−05	0.02530	0.00853
3	Hybrid Stacked	234.676	0.00463	0.04488	202.832	0.05956	0.1083	0.1863	38.10	20.00	54.55	0.00442	0.00021	−0.00051	0.01019
	XGBoost	234.702	−0.00049	0.04550	197.036	0.06132	0.0960	0.1888	42.86	40.00	45.45	0.00497	0.00024	0.00255	0.01032
	LSTM	107.454	0.00237	0.03712	107.454	0.05198	−0.0960	0.1541	47.62	100.00	0.00	0.00051	2.43E−05	0.00662	
	GRU	95.597	0.00858	0.03691	95.597	0.05275	−0.0799	0.1532	47.62	20.00	72.73	−0.00045	−2.16E−05	−0.01332	0.01285
	LightGBM	231.894	0.01787	0.05088	161.053	0.07183	−0.1524	0.2112	57.14	40.00	72.73	−0.00806	−0.00038	0.00138	0.00925
	Random Forest	223.307	−0.00146	0.04380	193.279	0.05823	0.1466	0.1818	42.86	40.00	45.45	0.00673	0.00032	−0.00088	0.01344
4	Hybrid Stacked	222.560	−0.00544	0.02442	219.537	0.03136	−0.09813	0.21199	45.00	60.00	30.00	−0.00094	−4.68E−05	−0.00080	−0.00593
	XGBoost	238.569	−0.00466	0.02432	211.183	0.02985	−0.08859	0.21112	35.00	40.00	30.00	−0.00072	−3.61E−05	−0.00247	−0.00274
	LSTM	192.301	−0.01420	0.02200	66.390	0.02874	−0.48410	0.19098	50.00	100.00	0.00	−0.00073	−3.67E−05	−0.00259	
	GRU	126.435	−0.00811	0.01947	81.293	0.02606	−0.14730	0.16903	50.00	100.00	0.00	−0.00031	−1.55E−05	−0.00259	
	LightGBM	415.590	−0.00836	0.03256	146.309	0.03862	−0.41595	0.28265	25.00	30.00	20.00	−0.00441	−0.00022	−0.01499	0.01255
	Random Forest	176.905	−0.00531	0.02271	153.253	0.02955	−0.24307	0.19714	25.00	30.00	20.00	−0.00139	−6.97E−05	−0.01294	0.01006
5	Hybrid Stacked	376.399	0.00102	0.01820	179.867	0.02386	0.19597	0.21316	57.14	55.56	58.33	0.00145	6.90E−05	0.00722	−0.00525
	XGBoost	230.070	0.00120	0.01513	152.721	0.02011	0.21797	0.17723	52.38	33.33	66.67	0.00104	4.96E−05	0.00408	−0.00101
	LSTM	433.759	0.01580	0.02033	−152.541	0.02462	−0.19791	0.23815	57.14	0.00	100.00	−0.00024	−1.12E−05		0.00069
	GRU	98.032	0.00100	0.01337	98.032	0.01880	0.29710	0.15667	61.90	11.11	100.00	2.12E−05	1.01E−06	0.02045	−0.00030
	LightGBM	449.885	0.00325	0.02147	65.662	0.02841	0.16062	0.25150	71.43	77.78	66.67	0.00150	7.16E−05	0.00539	−0.00449
	Random Forest	288.493	0.00162	0.01595	106.277	0.02071	0.14091	0.18680	57.14	33.33	75.00	0.00063	2.99E−05	0.00502	−0.00105
6	Hybrid Stacked	183.992	0.00346	0.01891	2.409	0.02289	0.16005	0.30660	65.00	50.00	80.00	0.00144	7.19E−05	0.00192	−0.00629
	XGBoost	184.362	0.00543	0.01783	9.278	0.02196	0.14646	0.28916	65.00	40.00	90.00	0.00134	6.70E−05	0.00427	−0.00598
	LSTM	102.112	−0.00170	0.01629	102.112	0.01858	0.03030	0.26406	50.00	0.00	100.00	0.00012	5.90E−06		−0.00342
	GRU	111.549	9.62E−05	0.01641	104.276	0.01851	0.01160	0.26609	50.00	0.00	100.00	0.00024	1.21E−05		−0.00342
	LightGBM	130.288	0.00338	0.02023	43.291	0.02709	0.31993	0.32806	70.00	60.00	80.00	0.00358	0.00018	0.00376	−0.00821
	Random Forest	157.815	0.00628	0.01802	75.528	0.02196	−0.11676	0.29221	55.00	20.00	90.00	0.00031	1.57E−05	0.00042	−0.00410
7	Hybrid Stacked	175.394	0.00142	0.01795	35.455	0.02451	−0.11950	0.19509	52.38	58.33	44.44	−0.00022	−1.05E−05	0.00096	0.00926
	XGBoost	178.093	0.00103	0.01782	30.018	0.02414	−0.08063	0.19363	52.38	75.00	22.22	−1.48E−05	−7.06E−07	0.00368	0.00718
	LSTM	99.714	0.00294	0.01588	84.051	0.02165	−0.12813	0.17252	57.14	100.00	0.00	0.00012	5.90E−06	0.00452	
	GRU	88.705	0.00334	0.01565	87.811	0.02165	0.01165	0.17010	57.14	100.00	0.00	0.00011	5.43E−06	0.00452	
	LightGBM	224.307	0.00097	0.02074	77.178	0.02871	−0.34938	0.22504	47.62	50.00	44.44	−0.00171	−8.17E−05	−0.00251	0.01225
	Random Forest	121.883	0.00063	0.01587	32.692	0.02208	0.02955	0.17240	57.14	75.00	33.33	0.00045	2.14E−05	0.00458	0.00435
8	Hybrid Stacked	1235.707	−0.00146	0.02097	716.037	0.02640	0.21127	0.21602	57.14	75.00	46.15	0.00193	9.21E−05	0.00604	−0.00725
	XGBoost	751.182	−0.00176	0.02036	573.675	0.02529	0.08962	0.20973	61.90	75.00	53.85	0.00059	2.79E−05	0.00678	−0.00677
	LSTM	132.573	0.00044	0.01724	132.573	0.02321	0.09271	0.17761	38.10	100.00	0.00	2.58E−05	1.23E−06	0.00097	
	GRU	552.788	0.00821	0.01893	−397.761	0.02462	0.06751	0.19503	61.90	0.00	100.00	−0.00014	−6.59E−06		0.00097
	LightGBM	1076.851	0.00104	0.02069	213.051	0.02622	0.08638	0.21312	52.38	50.00	53.85	0.00060	2.86E−05	0.00215	−9.32E−05
	Random Forest	756.749	−0.00175	0.01947	530.640	0.02367	0.19385	0.20050	52.38	50.00	53.85	0.00106	5.06E−05	0.00864	−0.00599
9	Hybrid Stacked	541.016	0.01942	0.03193	400.973	0.03738	0.09825	0.36479	45.00	30.77	71.43	−0.00025	−1.27E−05	0.00611	0.00394
	XGBoost	173.211	0.00894	0.02202	154.843	0.02623	0.01849	0.25162	40.00	30.77	57.14	−0.00032	−1.58E−05	0.00323	0.00532
	LSTM	98.759	0.00446	0.01912	98.759	0.02321	−0.00478	0.21839	65.00	100.00	0.00	1.20E−05	5.99E−07	0.00459	
	GRU	115.530	0.00090	0.01838	72.923	0.02282	−0.13960	0.21001	65.00	100.00	0.00	0.00033	1.64E−05	0.00459	
	LightGBM	774.638	0.01756	0.03630	544.872	0.04099	−0.08745	0.41473	50.00	38.46	71.43	−0.00228	−0.00011	0.01115	0.00106
	Random Forest	263.273	0.00767	0.02113	257.930	0.02511	0.03022	0.24142	40.00	30.77	57.14	−0.00017	−8.62E−06	0.00399	0.00491
10	Hybrid Stacked	120.280	−0.00502	0.01553	39.609	0.01855	0.27380	0.24057	71.43	91.67	44.44	0.00132	6.27E−05	0.00424	−0.00873
	XGBoost	84.242	−0.00360	0.01275	66.572	0.01584	0.41105	0.19748	71.43	91.67	44.44	0.00104	4.97E−05	0.00570	−0.01341
	LSTM	98.079	−0.00279	0.01420	98.079	0.01715	−0.21505	0.22002	57.14	100.00	0.00	9.03E−05	4.30E−06	0.00115	
	GRU	98.479	−0.00131	0.01437	98.479	0.01692	0.24159	0.22259	57.14	100.00	0.00	7.34E−05	3.50E−06	0.00115	
	LightGBM	132.303	−0.00654	0.01642	18.381	0.01931	0.30909	0.25432	71.43	75.00	66.67	0.00169	8.03E−05	0.00496	−0.00394
	Random Forest	92.522	−0.00399	0.01430	85.220	0.01809	−0.08802	0.22159	61.90	100.00	11.11	6.10E−06	2.90E−07	0.00170	−0.00992
11	Hybrid Stacked	386.748	0.01089	0.02829	34.503	0.03246	0.21257	0.26570	47.37	41.67	57.14	0.00200	1.05E−04	0.01414	0.00427
	XGBoost	193.449	0.00528	0.02157	69.999	0.02626	0.23194	0.20261	52.63	50.00	57.14	0.00218	1.15E−04	0.01296	0.00435
	LSTM	102.375	0.00679	0.01879	102.375	0.02534	0.22129	0.17651	63.16	100.00	0.00	2.71E−04	1.42E−05	0.00843	
	GRU	115.453	0.01166	0.02032	97.245	0.02715	−0.19077	0.19087	36.84	0.00	100.00	−0.00056	−2.93E−05		0.00843
	LightGBM	369.853	0.01221	0.03140	1.638	0.03733	−0.00937	0.29494	52.63	50.00	57.14	−0.00072	−3.77E−05	0.01296	0.00435
	Random Forest	158.395	0.00716	0.02031	78.007	0.02495	0.29565	0.19073	57.89	58.33	57.14	0.00190	1.00E−04	0.01649	−0.00053
12	Hybrid Stacked	329.981	0.02128	0.03634	322.471	0.05045	−0.17050	0.26136	53.33	60.00	40.00	−0.00316	−2.11E−04	0.00916	0.02710
	XGBoost	240.144	0.01789	0.03163	238.307	0.04503	−0.23916	0.22749	40.00	50.00	20.00	−0.00175	−1.17E−04	0.00684	0.03057
	LSTM	100.479	0.01623	0.02852	100.479	0.04094	0.29090	0.20516	60.00	70.00	40.00	1.07E−04	7.16E−06	0.01970	0.00959
	GRU	111.394	0.01804	0.02963	104.833	0.04219	−0.37294	0.21310	26.67	0.00	80.00	−0.00065	−4.30E−05	−0.03139	0.01974
	LightGBM	356.158	0.01836	0.03653	210.439	0.04965	−0.37176	0.26273	53.33	50.00	60.00	−0.00386	−2.58E−04	0.00894	0.02280
	Random Forest	236.693	0.02038	0.03235	236.693	0.04498	−0.19616	0.23266	33.33	30.00	40.00	−0.00189	−1.26E−04	0.00666	0.02278

Hybrid Stacked model shows competitive performance in error metrics for certain months but demonstrates variability across the dataset. For example, in October, the model achieves a relatively low MAPE of 0.01553 and RMSE of 0.01855, reflecting its ability to capture price movements accurately. However, in November, its MAPE increases to 0.02829 and RMSE to 0.03246, suggesting sensitivity to monthly volatility.

The XGBoost model exhibits consistent performance with low MAPE values across multiple months. In October, it records a MAPE of 0.01275 and an RMSE of 0.01584, the best among all models for that month. This consistency underscores its robustness in handling noisy and volatile market conditions.

The LSTM and GRU models excel in minimizing short-term prediction errors, with LSTM achieving a low MAPE of 0.02151 and RMSE of 0.02680 in February. Similarly, GRU performs well in March with a MAPE of 0.03691. However, these models display occasional spikes in error metrics, such as in July, where LSTM records a higher MAPE of 0.01588.

The LightGBM model performs well in specific months but struggles with error metric variability. For instance, in December, it achieves a MAPE of 0.03653 and RMSE of 0.04965, reflecting difficulties in capturing extreme fluctuations. Random Forest shows moderate performance, achieving a low MAPE of 0.01430 in October but experiencing higher error values in months like September.

Directional accuracy metrics provide further insights into the models' predictive robustness. LSTM and GRU demonstrate strong performance in predicting upward trends, achieving MDA+ of 100 % in February and July. However, their downward trend predictions (MDA-) remain inconsistent, with GRU recording as low as 11.11 % in May.

XGBoost and LightGBM show balanced directional accuracy across months. In December, XGBoost achieves an MDA of 50.00 %, while LightGBM records 53.33 %, reflecting their effectiveness in predicting both upward and downward trends. Random Forest achieves a high MDA of 61.90 % in August, highlighting its strength in relatively stable markets.

Financial metrics, including Return Long and Return Short, further contextualize the trading performance of the models. The Hybrid Stacked model records competitive returns in specific months, such as November, where it achieves a Return Long of 0.01414. However, it shows negative returns in volatile months, such as March. XGBoost delivers consistent returns across months, achieving a Return Short of 0.00678 in September. LightGBM performs well in some months, with a Return Short of 0.01115 in September.

Random Forest demonstrates resilience in financial performance, particularly in upward market conditions. In October, it records a Return Long of 0.00170 while maintaining a high MDA of 61.90 %. However, its returns fluctuate across months, reflecting sensitivity to volatility.

Fig. 6 illustrates the cumulative returns for Month 3 (March) of the Binance dataset, highlighting the performance of the models compared to true returns. The Hybrid Stacked model shows strong alignment with the true returns, outperforming other models in capturing cumulative returns during this month. XGBoost and LightGBM follow closely, demonstrating relatively consistent performance. In contrast, Random Forest and GRU exhibit more significant deviations from true returns, reflecting challenges in adapting to the month's volatility. This visualization complements the numerical analysis by showcasing the Hybrid Stacked model's capability to track market trends effectively.

While the LSTM and GRU models minimize short-term prediction errors and excel in upward trend predictions, they exhibit directional biases. XGBoost and LightGBM provide balanced and consistent performance across metrics and months. The Hybrid Stacked model demonstrates strong financial returns but is more sensitive to data variability, and Random Forest shows moderate performance, particularly in stable markets.

4.2.2. Evaluation on Bitcoin monthly dataset

The Hybrid Stacked model consistently demonstrates competitive performance across months, as highlighted in Table 8. For instance,
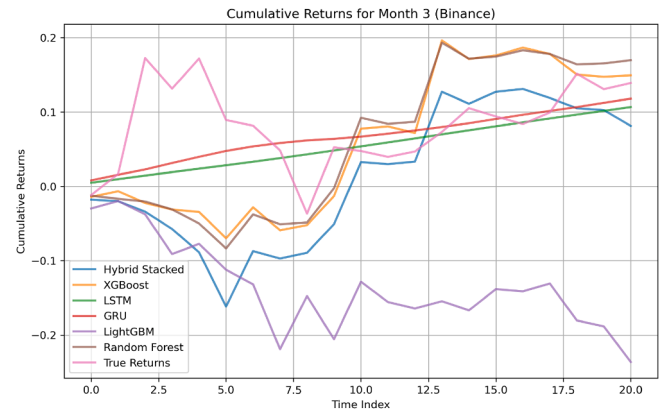

Fig. 6. Cumulative Returns for Month 3 (Binance Dataset). The figure shows the cumulative returns of various models compared to true returns, with the Hybrid Stacked model aligning closely to market trends, highlighting its predictive accuracy.
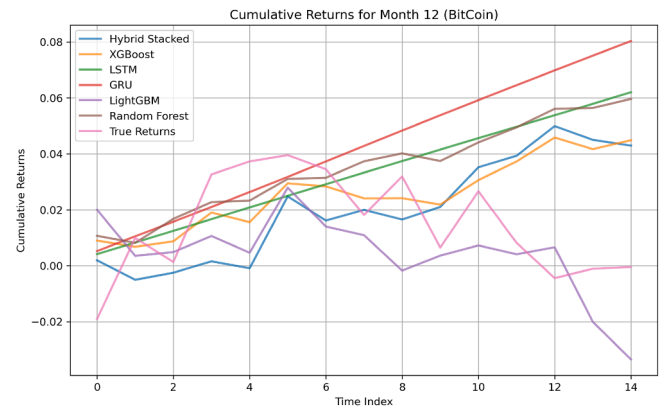

Fig. 7. Cumulative Returns for Month 12 (Bitcoin Dataset). The figure highlights the cumulative returns achieved by various models during Month 12. GRU exhibits the highest cumulative returns with a near-linear trajectory, while the Hybrid Stacked model closely follows market trends but slightly lags in performance.

it achieves strong directional accuracy, such as an MDA of 73.33 % in Month 1 and 65 % in Month 4, while delivering robust returns in certain months. However, the model exhibits variability, with a lower MDA (33.33 %) in Month 12, indicating sensitivity to monthly market dynamics and volatility. Notably, in Month 1, the Hybrid Stacked model achieves the highest Sharpe Ratio (0.00175) among all models, demonstrating its capability to capture risk-adjusted returns effectively. This performance is also evident in the cumulative returns plot for Month 12 which can be seen in Fig. 7, where the Hybrid Stacked model aligns closely with market trends, though slightly lagging behind GRU's linear trajectory.

The XGBoost model demonstrates consistent error minimization and stable directional accuracy across most months. For instance, it achieves a low RMSE of 0.02295 in Month 1 and maintains an MDA of 70 % in Month 4. However, the model's financial returns, particularly Return Short, are limited in some months, as evidenced by Month 12 in Fig. 7, where its cumulative returns lag behind GRU and LightGBM.

LSTM and GRU models excel in error minimization and upward trend predictions. For example, LSTM records the lowest MAPE (96.566) in Month 5, while GRU achieves an RMSE of 0.01757 in Month 11. Their upward directional accuracy (MDA+) is consistently high, reaching 100 % in several months, including Month 1 and Month 11. However, their ability to predict downward trends (MDA-) is often negligible or absent, as seen in Table 8, limiting their overall robustness. This behav-

Table 8
Monthly data evaluation metrics for Bitcoin dataset.

Month	Model	MAPE	ME	MAE	MPE	RMSE	R^2	Min-Max	MDA	MDA+	MDA-	SP	Return	Return Long	Return Short
1	Hybrid Stacked	197.172	−0.00496	0.01828	−14.648	0.02436	0.30334	0.18745	73.33	85.71	62.50	0.00175	1.17E−04	0.00471	−0.00911
	XGBoost	196.881	−0.00689	0.01790	78.834	0.02295	0.31846	0.18353	60.00	71.43	50.00	0.00135	9.02E−05	0.00399	−0.00804
	LSTM	103.722	−0.00225	0.01533	100.966	0.02216	0.03175	0.15719	46.67	100.00	0.00	−1.48E−05	−9.85E−07	−0.00082	
	GRU	195.017	−0.00915	0.01723	89.855	0.02387	0.06224	0.17669	46.67	100.00	0.00	−0.00010	−6.70E−06	−0.00082	
	LightGBM	196.835	−0.00200	0.01727	138.738	0.02200	0.30241	0.17707	60.00	71.43	50.00	0.00127	8.47E−05	0.00358	−0.00742
	Random Forest	151.811	−0.00503	0.01593	73.106	0.02116	0.37818	0.16330	53.33	57.14	50.00	0.00130	8.65E−05	0.00365	−0.00592
2	Hybrid Stacked	163.155	0.02337	0.02837	66.600	0.04245	−0.21707	0.25413	50.00	36.36	71.43	−0.00421	−2.34E−04	0.00878	0.01547
	XGBoost	134.566	0.02010	0.02507	70.127	0.03669	−0.11688	0.22456	55.56	45.45	71.43	−0.00220	−1.22E−04	0.01098	0.01467
	LSTM	97.209	0.01040	0.02040	88.311	0.02976	0.41008	0.18274	61.11	100.00	0.00	7.56E−04	4.20E−05	0.01324	
	GRU	186.780	0.00080	0.02191	43.937	0.02803	0.03651	0.19627	61.11	100.00	0.00	0.00297	1.65E−04	0.01324	
	LightGBM	259.172	0.02383	0.02964	146.064	0.03875	0.14253	0.26558	44.44	27.27	71.43	−0.00131	−7.30E−05	0.01447	0.01276
	Random Forest	174.124	0.01912	0.02601	45.021	0.03692	−0.24793	0.23302	38.89	18.18	71.43	−0.00255	−1.42E−04	0.00205	0.01643
3	Hybrid Stacked	202.461	0.00284	0.03612	18.712	0.04899	−0.01439	0.20186	47.62	40.00	54.55	−0.00038	−1.79E−05	−0.01090	0.01087
	XGBoost	293.360	0.01508	0.04415	118.878	0.08106	0.03282	0.24670	66.67	50.00	81.82	0.00149	7.10E−05	0.01243	−0.00391
	LSTM	132.652	−0.00678	0.03210	105.488	0.04035	0.16954	0.17939	47.62	100.00	0.00	3.06E−04	1.46E−05	0.00154	
	GRU	97.706	0.00381	0.03097	95.062	0.03990	0.11007	0.17307	52.38	0.00	100.00	2.06E−05	9.80E−07		0.00154
	LightGBM	151.585	0.00665	0.03623	46.691	0.05217	−0.28191	0.20248	61.90	52.38	40.00	−0.00575	−2.74E−04	−0.00825	0.00757
	Random Forest	182.656	−0.00084	0.03535	65.380	0.04677	0.05685	0.19753	61.90	60.00	63.64	0.00136	6.45E−05	0.00332	−7.32E−05
4	Hybrid Stacked	414.171	−0.01117	0.02310	−159.517	0.02703	0.41176	0.27870	65.00	66.67	63.64	0.00361	1.81E−04	0.00285	−0.01773
	XGBoost	374.116	−0.01047	0.01932	−192.231	0.02252	0.56720	0.23312	70.00	77.78	63.64	0.00334	1.67E−04	0.00337	−0.02066
	LSTM	106.062	−0.00945	0.02015	79.700	0.02592	0.43088	0.24312	45.00	100.00	0.00	−0.00026	−1.31E−05	−0.00744	
	GRU	97.493	−0.00748	0.01964	97.493	0.02519	0.23359	0.23694	65.00	66.67	63.64	7.15E−05	3.58E−06	−0.00345	−0.01144
	LightGBM	235.530	−0.00900	0.02163	4.970	0.02554	0.35413	0.26098	70.00	77.78	63.64	0.00255	1.27E−04	0.00471	−0.02230
	Random Forest	236.871	−0.00934	0.02010	−10.990	0.02319	0.48244	0.24254	70.00	66.67	72.73	0.00224	1.12E−04	0.00499	−0.01762
5	Hybrid Stacked	195.935	0.01670	0.02509	130.551	0.03484	0.19237	0.26469	52.38	33.33	66.67	0.00107	5.10E−05	0.01239	0.00125
	XGBoost	133.317	0.01537	0.02183	102.165	0.03225	0.11572	0.23030	52.38	33.33	66.67	−0.00016	−7.41E−06	0.01074	0.00207
	LSTM	96.566	0.00785	0.01865	78.844	0.02708	0.19084	0.19683	57.14	0.00	100.00	−0.00024	−1.17E−05		0.00496
	GRU	97.904	0.00561	0.01899	97.904	0.02672	−0.19485	0.20041	38.10	0.00	66.67	−0.00012	−5.88E−06	−0.01193	0.00894
	LightGBM	253.863	0.01897	0.02898	153.708	0.03618	0.30423	0.30575	42.86	22.22	58.33	0.00290	1.38E−04	0.00706	0.00391
	Random Forest	115.861	0.01184	0.02047	108.255	0.03065	−0.00032	0.21600	52.38	33.33	66.67	−0.00072	−3.42E−05	0.00945	0.00272
6	Hybrid Stacked	289.656	−0.01200	0.02229	183.259	0.02675	0.14639	0.28769	55.00	75.00	41.67	0.00028	1.39E−05	−0.00045	−0.01357
	XGBoost	212.023	−0.00264	0.01899	194.374	0.02440	−0.19998	0.24508	45.00	50.00	41.67	−0.00068	−3.38E−05	−0.00586	−0.00405
	LSTM	162.021	0.00223	0.01534	65.573	0.01855	0.29947	0.19799	60.00	0.00	100.00	0.00076	3.82E−05		−0.00504
	GRU	225.734	0.00699	0.01600	41.531	0.01974	0.22019	0.20643	60.00	0.00	100.00	0.00123	6.14E−05		−0.00504
	LightGBM	308.298	−0.01266	0.02321	131.059	0.02966	0.14301	0.29959	45.00	62.50	33.33	0.00041	2.04E−05	−0.00488	−0.00535
	Random Forest	149.850	−0.00343	0.01667	144.803	0.02101	−0.24962	0.21517	35.00	25.00	41.67	−0.00038	−1.88E−05	−0.01218	0.00079
7	Hybrid Stacked	218.630	−0.00088	0.02634	146.044	0.03210	−0.51512	0.30381	33.33	33.33	33.33	−0.00281	−1.34E−04	−0.00213	0.01226
	XGBoost	145.174	0.00146	0.02055	122.790	0.02630	−0.27562	0.23709	42.86	41.67	44.44	−0.00082	−3.91E−05	0.00323	0.00739
	LSTM	100.382	0.00540	0.01660	100.382	0.02223	−0.10636	0.19150	42.86	41.67	44.44	−5.03E−06	−2.39E−07	0.00143	0.00904
	GRU	104.671	0.00742	0.01684	104.671	0.02279	0.02611	0.19429	42.86	0.00	100.00	−0.00023	−1.08E−05		0.00541
	LightGBM	274.278	−0.00163	0.02919	114.661	0.03494	−0.57598	0.33676	38.10	50.00	22.22	−0.00382	−1.82E−04	−0.00239	0.01809
	Random Forest	126.097	0.00198	0.01869	117.754	0.02381	−0.28267	0.21555	38.10	50.00	22.22	−0.00032	−1.53E−05	0.00365	0.00827
8	Hybrid Stacked	186.778	−0.00041	0.02217	159.219	0.02840	−0.06035	0.19654	47.62	50.00	45.45	−0.00037	−1.78E−05	−0.00538	0.00273
	XGBoost	181.031	0.00086	0.02037	169.373	0.02619	−0.00488	0.18062	38.10	30.00	45.45	5.20E−05	2.47E−06	−0.00815	0.00346
	LSTM	103.125	0.00136	0.01856	103.125	0.02445	−0.20969	0.16453	52.38	0.00	100.00	5.07E−05	2.41E−06		−0.00152
	GRU	107.976	0.00240	0.01854	105.612	0.02440	0.12398	0.16438	52.38	0.00	100.00	0.00015	7.26E−06		−0.00152
	LightGBM	160.754	−0.00298	0.02078	114.848	0.02674	0.05432	0.18420	61.90	60.00	63.64	0.00029	1.38E−05	0.00411	−0.00663
	Random Forest	174.093	−0.00011	0.02101	162.594	0.02651	−0.03573	0.18627	42.86	40.00	45.45	−0.00013	−6.29E−06	−0.00444	0.00114
9	Hybrid Stacked	234.265	0.01286	0.02145	74.693	0.02676	0.24537	0.27379	55.00	50.00	62.50	0.00102	5.12E−05	0.00942	0.00081
	XGBoost	143.553	0.00851	0.01834	52.140	0.02471	0.21313	0.23412	70.00	83.33	50.00	0.00110	5.48E−05	0.00765	−0.00223
	LSTM	98.389	0.00403	0.01528	98.389	0.01943	0.06372	0.19509	60.00	100.00	0.00	6.51E−05	3.26E−06	0.00469	
	GRU	98.201	0.00421	0.01533	98.201	0.01951	−0.09866	0.19565	60.00	91.67	12.50	3.17E−05	1.58E−06	0.00418	0.00925
	LightGBM	240.776	0.01231	0.02234	89.062	0.02644	0.11973	0.28599	60.00	50.00	75.00	1.88E−05	9.38E−07	0.01202	−0.00021
	Random Forest	152.359	0.00854	0.01864	32.367	0.02639	0.09939	0.23801	65.00	75.00	50.00	0.00033	1.63E−05	0.00806	−0.00158
10	Hybrid Stacked	333.256	−0.00362	0.02214	283.411	0.02667	−0.28305	0.28972	52.38	83.33	11.11	0.00026	1.22E−05	0.00642	0.01380
	XGBoost	261.605	−0.00183	0.01882	233.173	0.02286	−0.11220	0.24619	52.38	83.33	11.11	0.00104	4.97E−05	0.00746	0.00751
	LSTM	202.897	−0.00126	0.01599	200.390	0.02012	−0.13122	0.20926	57.14	100.00	0.00	0.00136	6.48E−05	0.00747	
	GRU	209.800	−0.00188	0.01599	205.717	0.02013	0.27375	0.20923	57.14	100.00	0.00	0.00147	7.02E−05	0.00747	
	LightGBM	245.819	−0.00004	0.01726	243.373	0.02130	0.17614	0.22582	42.86	75.00	0.00	0.00203	9.68E−05	0.00635	0.01418
	Random Forest	241.549	−0.00250	0.01982	224.717	0.02440	−0.40458	0.25935	52.38	91.67	0.00	0.00024	1.14E−05	0.00720	0.01282
11	Hybrid Stacked	434.529	0.00523	0.02118	376.791	0.02811	−0.24511	0.25653	42.86	44.44	40.00	−0.00115	−8.18E−05	0.00479	0.00097
	XGBoost	286.641	0.00201	0.01613	272.653	0.02105	−0.02199	0.19537	35.71	33.33	40.00	−0.00036	−2.45E−05	0.00374	0.00224
	LSTM	100.324	0.00066	0.01148	61.942	0.01753	0.69450	0.13899	64.29	100.00	0.00	9.14E−05	6.53E−06	0.00288	
	GRU	89.495	0.00228	0.01149	89.495	0.01757	0.68502	0.13915	64.29	100.00	0.00	4.86E−05	3.47E−06	0.00288	
	LightGBM	429.574	0.00716	0.02065	283.337	0.02626	−0.21740	0.25010	35.71	22.22	60.00	−0.00096	−6.87E−05	0.00144	0.00346
	Random Forest	376.615	0.00436	0.01747	313.145	0.02266	−0.05816	0.21160	42.86	33.33	60.00	−0.00024	−1.72E−05	0.00612	0.00108
12	Hybrid Stacked	186.204	−0.00114	0.01684	101.485	0.01930	−0.08294	0.29690	33.33	37.50	28.57	−0.00017	−1.10E−05	−0.00267	0.00298
	XGBoost	166.218	−0.00302	0.01540	39.808	0.01763	0.10520	0.27152	53.33	62.50	42.86	0.00016	1.03E−05	0.00104	−0.00164
	LSTM	241.584	−0.00960	0.01704	−42.751	0.01948	0.24670	0.30034	53.33	100.00	0.00	2.07E−06	1.38E−07	−3.24E−05	
	GRU	244.693	−0.00973	0.01715	−44.481	0.01957	−0.00107	0.30234	53.33	100.00	0.00	−4.75E−06	−3.16E−07	−3.24E−05	
	LightGBM	387.330	0.00220	0.02144	240.857	0.02396	−0.25792	0.37802	40.00	37.50	42.86	−0.00085	−5.66E−05	−0.00171	0.00143
	Random Forest	144.332	−0.00401	0.01579	53.778	0.01853	−0.17650	0.27829	53.33	87.50	14.29	−0.00018	−1.19E−05	−0.00031	0.00177

ior is reflected in Fig. 7, where GRU's linear cumulative returns deviate from the true returns, demonstrating limited adaptability to market fluctuations.

Random Forest demonstrates moderate directional accuracy and error performance across months but struggles with financial metrics and variance explained (R^2). For instance, it achieves an MDA of 70 % in Month 4, yet its cumulative returns, as seen in Fig. 7, remain subdued compared to other models. This highlights the model's limited adaptability to the nonlinear complexities of cryptocurrency markets.

Overall, the monthly evaluation highlights the detailed strengths and limitations of each model. While LSTM and GRU excel in error minimization and upward trend predictions, their inability to capture downward trends hinders financial performance. Hybrid Stacked and LightGBM balance directional accuracy and financial returns effectively but exhibit sensitivity to monthly volatility. XGBoost provides consistent error minimization and directional accuracy, making it a reliable alternative, though its financial metrics require further optimization. Random Forest, while steady, lacks adaptability to volatile cryptocurrency price patterns.

4.2.3. Evaluation on Ethereum monthly dataset

Table 9 provides a detailed evaluation of the performance of various models on the Ethereum (ETH) dataset using monthly data. The Hybrid Stacked model demonstrates consistently strong performance across most metrics and months, effectively leveraging its ensemble structure to integrate the strengths of individual models. For instance, in Month 1, it achieves a solid Mean Directional Accuracy (MDA) of 68.75 %, indicative of reliable directional predictions, despite recording a slightly higher Mean Error (ME) of −0.00719. In Month 5, the Hybrid Stacked model significantly outperforms other models in cumulative returns, closely tracking the true returns and capturing the sharp upward trend during this period, showcasing its robustness in volatile conditions.

XGBoost consistently excels across multiple metrics, particularly in error-based measures and directional accuracy. In Month 1, it achieves a low Mean Absolute Percentage Error (MAPE) of 126.805 and an MAE of 0.01593, paired with an impressive MDA of 75 %. However, in Month 7, XGBoost exhibits a decline in performance, with an increase in MAPE to 165.556 and a decrease in cumulative returns, reflecting its sensitivity to market shifts during this period. Despite this, XGBoost remains a strong contender, particularly in stable market scenarios where it provides precise predictions and reliable directional performance.

Deep learning models, such as LSTM and GRU, exhibit strengths in capturing temporal dependencies but are inconsistent in directional accuracy across months. For example, in Month 3, LSTM records an RMSE of 0.04550, highlighting its ability to adapt to volatile conditions. However, its Mean Directional Accuracy (MDA) drops to 47.62 %, indicating limited success in predicting directional trends. GRU shows strong directional accuracy in certain months, such as an MDA+ of 100.00 % in Month 4, but struggles in others, like Month 8, where its cumulative returns lag behind those of the Hybrid Stacked model and XGBoost.

Tree-based models, LightGBM and Random Forest, show variable performance depending on market conditions. In Month 1, LightGBM achieves a high Min-Max Accuracy of 0.34763, reflecting strong alignment with observed data. However, in Month 6, it underperforms with a higher MAPE of 449.229 and a negative R^2, indicating difficulties in adapting to market volatility. Random Forest, on the other hand, consistently balances accuracy and stability. In Month 1, it achieves a low MAPE of 93.692 and an MDA of 56.25 %. However, its performance in Month 8 highlights its limitations, with cumulative returns falling significantly below those of the Hybrid Stacked model and XGBoost.

The cumulative returns for Month 5, visualized in Fig. 8, further emphasize the Hybrid Stacked model's superior ability to adapt to sharp market movements, outperforming other models in both upward and downward trends. While XGBoost tracks closely, Random Forest and
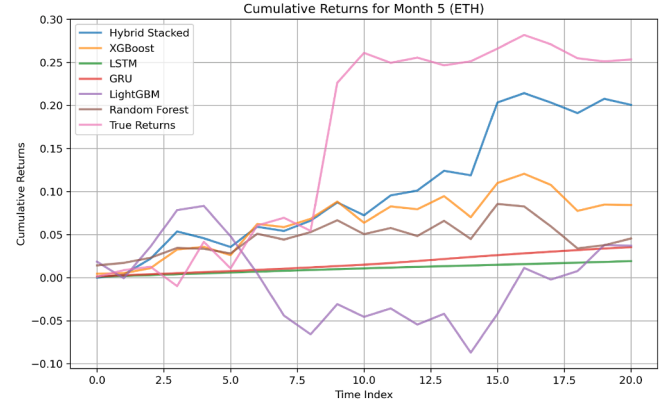

Fig. 8. Cumulative Returns for Month 5 (ETH Dataset). The Hybrid Stacked model outperforms other models by closely tracking true returns and achieving higher cumulative returns, demonstrating its robustness in capturing market trends.

LSTM exhibit moderate returns with fluctuations, and LightGBM shows notable underperformance due to its sensitivity to market volatility.

Overall, the results in Table 9 highlight the complementary strengths and weaknesses of different models across various metrics and months. Machine learning models such as XGBoost and Random Forest excel in error-based metrics and directional accuracy during stable market conditions, while deep learning models such as LSTM and GRU are better suited for capturing complex temporal dependencies in volatile scenarios. The Hybrid Stacked model integrates these strengths, demonstrating consistent and robust performance across diverse evaluation metrics. The monthly variations underscore the importance of adaptive modeling strategies in dynamic cryptocurrency markets. Ensemble approaches, such as the Hybrid Stacked model, prove particularly effective in achieving reliable and actionable predictions in these complex and volatile environments.

We compute $R^2_{\text{oos}} = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y}_{\text{test}})^2}$. Thus $R^2_{\text{oos}} < 0$ indicates the model underperforms the test-set mean predictor on that window. This can occur in short, high-volatility segments of return forecasting where $\bar{y}_{\text{test}} \approx 0$ (small denominator) and distribution shift increases level error. We therefore report R^2 for completeness but focus interpretation on RMSE/MAE and directional metrics (DA, MDA + , MDA -), which remain most relevant for trading decisions.

We clarify that references to CAPM were intended as *exploratory* rather than definitive. Some studies report partial links between crypto returns and systematic risk proxies, though evidence is mixed and regime-dependent ?.

4.3. Discussion

The results demonstrate that the Hybrid Stacked model consistently outperforms standalone models across multiple evaluation metrics, establishing it as the most balanced predictive framework in this study. These findings align with Pourrezaee and Hajizadeh (2024) and Ramos-Pérez et al. (2019), who applied the Hybrid Stacked model to Bitcoin volatility and S&P 500 volatility, respectively. However, our study differs in several key aspects.

First, we incorporate a comprehensive set of technical indicators and economic factors to assess model performance, offering a broader perspective on market dynamics. Second, we employ rigorous feature engineering to capture complex market patterns more effectively. Additionally, we conduct two separate experiments: one focusing on short-term trends through monthly performance analysis and another utilizing a cross-fold evaluation technique for long-term validation. Finally, by analyzing three different cryptocurrencies, we enhance the robustness and generalizability of our results.

Table 9
Monthly data evaluation metrics for ETH dataset.

Month	Model	MAPE	ME	MAE	MPE	RMSE	R^2	Min-Max	MDA	MDA+	MDA-	SP	Return	Return Long	Return Short
1	Hybrid Stacked	210.140	−0.00719	0.02075	−26.731	0.02673	0.33383	0.22777	68.75	66.67	70.00	0.00253	0.00016	−0.00169	−0.00926
	XGBoost	126.805	−0.00690	0.01593	4.7444	0.02251	0.38983	0.17487	75.00	83.33	70.00	0.00169	0.00011	0.00427	−0.01616
	LSTM	98.084	−0.00322	0.01794	86.050	0.02320	−0.10861	0.19694	62.50	0.00	100.00	0.00025	1.58E−05		−0.00594
	GRU	105.962	−0.00659	0.01861	105.962	0.02382	0.11806	0.20427	43.75	83.33	20.00	−3.58E−05	−2.24E−06	−0.00587	−0.00628
	LightGBM	524.091	−0.00852	0.03166	−105.872	0.03460	0.32462	0.34763	56.25	66.67	50.00	0.00369	0.00023	−0.00058	−0.01284
	Random Forest	93.692	−0.00732	0.01628	65.864	0.02355	0.29009	0.17870	56.25	66.67	50.00	0.00104	6.50E−05	−0.00690	−0.00471
2	Hybrid Stacked	535.978	0.02198	0.02747	−128.453	0.03294	0.22449	0.37427	42.11	25.00	71.43	−0.00033	−1.73E−05	0.01297	0.01609
	XGBoost	317.197	0.01784	0.02320	−10.819	0.02699	0.36840	0.31603	47.37	41.67	57.14	0.00103	5.42E−05	0.01727	0.01382
	LSTM	100.755	0.01708	0.02143	86.239	0.02728	0.37064	0.29195	36.84	0.00	100.00	−0.00048	−2.53E−05		0.01527
	GRU	115.637	0.01266	0.02011	113.074	0.02470	0.31879	0.27403	63.16	100.00	0.00	0.00083	4.36E−05	0.01527	
	LightGBM	430.407	0.01726	0.02411	−202.521	0.02968	0.28117	0.32853	68.42	66.67	71.43	0.00156	8.24E−05	0.02600	0.00335
	Random Forest	178.059	0.01659	0.02130	50.591	0.02694	0.18170	0.29027	47.37	41.67	57.14	0.00012	6.45E−06	0.01231	0.01742
3	Hybrid Stacked	590.015	−0.00199	0.04161	283.448	0.05251	0.05111	0.19794	52.38	40.00	63.64	0.00153	7.28E−05	−0.00139	−0.00407
	XGBoost	449.366	−0.00543	0.03401	284.132	0.04453	0.20586	0.16178	57.14	60.00	54.55	0.00348	0.00017	−0.00056	−0.00578
	LSTM	240.715	−0.01197	0.03510	163.685	0.04550	0.14636	0.16696	47.62	100.00	0.00	−0.00048	−2.29E−05	−0.00305	
	GRU	107.807	−0.00206	0.03391	90.909	0.04398	0.30049	0.16131	52.38	0.00	100.00	0.00011	5.27E−06		−0.00305
	LightGBM	423.190	0.00146	0.04225	407.601	0.05414	−0.17879	0.20098	28.57	20.00	36.36	−0.00378	−0.00018	−0.02601	0.01418
	Random Forest	417.936	−0.00888	0.03547	170.108	0.04502	0.16821	0.16874	61.90	70.00	54.55	0.00206	9.79E−05	0.00261	−0.01059
4	Hybrid Stacked	2758.524	−0.01184	0.02327	−2496.664	0.03043	0.53161	0.18933	60.00	55.56	63.64	0.00465	0.00023	0.00702	−0.02021
	XGBoost	1981.064	−0.00869	0.02240	−1527.987	0.02867	0.60937	0.18228	65.00	66.67	63.64	0.00476	0.00024	0.00797	−0.02388
	LSTM	288.302	−0.00661	0.02439	288.302	0.03355	0.34709	0.19846	55.00	0.00	100.00	0.00026	1.32E−05		−0.00796
	GRU	287.316	−0.00908	0.02495	−97.808	0.03426	−0.16239	0.20295	45.00	100.00	0.00	−0.00022	−1.12E−05	−0.00796	
	LightGBM	1927.099	−0.01197	0.02544	1917.716	0.03434	0.24585	0.20694	40.00	55.56	27.27	0.00130	6.50E−05	−0.00681	−0.01010
	Random Forest	2563.431	−0.00884	0.02453	−2387.077	0.03139	0.40924	0.19959	65.00	66.67	63.64	0.00307	0.00015	0.00646	−0.02238
5	Hybrid Stacked	279.181	0.00138	0.03150	53.709	0.04535	0.14821	0.15536	47.62	53.85	37.50	0.00602	0.00029	0.02055	0.00074
	XGBoost	165.556	0.00805	0.02579	100.782	0.04131	0.26092	0.12716	52.38	61.54	37.50	0.00507	0.00024	0.01956	−0.00013
	LSTM	104.609	0.01240	0.02360	104.609	0.04317	−0.11467	0.11637	33.33	7.69	75.00	−0.00011	−5.23E−06	−0.00484	0.01488
	GRU	106.615	0.01161	0.02360	106.615	0.04300	−0.04813	0.11637	42.86	53.85	25.00	0.00006	2.85E−06	0.01258	0.01123
	LightGBM	380.372	0.01029	0.03749	83.980	0.04874	0.15173	0.18490	42.86	46.15	37.50	0.00450	0.00021	0.01783	0.00572
	Random Forest	218.128	0.00990	0.02729	29.858	0.04211	0.21603	0.13460	47.62	53.85	37.50	0.00351	0.00017	0.01578	0.00710
6	Hybrid Stacked	185.951	−0.00257	0.01788	−3.760	0.02244	0.15839	0.26700	55.00	55.56	54.55	0.00094	4.70E−05	−0.00185	−0.00459
	XGBoost	155.050	5.60E−06	0.01699	32.063	0.02160	0.12399	0.25369	55.00	44.44	63.64	0.00081	4.03E−05	−0.00418	−0.00258
	LSTM	102.217	−0.00302	0.01669	102.217	0.01948	−0.53082	0.24934	55.00	0.00	100.00	−3.39E−06	−1.70E−07		−0.00322
	GRU	105.680	−0.00403	0.01687	87.107	0.01968	−0.32101	0.25191	45.00	100.00	0.00	−7.66E−05	−3.83E−06	−0.00322	
	LightGBM	449.229	0.00563	0.02221	434.561	0.02783	0.00091	0.33174	40.00	11.11	63.64	0.00064	3.18E−05	−0.00298	−0.00330
	Random Forest	95.385	−0.00106	0.01510	51.160	0.01876	0.26012	0.22556	65.00	44.44	81.82	0.00091	4.53E−05	0.00324	−0.00599
7	Hybrid Stacked	603.4262	−0.00702	0.02873	−263.9691	0.03557	−0.08100	0.24348	52.38	61.54	37.5	−0.00059	−2.83E−05	0.00246	0.00098
	XGBoost	130.3605	−0.00899	0.02209	73.5933	0.02901	0.11360	0.18718	61.90	76.92	37.5	0.00131	6.22E−05	0.00674	−0.01021
	LSTM	171.2633	−0.00287	0.01868	−5.3401	0.02529	0.11028	0.15833	61.90	100.00	0.00	0.00020	9.75E−06	0.00190	
	GRU	140.0564	−0.00102	0.01886	32.6131	0.02515	0.06447	0.15984	61.90	100.00	0.00	0.00013	6.13E−06	0.00190	
	LightGBM	787.5314	−0.00800	0.03165	−395.4503	0.03959	−0.30672	0.26825	47.62	61.54	25.0	−0.00003	−1.57E−04	−0.00069	0.00707
	Random Forest	244.2596	−0.00794	0.01979	−69.9672	0.02590	0.25278	0.16777	76.19	100.00	37.5	0.00176	8.40E−05	0.00833	−0.03670
8	Hybrid Stacked	7014.8756	−0.00049	0.02858	6473.2855	0.03731	−0.14756	0.19233	47.62	55.56	41.67	−0.00169	−8.04E−05	−0.00569	0.00346
	XGBoost	5161.4679	−0.00120	0.02783	4690.1987	0.03758	−0.24868	0.18726	47.62	55.56	41.67	−0.00262	−1.25E−04	−0.00667	0.00477
	LSTM	390.4140	−0.00065	0.01951	−207.3019	0.02952	−0.24924	0.13130	57.14	0.00	100.00	−3.12E−05	−1.49E−06		−0.00177
	GRU	1609.2614	0.00449	0.02001	−1363.9301	0.02957	0.35669	0.13467	57.14	0.00	100.00	0.00033	1.59E−05		−0.00177
	LightGBM	6313.4627	−0.00466	0.02112	6273.9512	0.02962	0.29536	0.14212	52.38	77.78	33.33	0.00297	1.42E−04	0.00199	−0.01117
	Random Forest	3945.7315	−0.00219	0.02549	3753.5579	0.03428	−0.22807	0.17152	33.33	33.33	33.33	−0.00171	−8.15E−05	−0.01322	0.01083
9	Hybrid Stacked	165.4852	0.01424	0.02888	148.5877	0.03706	−0.04397	0.34209	40.00	36.36	44.44	−0.00133	−6.64E−05	−0.00038	0.00782
	XGBoost	137.5095	0.01011	0.02485	125.3672	0.03036	0.02585	0.29435	40.00	36.36	44.44	−0.00029	−1.43E−05	−0.00201	0.00916
	LSTM	98.3380	0.00285	0.02021	98.3380	0.02331	0.05450	0.23933	55.00	100.00	0.00	0.00011	5.73E−06	0.00413	
	GRU	99.9769	0.00395	0.02028	99.9769	0.02343	0.16449	0.24015	50.00	72.73	22.22	4.12E−05	2.06E−06	0.00496	0.00165
	LightGBM	187.3419	0.00666	0.02942	50.8747	0.03735	−0.25005	0.34842	45.00	54.55	33.33	−0.00291	−1.46E−04	0.00092	0.00895
	Random Forest	129.0456	0.00755	0.02392	117.5008	0.02829	0.05031	0.28324	30.00	36.36	22.22	8.13E−05	4.07E−06	−0.00167	0.01122
10	Hybrid Stacked	176.78	−0.0117	0.02803	76.3221	0.03398	0.06583	0.23571	42.86	69.23	0.00	0.00148	7.07E−05	−0.00014	0.01443
	XGBoost	154.27	−0.00829	0.02544	84.1513	0.03113	0.04496	0.21396	42.86	69.23	0.00	0.00095	4.54E−05	−2.67E−05	0.01396
	LSTM	97.18	−0.00235	0.02080	83.6069	0.02748	0.14232	0.17493	61.90	100.00	0.00	0.00030	1.42E−05	0.00264	
	GRU	102.31	−0.00263	0.02107	83.6711	0.02760	−0.13791	0.17720	61.90	100.00	0.00	0.00026	1.25E−05	0.00264	
	LightGBM	186.85	−0.00687	0.02501	81.3226	0.03163	−0.05387	0.21034	47.62	69.23	12.50	0.00013	6.18E−06	0.00162	0.00591
	Random Forest	127.32	−0.00767	0.02303	80.8591	0.02964	0.09253	0.19369	47.62	76.92	0.00	0.00116	5.55E−05	0.00111	0.01179
11	Hybrid Stacked	131.61	0.01805	0.04341	85.9656	0.05403	−0.34810	0.27693	36.84	33.33	42.86	−0.00475	−0.00025	0.00521	0.02775
	XGBoost	128.20	0.01448	0.04280	105.0340	0.05183	−0.46914	0.27304	42.11	58.33	14.29	−0.00360	−0.00019	0.01026	0.03559
	LSTM	94.73	0.01158	0.03476	94.7345	0.04382	0.29839	0.22175	63.16	100.00	0.00	0.00234	0.00012	0.01826	
	GRU	94.73	0.01177	0.03487	94.7349	0.04400	−0.16834	0.22243	63.16	100.00	0.00	0.00218	0.00011	0.01826	
	LightGBM	138.26	0.01379	0.03969	77.2312	0.04652	0.13664	0.25322	52.63	66.67	28.57	0.00381	0.00020	0.02016	0.01414
	Random Forest	114.14	0.01433	0.03937	82.6703	0.04808	−0.19767	0.25116	36.84	50.00	14.29	−0.00043	−2.28E−05	0.00834	0.03527
12	Hybrid Stacked	184.36	0.00446	0.02021	136.7550	0.02414	0.35203	0.23791	53.33	66.67	44.44	0.00213	0.00014	0.00739	−0.00718
	XGBoost	228.21	2.13E−05	0.01938	100.1257	0.02363	0.23494	0.22811	53.33	66.67	44.44	0.00111	7.37E−05	0.00802	−0.00814
	LSTM	126.74	−0.00219	0.01764	104.8799	0.02301	0.07530	0.20758	40.00	100.00	0.00	9.56E−05	6.38E−06	0.00156	
	GRU	161.93	−0.00489	0.01867	100.5357	0.02325	0.37857	0.21978	40.00	100.00	0.00	0.00022	1.48E−05	0.00156	
	LightGBM	252.49	0.00019	0.02443	149.6056	0.02947	0.05599	0.28752	40.00	50.00	33.33	0.00041	2.76E−05	0.00143	0.00175
	Random Forest	169.67	0.00151	0.01778	169.0290	0.02341	0.14664	0.20932	46.67	50.00	44.44	0.00045	3.03E−05	0.00816	−0.00598

The application of machine learning and artificial intelligence in financial markets has attracted significant interest from both academia and industry. Gunnarsson et al. (2024) provide a comprehensive review of the literature, highlighting the superior performance of AI and ML models while acknowledging the continued relevance of traditional econometric approaches. Their study also proposes hybrid models as a promising direction for future research.

Follwing their suggestion, in this study, when compared across a range of models, the hybrid-stacked model demonstrates the lowest RMSE and MAE across all datasets. It confirms the ability of this model to minimize prediction errors effectively. In addition, it maintains a competitive MDA +, showing strong directional forecasting capabilities. While LightGBM and XGBoost also perform well in error minimization, they do not match the hybrid-stacked model in financial performance, particularly in cumulative returns.

Among standalone models:

- LightGBM emerges as a strong competitor in predictive accuracy, achieving high R-squared values and stable performance across different datasets. However, it struggles in extreme market conditions.
- XGBoost is the most consistent model in terms of error minimization, with relatively stable month-to-month performance. However, it falls short in tracking market movements as effectively as sequential models like LSTM and GRU.
- LSTM and GRU excel in capturing short-term trends, showing 100 % MDA + in certain folds, but their tendency to overestimate upward trends introduces directional bias. Additionally, their financial performance is inconsistent, particularly in volatile conditions.
- Random Forest performs moderately well in stable periods but struggles to handle nonlinear complexities, leading to suboptimal financial returns.

The strengths and limitations of these models have been widely acknowledged across various domains within finance and economics. These models have been successfully applied to predict key economic indicators such as inflation, mergers and acquisitions, and environmental quality (Cakici & Zaremba, 2024; Ghallabi et al., 2025; Mirza et al., 2024; Zhang et al., 2025b; Zhao et al., 2025) . Their ability to capture complex patterns and enhance predictive accuracy underscores their growing importance in financial decision-making and beyond.

The consistent outperformance of the Hybrid Stacked model in minimizing RMSE and MAE across multiple datasets highlights an important question in financial forecasting: Can financial markets, particularly cryptocurrency markets, be predicted with high accuracy? This question is deeply connected to the EMH, which posits that asset prices fully reflect all available information at any given time. It implies that it is impossible to consistently achieve returns greater than average market returns on a risk-adjusted basis through predictive modeling. According to EMH, any deviation in prices from their intrinsic value should be random and quickly corrected by the market participants, making prediction a highly uncertain endeavor (Fama, 1970).

However, the results of this study suggest that cryptocurrency markets, while volatile, exhibit predictable patterns that some models can capture (Khedr et al., 2021). The ability of the Hybrid Stacked model to minimize errors and generate reliable financial performance challenges the strict interpretation of the EMH, particularly in the case of cryptocurrencies, which are often considered less efficient than traditional markets. The predictive power of models such as XGBoost and the Hybrid Stacked model suggests that, at least in certain conditions, inefficiencies exist within cryptocurrency markets, opening the door for more sophisticated models to exploit these inefficiencies for forecasting.

This aligns with the findings of empirical studies on cryptocurrency markets, where non-random patterns, such as short-term price momentum and volatility clustering, are frequently observed (Chen et al., 2024; Ghosh & Jana, 2023). These findings support the view that cryptocurrency markets may not be perfectly efficient in the sense defined by EMH, particularly in the short term, where factors like investor sentiment, speculative trading, and macroeconomic events play a significant role in price dynamics (Le Tran & Leirvik, 2020; Urquhart, 2016). The performance of the hybrid-stacked model on datasets from Binance, Bitcoin, and Ethereum further illuminate how market inefficiencies can be leveraged for better predictive performance. The fact that the hybrid-stacked model outperforms in terms of both error reduction and cumulative returns, while also handling volatility well, speaks to the Adaptive Market Hypothesis (Lo, 2004). AMH posits that financial markets evolve over time based on the changing strategies of market participants, suggesting that markets are neither completely efficient nor completely inefficient, but adapt to new information, shifting investor behavior, and technological innovations. In cryptocurrency markets, where innovation and investor behavior are rapidly evolving, AMH suggests that predictive models can adapt to market changes. Especially if they integrate learning algorithms capable of identifying complex patterns and adjusting to new data over time. The success of the hybrid stack model in reducing error while maintaining strong financial returns could be seen as a reflection of an adaptive learning process (Chu et al., 2019; Ghazani & Jafari, 2021). furthermore, its superior performance in long-term cumulative returns suggests that while short-term price movements are subject to random noise and market inefficiencies (as seen with models like LSTM and GRU), longer-term trends in cryptocurrency markets may still be subject to broader market forces, such as technological developments, regulation, and macroeconomic factors, which can be predicted with reasonable accuracy (Jiang et al., 2023; Wang et al., 2023). The Capital Asset Pricing Model (CAPM), which helps to estimate the expected return of an asset based on its risk relative to the market, is less applicable in its traditional form to cryptocurrency markets, due to their distinct lack of correlation with broader financial markets. However, the Hybrid Stacked model's superior performance implies that such non-traditional markets may still follow some form of systematic risk that can be measured and predicted, challenging the notion that cryptocurrencies are entirely disconnected from traditional financial frameworks (Dunbar & Owusu-Amoako, 2022; Wang & Chong, 2021). The results from this study suggest that cryptocurrency markets, while exhibiting volatility and inefficiencies, may still follow predictable patterns that can be captured by sophisticated forecasting models (Borri & Shakhnov, 2022; Makarov & Schoar, 2020). The success of the Hybrid Stacked model challenges the strict interpretation of the Efficient Market Hypothesis, particularly in less liquid and more volatile markets like cryptocurrencies. Additionally, the results align with the Adaptive Market Hypothesis and CAPM, suggesting that predictive models must evolve alongside market dynamics to remain effective and the cryptocurrency markets also follow a some form of systematic risk.

5. Conclusion & future work

The comparative analysis of machine learning and deep learning models for cryptocurrency price prediction demonstrates the potential of hybrid stacked models to achieve superior performance across diverse metrics. This study underscores the hybrid model's capacity to effectively integrate the predictive strengths of standalone models, such as XGBoost, LightGBM, LSTM, and GRU, while addressing their individual shortcomings.

The hybrid stacked model exhibited strong performance in both error minimization and financial metrics. For instance, it consistently achieved competitive RMSE and MAE values across datasets, outperforming standalone models such as Random Forest and XGBoost, particularly in volatile market conditions. Notably, the hybrid model aligned closely with true market returns in cumulative returns analyses, reflecting its ability to capture nonlinearmarket dynamics.

However, the study also highlights why certain standalone models faltered in delivering reliable financial returns. Sequential models, such as LSTM and GRU, showed significant bias towards predicting upward trends (MDA + often reaching 100 %) but struggled with downward

trends (MDA- frequently close to 0 %). This directional imbalance limited their applicability in bearish markets. Tree-based models like Random Forest exhibited moderate performance in capturing returns but were hindered by their sensitivity to market volatility, leading to fluctuating directional accuracies.

Beyond methodological contributions, the implications of this research extend into strategic domains. For financial market participants, the hybrid stacked model offers a reliable tool for making informed decisions by providing accurate forecasts and capturing market dynamics effectively. Financial institutions can use these insights to enhance risk management strategies, particularly in the face of cryptocurrency market volatility. Regulators and policymakers can leverage the model's ability to integrate macroeconomic and sentiment features to monitor and assess market stability, supporting the development of regulations that mitigate systemic risks. This predictive capacity is crucial for shaping regulatory frameworks that are adaptive, data-driven, and responsive to technological disruptions in finance. Overall, the model provides a robust framework that addresses the unique challenges of cryptocurrency forecasting, benefiting a wide range of stakeholders in navigating this complex market.

The limitations observed in standalone models emphasize the necessity of hybrid approaches. The hybrid stacked model demonstrated resilience by balancing directional accuracies (e.g., maintaining a high MDA+ and MDA-) and delivering robust financial metrics such as risk-adjusted returns and Sharpe Ratios. This highlights its adaptability in dynamic market environments and its robustness against data variability.

While the results of this study are promising, several areas remain open for further exploration and enhancement. Future research can build upon the foundation established here to address some of the limitations and extend the applicability of the proposed model. The following directions are particularly noteworthy:

One of the primary limitations of complex hybrid models is their opacity. Integrating explainable AI (XAI) methods, such as SHapley Additive exPlanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME), could enhance the interpretability of the predictions. This would allow practitioners to better understand the drivers of price changes and improve trust in the model's outputs.

Deploying the hybrid model in real-time trading systems poses challenges related to latency, data preprocessing, and adaptive learning. Future work could focus on optimizing the computational efficiency of the model and incorporating online learning techniques to enable real-time predictions and automated decision-making.

Extending the analysis to include a broader range of cryptocurrencies, including those with smaller market capitalizations or emerging tokens, could provide valuable insights into the model's scalability. Furthermore, applying the model to other financial instruments, such as stock indices or commodity prices, may help generalize its applicability beyond cryptocurrencies.

Cryptocurrencies are prone to abrupt price shocks caused by factors such as regulatory announcements or cyberattacks. Future work could investigate the robustness of the hybrid model under such adversarial scenarios, potentially incorporating techniques for anomaly detection and adaptive risk mitigation.

In addition to exploring model behavior under extreme market conditions, it is also crucial to recognize the current study's methodological boundaries. While this study demonstrates the robustness of a hybrid stacked framework for cryptocurrency price prediction, it is important to acknowledge certain limitations. In particular, on-chain metrics, such as transaction volume, active addresses, and network hash rates, represent valuable predictors of market dynamics but were excluded from the present analysis due to incomplete historical coverage and inconsistent data quality across the 2020–2024 period. Many publicly available blockchain datasets exhibit significant temporal gaps or are restricted to specific exchanges or assets, especially prior to mid-2022, limiting their suitability for a full-horizon analysis. Future research could incorporate these features once more comprehensive and standardized datasets become available, potentially enhancing the predictive power of ensemble models through deeper integration of network-level behavioral signals.

In addition to data-related constraints, several methodological considerations should also be noted. Overfitting risks, although mitigated through techniques such as early stopping, dropout regularization, and validation-based hyperparameter tuning, cannot be entirely ruled out, particularly under rapidly shifting market conditions. To prevent data leakage, all experiments employ strictly chronological train-validation-test splits and a walk-forward evaluation design, ensuring that no future information influences the training process. Finally, while the sensitivity of predictive performance to hyperparameter configurations was systematically addressed through extensive search procedures, slight variations may still occur if the search space or optimization criteria are modified. These factors highlight opportunities for further robustness testing and methodological refinement in future research.

In conclusion, this study demonstrated the potential of hybrid stacked models in addressing the challenges of cryptocurrency price prediction, offering a significant step forward in this domain. The avenues outlined for future research highlight the dynamic nature of this field and the opportunities for continued innovation. By advancing both the methodological and practical dimensions of cryptocurrency forecasting, future studies can contribute to the development of more robust, scalable, and interpretable predictive systems.

CRediT authorship contribution statement

Khushbakht Kamal: Conceptualization, Project Administration, Writing – original draft; **Kainat Mustafa:** Conceptualization, Project Administration, Writing – original draft; **Rashid Kamal:** Conceptualization, Data curation, Methodology, Formal analysis, Software, and Writing – original draft; **Yasir Riaz:** Conceptualization, Investigation, Methodology, Writing – original draft; **Chris Nugent:** Investigation, Writing – review and editing; **Fouzia Jumani:** Supervision, Investigation, Project administration, Validation; **Sheraz Aslam:** Project Administration, Validation, Writing – review and editing; **Nadeem Javaid:** Supervision, Investigation, Writing – review and editing. Each author has significantly contributed to the development of the manuscript, and all authors have read and approved the final version.

Originality

This manuscript is an original work and has not been published previously nor is it under consideration for publication elsewhere.

Ethical considerations

The research adheres to ethical guidelines and standards, with all data collected and processed responsibly and ethically.

Responsibility

We collectively take responsibility for the content and integrity of this manuscript and are available to address any questions regarding the accuracy of the work.

Data availability

The dataset used in this study is openly available. Foreign exchange rates and gold price data were obtained from Yahoo Finance, while cryptocurrency data were collected from publicly accessible sources such as CoinGecko and CoinMarketCap.

Declaration of competing interest

The authors declare that there are no known financial conflicts of interest or personal relationships that could have influenced the work presented in this paper.

References

Agbemenou, A. K. H., Motamed, R., & Talaei-Khoei, A. (2023). A predictive analytics model for designing deep underground foundations using artificial neural networks. *Decision Analytics Journal*, *7*, 100220.

Almeida, J., Gaio, C., & Gonçalves, T. C. (2024). Crypto market relationships with bric countries' uncertainty–a wavelet-based approach. *Technological Forecasting and Social Change*, *200*, 123078.

Antonopoulos, A.M. (2014). Mastering Bitcoin: Unlocking digital cryptocurrencies. "O'Reilly Media, Inc.".

Bedoui, R., Benkraiem, R., Guesmi, K., & Kedidi, I. (2023). Portfolio optimization through hybrid deep learning and genetic algorithms vine copula-GARCH-EVT-CVar model. *Technological Forecasting and Social Change*, *197*, 122887.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, 1937–1967.

Borri, N., & Shakhnov, K. (2022). The cross-section of cryptocurrency returns. *The Review of Asset Pricing Studies*, *12*(3), 667–705.

Bouteska, A., Mefteh-Wali, S., & Dang, T. (2022). Predictive power of investor sentiment for bitcoin returns: Evidence from COVID-19 pandemic. *Technological Forecasting and Social Change*, *184*, 121999.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. *Expert Systems with Applications*, *156*, 113464.

Buterin, V. et al. (2014). A next-generation smart contract and decentralized application platform. *white Paper*, *3*(37), 1–2.

Cakici, N., Shahzad, S.J.H., Będowska-Sójka, B., & Zaremba, A. (2024). Machine learning and the cross-section of cryptocurrency returns. *International Review of Financial Analysis*, *94*, 103244.

Cakici, N., & Zaremba, A. (2024). What drives stock returns across countries? insights from machine learning models. *International Review of Financial Analysis*, *96*, 103569.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794).

Chen, Y., Liu, Y., & Zhang, F. (2024). Coskewness and the short-term predictability for bitcoin return. *Technological Forecasting and Social Change*, *200*, 123196.

Cheng, J., Tiwari, S., Khaled, D., Mahendru, M., & Shahzad, U. (2024). Forecasting bitcoin prices using artificial intelligence: Combination of ML, SARIMA, and facebook prophet models. *Technological Forecasting and Social Change*, *198*, 122938.

Cho, K. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv:1406.1078.

Chu, J., Zhang, Y., & Chan, S. (2019). The adaptive market hypothesis in the high frequency cryptocurrency market. *International Review of Financial Analysis*, *64*, 221–231.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223.

Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, *62*, 182–199.

Dovilė, K., Aušra, V., & Krilavičius, T. (2019). Cryptocurrencies short-term forecast: Application of ARIMA, GARCH and SVR models. In *Ceur workshop proceedings [electronic resource]: Ivus 2019, international conference on information technologies, Kaunas, Lithuania, 25 April, 2019, vol. 2470*. Aachen: CEUR-WS.

Dunbar, K., & Owusu-Amoako, J. (2022). Cryptocurrency returns under empirical asset pricing. *International Review of Financial Analysis*, *82*, 102216.

Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*(2), 383–417.

Ghallabi, F., Souissi, B., Du, A.M., & Ali, S. (2025). Esg stock markets and clean energy prices prediction: insights from advanced machine learning. *International Review of Financial Analysis*, *97*, 103889.

Ghazani, M.M., & Jafari, M.A. (2021). Cryptocurrencies, gold, and WTI crude oil market efficiency: A dynamic analysis based on the adaptive market hypothesis. *Financial Innovation*, *7*(1), 29.

Ghosh, I., & Jana, R.K. (2023). A granular machine learning framework for forecasting high-frequency financial market variables during the recent black swan event. *Technological Forecasting and Social Change*, *194*, 122719.

Gordon, R.A. (2012). Applied statistics for the social and health sciences. Routledge.

Gunnarsson, E.S., Isern, H.R., Kaloudis, A., Risstad, M., Vigdel, B., & Westgaard, S. (2024). Prediction of realized volatility and implied volatility indices using AI and machine learning: A review. *International Review of Financial Analysis*, *93*, 103221.

Gupta, H., & Chaudhary, R. (2022). An empirical study of volatility in cryptocurrency market. *Journal of Risk and Financial Management*, *15*(11), 513.

Hao, R., & Yang, X. (2024). Multiple-output quantile regression neural network. *Statistics and Computing*, *34*(2), 89.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Huang, X., Tang, J., & Shen, Y. (2024). Long time series of ocean wave prediction based on patchTST model. *Ocean Engineering*, *301*, 117572.

Hyndman, R.J., & Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.

Jiang, X., Rodríguez Jr, I.M., & Zhang, Q. (2023). Macroeconomic fundamentals and cryptocurrency prices: a common trend approach. *Financial Management*, *52*(1), 181–198.

Kan, R., Wang, X., & Zheng, X. (2024). In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models. *Journal of Financial Economics*, *155*, 103837.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.Y. (2017). Light-GBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30* 1–9.

Khedr, A.M., Arif, I., El-Bannany, M., Alhashmi, S.M., & Sreedharan, M. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. *Intelligent Systems in Accounting, Finance and Management*, *28*(1), 3–34.

Kumar, R., Kumar, P., & Kumar, Y. (2021). Analysis of financial time series forecasting using deep learning model. In *2021 11th International conference on cloud computing, data science & engineering (confluence)* (pp. 877–881). IEEE.

Le Tran, V., & Leirvik, T. (2020). Efficiency in the markets of crypto-currencies. *Finance Research Letters*, *35*, 101382.

Lea, C., Flynn, M.D., Vidal, R., Reiter, A., & Hager, G.D. (2017). Temporal convolutional networks for action segmentation and detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 156–165).

Liu, Y., Li, Z., Nekhili, R., & Sultan, J. (2023). Forecasting cryptocurrency returns with machine learning. *Research in International Business and Finance*, *64*, 101905.

Liu, Y., & Tsyvinski, A. (2021). Risks and returns of cryptocurrency. *The Review of Financial Studies*, *34*(6), 2689–2727.

Lo, A.W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management*. *30*(5), 15–29.

Lu, L., Bu, C., Su, Z., Guan, B., Yu, Q., Pan, W., & Zhang, Q. (2024). Generative deep-learning-embedded asynchronous structured light for three-dimensional imaging. *Advanced Photonics*, *6*(4), 046004.

Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, *135*(2), 293–319.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111–153.

Mirza, N., Rizvi, S.K.A., Naqvi, B., & Umar, M. (2024). Inflation prediction in emerging economies: Machine learning and FX reserves integration for enhanced forecasting. *International Review of Financial Analysis*, *94*, 103238.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf. White paper.

Ozbayoglu, A.M., Gudelek, M.U., & Sezer, O.B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, *93*, 106384.

Pourrezaee, A., & Hajizadeh, E. (2024). Forecasting bitcoin volatility and value-at-risk using stacking machine learning models with intraday data. *Computational Economics*, *66*(1), 485–515.

Ramos-Pérez, E., Alonso-González, P.J., & Núñez-Velázquez, J.J. (2019). Forecasting volatility with a stacked model based on a hybridized artificial neural network. *Expert Systems with Applications*, *129*, 1–9.

Smales, L.A. (2019). Bitcoin as a safe haven: Is it even worth considering? *Finance Research Letters*, *30*, 385–393.

Song, Y., Zhang, D., Wang, J., Wang, Y., Wang, Y., & Ding, P. (2025). Application of deep learning in malware detection: A review. *Journal of Big Data*, *12*(1), 99.

Urquhart, A. (2016). The inefficiency of bitcoin. *Economics Letters*, *148*, 80–82.

Wang, J., Ma, F., Bouri, E., & Guo, Y. (2023). Which factors drive bitcoin volatility: macroeconomic, technical, or both? *Journal of Forecasting*, *42*(4), 970–988.

Wang, Q., & Chong, T.T.L. (2021). Factor pricing of cryptocurrencies. *The North American Journal of Economics and Finance*, *57*, 101348.

Wang, Y., Lucey, B.M., Vigne, S.A., & Yarovaya, L. (2022). The effects of central bank digital currencies news on financial markets. *Technological Forecasting and Social Change*, *180*, 121715.

Willmott, C.J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79–82.

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2022). Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv:2210.02186.

Yang, X., Han, Q., Ni, J., & Li, L. (2025). Research on the expansion of deposit insurance pricing model based on the merton option pricing framework. *Computational Economics, Forthcoming*, (pp. 1–23).

Yu, D., & Huang, D. (2023). Cross-sectional uncertainty and expected stock returns. *Journal of Empirical Finance*, *72*, 321–340.

Zhang, C., Sjarif, N.N.A., & Ibrahim, R. (2024). Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020–2022. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *14*(1), e1519.

Zhang, K., Wang, H., Chen, M., Chen, X., Liu, L., Geng, Q., & Zhou, Y. (2025a). Leveraging machine learning to proactively identify phishing campaigns before they strike. *Journal of Big Data*, *12*(1), 124.

Zhang, Z., Chang, Z., Gan, Y., & Li, J. (2025b). Renewable energy, innovation, and stock markets: Machine learning perspectives on environmental sustainability. *International Review of Financial Analysis*, *97*, 103880.

Zhao, Y., Xiaogang, B., & Qing-Ping, M.A. (2025). Predicting mergers & acquisitions: A machine learning-based approach. *International Review of Financial Analysis*, *99*, 103933.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, *35*, 11106–11115.

Zhu, J., Jin, J., Chen, C., Wu, L., Lu, M., & Ouyang, A. (2024). A new-type zeroing neural network model and its application in dynamic cryptography. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *9*(1), 176–191.