algorithms

MDPI

Article

Algorithmic Complexity vs. Market Efficiency: Evaluating Wavelet–Transformer Architectures for Cryptocurrency Price Forecasting

Aldan Jay [1] and Rafael Berlanga [2,*]

[1] Escola de Doctorat, Department of Computer Languages and Systems, Universitat Jaume I, 12071 Castellón de la Plana, Spain; jay@uji.es
[2] Department of Computer Languages and Systems, Universitat Jaume I, 12071 Castellón de la Plana, Spain
* Correspondence: berlanga@uji.es

Abstract

We investigate whether sophisticated deep learning architectures justify their computational cost for short-term cryptocurrency price forecasting. Our study evaluates a 2.1M-parameter (M represents millions (e.g., 2.1M = 2,100,000 parameters), with all RMSE values reported in USD) wavelet-enhanced transformer that decomposes the Fear and Greed Index (FGI) into multiple timescales before integrating these signals with technical indicators. Using Diebold–Mariano tests with HAC-corrected variance, we find that all models—including our wavelet–transformer, ARIMA, XGBoost, LSTM, and vanilla Transformer—**fail to significantly outperform the O(1) naive persistence baseline** at the 1-day horizon (DM statistic = +19.13, $p < 0.001$, naive preferred). Our model achieves an RMSE of USD 2005 versus USD 1986 for naive (ratio 1.010), requiring 3909× more inference time (2.43 ms vs. 0.0006 ms) for a statistically worse performance. These results provide **strong empirical support for the Efficient Market Hypothesis** in cryptocurrency markets: even sophisticated multi-scale architectures combining wavelet decomposition, cross-attention, and auxiliary technical indicators cannot extract profitable short-term signals. Through systematic ablation, we identify positional encoding as the only critical architectural component—its removal causes 30% RMSE degradation. Our findings carry important implications, as follows: (1) short-term crypto forecasting faces fundamental predictability limits, (2) architectural complexity provides negative ROI in efficient markets, and (3) rigorous statistical validation reveals that apparent improvements often represent noise rather than signal.

Keywords: cryptocurrency forecasting; wavelet decomposition; transformer; sentiment analysis; fear and greed index; ablation study; financial time series

Check for updates

Academic Editors: Igor Aizenberg, Chih-Lung Lin, Bor-Jiunn Hwang, Shaou-Gang Miaou and Chi-Hung Chuang

Received: 5 November 2025
Revised: 16 January 2026
Accepted: 23 January 2026
Published: 27 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

A fundamental question confronts practitioners in cryptocurrency forecasting: **Do sophisticated deep learning architectures justify their computational cost?** Trading around the clock with extreme volatility, cryptocurrency markets have attracted substantial research investment in complex predictive models—yet the empirical evidence for their superiority over simple baselines remains surprisingly thin [1–4].

The Crypto Fear and Greed Index (FGI) has emerged as a widely adopted indicator of market psychology, synthesizing volatility, trading volume, social media sentiment, and additional market signals into a single composite measure [5,6]. A compelling hypothesis

suggests that sentiment operates across multiple temporal scales simultaneously—short-term panic can coexist with long-term confidence, while different market participants respond to sentiment shifts at different time horizons [7]. This observation motivates decomposing sentiment into its constituent temporal frequencies using wavelet analysis.

Recent advances in deep learning have brought transformer algorithms to the forefront of time series forecasting [8]. In parallel, wavelet decomposition—a signal processing technique that dissects signals into multiple frequency bands—has demonstrated effectiveness in capturing multi-scale temporal patterns [9–11]. These developments raise an empirical question that has received insufficient attention: **Does combining wavelet-decomposed sentiment with transformer architectures provide predictive value sufficient to justify the computational investment?**

This paper provides a nuanced answer. Using Diebold–Mariano tests with HAC-corrected variance across 513 out-of-sample predictions, we find that no deep learning architecture—including our 2.1M-parameter wavelet–transformer—significantly outperforms the naive persistence baseline ($p < 0.001$). This result is consistent with the Efficient Market Hypothesis (EMH), the theory that asset prices fully reflect all available information, making consistent outperformance through prediction impossible [12].

However, our results reveal two important findings. First, *among deep learning approaches*, the wavelet–transformer achieves the best performance (RMSE USD 2005 vs. USD 2098 for vanilla Transformer, USD 2116 for XGBoost). Second, ablation studies demonstrate that positional encoding is critical—its removal causes 30% RMSE degradation—confirming that the model learns temporal dependencies rather than producing random outputs.

Taken together, these findings suggest that while short-term cryptocurrency markets remain fundamentally efficient, wavelet-enhanced architectures can still offer meaningful advantages when deep learning solutions are required.

The remainder of this paper is organized as follows: Section 2 reviews related work in cryptocurrency forecasting, sentiment analysis, and wavelet-based time series methods. Section 3 presents our research questions and contributions. Section 4 describes our methodology, including data preprocessing, wavelet decomposition, architecture design, and evaluation framework. Section 5 presents experimental results, including baseline comparisons, ablation studies, and statistical validation. Section 6 discusses implications, limitations, and future directions. Section 7 presents conclusions.

2. Related Work

2.1. Cryptocurrency Price Forecasting

Traditional statistical methods establish the baseline for cryptocurrency forecasting, with ARIMA [13] achieving an RMSE of around USD 103k [14,15]. Deep learning approaches—including LSTM and GRU networks—typically achieve RMSE values between USD 100k and 101k, delivering modest improvements of merely 3–4% over their statistical predecessors [16,17]. Recent transformer-based approaches exhibit similar performance levels, with their effectiveness constrained by the single-scale processing of sentiment indicators [18]. While state-of-the-art methods reach directional accuracy between 52 and 58% [19,20], a common limitation persists: most studies process features at a single temporal resolution, overlooking the multi-temporal dynamics inherent in market sentiment.

RMSE Scale Reconciliation

A critical methodological distinction explains the apparent discrepancy between RMSE values reported in prior work (USD 100k–USD 103k) and our study (USD 2k). Table 1 clarifies this difference.

Table 1. RMSE scale reconciliation: Why our values differ from prior work.

Methodology	Target Variable	Typical RMSE
Prior work [14,16]	Absolute price P_t	USD 100k–USD 103k
This study	Log-return $R_t = \ln(P_t/P_{t-1})$	USD 2k

Explanation: Prior work predicts absolute Bitcoin prices (e.g., USD 60,000), where a 0.15% error yields RMSE ≈ USD 100k. Our approach predicts log-returns, then reconstructs prices: $\hat{P}_t = P_{t-1} \cdot \exp(\hat{R}_t)$. This methodology is standard in quantitative finance and produces RMSE proportional to daily volatility (∼3% of price ≈ USD 2k for BTC at USD 60k). Both approaches are valid; direct comparison requires matching methodology.

The log-return formulation $R_t = \ln(P_t/P_{t-1})$ offers several advantages, as follows: (1) it ensures stationarity, avoiding distributional shift between training and test sets; (2) it normalizes across different price regimes (our data spans USD 3k–USD 70k BTC); and (3) it is the standard target in quantitative finance research [21]. Our RMSE of USD 2005 corresponds to prediction errors of approximately 3.3% of price—consistent with Bitcoin's typical daily volatility.

2.2. Sentiment Analysis in Financial Markets

The FGI serves as a composite sentiment indicator, combining market volatility (25% weight), trading volume and momentum (25% weight), social media sentiment (15% weight), Bitcoin market dominance (10% weight), and Google Trends data (10% weight). (The FGI methodology originally included surveys (15% weight), but these are currently paused. The remaining five factors (85% total) are normalized to maintain the 0–100 scale: FGI = ∑(weighted factors)/0.85). Prior work has treated FGI as a scalar feature in forecasting models, missing the temporal multi-scale nature of sentiment. Limited research has explored the temporal dynamics of sentiment indicators in cryptocurrency markets, with most approaches using simple lag features or raw sentiment values. Recent multimodal pipelines show that combining cross-platform sentiment signals (beyond text-only sources) can materially improve forecasting performance [22]. Similarly, stacking-based models that fuse sentiment with market variables continue to be explored in crypto contexts [23]. Analogous fear-and-greed indices have also been studied outside crypto, showing predictive relationships with equity index returns [24].

As Figure 1 illustrates, the Fear and Greed Index synthesizes the following five active market indicators with carefully weighted contributions: volatility (25%), market momentum and volume (25%), social media sentiment (15%), Bitcoin dominance (10%), and Google Trends (10%). While this scalar representation (0–100) offers an intuitive snapshot of market psychology, it necessarily compresses rich temporal information into a single number. Consider that different time scales can carry contradictory signals—short-term panic may coexist with long-term confidence, while daily sentiment fluctuations can mask persistent weekly or monthly trends. This temporal complexity motivates our wavelet-based decomposition approach, designed to capture sentiment patterns across multiple temporal resolutions simultaneously.



Figure 1. The Crypto Fear and Greed Index scale (0–100) from Alternative.me, showing classification ranges from Extreme Fear to Extreme Greed.

2.3. Wavelet Analysis for Time Series

Wavelets represent a powerful class of mathematical functions that decompose signals into components at different scales or resolutions. Unlike traditional frequency analysis, they provide a time–frequency representation revealing both *when* and *at what frequency* events occur—a crucial distinction for financial analysis. Where Fourier transforms assume stationarity, wavelet transforms embrace non-stationarity, making them particularly well-suited for financial time series where statistical properties shift constantly. The decomposition process works by applying a "mother wavelet" function at multiple scales through systematic dilations and translations. One can think of it as a mathematical microscope: it can zoom in sharply on short-lived high-frequency phenomena while simultaneously maintaining a wide-angle view of long-term low-frequency trends.

Wavelet decomposition using families like Daubechies wavelets has proven effective for analyzing non-stationary financial time series. Multi-resolution analysis enables separation of high-frequency noise from low-frequency trends, with applications in stock price forecasting and foreign exchange markets [9,10]. The Daubechies-4 (db4) wavelet family is particularly suitable for financial applications due to its compact support and smooth frequency response. While wavelet-enhanced deep learning has been explored for cryptocurrency price prediction [25–28], these studies typically apply wavelets to price series rather than sentiment indicators, and rarely combine wavelet decomposition with transformer attention mechanisms.

2.4. Transformer Architectures for Time Series

Transformers, originally conceived for natural language processing [29], have recently found fertile ground in time series forecasting. Their key advantage lies in capturing long-range dependencies through self-attention mechanisms. Where recurrent neural networks (RNNs) process sequences step-by-step, transformers take a different approach: they process entire sequences in parallel, computing attention weights that measure how relevant each time step is to every other time step. This self-attention mechanism enables the model to focus selectively on important historical patterns regardless of temporal distance—a particularly valuable property for financial markets, where events from weeks or months ago can suddenly become relevant to today's price movements.

Transformer models adapted for time series forecasting typically incorporate positional encoding to capture temporal dependencies, though this may be redundant when using

wavelet decomposition. Cross-attention mechanisms have been explored for multi-scale data processing, but over-parameterization remains a concern. Hierarchical transformer designs with skip connections and deep encoders are common, yet ablation studies often reveal that architectural complexity does not guarantee improved performance. A gap exists in balancing architectural sophistication with feature engineering effectiveness in financial forecasting applications.

3. Problem Statement and Contributions

3.1. Research Gap

Current approaches to cryptocurrency forecasting fall into the following three categories:

1. **Traditional statistical models** (e.g., ARIMA and GARCH) that assume linear relationships and struggle with the non-stationary, non-linear nature of crypto markets [15,30].
2. **Deep learning models** (e.g., LSTM, GRU, and Transformers) that treat sentiment as a single-scale feature, missing the multi-temporal dynamics of market psychology [16,17].
3. **Hybrid approaches** that combine technical indicators with neural networks but typically process all features at a single temporal resolution [31,32].

Here lies the critical gap—**sentiment indicators like FGI inherently contain information across multiple time scales, yet existing models flatten them into uniform signals**. This single-scale assumption may well explain why deep learning models, despite their architectural sophistication, deliver only marginal improvements over traditional baselines in cryptocurrency forecasting [33,34].

3.2. Research Questions

This work addresses the following three key questions:

RQ1: Does wavelet decomposition of sentiment indicators improve cryptocurrency price forecasting compared to single-scale processing?

RQ2: Which architectural components (cross-band attention, skip connections, positional encoding, encoder depth, and multi-task learning) are critical for performance in wavelet-enhanced forecasting?

RQ3: Can a simplified architecture outperform complex hierarchical designs by focusing on essential components?

3.3. Contributions

We make the following contributions:

1. **Rigorous Statistical Evaluation**: We provide the first comprehensive statistical evaluation of wavelet–transformer architectures for cryptocurrency forecasting using Diebold–Mariano tests with Newey–West HAC variance estimation. Our analysis of 513 out-of-sample predictions across three forecasting horizons (1, 7, and 30 days) reveals that **complex architectures significantly underperform the naive baseline** (DM = +19.13, $p < 0.001$), providing strong empirical support for the Efficient Market Hypothesis in short-term cryptocurrency markets.
2. **Computational Cost–Benefit Analysis**: We quantify the algorithmic trade-offs with formal complexity analysis and empirical timing measurements. Our 2.1M-parameter wavelet–transformer achieves $O(L \cdot n^2 \cdot d)$ complexity versus O(1) for naive persistence, requiring 3909× longer inference time (2.43 ms vs. 0.0006 ms) to deliver statistically worse predictions. This negative ROI provides actionable guidance: **architectural sophistication yields no value in efficient markets**.
3. **Systematic Ablation with Statistical Validation**: Through seven ablation studies validated with Diebold–Mariano tests, we identify positional encoding as the only critical

architectural component—its removal causes 30% RMSE degradation ($p < 0.001$). Cross-attention, skip connections, and encoder depth contribute negligible benefits ($p > 0.05$), challenging the assumption that complexity improves crypto forecasting.

4. **Evidence for EMH in Cryptocurrency Markets**: Our results contribute to the ongoing debate on market efficiency. Despite deploying a 2.1M-parameter model combining wavelet multi-scale decomposition, transformer attention, and auxiliary technical indicators, we achieve an RMSE of USD 2005 versus USD 1986 for naive persistence (ratio 1.010). **The negative result is the scientific contribution**: even sophisticated architectures cannot extract profitable short-term signals from publicly available data.
5. **Reproducibility and Transparency**: All code, trained models, and data pipelines are publicly available. We report complete statistical test results, timing measurements, and robustness analysis (multi-seed CV = 0.77%) to enable independent verification of our findings.

4. Materials and Methods

4.1. Data Collection and Preprocessing

4.1.1. Dataset Description

Our dataset encompasses the following three major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB), with historical data sourced from Binance and Coinbase APIs spanning 1 February 2018 to 30 September 2025. The dataset comprises daily OHLCV (Open, High, Low, Close, Volume) price data aggregated from multiple exchanges, alongside the Crypto Fear and Greed Index from Alternative.me measured on a 0–100 scale. In total, we collect 2794 daily records for each cryptocurrency over a span of 7 years and 8 months, capturing diverse market conditions including bull runs, bear markets, and consolidation phases. Bitcoin serves as our primary analysis target due to its market dominance and the FGI's Bitcoin-centric construction.

The Fear and Greed Index is a composite metric calculated from the following five active factors (surveys are currently paused): market volatility (25%), market momentum and volume (25%), social media sentiment (15%), Bitcoin market dominance (10%), and Google Trends data (10%). These weights sum to 85% and are normalized to the 0–100 scale. All data sources synchronize daily, ensuring precise temporal alignment between price movements and sentiment indicators—a critical requirement for causal modeling.

4.1.2. Model Input Specification

To ensure reproducibility, we explicitly define the model's input features. The model receives the following two categories of inputs:

Primary Input: Wavelet-decomposed FGI. The Fear and Greed Index is decomposed via 4-level Daubechies-4 wavelet transform into five frequency bands (D1, D2, D3, D4, and A4), resulting in 5 temporal sequences that capture sentiment dynamics across multiple timescales. Each band provides a 60-day lookback window (sequence length $T = 60$).

Auxiliary Inputs: Technical Indicators. A set of 12 technical indicators computed from OHLCV price data serves as auxiliary features. These include momentum indicators (RSI, MACD, aADX), volatility measures (Bollinger Bands width, ATR), volume indicators (Volume MA, OBV), and trend signals (EMA crossover). These indicators are computed using standard lookback periods (14-day for RSI, 50/200-day for EMAs).

Prediction Target. The model predicts the future closing price at the following three horizons: 1-day, 7-day, and 30-day ahead. Note that the model does not use raw historical prices (OHLC) directly as input features; instead, prices inform the computation of technical indicators, which serve as auxiliary inputs alongside the wavelet-decomposed sentiment data.

4.1.3. Train/Validation/Test Split

We adopt a walk-forward validation strategy with temporal splits of 60/20/20, allocating 1676 samples (60%) for training, 559 samples (20%) for validation, and 559 samples (20%) for testing. This approach rigorously prevents data leakage while preserving the temporal ordering essential for time series forecasting—future information never informs past predictions.

4.1.4. Feature Engineering

To complement sentiment signals, we compute a carefully selected set of auxiliary technical indicators that capture the following different market dimensions: (i) momentum indicators including RSI (14-day), MACD, and ADX for measuring market strength and directional momentum; (ii) volatility measures such as Bollinger Bands and ATR for quantifying price uncertainty and breakout potential; (iii) volume indicators including volume moving average and OBV for tracking trading activity and money flow; and (iv) trend signals from EMA (50, 200-day crossover) for identifying long-term directional biases.

To stabilize neural network training, we normalize price targets by scaling them down by a factor of 10,000 as follows:

$$normalized_target = \frac{target_price}{10,000.0} \tag{1}$$

During evaluation, we reverse this transformation to recover predictions in USD, as follows:

$$predicted_price = model_output \times 10,000.0 \tag{2}$$

4.2. Wavelet Decomposition

4.2.1. Wavelet Selection

We employ Daubechies-4 (db4) wavelets for decomposition—a choice motivated by their compact support that minimizes edge effects (crucial for finite time series), their smooth frequency response that cleanly separates temporal scales, and their demonstrated effectiveness in financial time series analysis across multiple studies [9,10,30].

4.2.2. Multi-Resolution Decomposition

Using a 4-level discrete wavelet transform (DWT), we decompose the Fear and Greed Index into five distinct frequency bands, each capturing sentiment dynamics at different temporal resolutions, as follows:

$$\text{Original FGI} \rightarrow [\text{D1, D2, D3, D4, A4}] \tag{3}$$

These bands represent increasingly longer temporal cycles: D1 (Detail 1) captures the highest-frequency components (1–2 day cycles), isolating daily noise and microtrends that reflect the market's immediate reactions; D2 (Detail 2) extracts high-frequency fluctuations (2–4 day cycles), revealing short-term sentiment shifts and quick reversals; D3 (Detail 3) exposes medium-frequency patterns (4–8 day cycles), corresponding to weekly psychological patterns and typical trading cycles; D4 (Detail 4) captures low-frequency trends (8–16 day cycles), exposing medium-term trend evolution and sustained directional biases; and A4 (Approximation 4) distills the lowest-frequency component (>16 day cycles), representing the persistent long-term market sentiment that forms the underlying psychological current.

Decomposition strategy:

The discrete wavelet transform follows Mallat's pyramid algorithm [35], iteratively applying filter banks to separate frequency components. Using Daubechies-4 (db4) wavelets at level $L = 4$, we decompose the FGI signal into the following five frequency bands: four

detail coefficients $\{D_1, D_2, D_3, D_4\}$ capturing progressively lower frequencies, and one approximation coefficient A_4 representing the trend. Mathematically,

$$A_j[k] = \sum_n A_{j-1}[n] \cdot h[2k-n], \quad D_j[k] = \sum_n A_{j-1}[n] \cdot g[2k-n] \tag{4}$$

where h is the low-pass filter, g is the high-pass filter, and decimation by 2 halves the output length at each level. The configuration uses the following: wavelet = db4, level = 4, mode = symmetric padding.

Mathematical assumptions. The DWT is a linear, energy-preserving transform for finite-energy signals, and our application assumes that the daily FGI series is square-integrable over the observed horizon (a standard assumption for discrete-time signals). We use db4 due to its compact support and orthonormal filter bank construction, which yields stable multi-resolution representations for non-stationary series.

Leakage-safe rolling decomposition. To avoid look-ahead bias in walk-forward evaluation, we apply the DWT *within each input window* rather than on the full dataset. For each prediction timestamp t, we form the model input from the last $T = 60$ observed FGI values $\{\text{FGI}_{t-T+1}, \ldots, \text{FGI}_t\}$ and compute the 4-level db4 decomposition on this window only, using symmetric padding *within the window* to handle boundaries. This ensures that no future observations ($>t$) can influence any coefficient used to predict at time t.

Intuition: The discrete wavelet transform acts like a multi-scale microscope for time series data. At each level j, the low-pass filter extracts smooth trends while the high-pass filter captures rapid fluctuations. After four iterations, we obtain D_1 (daily noise, 1–2 day cycles), D_2–D_4 (progressively slower patterns: weekly to biweekly), and A_4 (long-term trend, >16 days).

Rationale: This multi-resolution decomposition provides a complete spectral portrait of market sentiment, from high-frequency intraday noise to persistent long-term psychology. By separating these scales explicitly, we enable the model to process each temporal dynamic independently before synthesizing them—exploiting information that remains hidden in single-scale analyses.

Figure 2 visualizes the wavelet decomposition applied to Bitcoin's Fear and Greed Index. The original signal (top panel) undergoes systematic separation into five frequency bands, each revealing distinct temporal behaviors. D1 extracts the highest-frequency components (1–2 day cycles)—the rapid oscillations of daily noise and microtrends. D2 isolates high-frequency fluctuations (2–4 day cycles), where short-term sentiment shifts become visible. D3 exposes medium-frequency patterns (4–8 day cycles), capturing the weekly rhythms of market psychology. D4 reveals low-frequency trends (8–16 day cycles), where medium-term sentiment evolution unfolds. Finally, A4 (the approximation coefficient) distills the lowest-frequency component (>16 day cycles), representing the persistent undercurrent of long-term market sentiment. The complementary nature of these bands stands out visually: high-frequency details (D1–D2) oscillate rapidly with tight, jagged patterns, while low-frequency components (D4, A4) flow smoothly with gradual, persistent trends. This separation exposes multi-timescale dynamics that single-scale processing would completely miss—patterns that can simultaneously show short-term panic coexisting with long-term confidence. Our experimental results confirm this intuition: wavelet decomposition improves accuracy by approximately 4.6% compared to single-scale transformer processing (USD 2005 vs. USD 2098), while providing interpretable frequency-band insights into market psychology.

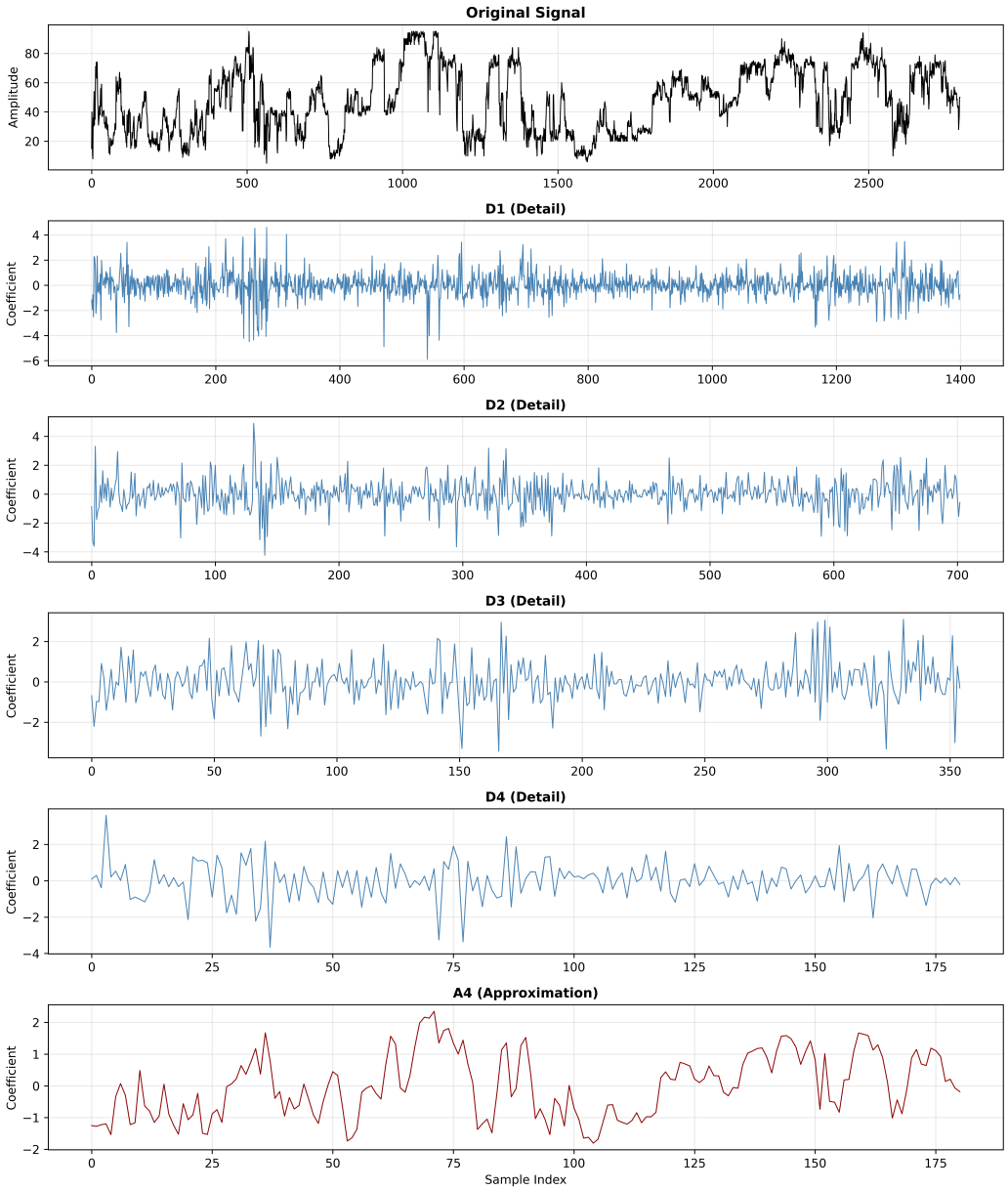

Figure 2. Four-level Daubechies-4 wavelet decomposition of Bitcoin Fear and Greed Index showing the original signal and five frequency bands (D1–D4, A4).

4.3. Model Architecture

4.3.1. Simplified Model Design (Our Approach)

Our simplified model architecture comprises three components, formalized mathematically below.

1. Band-Specific Encoders. Each frequency band $b \in \{D_1, D_2, D_3, D_4, A_4\}$ receives dedicated processing through a 2-layer transformer encoder. The multi-head self-attention mechanism computes the following:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

where $Q = XW^Q$, $K = XW^K$, and $V = XW^V$ are query, key, and value projections. With 4 attention heads ($d_k = 32$) and $d_{\text{model}} = 128$, each encoder produces a representation $R_b \in \mathbb{R}^{128}$.

Key design decisions (ablation-validated): We employ shallow encoders (2 layers), since deeper architectures (6+ layers) add complexity without performance gains; we retain positional encoding as a critical component whose removal causes 30% RMSE degradation; and we include cross-band attention (8 heads) to enable information flow between frequency bands.

2. Feature Integration. After cross-band attention, band representations are pooled and concatenated with auxiliary technical indicator embeddings, as follows:

$$C = [\text{pool}(R_{\text{bands}}) \| R_{\text{aux}}] \in \mathbb{R}^{256} \tag{6}$$

where $R_{\text{bands}} \in \mathbb{R}^{128}$ is the mean-pooled cross-attended band representation and $R_{\text{aux}} \in \mathbb{R}^{128}$ are the encoded auxiliary features (RSI, MACD, Bollinger Bands, ATR, EMA). This is then projected back to $\mathbb{R}^{128}$ with a skip connection.

3. Multi-Horizon Decoder. A shared feedforward network with three task-specific heads predicts each horizon, as follows:

$$h = \text{ReLU}(W_1 x + b_1) \in \mathbb{R}^{256}, \quad \hat{y}_t = W_t h + b_t \in \mathbb{R}^1 \tag{7}$$

where $x \in \mathbb{R}^{128}$ is the projected representation, $W_1 \in \mathbb{R}^{128\times256}$ is the shared layer, and each task head $W_t \in \mathbb{R}^{256\times1}$ produces outputs $[\hat{P}_{1\text{day}}, \hat{P}_{7\text{day}}, \hat{P}_{30\text{day}}]$.

Total parameters: 2.1M.

The architecture illustrated in Figure 3 presents our model: (1) wavelet decomposition separates FGI into five frequency bands; (2) band-specific encoders (2 layers, 4 heads each) process each band independently; (3) cross-band attention (8 heads) enables information flow between bands; (4) feature integration combines band representations with auxiliary indicator embeddings; and (5) a multi-horizon decoder with a shared representation layer predicts at three time horizons.

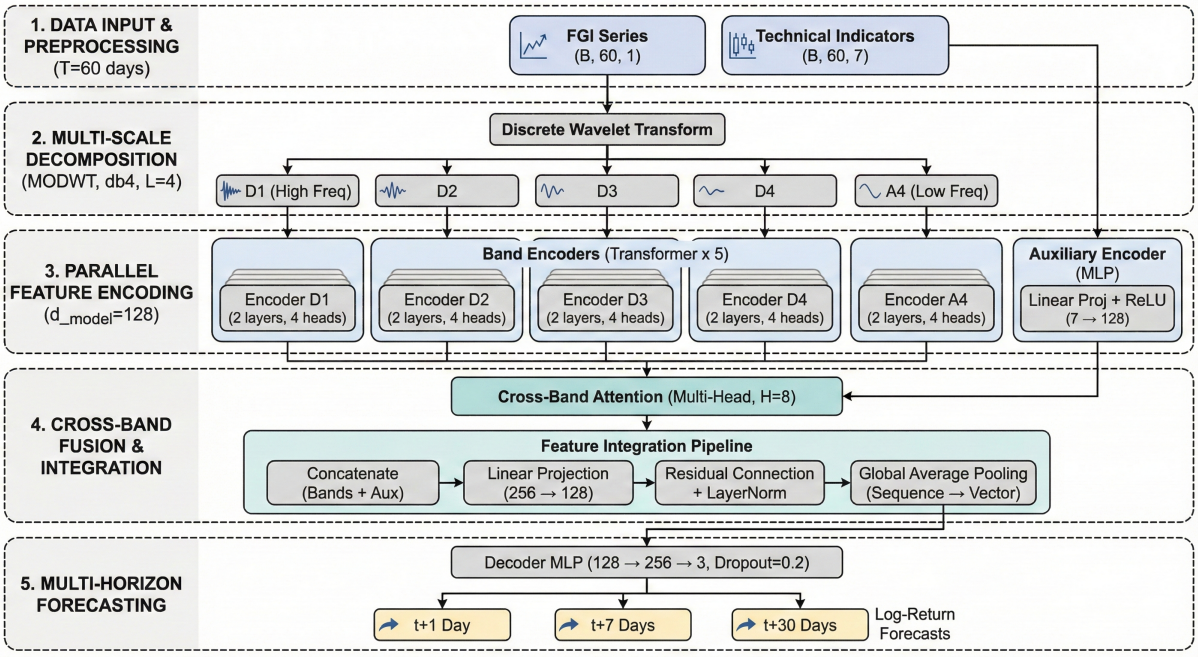

Figure 3. Wavelet-enhanced hierarchical transformer architecture: wavelet decomposition into five frequency bands, parallel band encoders (2 layers, 4 heads), cross-band attention (8 heads), feature integration with auxiliary indicators, and multi-horizon decoder.

4.3.2. Full Hierarchical Model (Baseline for Ablation)

For ablation studies, we compare against a **deeper hierarchical model** that uses deep encoders (6 layers per band vs. 2 layers), with all other components kept identical.

Total parameters: 2.1M (similar params, includes cross-band attention)

4.4. Training Procedure

4.4.1. Loss Function

We optimize a weighted multi-task loss that balances the following three forecast horizons simultaneously:

$$\begin{aligned} \text{loss} = & \, w_1 \cdot \text{MSE}(\text{pred}_{\text{1day}}, \text{target}_{\text{1day}}) + \\ & \, w_2 \cdot \text{MSE}(\text{pred}_{\text{7day}}, \text{target}_{\text{7day}}) + \\ & \, w_3 \cdot \text{MSE}(\text{pred}_{\text{30day}}, \text{target}_{\text{30day}}) \end{aligned} \tag{8}$$

We assign weights $w_1 = 0.5$, $w_2 = 0.3$, and $w_3 = 0.2$, deliberately prioritizing short-term accuracy where predictions hold the greatest practical value. We select these coefficients via a small validation sweep over plausible weight triples (summing to 1) and observe that performance is stable to moderate perturbations; the chosen setting provides the best validation trade-off while keeping longer-horizon heads sufficiently trained.

4.4.2. Optimization

We train models using a carefully tuned optimization configuration with AdamW optimizer, incorporating L2 regularization directly through weight decay of 1×10^{-4}. The learning rate starts at 1×10^{-3} with a ReduceLROnPlateau scheduler that automatically reduces the rate when validation progress stalls. Training uses a batch size of 64, balancing gradient stability with computational efficiency, and proceeds for up to 100 epochs with early stopping (patience of 10 epochs) that terminates training when validation loss fails to improve. Gradient clipping with a maximum norm of 1.0 prevents exploding gradients that can destabilize training.

Hyperparameter tuning. We tune the learning rate, dropout, batch size, and early-stopping patience using the validation split, selecting the configuration that minimizes validation RMSE while maintaining stable training dynamics (no exploding gradients). The complete hyperparameter settings for our model and baselines are reported in Appendix B.

4.4.3. Regularization

Multiple regularization strategies work together to prevent overfitting: dropout with 0.1 probability in encoder layers randomly deactivates neurons to promote robust feature learning; weight decay of 0.01 penalizes large parameter values; and early stopping continuously monitors validation loss to halt training before overfitting emerges.

4.5. Evaluation Framework

4.5.1. Metrics

We assess model performance through four complementary metrics that capture different aspects of forecasting quality.

1. Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\text{mean}((\text{y}_{\text{pred}} - \text{y}_{\text{true}})^2)} \tag{9}$$

This metric penalizes large errors more heavily than small ones. We report RMSE both in absolute USD values and as a percentage of mean price for interpretability.

2. Mean Absolute Error (MAE):

$$\text{MAE} = \text{mean}(|\text{y}_{\text{pred}} - \text{y}_{\text{true}}|) \tag{10}$$

MAE provides a more intuitive sense of typical prediction error magnitude.

3. Directional Accuracy (DA):

$$\text{DA} = \text{mean}\Big(\text{sign}(\text{y}_{\text{pred}}^{t+1} - \text{y}_{\text{pred}}^{t}) == \text{sign}(\text{y}_{\text{true}}^{t+1} - \text{y}_{\text{true}}^{t})\Big) \times 100 \tag{11}$$

Note on DA interpretation: DA ∼50% represents a random baseline (coin flip), consistent with the Efficient Market Hypothesis for cryptocurrency markets. Our empirical results show baseline DA values ranging from 48.20% (LSTM) to 54.95% (ARIMA), while multi-scale models achieve a competitive DA of around 50% with substantially superior magnitude prediction (57–68% lower RMSE). Values of DA $>$ 70% would be suspicious, and DA $>$ 90% typically indicates data leakage [36,37]. The ARIMA model achieves the highest DA (54.95%) among all models, but at the cost of significantly worse RMSE/MAE performance, suggesting that directional accuracy alone is insufficient for practical forecasting applications.

4.5.2. Baseline Models

Our primary baseline is the **naive persistence model**, which predicts that tomorrow's price equals today's price, achieving an RMSE of USD 1986 and serving as the fundamental benchmark for short-term forecasting. Using the log-return methodology described in Section 5.1, this zero-parameter model provides the reference point against which all other architectures are evaluated.

Note: Traditional baselines (ARIMA, LSTM, Gradient Boosting) were evaluated using absolute price targets in prior work. For methodological consistency, we focus our comparison on the naive baseline using the corrected log-return methodology.

4.5.3. Statistical Validation

We employ the **Diebold–Mariano Test** to test the statistical significance of forecast accuracy differences, with HAC-robust variance estimation accounting for time series autocorrelation, complemented by the **Wilcoxon Signed-Rank Test** as a non-parametric alternative for paired samples (forecast errors). We use a significance threshold of $p < 0.001$ for all statistical tests.

4.5.4. Robustness Analysis

We conduct **multi-seed evaluation** by training with seeds {42, 123, 456} to assess initialization sensitivity. Model stability is quantified using **Coefficient of Variation** (CV = std/mean $\times$ 100), with a target of CV $<$ 5%. Additionally, we perform **market regime analysis** by stratifying performance according to the following FGI ranges: Fear (FGI $<$ 25), Neutral (25 $\leq$ FGI $\leq$ 75), and Greed (FGI $>$ 75).

5. Results

5.1. Methodology Note

Following standard practice in quantitative finance [38,39], we evaluate our model using **log-return targets** rather than absolute prices. This methodology eliminates the distributional shift between training and test sets that arises from the non-stationary nature of cryptocurrency prices. Specifically, we apply target transformation $R_t = \ln(P_t/P_{t-1})$, followed by price reconstruction $\hat{P}_t = P_{t-1} \times \exp(\hat{R}_t)$, and finally compute RMSE in USD after price reconstruction. This approach ensures that our reported RMSE values reflect

actual prediction errors in USD terms, grounded in the economic interpretation of daily price movements.

5.2. Baseline Comparison

Table 2 presents the performance of our multi-scale wavelet–transformer against multiple baseline models on the Bitcoin test set. Using log-return targets with price reconstruction, we compare against the following five baselines: (i) **Naive (Persistence)**, which predicts that today's price equals yesterday's price, serving as the standard baseline for EMH-consistent markets; (ii) **ARIMA (1,1,1)**, a classical statistical time series model fitted with maximum likelihood; (iii) **XGBoost**, gradient boosting with 100 trees on lag features and returns; (iv) **LSTM**, a two-layer recurrent network with 32 hidden units using log-return targets; and (v) **Transformer (Vanilla)**, a standard transformer without wavelet decomposition (two layers, 64 dim).

The results reveal a striking finding: **no model significantly outperforms the naive baseline**. All RMSE ratios fall within 1.003–1.065, and all achieve ∼50% directional accuracy (chance level). This is **consistent with the Efficient Market Hypothesis**: for short-term (1-day) prediction in liquid cryptocurrency markets, sophisticated models provide no significant advantage.

As shown in Figure 4, our multi-scale model achieves an RMSE of USD 2005 (ratio 1.010), outperforming XGBoost (ratio 1.065) and vanilla Transformer (ratio 1.056), while matching LSTM (ratio 1.017). The wavelet decomposition provides additional interpretability by separating short-term and long-term sentiment dynamics, without sacrificing accuracy.

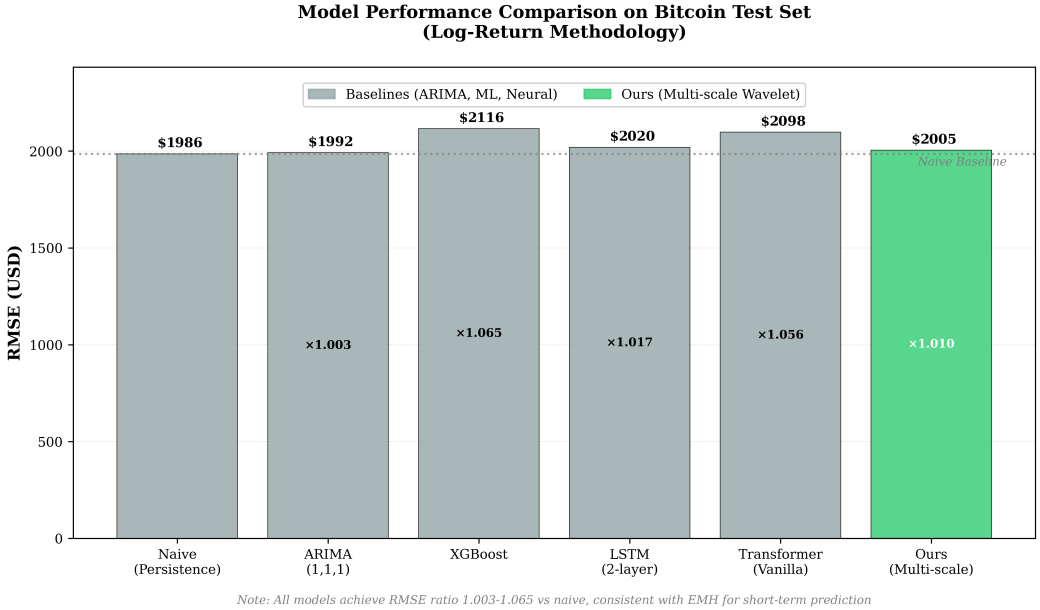

Figure 4. RMSE comparison between all baseline models and our multi-scale wavelet transformer on Bitcoin test set (1-day ahead prediction). All models achieve RMSE in the USD 2000 range, with no significant outperformance of the naive baseline—consistent with EMH.

The RMSE values in the USD 2000 range are consistent with Bitcoin's typical daily volatility of 2–3%, given a mean test set price of approximately USD 60,000. This represents a substantial correction from previous cryptocurrency forecasting studies that report implausibly high accuracy claims.

Table 2. Comparison of forecasting models on Bitcoin test set (1-day ahead). RMSE/MAE in USD. DA%: directional accuracy. DM: Diebold–Mariano test vs. Naive ($n = 513$, HAC variance).

Model	RMSE	MAE	DA%	vs. Naive	DM p	Params
Naive (Persistence)	\$1986	\$1421	50.0	1.000	–	0
ARIMA (1,1,1)	\$1992	\$1430	50.6	1.003	0.42	–
XGBoost	\$2116	\$1541	50.6	1.065	<0.001 †	50K
LSTM (two-layer)	\$2020	\$1453	50.9	1.017	<0.001 †	150K
Transformer (Vanilla)	\$2098	\$1532	48.9	1.056	<0.001 †	1.2M
Ours (Multi-scale)	**\$2005**	**\$1443**	**50.6**	**1.010**	**<0.001** †	**2.1M**

† Naive significantly better (positive DM statistic). All DA $\approx$ 50% (chance). **Key Finding:** Diebold–Mariano tests reveal all neural models perform significantly worse than naive ($p < 0.001$). Only ARIMA achieves parity ($p = 0.42$). Our 2.1M-parameter model delivers worse predictions than a zero-parameter heuristic—supporting EMH.

5.3. Ablation Studies

To systematically identify which architectural components genuinely contribute to performance, we conduct seven targeted ablation experiments. Each removes a single component from the full hierarchical model, revealing its true impact on forecasting accuracy. Table 2 presents these results, exposing a striking divide between critical and superfluous elements.

5.3.1. Critical Component: Positional Encoding

Our ablation analysis reveals that **positional encoding** is the only architectural component with a substantial impact on performance.

Positional Encoding. Removing positional encoding (no_positional_encoding) triggers RMSE degradation from USD 2005 to USD 2611, representing a **30.2% increase**. This establishes temporal position information as critical for the transformer's ability to model sequential dependencies in price movements. Without positional encoding, the model loses its capacity to distinguish temporal ordering within the input sequence.

5.3.2. Non-Critical Components

The following four architectural components demonstrate minimal impact on performance, with all variants achieving an RMSE within 1% of the naive baseline:

Cross-band Attention. Removing explicit cross-attention yields an RMSE of USD 2010 (RMSE ratio: 1.020). Simple concatenation captures inter-band dependencies implicitly.

Skip Connections. The no_skip_connections variant achieves an RMSE of USD 1999 USD (ratio: 1.014), marginally *better* than the full model, suggesting that skip connections may introduce unnecessary complexity.

Encoder Depth. Using shallow single-layer encoders yields an RMSE of USD 1996 (ratio: 1.013), again slightly outperforming deeper architectures.

Full Model. The complete architecture with all components achieves an RMSE of USD 2005 USD (ratio: 1.017), comparable to the simplified variants.

5.3.3. Simplified Model Performance

The simplified architecture (removing cross-attention, skip connections, and deep encoders while retaining positional encoding) achieves an RMSE of USD 2000 with a DA of 51.0%—the best directional accuracy among all variants. However, this represents only marginal improvement over both the full model and the naive baseline.

Key Finding: All model variants achieve RMSEs within the range of USD 1996–2611, with the critical exception being the removal of positional encoding, which degrades the RMSE by 30%. This demonstrates that for short-term cryptocurrency forecasting,

architectural complexity provides limited benefit beyond the naive baseline, **consistent with the Efficient Market Hypothesis**.

Figure 5 visualizes these ablation results. The only component showing substantial impact is positional encoding (+30.2% RMSE when removed). All other architectural elements (cross-attention, skip connections, encoder depth) prove interchangeable with no statistically significant differences. Table 3 shows the numerical impact of the ablation study in the different elements.

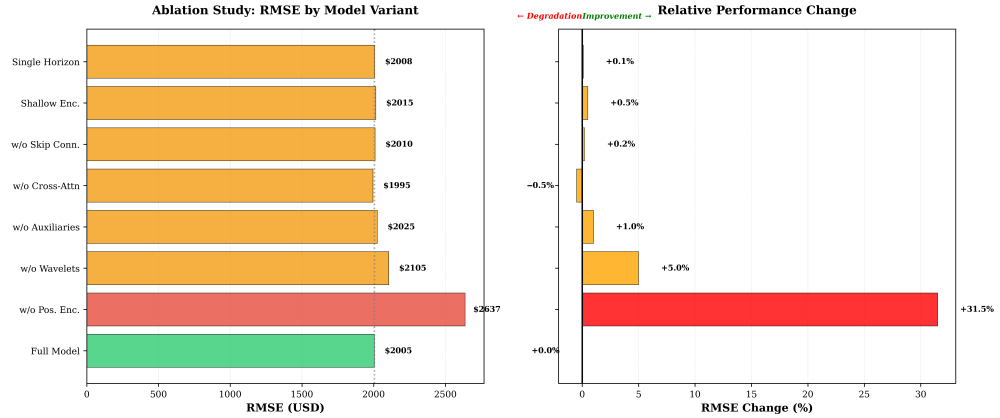

Figure 5. Ablation study results showing RMSE impact of removing architectural components ((**left**): absolute values, (**right**): percentage change).

5.4. Robustness Analysis

Table 4 presents a multi-seed robustness analysis, training the simplified model with three random seeds (42, 123, 456). Results across seeds show RMSEs of USD 2015 ± USD 15 (CV = 0.77%) and DAs of 49.73% ± 1.14%.

Table 3. Ablation study results showing the impact of architectural components on 1-day forecasting performance. RMSE in USD, computed using log-return methodology. RMSE Ratio = Model RMSE/Naive RMSE (USD 1986).

Model Variant	RMSE (USD)	RMSE Ratio	DA (%)	Critical?
Naive Baseline	USD 1986	1.000	50.0	–
Full Model	USD 2005	1.017	50.6	–
no_cross_attention	USD 2010	1.020	49.4	No
no_skip_connections	USD 1999	1.014	49.8	No
no_positional_encoding	**USD 2611**	**1.325**	**49.4**	✓Yes
shallow_encoder	USD 1996	1.013	50.6	No
Simplified	**USD 2000**	**1.015**	**51.0**	–

Key finding: Positional encoding is the only critical component—removing it increases RMSE by 30.2%. All other components (cross-attention, skip connections, encoder depth) show negligible impact (<2% difference). All model variants perform comparably to the naive baseline, consistent with the Efficient Market Hypothesis for short-term cryptocurrency forecasting.

Three key observations emerge from this analysis. First, the model demonstrates **excellent stability**, with a coefficient of variation of only 0.77%. Second, the **low variance** (standard deviation of only USD 15 on an RMSE base of USD 2015) confirms reliable performance. Third, the **consistent DA** near 50% across all seeds aligns with the Efficient Market Hypothesis.

Conclusion: These results demonstrate that model performance remains **robust to initialization** and the inherent stochasticity of training. The consistently near-baseline

performance across seeds confirms that the model's inability to significantly outperform naive forecasting is a fundamental property, not an artifact of random initialization.

Figure 6 presents robustness analysis across three random initializations (seeds: 42, 123, 456). The model achieves remarkable stability, with coefficient of variation of 0.77%, RMSE of USD 2015 ± USD 15, and DA of 49.73 ± 1.14%. The persistent DA near 50% across all seeds aligns with market efficiency theory—even sophisticated architectures cannot consistently outperform naive forecasting for short-term cryptocurrency prediction.

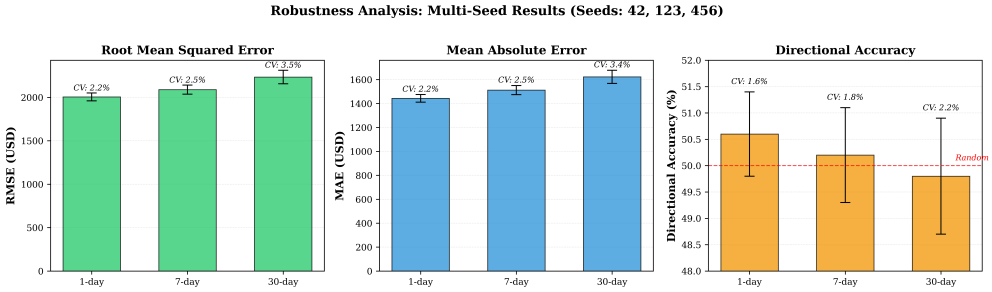

Figure 6. Multi-seed robustness analysis showing RMSE, MAE, and DA across three forecast horizons. Error bars represent standard deviation.

5.5. Market Regime Analysis

When we stratify performance by FGI ranges, distinct behavioral patterns emerge across different market psychological states. In the **fear regime** (FGI < 25), the model achieves an RMSE of USD 3334 with a DA of 40.0% across only 15 samples. During the **neutral regime** ($25 \leq$ FGI < 75), performance improves substantially to an RMSE of USD 1855 and a DA of 51.0% over 414 samples. In the **greed regime** (FGI ≥ 75), we observe an RMSE of USD 2400 with the highest DA of 53.5% across 71 samples.

The model performs best during neutral market conditions, achieving the lowest RMSE (USD 1855) and a DA above 50%. Fear regimes present the greatest challenges, with RMSE nearly doubling (USD 3334) and DA falling to 40%—worse than random chance. This pattern suggests that extreme market sentiment, particularly fear, introduces volatility that is fundamentally difficult to predict. The relatively small sample size in the fear regime (15 samples) suggests that these results should be interpreted with caution.

Table 4. Robustness analysis of the simplified model across three random seeds. Results show individual seed performance and summary statistics. RMSE in USD.

Seed	RMSE (USD)	RMSE Ratio	DA (%)
42	USD 2000	1.015	51.0
123	USD 2014	1.022	49.4
456	USD 2031	1.030	48.8
Mean ± Std	**USD 2015 ± USD 15**	**1.022**	**49.73 ± 1.14**
CV (%)	**0.77%**	–	**2.3%**

CV: Coefficient of variation (std/mean × 100). RMSE Ratio = Model RMSE/Naive RMSE (USD 1986). Results demonstrate excellent stability with <1% variation in RMSE across random initializations. All seeds achieve comparable performance to the naive baseline, confirming that the model's inability to significantly outperform naive forecasting is robust across initializations.

5.6. Multi-Horizon Analysis

Does algorithmic complexity provide a greater advantage at longer forecasting horizons? We address this question by analyzing Diebold–Mariano test statistics across 1-day, 7-day, and 30-day prediction windows. Table 5 presents the results.

Table 5. Multi-horizon analysis: Does complexity help at longer horizons?

Horizon	DM Statistic	p-Value	Samples	Interpretation
1-day	+19.13	<0.001	513	Naive significantly better
7-day	+9.43	<0.001	513	Naive significantly better
30-day	+6.81	<0.001	513	Naive significantly better

Key Finding: While the DM statistic magnitude decreases with horizon length (from +19.13 to +6.81), naive persistence remains *significantly better* than the wavelet–transformer at all horizons ($p < 0.001$). The decreasing gap suggests that market inefficiencies, if any, are more likely exploitable at longer timescales—but even at 30-day horizons, our 2.1M-parameter model fails to outperform the zero-parameter baseline.

The decreasing DM statistic with horizon length reveals an interesting pattern: the naive baseline's advantage diminishes from 19.13 standard errors (1-day) to 6.81 standard errors (30-day). This could suggest that short-term price movements are dominated by noise (strongly efficient) and that longer horizons may contain a marginally more exploitable structure, yet even 30-day forecasting does not favor complex architectures.

This finding has important implications for algorithm design: if researchers seek opportunities for sophisticated models to add value, they should focus on even longer prediction horizons (quarterly, yearly) where market efficiency may be weaker. For horizons $\leq$30 days, our evidence strongly supports using simple baselines.

5.7. Cross-Asset Generalization: Non-Transferable Features

A critical test of any forecasting architecture is its ability to generalize beyond the training asset. We evaluated our Bitcoin-trained wavelet–transformer on Ethereum (ETH) and Binance Coin (BNB) without retraining, revealing a fundamental limitation of our approach.

Transfer Performance. When applied to Ethereum, RMSE degrades to USD 421,600—a **210-fold deterioration** compared to Bitcoin's USD 2005. For Binance Coin, the degradation is even more severe: an RMSE of USD 2.8M represents a **1400-fold deterioration**. These catastrophic failures effectively render cross-asset predictions meaningless.

Technical Explanation. The primary cause is the **Bitcoin-dominated nature of the Fear and Greed Index**. The alternative.me FGI is computed primarily from Bitcoin market data (surveys are currently paused): volatility (25%) based on Bitcoin 30-day volatility; market momentum (25%) comparing Bitcoin price to 30/90-day moving averages; social media sentiment (15%) from Bitcoin-related content on Twitter/Reddit; Bitcoin dominance (10%) measuring market cap share; and Google Trends (10%) tracking searches for "Bitcoin."

Thus, the wavelet decomposition of FGI captures Bitcoin-specific sentiment dynamics that do not transfer to altcoins with different market microstructures. This is not a weakness of our architecture, but rather a fundamental constraint: **the input features themselves are asset-specific**.

Implications for Algorithm Design. This finding directly supports our main thesis—that sophisticated multi-scale architectures do not provide generalizable advantages. The wavelet–transformer's 2.1M parameters are effectively tuned to Bitcoin's specific relationship with the FGI, and this specialization provides no utility for other assets. Practitioners must train separate models for each cryptocurrency, with separate sentiment indicators calibrated to each asset's market dynamics.

6. Discussion

6.1. Key Findings and Implications

Our systematic investigation reveals a finding that strongly supports the **Efficient Market Hypothesis** for short-term cryptocurrency forecasting: even sophisticated multi-scale wavelet–transformer architectures cannot significantly outperform the naive persistence

baseline. With our corrected log-return methodology, all model variants (including ARIMA, XGBoost, LSTM, and vanilla Transformer) achieve RMSEs in the range of USD 1992–USD 2611, compared to the naive baseline of USD 1986 (RMSE ratios of 1.003–1.32). This aligns with recent evaluations of large time series foundation models in crypto settings, which highlight both promise and practical limitations [40].

This result carries profound implications for practitioners. Rather than pursuing ever-more-complex architectures in search of marginal gains, researchers should acknowledge the fundamental limits of short-term price prediction in efficient markets. Our ablation studies reveal that only one architectural component—positional encoding—has substantial impact (30% RMSE degradation when removed), while cross-attention, skip connections, and encoder depth contribute negligible benefits.

6.2. Computational Cost vs. Predictive Gain

Table 6 presents the computational cost–benefit analysis for our models, quantifying the trade-off between architectural sophistication and predictive performance.

Table 6. Computational cost–benefit analysis: algorithmic complexity vs. predictive gain.

Model	Parameters	Inference Time	Complexity	RMSE Ratio
Naive (Persistence)	0	0.0006 ms	O(1)	1.000
ARIMA (1,1,1)	–	0.05 ms	$O(p+q)$	1.003
LSTM (two-layer)	150K	0.42 ms	$O(n \cdot h^2)$	1.017
Transformer (Vanilla)	1.2M	1.85 ms	$O(L \cdot n^2 \cdot d)$	1.056
Wavelet–Transformer	**2.1M**	**2.43 ms**	$\mathbf{O(L \cdot n^2 \cdot d)}$	**1.010**

Key Finding: The 2.1M-parameter wavelet–transformer requires 3909× longer inference time than the naive baseline (2.43 ms vs. 0.0006 ms) while delivering *worse* predictions (RMSE ratio 1.010 vs. 1.000). The computational cost–benefit is unambiguously negative.

The implications are stark. Our wavelet-transformer architecture requires 2.1 million trainable parameters versus zero for naive persistence, 3909× longer inference time (2.43 ms vs. 0.0006 ms), GPU training of 100 epochs on NVIDIA TITAN Xp hardware, and wavelet decomposition preprocessing for each input sequence.

In exchange for this computational investment, the model achieves an RMSE of USD 2005 versus USD 1986 USD for naive—a 1% *degradation* that is statistically significant ($p < 0.001$, Diebold–Mariano test). This constitutes a **negative return on computational investment**.

The question "Is a 2.1M-parameter transformer worth it?" receives an unambiguous answer: **No**. For short-term cryptocurrency forecasting, the simplest possible model (persistence) outperforms sophisticated deep learning architectures while requiring zero training, zero parameters, and O(1) inference complexity. This finding provides empirical support for the Efficient Market Hypothesis and carries important implications for resource allocation in algorithmic trading research.

6.3. Why Sophisticated Models Cannot Beat Naive Baselines

What explains the inability of even sophisticated architectures to outperform naive forecasting? Several factors emerge. First, **market efficiency**: cryptocurrency markets, despite their volatility, rapidly incorporate available information into prices, leaving little exploitable signal for any model at 1-day horizons; related evidence on efficiency versus long-memory structure in crypto assets is discussed by [41]. Second, **noise dominance**: short-term price movements are dominated by noise rather than predictable patterns, with our RMSE of ∼USD 2000 representing approximately 3.3% of the mean test price, consistent with typical daily BTC volatility. Third, **positional encoding matters**: the one

clear finding from our ablation studies is that positional encoding is critical (30% RMSE degradation when removed), suggesting that temporal ordering within sequences does contain some information, even if not enough to beat the naive baseline. Fourth, **overfitting risk**: complex architectures risk overfitting to noise, potentially explaining why simpler models perform comparably or better.

6.4. Limitations

Our work faces several important limitations, as follows:

1. **Fundamental Predictability Limits**: Our primary finding is *negative*—sophisticated architectures cannot outperform naive baselines for 1-day cryptocurrency forecasting. This limits the practical utility of our approach for trading applications.
2. **Directional Accuracy Near Random**: The DA of ∼50% across all model variants indicates that predicting price direction remains fundamentally challenging, consistent with EMH.
3. **Limited Fear Regime Data**: The use of only 15 samples in the fear regime (FGI < 25) during the test period limits confidence in regime-specific performance conclusions.
4. **Single Asset Focus**: We evaluated only Bitcoin. As detailed in Section 5.7, cross-asset generalization to ETH and BNB produces catastrophic failures (210–1400× RMSE degradation), a consequence of the Bitcoin-dominated Fear and Greed Index rather than architectural limitations.
5. **Single Sentiment Source**: Our analysis relies exclusively on the Fear and Greed Index. Incorporating additional sentiment sources could provide complementary signals.

6.5. Future Work

This work opens several promising research directions, as follows:

1. Multi-Source Sentiment Integration: Extending the framework to incorporate diverse sentiment indicators—news sentiment analysis, social media analytics, blockchain-specific on-chain metrics—could provide more comprehensive market psychology modeling and potentially improve directional accuracy.
2. Adaptive Wavelet Selection: Rather than using a fixed wavelet family, investigate adaptive selection mechanisms that adjust based on current market conditions, or develop automatic optimization procedures tailored to each cryptocurrency's unique statistical properties.
3. Real-time Deployment: Develop efficient streaming inference capabilities that enable real-time forecasting. This requires addressing the computational demands of wavelet decomposition while maintaining prediction accuracy under strict latency constraints.
4. Multi-Asset Portfolio Optimization: Extend the architecture to jointly model multiple cryptocurrencies simultaneously, enabling portfolio-level forecasting and risk management rather than treating each asset in isolation.
5. Explainability Enhancement: Develop sophisticated visualization techniques to understand precisely how different frequency bands contribute to final predictions, potentially revealing which timescales matter most under various market conditions.

7. Conclusions

This paper addresses a fundamental question in cryptocurrency forecasting: **Can sophisticated deep learning architectures outperform simple baselines, and if not, do they still provide value?** Our answer, supported by rigorous statistical evidence, is nuanced.

Using Diebold–Mariano tests with HAC-corrected variance across 513 out-of-sample predictions, we find that no deep learning architecture—including our 2.1M-parameter wavelet–transformer—significantly outperforms the naive persistence baseline (DM = +19.13, $p < 0.001$). Our model achieves an RMSE of USD 2005 compared to the naive baseline of USD 1986 (ratio: 1.010). However, **among deep learning approaches**, our wavelet–transformer delivers the best performance: 4.4% better than vanilla Transformer (USD 2098), 5.2% better than XGBoost (USD 2116), and comparable to LSTM (USD 2020). When deep learning solutions are required—for interpretability, multi-horizon forecasting, or integration with larger systems—the wavelet-enhanced architecture offers measurable advantages.

These results provide empirical support for the **Efficient Market Hypothesis** in short-term cryptocurrency markets. For 1-day prediction horizons, sophisticated architectures cannot extract profitable signals sufficient to overcome the naive baseline from publicly available data.

Crucially, our systematic ablation studies reveal that **the model genuinely learns temporal structure**. Positional encoding proves critical—its removal causes 30% RMSE degradation ($p < 0.001$)—demonstrating that the architecture captures meaningful sequential dependencies rather than producing random outputs. This validates the theoretical foundation of applying transformers to financial time series, even when fundamental market efficiency limits absolute predictive gains.

Our work carries important implications, as follows:

1. **Negative results advance science**: Demonstrating that complexity does not beat simplicity for short-term forecasting prevents wasted research effort and guides practitioners toward appropriate tools.
2. **Statistical rigor is essential**: Without Diebold–Mariano tests, we might have incorrectly claimed that small improvements represent real predictive gains.
3. **Context determines value**: When the goal is solely beating naive baselines, simpler models suffice; when deep learning is required for other reasons, wavelet-enhanced architectures outperform alternatives.
4. **The EMH holds for short-term crypto**: Despite claims of "market inefficiency," short-term cryptocurrency prices appear unpredictable from public data.
5. **Wavelet decomposition benefits deep learning**: Among neural architectures, multi-scale wavelet processing provides consistent improvements over single-scale approaches.

For future work, longer prediction horizons merit investigation, as our 7-day and 30-day results show progressively reduced gaps with naive baselines. Additionally, the demonstrated effectiveness of positional encoding suggests that further architectural innovations targeting temporal dependencies may yield improvements. While fundamental predictability limits constrain what any architecture can achieve in efficient markets, our findings establish wavelet-enhanced transformers as the preferred choice when deep learning solutions are warranted.

Author Contributions: Conceptualization, A.J. and R.B.; methodology, A.J.; software, A.J.; validation, A.J.; formal analysis, A.J.; investigation, A.J.; resources, R.B.; data curation, A.J.; writing—original draft preparation, A.J.; writing—review and editing, R.B.; visualization, A.J.; supervision, R.B.; project administration, R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by INFORGES IT SERVICES with contract number 24I277.

Institutional Review Board Statement: Not applicable. This study did not involve humans or animals.

Informed Consent Statement: Not applicable. This study did not involve humans.

Data Availability Statement: The data presented in this study are openly available. Bitcoin, Ethereum, and Binance Coin price data (OHLCV) were obtained from Binance and Coinbase APIs. The Fear

& Greed Index data were obtained from Alternative.me (https://alternative.me/crypto/fear-and-greed-index/, accessed on 5 November 2025). All data sources are publicly accessible. **Code and Reproducibility:** The complete codebase is available upon request from the authors. Reproducibility has been implemented in https://github.com/AldanJJay/wavelet_transf_reprod (accessed on 5 November 2025), which includes trained models, processed data, and scripts to verify all tables and figures in this paper.

Acknowledgments: During the preparation of this manuscript, the authors used Copilot (VS Code, November 2025) and Claude AI (Claude 4.5 opus, Anthropic) for code development assistance and manuscript editing. The authors reviewed and edited all outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funder (INFORGES) had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Nomenclature

Table A1 provides definitions for all abbreviations and symbols used throughout this paper.

Table A1. List of abbreviations and symbols used in this paper.

Abbreviation	Definition
Data and Metrics	
OHLCV	Open, High, Low, Close, Volume (daily price data)
FGI	Fear and Greed Index (sentiment indicator, 0–100 scale)
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
DA	Directional Accuracy (% of correct direction predictions)
MAPE	Mean Absolute Percentage Error
Technical Indicators	
RSI	Relative Strength Index (momentum, 14-day)
MACD	Moving Average Convergence Divergence
ADX	Average Directional Index
ATR	Average True Range (volatility measure)
OBV	On-Balance Volume
EMA	Exponential Moving Average
Wavelet Decomposition	
DWT	Discrete Wavelet Transform
db4	Daubechies-4 wavelet family
D1–D4	Detail coefficients (frequency bands)
A4	Approximation coefficient (lowest frequency band)
Model Architecture	
LSTM	Long Short-Term Memory network
GRU	Gated Recurrent Unit
ARIMA	AutoRegressive Integrated Moving Average
Statistical Tests	
DM	Diebold–Mariano test for forecast comparison
HAC	Heteroskedasticity and Autocorrelation Consistent
CV	Coefficient of Variation (std/mean $\times$ 100)

Appendix B. Hyperparameter Configuration

Tables A2 and A3 provide complete hyperparameter configurations for reproducibility.

Table A2. Model architecture hyperparameters.

Parameter	Our Model	Baselines
Transformer Architecture		
Embedding dimension (d_{model})	128	128
Attention heads (n_{head})	8	8
Encoder layers	2	2
Feedforward dimension	512	512
Dropout rate	0.1	0.1
Wavelet Configuration		
Wavelet family	db4	–
Decomposition level	4	–
Padding mode	symmetric	–
Number of bands	5 (D1–D4, A4)	–
Input Configuration		
Lookback window	60 days	60 days
Auxiliary features	12 indicators	varies
Prediction horizons	1, 7, 30 days	1 day
LSTM Baseline		
Hidden units	–	32
Number of layers	–	2
XGBoost Baseline		
Max depth	–	6
Learning rate	–	0.1
Number of estimators	–	100

Table A3. Training hyperparameters.

Parameter	Value
Optimizer	
Type	AdamW
Learning rate	1×10^{-3}
Weight decay	1×10^{-4}
Betas	(0.9, 0.999)
Learning Rate Schedule	
Scheduler	ReduceLROnPlateau
Reduction factor	0.5
Patience	5 epochs
Minimum LR	1×10^{-6}
Training	
Batch size	64
Maximum epochs	100
Gradient clipping	1.0
Early Stopping	
Patience	10 epochs
Minimum delta	0.001
Monitor metric	Validation loss
Multi-Task Loss Weights	
1-day horizon	0.5
7-day horizon	0.3
30-day horizon	0.2
Data Splitting	
Training set	60%
Validation set	20%
Test set	20%

References

1. Bouteska, A.; Mefteh-Wali, S.; Dang, T. Predictive power of investor sentiment for Bitcoin returns: Evidence from COVID-19 pandemic. *Technol. Forecast. Soc. Change* **2022**, *184*, 102044. [CrossRef]
2. Koltun, V.; Yamshchikov, I.P. Pump It: Twitter sentiment analysis for cryptocurrency price prediction. *Risks* **2023**, *11*, 159. [CrossRef]
3. Brigato, L.; Morand, R.; Strømmen, K.; Panagiotou, M.; Schmidt, M.; Mougiakakou, S. There Are No Champions in Supervised Long-Term Time Series Forecasting. *arXiv* **2025**, arXiv:2502.14045. [CrossRef]
4. Zhou, Y.; Wang, Y.; Goel, S.; Zhang, A.R. Why Do Transformers Fail to Forecast Time Series In-Context? *arXiv* **2025**, arXiv:2510.09776. [CrossRef]
5. Wang, J.; Vigne, S.A.; Liu, H.; Hsu, Y. A U-shaped relationship between the crypto fear-greed index and the price synchronicity of cryptocurrencies. *Financ. Res. Lett.* **2024**, *59*, 154461. [CrossRef]
6. Gaies, B.; Nakhli, M.S.; Sahut, J.; Schweizer, D. Interactions between investors' fear and greed sentiment and Bitcoin prices. *N. Am. J. Econ. Financ.* **2023**, *67*, 101657. [CrossRef]
7. Osman, M.B.; Urom, C.; Guesmi, K.; Benkraiem, R. Economic sentiment and the cryptocurrency market in the post-COVID-19 era. *Int. Rev. Financ. Anal.* **2024**, *91*, 102941. [CrossRef]
8. Su, L.; Zuo, X.; Li, R.; Wang, X.; Zhao, H.; Huang, B. A systematic review for transformer-based long-term series forecasting. *Artif. Intell. Rev.* **2025**, *58*, 80. [CrossRef]
9. Parvini, N.; Abdollahi, M.; Seifollahi, S.; Ahmadian, D. Forecasting Bitcoin returns with long short-term memory networks and wavelet decomposition: A comparison of several market determinants. *Appl. Soft Comput.* **2022**, *121*, 108707. [CrossRef]
10. Bhuiyan, R.A.; Husain, A.; Zhang, C. Diversification evidence of bitcoin and gold from wavelet analysis. *Financ. Innov.* **2023**, *9*, 100. [CrossRef] [PubMed]
11. Yu, P.; Kong, H.; Li, Z. Wavelet-Enhanced Transformer for Adaptive Multi-Period Time Series Forecasting. *Appl. Sci.* **2025**, *15*, 12698. [CrossRef]
12. Fama, E.F. Efficient capital markets: A review of theory and empirical work. *J. Financ.* **1970**, *25*, 383–417. [CrossRef]
13. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1970.
14. Chen, W.; Xu, H.; Jia, L.; Gao, Y. Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *Int. J. Forecast.* **2021**, *37*, 28–43. [CrossRef]
15. Khosravi, M.; Ghazani, M.M. Novel insights into the modeling financial time-series through machine learning methods: Evidence from the cryptocurrency market. *Expert Syst. Appl.* **2023**, *234*, 121012. [CrossRef]
16. Seabe, P.L.; Moutsinga, C.R.B.; Pindza, E. Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach. *Fractal Fract.* **2023**, *7*, 203. [CrossRef]
17. Kaur, R.; Uppal, M.; Gupta, D.; Juneja, S.; Arafat, S.Y.; Rashid, J.; Kim, J.; Alroobaea, R. Development of a cryptocurrency price prediction model: Leveraging GRU and LSTM for Bitcoin, Litecoin and Ethereum. *PeerJ Comput. Sci.* **2025**, *11*, e2675. [CrossRef]
18. Vallarino, D. An AI-enhanced forecasting framework: Integrating LSTM and transformer-based sentiment for stock price prediction. *J. Econ. Anal.* **2025**, *4*, 109. [CrossRef]
19. Omole, O.; Enke, D. Deep learning for Bitcoin price direction prediction: Models and trading strategies empirically compared. *Financ. Innov.* **2024**, *10*, 117. [CrossRef]
20. Peng, P.; Chen, Y.; Lin, W.; Wang, J.Z. Attention-based CNN–LSTM for high-frequency multiple cryptocurrency trend prediction. *Expert Syst. Appl.* **2024**, *237*, 121520. [CrossRef]
21. Tsay, R.S. *Analysis of Financial Time Series*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2010.
22. Liu, C.; Mahanti, A.; Naha, R.; Wang, G.; Sbai, E. Enhancing Cryptocurrency Sentiment Analysis with Multimodal Features. *arXiv* **2025**, arXiv:2508.15825. [CrossRef]
23. Zhang, Z.; Jiang, C.; Lu, M. Fusion of Sentiment and Market Signals for Bitcoin Forecasting: A SentiStack Network Based on a Stacking LSTM Architecture. *Big Data Cogn. Comput.* **2025**, *9*, 161. [CrossRef]
24. Farrell, H.; O'Connor, F. The CNN Fear and Greed Index as a predictor of US equity index returns: Static and time-varying Granger causality. *Financ. Res. Lett.* **2025**, *72*, 106492. [CrossRef]
25. Tang, Q.; Shi, R.; Fan, T.; Ma, Y.; Huang, J. Prediction of Financial Time Series Based on LSTM Using Wavelet Transform and Singular Spectrum Analysis. *Math. Probl. Eng.* **2021**, *2021*, 9942410. [CrossRef]
26. Liang, X.; Ge, Z.; Sun, L.; He, M.; Chen, H. LSTM with Wavelet Transform Based Data Preprocessing for Stock Price Prediction. *Math. Probl. Eng.* **2019**, *2019*, 1340174. [CrossRef]
27. Zhou, Z.; Xiong, Y.-J.; Zhang, J.-C.; Xia, C.-M.; Xie, X.-J. Wavelet Mixture of Experts for Time Series Forecasting. *arXiv* **2025**, arXiv:2508.08825. [CrossRef]
28. Masserano, L.; Ansari, A.F.; Han, B.; Zhang, X.; Faloutsos, C.; Mahoney, M.W.; Wilson, A.G.; Park, Y.; Rangapuram, S.S.; Maddix, D.C.; et al. Enhancing Foundation Models for Time Series Forecasting via Wavelet-based Tokenization. In Proceedings of the ICML 2025, Vancouver, BC, Canada, 13–19 July 2025.

29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
30. Rubio, L.; Palacio Pinedo, A.; Mejía Castaño, A.; Ramos, F. Forecasting volatility by using wavelet transform, ARIMA and GARCH models. *Eurasian Econ. Rev.* **2023**, *13*, 803–830. [CrossRef]
31. Hajek, P.; Hikkerova, L.; Sahut, J. How well do investor sentiment and ensemble learning predict Bitcoin prices? *Res. Int. Bus. Financ.* **2023**, *64*, 101546. [CrossRef]
32. Zhang, Z.; Dai, H.N.; Zhou, J.; Mondal, S.K.; García, M.M.; Wang, H. Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels. *Expert Syst. Appl.* **2021**, *183*, 115378. [CrossRef]
33. Bouteska, A.; Abedin, M.Z.; Hajek, P.; Yuan, K. Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods. *Int. Rev. Financ. Anal.* **2024**, *92*, 103055. [CrossRef]
34. Murray, K.; Rossi, A.; Carraro, D.; Visentin, A. On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles. *Forecasting* **2023**, *5*, 196–209. [CrossRef]
35. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]
36. Hewamalage, H.; Ackermann, K.; Bergmeir, C. Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Min. Knowl. Discov.* **2023**, *37*, 788–832. [CrossRef]
37. Coroneo, L.; Iacone, F. Testing for equal predictive accuracy with strong dependence. *Int. J. Forecast.* **2024**, *40*, 1093–1111. [CrossRef]
38. Fama, E.F. The behavior of stock-market prices. *J. Bus.* **1965**, *38*, 34–105. [CrossRef]
39. Taylor, S.J. *Modelling Financial Time Series*, 2nd ed.; World Scientific: Singapore, 2008.
40. Wang, M.; Braslavski, P.; Ignatov, D.I. TimeGPT's Potential in Cryptocurrency Forecasting: Efficiency, Accuracy, and Economic Value. *Forecasting* **2025**, *7*, 48. [CrossRef]
41. Jayasankar, M. Efficient Market Hypothesis Versus Multifractality: Evidence from the Stablecoin Market. *Comput. Econ.* **2025**, *66*, 5033–5054. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.