An Explainable Multimodal Framework for Real-Time Bitcoin Forecasting

Dipesh Badal
Faculty of Computing, Digital and
Data
Technological University Dublin
Dublin, Leinster, Ireland
x00229547@mytudublin.ie

Abdelsalam Busalim
Faculty of Computing, Digital and
Data
Technological University Dublin
Dublin, Leinster, Ireland
Abdelsalam.Busalim@TUDublin.ie

Donghyeok Lee
CeADAR - Ireland's Centre for AI
Dublin, Leinster, Ireland
donghyeok.lee@ucd.ie

Abstract

High-frequency crypto forecasting requires systems that are accurate, explainable, and designed for human decision-making. Bitcoin presents a unique challenge for Human-Centred AI (HCAI) due to its volatility and sensitivity to heterogeneous technical, fundamental, and sentiment signals. This paper presents an explainable multimodal framework for Bitcoin forecasting at 15-minute resolution. We align five modalities—market data, on-chain metrics, the Fear & Greed Index (FGI), news, and Reddit—onto a unified, leakage-safe 15-minute grid. We evaluate tree-based, sequential, and Multimodal Fusion Block (MFB) models for next-interval log-return prediction using chronological splits. Results show that while short-horizon prediction remains challenging, multimodal features consistently improve over structured baselines, particularly during event-driven periods. To ensure transparency, the framework integrates a dual-layer explanation system: SHapley Additive exPlanations (SHAP) attributions combined with large language model (LLM) narratives, ensuring outputs are both technically faithful and human-accessible. This work unlocks the "black box" of complex predictive architectures, transforming opaque multimodal signals into transparent, actionable decision support for high-frequency trading.

CCS Concepts

• **Computing methodologies** → **Machine learning**; • **Applied computing** → *Economics*; • **Information systems** → *Sentiment analysis*; • **Human-centered computing** → Visualization design and evaluation methods.

Keywords

Bitcoin Forecasting, Multimodal Data, Machine Learning, Deep Learning, Explainable AI, SHAP

1 Introduction

Cryptocurrencies have evolved from niche experiments into a trillion-dollar financial ecosystem, with **Bitcoin (BTC)** as the dominant asset. Unlike traditional securities, BTC trades continuously, experiences extreme volatility, and is influenced not only by technical indicators but also by blockchain fundamentals, regulatory shocks, and social sentiment [9, 15, 30]. Conventional econometric models, which assume stable and linear relationships, are poorly suited to these regime shifts and nonlinear behaviors [5, 9]. In contrast, machine learning (ML) and deep learning (DL) approaches offer greater flexibility, particularly when paired with multimodal signals such as financial news, Reddit discussions, and on-chain metrics [26, 27, 29]. Yet these models are often criticized as "black boxes." In high-stakes financial domains, where predictions directly guide investment and risk decisions, interpretability is critical [3, 4, 31]. Recent advances in **Explainable Artificial Intelligence (XAI)** provide tools to address this challenge: feature attribution methods such as SHAP can reveal the drivers of model outputs, while large language models (LLMs) can transform such technical attributions into concise, human-readable narratives. Together, these complementary techniques enable predictive systems that are not only accurate but also transparent and accessible [16, 32].

Real-time BTC forecasting, however, presents several interlinked challenges. Heterogeneous signals from markets, blockchain activity, financial news, and social platforms must be integrated, despite operating on different timescales and formats [1, 17]. Most prior studies examine daily or hourly horizons, yet real-world trading requires much finer granularity, such as 15-minute intervals, to support timely decisions [7, 19]. Although advanced ML and DL models can capture hidden nonlinear dependencies, their lack of interpretability reduces practical adoption [3, 31]. Furthermore, the evaluation practices in all studies remain inconsistent, limiting comparability and reproducibility [29]. These challenges motivate the central question of this work: ***how can multimodal data integration and explainability techniques be combined to create a reliable, interpretable system for real-time BTC forecasting?***

From this problem statement, four guiding research questions emerge:

RQ1: How can multimodal signals—prices, on-chain metrics, news, Reddit, and sentiment indices—be aligned on a 15-minute grid to predict next-interval BTC log-returns, with forecasts mapped back to a price path?

RQ2: What is the comparative performance of tree-based, sequential, hybrid, and multimodal fusion block (MFB) architectures?

RQ3: How can reliable explanations for short-horizon forecasts be generated without degrading accuracy?

RQ4: How does explainable multimodal AI influence trading outcomes by mitigating over-trust and enabling abstention in low-confidence scenarios?

To answer these questions, the study pursues three objectives: (i) to integrate heterogeneous sources into a unified 15-minute dataset with advanced sentiment analysis, (ii) to design and evaluate a range of forecasting architectures including an adapted MFB, and (iii) to embed explainability through feature attribution and natural-language outputs. The resulting contributions are sixfold. First, the study constructs one of the most extensive multimodal BTC datasets to date, aligned at 15-minute resolution and integrating market, on-chain, news, Reddit, and Fear & Greed Index (FGI) signals [1, 17]. Second, it develops a Modern FinBERT-based sentiment framework enriched with dictionary-based event tagging, producing interpretable text-derived features [6, 12]. Third, it adapts the novel MFB architecture for high-frequency BTC log-return prediction, employing modality-specific towers that support attribution-friendly analysis [14]. Fourth, it introduces a dual-layer explainability framework that combines SHAP-based attributions with LLM-generated narratives to deliver both technical fidelity and human-centred interpretability [3, 16]. Fifth, it enforces methodological rigor via strict leakage-safe protocols and a modular, reproducible pipeline [13, 29]. Sixth, it advances return-based forecasting by targeting log-returns rather than raw closing prices at 15-minute resolution, supported by a side experiment demonstrating that price forecasting yields artificially inflated R^2 values.

Collectively, these contributions establish a transparent, reproducible, and human-centred framework for explainable real-time cryptocurrency forecasting. By uniting multimodal data integration, advanced predictive architectures, and dual-layer explanations, this work offers both methodological innovation and practical value: a system capable of transforming weak, noisy signals into interpretable decision support for analysts, traders, and regulators in volatile crypto markets.

2 Literature Review

The expansion of cryptocurrency markets has stimulated research at the intersection of financial forecasting, ML, and behavioral analysis. This section reviews prior work most relevant to this study: BTC price prediction, multimodal data integration, sentiment analysis and financial NLP, XAI in finance, and methodological considerations. The aim is to situate this study within current scholarship, highlight limitations, and identify the gaps it addresses [2, 17, 29].

2.1 Evolution of Cryptocurrency Price Prediction

Early work used econometric models such as ARIMA and GARCH, which provided volatility baselines but proved inadequate for BTC's 24/7 trading, regime shifts, and nonlinear dynamics [9]. ML methods introduced flexibility, with random forests and gradient boosting capturing complex dependencies [30]. DL approaches, particularly LSTMs, improved sequential modeling [26], while CNNs, CNN–LSTM hybrids, and Transformers modeled both short- and long-range dependencies [8, 19]. Despite strong in-sample accuracy,

many models overfit, show regime-dependent performance, and lack interpretability [22, 29].

2.2 Multimodal Data Integration

Forecasting has moved beyond OHLCV prices to include blockchain fundamentals (e.g., active addresses, fees, miner revenue) and macro-financial indicators [1, 9]. Unstructured sources such as news, Reddit, Twitter, and Google Trends capture sentiment and attention, though they introduce noise, alignment issues, and susceptibility to manipulation [6, 23]. Most multimodal studies remain daily or hourly [20, 30]; sub-hourly integration is rare despite its importance for trading [7]. The Multimodal Fusion Block (MFB) [14] combines recurrent layers, feature selection, and attention to fuse lagged sentiment with technical indicators, achieving strong next-hour accuracy. However, such work remains price-focused; this study extends MFB to log-return forecasting, which is scale-free and finance-appropriate.

2.3 Sentiment Analysis and Financial NLP

Sentiment analysis progressed from lexicon methods to supervised ML and now transformers. FinBERT and crypto-specific variants capture financial slang and memes [12, 33]. These improve forecasting but face regime-dependent sentiment–return links, evolving vocabularies, and limited multilingual coverage. Combining transformer embeddings with event dictionaries has proven effective [6] and is adopted here.

2.4 Explainable AI in Finance

XAI has become critical for compliance and adoption. Attribution methods such as SHAP decompose predictions into feature contributions and are widely used in trading, portfolio allocation, and risk management [3, 28]. Case studies show SHAP can both increase investor trust [3] and expose dominant BTC drivers [28]. Beyond attribution, dashboards integrating SHAP into investor tools [4, 11] and hybrid models applying SHAP to volatility–LSTM systems [22] illustrate the growing trend toward interpretable finance AI. More recently, LLMs augment SHAP by translating numerical attributions into textual narratives, improving usability and decision quality [16, 32]. This dual-layer perspective underpins the explainability framework here.

2.5 Methodological Considerations and Gaps

Financial ML faces three recurring pitfalls: (i) **data leakage**, often from random cross-validation, requiring chronological splits [13]; (ii) **metrics**, where MAE/RMSE overlook trading utility, making DA and DA-DB more relevant [15]; and (iii) **non-stationarity**, where regime shifts undermine static models, motivating rolling windows [17, 29].

Despite progress, several gaps remain:

- Few studies attempt high-frequency multimodal integration [1, 7].
- Most work predicts prices rather than log-returns, which are scale-free and finance-appropriate [14].
- Benchmarks and datasets remain non-standardized, limiting comparability [29].
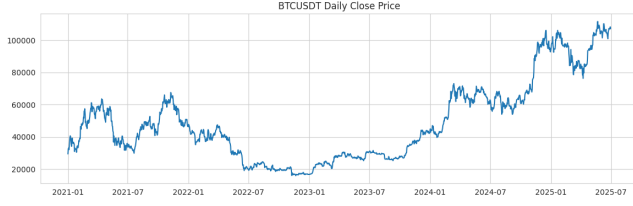
Figure 1: BTC daily close price history (15m granularity, 2021–2025).

- Interpretability is limited, with little user-centered explanation design for traders or regulators [11].

This study addresses these gaps by constructing a 15m multimodal dataset, extending MFB to return forecasting, integrating FinBERT with event tags, enforcing leakage-safe rolling evaluation, and implementing dual-layer XAI through SHAP and LLMs.

3 Data Collection and Preparation

Reliable short-term BTC forecasting requires a dataset that reflects heterogeneous market drivers. Unlike equities or commodities, BTC is shaped simultaneously by trading microstructure, on-chain fundamentals, regulatory/macro news, and online sentiment; capturing these forces necessitates a *multimodal* approach that unifies structured and unstructured signals [1, 9].

We construct a five-modality dataset aligned to a strict 15-minute (15m) UTC grid spanning **2021-01-01** to **2025-06-30** (Figure 1). Market price data define the canonical timeline; on-chain metrics, the Fear & Greed Index (FGI), financial/crypto news, and Reddit discussions are resampled/aggregated onto this backbone. Feature engineering enriches each source with domain-specific indicators under leakage-safe, reproducible policies [17, 24, 29].

3.1 Data Sources and Processing

Market (Binance BTC/USDT, 15m). *Rationale:* backbone for price/flow context [6, 19]. Klines standardized to UTC; missing 15m bins inserted to ensure **96 records/day**; rare OHLC gaps forward-filled; trade-count zeros where appropriate. Features: (i) technical indicators (SMA/EMA, RSI, MACD, Bollinger Bands, OBV), (ii) cyclical encodings (sine–cosine hour/day), (iii) lags (1–13 for close/volume). Target: next-interval *log-return* [6, 19].

On-chain (Dune). *Rationale:* adoption, demand, and cost [2, 20]. Metrics aggregated to 15m and aligned to Binance: `tx_count`, `active_addresses`, `total_btc_transferred`, `miner_revenue_btc`, `avg_tx_fee_btc`, `avg_block_size_bytes`, `avg_block_interval_secs`.

Fear & Greed Index (FGI). *Rationale:* macro sentiment anchor [23]. Daily values upsampled to 96 bins/day; features include `fgi_value`, `fgi_delta` (D/D), `fgi_7d_avg`, and categorical encodings.

News (FMP). *Rationale:* institutional narratives and shocks [8, 21, 30]. ~511k articles deduplicated, UTC-normalized, scored by Modern FinBERT for sentiment probabilities/labels [12, 33]; a dictionary applies 19 event tags (e.g., regulation, adoption, ETF, hacks). Aggregated to 15m: `news_sentiment_mean`; bullish/bearish/neutral counts; event flags; `news_count` (flow).

Reddit (PRAW). *Rationale:* retail narratives and early community reactions [6, 18]. ~189k posts/comments (major crypto subreddits); non-English/very short/deleted removed; PII stripped; processed with Modern FinBERT + same 19-event dictionary. 15m aggregates: `reddit_sentiment_mean`; bullish/bearish/neutral counts; event flags; engagement (`reddit_upvotes`); volume (`reddit_count`); `reddit_is_post`.

3.2 Alignment, Imputation, and Validation

All modalities are synchronized to the canonical 15m Binance grid to prevent drift [25]. *Modality-aware imputation:* counts/flags → zero-fill; continuous metrics (price, on-chain, FGI) → forward-fill with a *single initial backfill* if needed. Validation ensures **96 records/day**, no duplicate/missing bins after resampling, and cross-modality coherence (e.g., event surges align with news/Reddit volumes) [5, 7, 27].

3.3 Dataset Summary and Properties

The master table spans ~**157,000** aligned 15m intervals (2021-01-01 12:15–2025-06-30 23:30), integrating technical, fundamental, macro-sentiment, and narrative signals as summarised in Table 1:

Table 1: Modalities and representative features at 15m resolution.

Modality	Feature Groups	Examples	Prefix
Market	Technical, cyclical, lags, target	`SMA_20`, `RSI_14`, `lag_close_3`	—
On-chain	Levels, deltas, momentum	`active_addresses`, `tx_count_roc`, `fee_per_tx`	—
FGI	Daily macro sentiment	`fgi_value`, `fgi_delta`, `fgi_7d_avg`	`fgi_`
News	Sentiment, events, volume	`news_sentiment_mean`, `news_has_regulation`, `news_count`	`news_`
Reddit	Sentiment, events, engagement	`reddit_sentiment_mean`, `reddit_upvotes`, `reddit_has_etf`	`reddit_`

Key properties. *Comprehensive* (technical, fundamental, behavioral); *leakage-safe* (features use only past/present information) [13]; *interpretable* (clear prefixes/event flags); and *reproducible* (canonical 15m grid, documented pipeline) [29]. This multimodal dataset (`final_master_df`) underpins the pre-model pipeline in Section 4.

4 Prediction Models and Pipeline

With the multimodal dataset aligned on a strict 15-minute (15m) UTC grid (Jan 2021–Jun 2025), we convert raw signals into model-ready artifacts and train four forecasting families: tree ensembles, sequence models, a residual hybrid, and a modality-aware fusion block [2, 14, 17, 29]. The pipeline prevents leakage, stabilises statistics, and ensures reproducibility. The complete methodology is visualised in Figure 2.
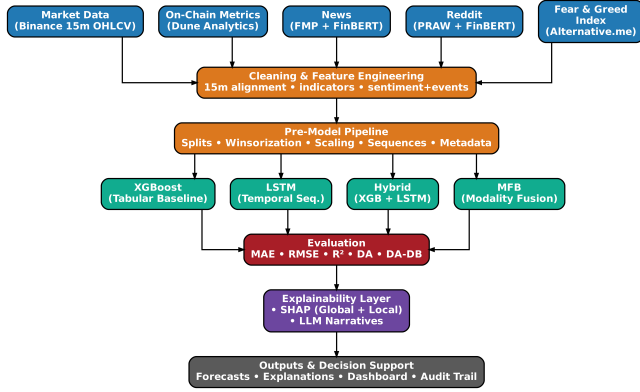
Figure 2: Leakage-safe pipeline and model families.

4.1 Pipeline Objectives

Five rules guide preprocessing: (i) **no leakage**—fit all transforms on *train only*; (ii) produce both *tabular* and *sequential* artifacts from the same index; (iii) **winsorize** heavy tails on train at **0.5%/99.5%** and reuse the caps; (iv) **prune** near-constant features and pairs with $|\rho| \geq 0.995$; (v) **log** split boundaries, scalers, caps, feature lists, and hashes for exact reruns.

4.2 Target and Leakage Controls

We forecast the next-interval **log-return**:

$$y_{t+1} = \ln\left(\frac{Close_{t+1}}{Close_t}\right),$$

a standard, scale-free target aiding stationarity [2, 10]. To avoid leakage we drop columns resembling future labels (y, next_*, future_*, target_*), fit winsorization/scaling/imputation on *train* and reuse on val/test, and remove NaN/Inf rows *after* chronological splitting.

4.3 Temporal Encodings and Feature Sets

Time-of-day and day-of-week use sine/cosine encodings to preserve circularity. We evaluate: (i) **Structured** (market/technical, on-chain, FGI), and (ii) **Multimodal** (Structured + news + Reddit).

4.4 Splits, Noise, Sequences, and Artifacts

We adopt a **70/15/15** chronological split (train/val/test, 2021–2025) and rolling **6→6 month** windows to probe regime effects [17, 25]. Heavy-tailed variables (e.g., volumes, news_count, reddit_count) are winsorized on train; features are scaled with train-fitted scalers (robust where appropriate). Sequential datasets use a **lookback of 64** steps (≈16h): inputs $[t, \ldots, t+63]$ predict y_{t+64}. *Recorded from train only:* the **deadband** threshold ϵ (60th percentile of $|y|$) for DA-DB and the **mean/std** of y. *Artifacts:* (a) unscaled tabular splits ($X_{\text{train/val/test}}$, aligned y); (b) scaled NPZ sequence bundles; (c) meta.json (hashes, splits, ϵ, caps, pruned lists, dtypes, scalers); and (d) modality maps (e.g., mfb_index_maps.json).

4.5 Model Families

XGBoost (XGB). Strong tabular baseline for mixed-scale data [27, 29]. Tuning primarily via n_estimators; refit on train + val; report metrics, prediction plots, and gain importances.

LSTM. Two stacked LSTMs with dropout and dense head trained on 64-step sequences (Adam 10^{-3}, batch 256, early stopping) [6, 33].

Residual Hybrid (XGB+LSTM). Train XGB, form leakage-safe residuals $r = y - \hat{y}_{\text{XGB}}$, fit LSTM on residual sequences, combine with shrinkage α from validation [12, 24].

Multimodal Fusion Block (MFB). Five towers (market, on-chain, FGI, news, Reddit): SpatialDropout1D → BiLSTM → BiGRU → attention → Dense(32); concatenate embeddings and map to output (Adam 10^{-3}, batch 256, early stopping) [14, 20].

4.6 Evaluation Metrics

We report **MAE**, **RMSE**, and R^2:

$$\text{MAE} = \frac{1}{N}\sum|y_i - \hat{y}_i|, \quad \text{RMSE} = \sqrt{\frac{1}{N}\sum(y_i - \hat{y}_i)^2}, \quad R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}.$$

We also use **DA** and **DA-DB**, ignoring $|y| < \epsilon$ (train-derived) to reflect trading relevance [15].

The pipeline aligns all data on a 15m grid, prevents leakage, stabilizes heavy tails, and produces reproducible tabular/sequence artifacts. On this base, we train XGB, LSTM, Hybrid, and MFB and evaluate them on global and rolling splits, setting up the results (Section 5) and explainability analyses [13, 23, 31].

5 Experimental Results

We evaluate XGB, LSTM, Hybrid (XGB+LSTM), and MFB for 15m log-return forecasting. Splits: global chronological 70/15/15 (2021–2025) and rolling 6→6-month windows. Direction uses deadband-adjusted accuracy (DA-DB) with ϵ=0.001755 (the train 60th percentile of $|y|$). Negative R^2 at 15m is common due to low explained variance [17].

5.1 Global Results

Table 2 shows global test results. MFB attains the lowest RMSE, closely followed by LSTM; XGB is competitive. DA-DB stays near 50–51%, consistent with short-horizon efficiency. Adding news/Reddit gives a modest, consistent lift over structured inputs.

Table 2: Global test performance (15m log-returns). Lower RMSE/MAE better; higher DA-DB better.

Model (Input)	RMSE	MAE	R^2	DA-DB
XGB (Struct.)	0.002778	0.001842	-0.064	0.503
XGB (Multi.)	0.002772	0.001836	-0.059	**0.504**
LSTM (Struct.)	0.002719	0.001801	-0.017	0.491
LSTM (Multi.)	0.002705	0.001785	-0.007	0.493
Hybrid (Multi., α=0.211)	0.002778	0.001842	-0.062	0.504
MFB (Multi.)	**0.002699**	**0.001778**	≈ -0.002	0.491

5.2 Rolling Windows (Regime Sensitivity)

Table 3 summarizes test results across rolling 6→6-month windows. Errors peak in 2022 (turbulence) and are smallest in 2023. DA-DB remains near chance (49–51.5%) regardless of model.

Table 3: Rolling-window test performance summary.

Model	RMSE (min)	RMSE (median)	RMSE (max)	DA-DB (median)
XGB (Multi.)	0.002301	0.002987	0.004403	0.501
MFB (Multi.)	0.002395	0.003204	0.005233	0.499

5.3 Insights & Robustness

Drivers: price/trend indicators dominate (EMA/SMA, Bollinger, short lags); on-chain fees and select text features contribute episodically. **Multimodal lift:** removing news/Reddit closes the small gap to structured inputs. **Stability:** results robust to deadband perturbations and random seeds (ΔRMSE $\sim 10^{-5}$). **Hybrid:** no gain—little residual signal after XGB.

5.4 Model Choice for Explainability

We adopt **XGB (multimodal)** for XAI due to (i) *exact* TreeSHAP for fast, reliable attributions [31]; (ii) stable global/local explanations; (iii) clear multimodal splits; and (iv) similar accuracy to LSTM/MFB at lower cost [2].

5.5 Price vs. Return

Forecasting next-price (instead of return) yields much higher scores (test $R^2 \approx 0.827$) but is scale-driven and less actionable [22]; log-returns remain the finance-appropriate target.

6 Explainability

We explain the final **XGB (multimodal)** model using a dual-layer framework: (i) *global attributions* with TreeSHAP to show which features matter most; and (ii) *local narratives* that translate individual 15m forecasts into plain-language summaries. This combines SHAP's fidelity [28] with LLM-generated rationales for accessible, human-centred insights [4, 16, 31, 32]. All explanations are computed on the held-out **test** split.

6.1 Methods

Global SHAP. SHAP values are computed across the test set and summarised with beeswarm plots as shown in Figure 3 (Top), capturing both feature importance and directional effects. **Local SHAP.** Figure 3 (Bottom) represents force plots that decompose single predictions into positive/negative contributions relative to the model baseline [28]. **Narratives.** Top drivers are automatically described by an LLM (direction, key contributors, their push up/down). The LLM does not alter SHAP, only translates it [16, 24, 32].

6.2 Results

Global attributions. Price/trend indicators (EMA/SMA, Bollinger, lags) dominate; momentum (MACD), trading activity (volume, trades), and macro mood (FGI level/change) provide incremental stability. Rankings are consistent with XGB gain importances and LSTM permutation insights. Findings mirror prior studies highlighting technical drivers with episodic sentiment lift [6, 15, 30]. **Local narratives.** At short horizons, opposing drivers often cancel. Examples: (A) flat—FGI↑, EMA(50)↑ vs. SMA(20)↓; (B) slight down—SMA(20), MACD, FGI drop; (C) slight up—EMA(50)↑ offset by SMA(20), Bollinger mid.

6.3 Narratives Dashboard

Figure 4 shows a prototype dashboard where each forecast is paired with its SHAP plot and a short narrative, turning quantitative attributions into actionable, audit-ready insights for analysts and regulators [11, 23, 24].
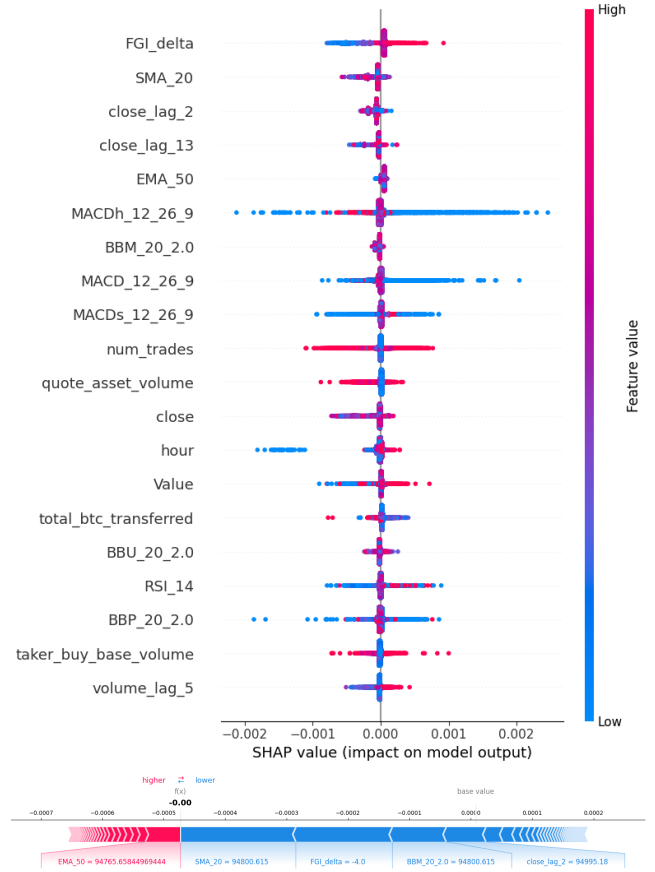

Figure 3: (Top) Global SHAP beeswarm; (Bottom) Local SHAP force plot (example instance).
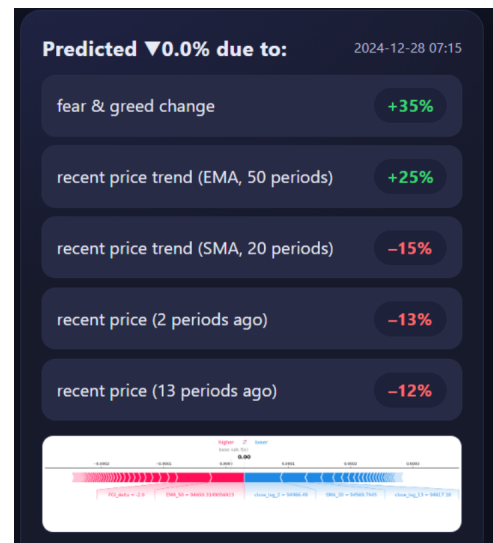

Figure 4: Prototype dashboard: SHAP with LLM narratives.

6.4 Checks and Use

TreeSHAP is deterministic and consistent with other importance measures. Explanations describe *model behaviour*, not causality; correlated variables may share ranks [31]. In live use, forecasts can surface as: (i) a compact *driver card* (top contributors + signs), (ii) a one-line narrative, and (iii) an abstain cue when signals conflict.

Price/trend and momentum dominate; sentiment adds modest stability. Local SHAP + LLM narratives show many 15m forecasts balance out, yielding small moves. The dual-layer approach—technical attributions plus human-readable summaries—makes outputs transparent, accessible, and usable for human-in-the-loop trading and oversight [3, 4, 11, 31].

7 Discussion and Limitations

7.1 Forecasting Performance

At the 15-minute horizon, all models achieved low errors (RMSE ≈ 2.7–2.9×10^{-3}) but only $\approx 50\%$ directional accuracy, consistent with high-frequency market efficiency [7, 29]. Predictions are thus weak statistical edges requiring filters or abstention logic. LSTM and MFB slightly reduced RMSE relative to XGB, but directional accuracy stayed near chance. The residual hybrid offered no benefit, confirming that naïve ensembles cannot exploit residual signal. Regime effects were stronger than model differences: 2022 produced the weakest forecasts, while calmer 2023 periods yielded lower errors. A side experiment with next-close prices inflated R^2, underscoring log-returns as the scale-free and finance-appropriate target [14, 17].

7.2 Multimodal Signals

Global SHAP confirmed that technical indicators—moving averages, Bollinger bands, MACD, and short lags—dominated consistently. On-chain activity and sentiment contributed situationally, mainly during shocks or news-driven periods, echoing prior evidence that multimodality helps chiefly in event regimes [6, 20]. Simply adding modalities cannot overcome structural noise; regime-aware adaptation remains essential.

7.3 Value of Explainability

SHAP revealed shifting feature importance across regimes, clarifying why accuracy plateaued near 52%. The LLM layer translated attributions into concise narratives, surfacing fragile forecasts and supporting abstention. This dual-layer approach aligns with prior XAI-in-finance work [16, 32] and functions less as a post-hoc justification and more as a safeguard against over-trust.

7.4 Practical Implications

For deployment: (i) forecasts should be combined with thresholds, deadbands, or volatility filters; (ii) multimodal inputs add value mainly during event-driven periods; (iii) regime awareness is critical, suggesting classifiers or probabilistic outputs to adapt to volatility phases; and (iv) dual-layer explainability is indispensable for human-in-the-loop use, reducing automation bias and aligning with human-centred AI principles [11, 31].

7.5 Limitations and Ethics

Limitations span: **data noise** (*timestamp misalignments*, vulnerability to intentional manipulation via inauthentic or coordinated social media content, winsorization trimming extremes), **modeling scope** (*relatively simple LSTM/MFB*; no transaction-cost-aware backtests), and **explainability workflow** (SHAP reflects model logic not causality; correlated features can swap ranks; LLM narratives risk oversimplification or prompt sensitivity; SHAP+LLM adds computational cost for real-time use, demanding a resource-footprint versus value assessment. We partially mitigate this by employing **XGBoost** and **TreeSHAP**, which are computationally efficient, but a formal **cost-benefit analysis** (including environmental footprint) remains a necessary step for deployment.). Ethical concerns include **automation bias** (*users over-trusting simplified narratives*), **data rights** (compliance with Reddit/news platforms), and **transparency** (*model cards with splits, scalers, caps, ϵ are needed for accountability*). Despite these constraints, results show that explainability transforms weak predictive signals into safer, more actionable decision aids. Furthermore, The framework is also subject to **market reflexivity**: it was tested on historical data in which its own signals were absent. If deployed, its predictive advantage would likely fade as other market participants detect and adapt to its patterns, eroding its returns.

8 Conclusion and Future Work

This study introduced an **explainable multimodal framework** for real-time Bitcoin (BTC) forecasting at **15-minute resolution**. It addressed high-frequency prediction challenges by constructing a comprehensive multimodal dataset integrating market data, on-chain metrics, the Fear & Greed Index (FGI), and sentiment from news and Reddit. A leakage-safe pipeline ensured reproducibility and fair comparison across models, including XGBoost, LSTM, a residual hybrid, and an adapted Multimodal Fusion Block (MFB). Multimodal integration delivered modest but consistent improvements, particularly in **event-driven regimes** where sentiment and news complement technical indicators. LSTM and MFB models achieved slightly lower error rates, though none surpassed the persistent ceiling on directional accuracy.

The study advanced explainability via a **dual-layer approach** combining **SHAP-based attributions** with **LLM narratives**. This delivers rigorous feature importance and intuitive, human-readable insights, enabling informed and cautious decision-making. While short-term BTC forecasting remains constrained by market efficiency, the framework demonstrates that careful multimodal integration and explainability transform weak predictive signals into practical decision support tools.

Future research should investigate regime- and risk-aware models, cross-modal attention, multilingual sentiment, execution-aware evaluation with transaction costs, and a **formal study of computational and environmental resource consumption**. Developing scalable SHAP variants for real-time use, adaptive online learning, and interactive dashboards offers a promising avenue for advancing human-centred financial AI. To support future work and reproducibility, the source code is publicly available at: https://github.com/iamdbadal/Explainable-Bitcoin-Forecasting.git.

References

[1] R. Aarthi, P. Vanitha, S. Reshma, C. Mounisha, and G. Pugalenthi. 2025. Temporal data integration and forecasting prices for Bitcoin and Ethereum using machine learning and deep learning techniques. *SSRN Electronic Journal* (2025). doi:10.2139/ssrn.5091385

[2] Pierre-Daniel Arsenault, Shengrui Wang, and Jean-Marc Patenaude. 2025. A survey of explainable artificial intelligence (XAI) in financial time series forecasting. *Comput. Surveys* 57, 10 (2025). doi:10.1145/3729531

[3] Sahar Arshad, Seemab Latif, Ahmad Salman, and Rabia Latif. 2024. Enhancing profitability and investor confidence through interpretable AI models for investment decisions. *arXiv preprint arXiv:2312.16223* (2024). doi:10.48550/arXiv.2312.16223 v2 submitted Jul 2024.

[4] Golnoosh Babaei, Paolo Giudici, and Emanuela Raffinetti. 2022. Explainable artificial intelligence for crypto asset allocation. *Finance Research Letters* 47 (2022), 102941. doi:10.1016/j.frl.2022.102941

[5] Adela Bâra and Simona-Vasilica Oprea. 2024. An ensemble learning method for Bitcoin price prediction based on volatility indicators and trend. *Engineering Applications of Artificial Intelligence* 133 (2024), 107991. doi:10.1016/j.engappai.2024.107991

[6] Saachin Bhatt, Mustansar Ghazanfar, and Mohammad Hossein Amirhosseini. 2023. Sentiment-driven cryptocurrency price prediction: A machine learning approach utilizing historical data and social media sentiment analysis. *Machine Learning and Applications: An International Journal (MLAIJ)* 10, 2/3 (Sept. 2023). doi:10.5121/mlaij.2023.10301

[7] Omkar Bhattarai, Raj Chaudhary, and Ganesh Gupta. 2025. Time-series real-time Bitcoin price prediction using LSTM. In *International Conference on Artificial Intelligence and Data Science Applications (ICAIDSA)*. Foundation of Computer Science (FCS), 1–6. doi:10.5120/icaidsc202411 ISSN: 0975-8887.

[8] Şükrü Selim Çalık, Andaç Akyüz, Zeynep Hilal Kilimci, and Kerem Çolak. 2025. Explainable-AI powered stock price prediction using time series transformers: A case study on BIST100. *arXiv preprint arXiv:2506.06345* (2025). doi:10.48550/arXiv.2506.06345 Preprint submitted to Elsevier.

[9] José Manuel Carbó and Sergio Gorjón. 2024. Determinants of the price of Bitcoin: An analysis with machine learning and interpretability techniques. *International Review of Economics and Finance* 92 (2024), 123–140. doi:10.1016/j.iref.2024.01.070

[10] Saeed Mohammadi Dashtaki, Mehdi Hosseini Chagahi, Behzad Moshiri, and Md. Jalil Piran. 2024. Multi-source hard and soft information fusion approach for accurate cryptocurrency price movement prediction. *arXiv preprint arXiv:2409.18895* (2024). doi:10.48550/arXiv.2409.18895

[11] Jacopo Fior, Luca Cagliero, and Paolo Garza. 2022. Leveraging explainable AI to support cryptocurrency investors. *Future Internet* 14, 9 (2022). doi:10.3390/fi14090251

[12] Vincent Gurgula, Stefan Lessmann, and Wolfgang Karl Härdle. 2024. Deep learning and NLP in cryptocurrency forecasting: Integrating financial, blockchain, and social media data. *arXiv preprint arXiv:2311.14759* (2024). doi:10.48550/arXiv.2311.14759

[13] Branka Hadji Misheva and Joerg Osterrieder. 2023. A hypothesis on good practices for AI-based systems for financial time series forecasting: Towards domain-driven XAI methods. *arXiv preprint arXiv:2311.07513* (2023). https://arxiv.org/abs/2311.07513

[14] Ping Han, Hui Chen, Abdur Rasool, Qingshan Jiang, and Min Yang. 2025. MFB: A generalized multimodal fusion approach for Bitcoin price prediction using time-lagged sentiment and indicator features. *Expert Systems with Applications* 261 (2025), 125515. doi:10.1016/j.eswa.2024.125515

[15] Hla Soe Htay, Mani Ghahremani, and Stavros Shiaeles. 2025. Enhancing Bitcoin price prediction with deep learning: Integrating social media sentiment and historical data. *Applied Sciences* 15, 3 (2025), 1554. doi:10.3390/app15031554

[16] Yushan Jiang, Wenchao Yu, Geon Lee, Wei Cheng, and Yanchi Liu. 2025. TimeXL: Explainable multi-modal time series prediction with LLM-in-the-loop. *arXiv preprint arXiv:2503.01013* (2025). doi:10.48550/arXiv.2503.01013

[17] David L. John, Sebastian Binnewies, and Bela Stantic. 2024. Cryptocurrency price prediction algorithms: A survey and future directions. *Forecasting* 6, 3 (2024), 637–671. doi:10.3390/forecast6030034

[18] Furkan Karadaş, Bahaeddin Eravcı, and Ahmet Murat Özbayoğlu. 2025. Multimodal stock price prediction. In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART)*, Vol. 3. SCITEPRESS, 687–694. doi:10.5220/0013174500003890

[19] Saman Kazeminia, Hedieh Sajedi, and Masoud Arjmand. 2023. Real-time Bitcoin price prediction using hybrid 2D-CNN LSTM model. In *Proceedings of the 9th International Conference on Web Research (ICWR)*. IEEE, 188–193. doi:10.1109/ICWR57742.2023.10139275

[20] Amit Kumar and Taoran Ji. 2025. CryptoPulse: Short-term cryptocurrency forecasting with dual-prediction and cross-correlated market indicators. *arXiv preprint arXiv:2502.19349* (2025). doi:10.48550/arXiv.2502.19349

[21] Jutur Manogna, Gogineni Sravan Chowdary, Gogineni Meghana, and Priyanka C. Nair. 2023. Bitcoin price prediction based on sentiment analysis. In *2023 IEEE 20th India Council International Conference (INDICON)*. IEEE, 233–238. doi:10.1109/INDICON59947.2023.10440877

[22] Ashraf Nidhal Mohammed. 2024. Hybrid cryptocurrency price prediction: Integrating EGARCH and LSTM with explainable AI. *International Journal of Innovative Science and Research Technology (IJISRT)* 9, 12 (Dec. 2024), 731–738. ISSN: 2456-2165.

[23] Hamid Moradi-Kamali, Mohammad-Hossein Rajabi-Ghozlou, Mahdi Ghazavi, Ali Soltani, Amirreza Sattarzadeh, and Reza Entezari-Maleki. 2025. Market-derived financial sentiment analysis: Context-aware language models for crypto forecasting. *arXiv preprint arXiv:2502.14897* (2025). doi:10.48550/arXiv.2502.14897

[24] Kinza Muneer and Ubaida Fatima. 2025. Cryptocurrencies analytics with machine learning and human-centered explainable AI: Enhancing decision-making in dynamic market. *International Journal of Computer Applications* 186, 62 (Jan. 2025), 52–67. doi:10.5120/ijca2025924418

[25] S. M. Raju and Ali Mohammad Tarif. 2020. Real-time prediction of Bitcoin price using machine learning techniques and public sentiment analysis. *arXiv preprint arXiv:2006.14473* (2020). doi:10.48550/arXiv.2006.14473

[26] Phumudzo Lloyd Seabe, Claude Rodrigue Bambe Moutsinga, and Edson Pindza. 2023. Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: A deep learning approach. *Fractal and Fractional* 7, 2 (Feb. 2023), 203. doi:10.3390/fractalfract7020203

[27] Mohammad Shahab Sepehri, Asal Mehradfar, Mahdi Soltanolkotabi, and Salman Avestimehr. 2025. CryptoMamba: Leveraging state space models for accurate Bitcoin price prediction. *arXiv preprint arXiv:2501.01010* (2025). doi:10.48550/arXiv.2501.01010

[28] Danting Tao, Yitong Wu, and Yutong Liu. 2024. Prediction of Bitcoin pricing by SHapley Additive exPlanations. In *Proceedings of the 2024 5th International Conference on Economics, Education and Social Research (ICEESR)*. Francis Academic Press, London, UK, 290–297. doi:10.25236/iceesr.2024.047

[29] Jingyang Wu, Xinyi Zhang, Fangyixuan Huang, Haochen Zhou, and Rohitash Chandra. 2024. Review of deep learning models for crypto price prediction: Implementation and evaluation. *arXiv preprint arXiv:2405.11431* (2024). doi:10.48550/arXiv.2405.11431

[30] Zi Ye, Yinxu Wu, Hui Chen, Yi Pan, and Qingshan Jiang. 2022. A stacking ensemble deep learning model for Bitcoin price prediction using Twitter comments on Bitcoin. *Mathematics* 10, 8 (2022), 1307. doi:10.3390/math10081307

[31] Wei Jie Yeo, Wihan Van Der Heever, Rui Mao, Erik Cambria, Ranjan Satapathy, and Gianmarco Mengaldo. 2025. A comprehensive review on financial explainable AI. *Artificial Intelligence Review* 58 (2025), 189. doi:10.1007/s10462-024-11077-7

[32] Xinli Yu, Zheng Chen, Yuan Ling, Shujing Dong, Zongyi Liu, and Yanbin Lu. 2023. Temporal data meets LLM: Explainable financial time series forecasting. *arXiv preprint arXiv:2306.11025* (2023). doi:10.48550/arXiv.2306.11025

[33] Yanzhao Zou and Dorien Herremans. 2023. PreBit: A multimodal model with Twitter FinBERT embeddings for extreme price movement prediction of Bitcoin. *Expert Systems with Applications* 233 (2023), 120838. doi:10.1016/j.eswa.2023.120838