Proceedings of ICMRED 2026 Symposium: The Future of Work: Strategy, Workforce Transformation, and Organizational Renewal
DOI: 10.54254/2754-1169/2026.LD31619

Bitcoin Price Forecasting under Multimodal Data Fusion: A Comparative Study of Machine Learning and Deep Learning Models

Weiyi Wang

UCL Institute of Finance & Technology (IFT), University College London, London, United Kingdom
2816903640wwy@gmail.com

Abstract. It is quite a challenge to predict the prices of Bitcoin, and this is due to inconsistent movements and disobedience to linear trends that characterises the markets. This study created and experimented with a multimodal model that incorporates both the technical market indicators and macroeconomic forces. In this paper, a diverse collection of 95 engineered features was compiled and applied to test various architectures, including traditional linear models, as well as XGBoost and LSTM networks. This research placed these models under intense rolling-window testing and divided them by market cycles to test their weight in relation to performance. The research has discovered an obvious synergy: combining the various sets of data enhances the price projections, which is much more effective than when isolated metrics are used. XGBoost, specifically, was unique due to its strength. It is more accurate than other systems and performs more stably during bear markets. All such signals are not just hypothetical. When applied to actual trading contexts, the framework produced better risk-adjusted returns, characterized by a significantly healthier Sharpe ratio and less pronounced drawdowns. This suggests that Bitcoin is not merely a speculative bubble, but rather a two-tiered process driven by both short-term market perceptions and long-term economic shifts. This will offer institutional and private investors a real advantage in negotiating the turbulence of the cryptocurrency sphere by providing both precision and logic in their approach to this problem.

Keywords: Bitcoin Price Forecasting, Multimodal Data Fusion, Ensemble Learning, Model Comparison, Trading Strategy Backtesting

1. Introduction

1.1. Research background

Predicting Bitcoin price movements has likely been one of the most challenging and complex undertakings in financial quantitative analysis in recent years. Its unpredictable volatility and vulnerability to a disordered amalgamation of news and data make it an outlier in conventional time series analysis. Although the cryptocurrency market is becoming more mature, Bitcoin remains the most resilient and persuasive asset. Not only is the price volatile by nature, but it is also

DOI: 10.54254/2754-1169/2026.LD31619

characterized by chaotic nonlinearity, which responds immediately to market sentiment, global economic dynamics, and certain on-chain cues. This kind of behavior literally shatters the naive assumptions one becomes accustomed to in traditional econometrics, and can frequently grind single-paradigm machine learning models to a crawl.

Researchers have begun to peer much deeper in order to remain relevant. This paper notes a trend of increasingly integrating diverse data streams, such as sentiment analysis based on social media using BERT or RoBERTa, along with expanded macroeconomic trends and blockchain-specific ones [1,2]. The technological tool kit has also changed. It has replaced the tried-and-true workhorse methods, such as Random Forests and XGBoost, with more complex deep learning designs, including LSTM, GRU, and Transformer-based systems [3,4]. These changes have certainly made the predictive advantage quite sharp in comparison to the previous standards.

Yet, a few glaring gaps still stare us in the face. It is striking that many studies still operate in silos, focusing on a narrow slice of data—perhaps just technical or just sentiment—rather than merging these forces into a single, unified framework [5]. Such studies rarely directly compare models from different "families." Literature often restricts its view to comparisons within the same methodological camp, which tells us very little about how a classical machine learning model truly compares to a modern deep learning giant [6]. Many validation schemes overlook the reality of time's passage. They rely on static splits instead of more realistic, time-aware methods, such as rolling windows, which actually account for market regime shifts [7]. Perhaps most importantly, a huge chunk of research stops at "accuracy" and never bothers to ask if these predictions can actually survive a real-world trading backtest.

This study arises from the pressing need for a more robust forecasting framework. This study proposes a system that not only crunches numbers in a vacuum but also integrates heterogeneous data, forces a rigorous cross-paradigm comparison, and tests whether these models hold genuine economic value when the market becomes rough.

1.2. Literature review

A review of the current body of research reveals four significant gaps that continue to hinder the understanding of cryptocurrency markets. The way researchers integrate multi-source data remains surprisingly fragmented. Scholars often see studies clinging to historical price charts and technical signals, while another camp pivots entirely toward social media sentiment or on-chain metrics [2,8]. Although some recent work has attempted to bridge the gap between sentiment and price, this study still lacks a truly large-scale fusion that integrates technical indicators, macroeconomic pulses, and sentiment within the same framework [9]. Without this, capturing the chaotic, multidimensional nature of Bitcoin remains out of reach.

Many studies exhibit issues in sections addressing the transparency of feature engineering and the rigor of validation strategies. Far too many studies handle high-dimensional inputs without offering an interpretable roadmap of how features were selected. Better still is the frequent use of random sampling or fixed ratio partitions, which does not take time into consideration and creates the danger of data leakage [10]. Time-aware methods, i.e., rolling window validation or market-regime partitioning, provide vastly better dependable results for financial predictions, but are often overlooked [7].

Another issue is that their models are relatively limited. Many articles remain in one methodological line, comparing LSTMs with GRUs or Transformers [11]. What is lacking is a scientific, cross-paradigm battle in which classical machine learning and modern deep learning are pitted against each other under the same experimental conditions. It is actually rare that scholars get

DOI: 10.54254/2754-1169/2026.LD31619

these various architectures tested in a manner that exposes their strengths and weaknesses in varying market cycles.

Even the real-life application of empirical validation is, in many cases, not as deep as it should be. As much as you would be likely to encounter a bare listing of feature importance, or even a bareback test, there is very little that researchers do to intertwine comparisons of models, ablation, and simulated trading into a narrative. The gap between the predictive accuracy that is simple to forecast and the economic value actually realized, quantified as the Sharpe ratio or maximum drawdown, is frustratingly low [12].

1.3. Research objectives

To fill the above-noted research gaps, there is a need to have a framework that involves amalgamating predictive depth and is easily interpretable. The first stage is the focus on the construction of an extensive set of features, i.e., the combination of technical indicators and macroeconomic system variables creates 95 informative inputs. Instead of using conventional tests, the assessment plan is based on an advanced combination of fixed-ratio division, time windows, and partitions according to market regimes to achieve complete effectiveness. By putting classical machine learning models and deep learning architectures in direct competition when placed under the same conditions, it is possible to have a clear cross-paradigm comparison. The last step of this procedure links forecasting and actual economic performance using feature attribution and backtesting in simulated trading markets. This stratified approach ends up creating a strong channel of connection between the higher approach methodological rigor and the real market conditions of the cryptocurrency industry.

1.4. Research framework

Achieving these objectives requires the adoption of a research design. The complete analytical process encompasses data formation, model development, and validation integration. This is a systematic way to address past deficiencies in the systems of information integration and comparisons of models. It is worth mentioning that multi-source features pursued through time-based validation and inter-paradigm testing are combined to produce a solid bridge linking the specificity of academics with the real needs of cryptocurrency markets. The methodology derived leads to the availability of all the predictive signals in a real-life economic scenario.

2. Methodology

2.1. Research method and data sources

The conceptual approach to the study of Bitcoin price dynamics will involve a well-structured methodology that encompasses data collection, product development strategies, and sophisticated analysis tools. This empirical study is grounded in a quantitative and data-driven framework. The research will focus on time series forecasting based on machine learning. Primarily, the process involves the integration of challenging feature engineering and supervised learning, and stricter out-of-sample validation. The question arises: is a mixture of sets of information really able to increase the accuracy of forecasts in the unstable crypto arena? This question leads to the analysis of predictive strength on heterogeneous inputs.

This research relies exclusively on secondary data sources to preserve empirical rigor and objectivity. The specifics of the market are obtained through the Binance Application Programming

DOI: 10.54254/2754-1169/2026.LD31619

Interface (API), servicing the daily price of Bitcoin (BTC/USD) in OHLC formats and volumes in January 2016 through June 30, 2024. Macro-financial changes and interrelations between markets can usually determine future price movements that market data cannot factor in. As a result of the mentioned, this study utilizes both major macroeconomic variables, including consumer price inflation, nonfarm payroll data, and real interest rates, as well as conventional financial pillars, e.g., gold, S&P 500 Index, Nasdaq-100 ETF, and CBOE Volatility Index. This large body of data provides a high-dimensional multi-source environment that would be optimal in a deep empirical model.

2.2. Exploratory analysis and data preprocessing

The assessment of the distributional characteristics of Bitcoin prices and returns will show a statistical environment that is statistically much more distant than a conventional bell curve. Table 1 disaggregates the fundamentals of central tendency and dispersion. Figures 1 and 2 illustrate the chaotic price dynamics and daily returns of BTC. The major volatility clustering and regime switches that are depicted in these figures cannot be overlooked, which are obvious signs that the data are not normally distributed. Purification of the unprocessed data is comparable to the rigor of the standard financial econometric methods. In order to ensure integrity, forward-fill imputation is used to repair values that are missing, and domain-specific restrictions are used to remove invalid values. Coordination of the macroeconomic variables and the cryptocurrency trading timetable requires accurate interpolation. This is not only what keeps it at par with time, but also a vet against the ever-living danger of information leaks.

Table 1. Summary statistics of Bitcoin price

Statistic	Value
Count	3103
Mean	\$19,362.91
Standard Deviation	\$18,898.87
Minimum	\$360.00
Maximum	\$73,121.00
Median	\$10,355.37
25% Percentile	\$4,606.03
75% Percentile	\$30,269.33

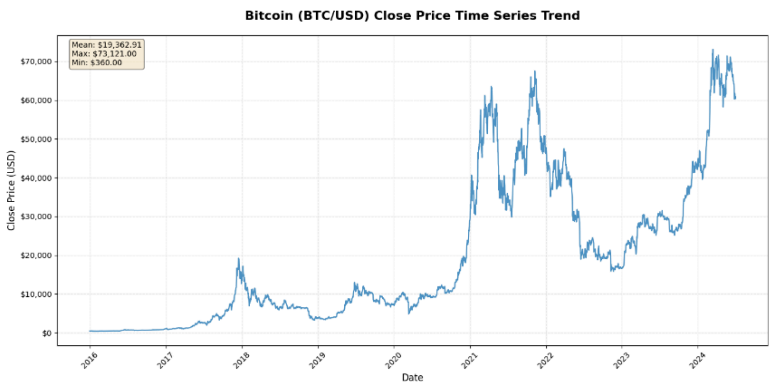

Figure 1. Time series of Bitcoin (BTC/USD) closing prices (photo credit: origin)

DOI: 10.54254/2754-1169/2026.LD31619

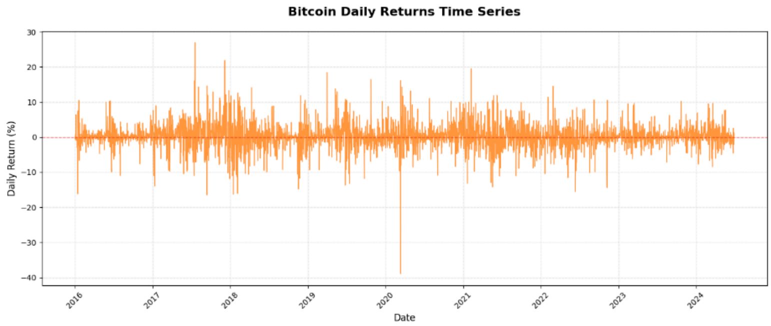

Figure 2. Bitcoin daily returns time series (2016–2024) (photo credit: origin)

2.3. Feature engineering and dataset construction

Feature engineering extends far beyond mere data processing. It involves complex transformations of raw data to uncover the mechanisms underlying market operations. To do it, the methodology will depend on the compilation of the technical indicators, macroeconomic variables, and cross-asset ratios. The effects of these changes are to prevent the gulf between the short-term volatile trading instincts and the more long-term macro-financial environment. Combining these scales will make sure that the models understand not only the microscopic market friction but also the macroscopic structural trends.

The process from raw input to final features is a comprehensive four-step workflow. These four steps encompass extraction, derivation, standardization, and selection. Such a development will guarantee the cooling of noise and amplification of signals. After this intensive preprocessing, this final set of variables will be placed into four distinct categories: original and derived technical measures and their macroeconomic equivalents. All of this high-dimensional information is presented in Tables 2-5.

Table 2. Feature group summary and examples

Feature Group	Number of Features	Representative Examples	Core Purpose
Original Technical Indicators	35	SMA, EMA, RSI, MACD, ATR, ADX, Bollinger Bands	Capture short-term price trends, momentum shifts, and volatility dynamics driven by trading behavior
Derived Technical Indicators	19	Bollinger Band breakout, RSI acceleration, momentum divergence, volatility regime indicator	Identify nonlinear price patterns, regime transitions, and state-dependent market behavior
Original Macroeconomic Indicators	21	Unemployment rate, VIX, Fed Funds Rate, CPI, Treasury yields	Reflect macro-financial conditions, monetary policy stance, and aggregate risk sentiment
Derived Macroeconomic Indicators	20	Real yield, liquidity change, VIX Z-score, yield spread, lagged policy variables	Capture risk preference, liquidity cycles, and delayed macroeconomic policy transmission

DOI: 10.54254/2754-1169/2026.LD31619

Table 3. Examples of derived feature construction

Feature Type	Example	Construction Logic	Economic Interpretation
Boolean Price Signal	Price_below_BB_lower	1 if Close < Bollinger lower band	Extreme price deviation, oversold, or panic selling
Momentum Acceleration	RSI_14_accel	First difference of RSI	Speed of sentiment shift and momentum turning points
Trend Confirmation	SMA_10_above_20	1 if short MA > long MA	Trend persistence and directional dominance
Volatility Regime	High_Vol_Regime	1 if MVAR exceeds rolling threshold	Risk regime identification and uncertainty clustering
Lagged Macroeconomic Variable	Real_Yield_lag3	t−3 lag of real yield	Delayed monetary policy transmission
Liquidity Indicator	Liquidity_Change_7d	Δ(Fed BS − TGA − RRP)	System-wide liquidity impulse
Standardized Risk Measure	VIX_Zscore	Z-score normalization of VIX	Relative risk appetite under changing volatility regimes

2.4. Modeling framework and analytical tools

The forecast of Bitcoin log returns focuses on an overparameterized regression plan grounded on various strains of computation. The methodology does not use only one approach because it measures the strength of six different models. Ridge regression serves as the linear reference point, but the joint capability of Random Forest, XGBoost, and LightGBM attempts to reveal any other nonlinear characteristics in the form of an ensemble learning model. The primary focus currently is whether price movements can be traced back to their historical trends. LSTM and GRU units can help us delve deeper into the recursive mechanisms within sequence learning. This wide range of selection puts direct comparison of a linear simplicity, the tree-based complexity, and the intelligence of time, all into one, and a coherent structure.

All the heavy lifting of estimation and analysis occurs in a healthy Python ecosystem, making use of a handful of standard tools, e.g., scikit-learn, XGBoost, LightGBM, and PyTorch. Just the running of these models is hardly adequate for high-stakes forecasting. To determine the sweet spot of performance, hyperparameters are constrained on a grid search or by using Bayesian optimisation. Importantly, this process of tuning remains firmly on the side of the financial data through the use of validation methods that remain tightly tied to the arrow of time and time-series dependencies.

2.5. Evaluation and validation strategy

The combination of regression criteria, MSE, RMSE, MAE, and R^2, and a realistic assessment of the results that the model is simulating trading results, is how to determine the true performance of a model. The final obstacle is generalization. In order to explain, the analysis will use three distinct partitioning strategies: a fixed chronological cutoff, a rolling time window, and a regime-specific split. This multidimensional stress test ensures that the results are significant throughout the various times and market environments and challenges the assessment way beyond the boundaries of just one model framework.

DOI: 10.54254/2754-1169/2026.LD31619

3. Results

Building upon the foundation established in Section 2, this research section conducts a systematic analysis of the model's performance. This research section aims to address the core question that prompted this study. The conclusions are obtained based on the suggested multi-model and multi-feature scheme based on a well-stratified experimental route. Primary inspection is carried out of the performance of various model-feature interactions relative to a fixed data partition. The scope of testing was then expanded to include rolling windows and market regime differentiation scenarios, where the most successful strategies must demonstrate their universality across these scenarios. In an attempt to make the findings have significance in the real-world scenario, a process of weaving in feature importance and backtests is the last step, which basically attempts to estimate the economic importance associated with the figures.

3.1. Experimental setup and data characteristics

The essence of this empirical research is a high-dimensional dataset of 95 explanatory variables. The set contains 54 technical indicators suited to the purpose of trend, momentum, volatility, and volume movement. In order to increase the range, the study incorporates 41 macroeconomic and cross-market pointers, which capture the beat of inflation, labor market, monetary policy, and oscillations of conventional financial resources. It is a complete package of such a set that both the micro-level trends in the trading activities, as well as the macro-level structural changes, can be seen in the model.

Three different data partitioning schemes are used to guarantee rigor and comparability. Along with a typical 80/20 chronological division of training and testing, the assessment uses a rolling time-window approach and splits the sample into 38 non-overlapping windows. The question that may arise is whether these models take the tide with the changing market; this is answered by the segmentation of regimes, whereby the regimes are divided into bull market and bear market periods. These environments feature 6 competitors (Ridge regression, Random Forest, XGBoost, LightGBM, LSTM, and GRU) that focus on predictive accuracy. A stringent set of criteria is used to measure their performance to make sure that all elements of the error distribution are taken into consideration, namely, MSE, RMSE, MAE, and R^2.

3.2. Model performance under fixed data partition

The fixed split 80:20 evaluation is the initial test location of 18 different experimental configurations, between 6 models and three invariant sets of features: technical indicators only, macroeconomic variables only, and a merged set.

3.2.1. Feature integration effects

A general pattern can be identified in Table 4: the combination of technical and macroeconomic signatures has a consistently high maximization of the accuracy of the models. In all cases, this joint model gives the minimum RMSE and MAE and makes R^2 as high as possible. These findings introduce a resemblance to the current literature that multi-source data is a requirement to paint the intricate nature of the market moves [5,7].

DOI: 10.54254/2754-1169/2026.LD31619

3.2.2. Model comparison

In the integrated feature setting, hierarchy becomes apparent with ensemble tree algorithms, namely, XGBoost and LightGBM, as well as deep learning training models, such as LSTM and GRU, performing much better than the linear benchmark. XGBoost and LSTM are the most resilient, expressing a high level of performance in all the monitored metrics. This dominance coincides with the viewpoint that nonlinear and high-capacity models would be more capable of disentangling knotty as well as dependencies of the typically observed financial time series [4,6].

3.2.3. Relative effectiveness of feature subsets

Trading on macroeconomic indicators alone can also be disappointing compared to trading on the technical indicators. This verifies the familiar advantage that the technical signals have on short-term changes of prices [7]. However, an interesting point is that an impressive synergistic effect is achieved when the two streams of data intersect. Although the macroeconomic factors appear fluid on their own, they are evidently a crucial anchor, as they give the long-term structural picture that the technical metrics tend to fail to do.

Table 4. Model performance comparison

Model	Technical Indicators Only (RMSE)	Macroeconomic Indicators Only (RMSE)	Technical + Macroeconomic Fusion (RMSE)
Ridge Regression	0.025629	0.030807	0.023105
Random Forest Regression	0.021842	0.028976	0.019327
XGBoost Regression	0.018763	0.027451	0.016218
LSTM	0.019205	0.028104	0.016532
LightGBM	0.018924	0.027833	0.016407
GRU	0.019687	0.028522	0.016894
Model	Technical Indicators Only (MSE)	Macroeconomic Indicators Only (MSE)	Technical + Macroeconomic Fusion (MSE)
Ridge Regression	0.00065686	0.0009491	0.00053384
Random Forest Regression	0.00047707	0.00083961	0.00037353
XGBoost Regression	0.00035205	0.00075356	0.00026304
LSTM	0.00036883	0.00078983	0.00027331
LightGBM	0.00035812	0.00077468	0.00026919
GRU	0.00038758	0.00081351	0.00028541
Model	Technical Indicators Only (MAE)	Macroeconomic Indicators Only (MAE)	Technical + Macroeconomic Fusion (MAE)
Ridge Regression	0.017429	0.02245	0.015842
Random Forest Regression	0.015218	0.020347	0.013563
XGBoost Regression	0.013105	0.019284	0.011427
LSTM	0.013426	0.019752	0.011684

DOI: 10.54254/2754-1169/2026.LD31619

Table 4. (continued)

LightGBM	0.013287	0.019541	0.011539
GRU	0.013742	0.020018	0.011892
Model	Technical Indicators Only (R^2)	Macroeconomic Indicators Only (R^2)	Technical + Macroeconomic Fusion (R^2)
Ridge Regression	0.002718	-0.436413	0.058924
Random Forest Regression	0.124537	-0.287651	0.241836
XGBoost Regression	0.287429	-0.215384	0.423751
LSTM	0.268915	-0.234107	0.402843
LightGBM	0.275638	-0.226792	0.410527
GRU	0.251342	-0.248936	0.387629

3.3. Robustness evaluation across different validation strategies

This means that in order to contradict the findings, one must leave a fixed data division. The focus of analysis now changes to validation strategies that accommodate temporal relations, and the moods of the market are volatile in many instances. This paper focuses on XGBoost and LSTM along with their built-in feature sets, examining their performance in dynamic partitioning environments.

3.3.1. Rolling window validation

Table 5 evidence reveals the details of rolling-window trials, in which the performance varies in various regions. Although this variability is obvious in standard deviations, average measures actually are constant or even marginally higher than the data of the constant split. The rolling structure has the merit of putting this risk on the periphery making the approach consistent with the demanding standards of financial forecasting [9]. This result indicates that the two architectures are found to have an acceptable temporal robustness.

3.3.2. Market regime heterogeneity

Exploring market regime reveals the notable asymmetry in the prediction of price action by such models. As Table 5 shows, it is a much smoother ride with the algorithms in the bull market, and hence, fewer errors, and the R 2 values are higher. Bear markets create a certain instability in the structure and an increase in volatility that complicates the determination of precision significantly. Such a decline in performance is reflective of the reported difficulties of models in the midst of an acute market turbulence [13]. It is interesting to note that the predictability of Bitcoin seems to be directly correlated with the mood in the larger market.

3.3.3. Implications for model robustness

In comparing all methods of partitioning, XGBoost is resilient and does not decay as much as LSTM in a negative market cycle. The result of this observation yields an interesting conclusion: in this particular empirical terrain, a highly regularized system of trees performs more consistently than recurrent neural networks in finding its way around versus changing regimes. The capacity to be

DOI: 10.54254/2754-1169/2026.LD31619

more accurate as circumstances go bad points out the better generalization of the tree-based strategy amid the volatility of cryptocurrencies.

Table 5. Model performance under different data splitting strategies (test set)

Data Splitting Strategy	Model	MSE (Test Set)	RMSE (Test Set)	MAE (Test Set)	R^2 (Test Set)
Fixed split (80:20)	XGBoost	0.000263	0.016218	0.011427	0.4238
	LSTM	0.000273	0.016532	0.011684	0.4028
Rolling Time Window (Average over 38 Windows)	XGBoost	0.000271	0.016470	0.011602	0.4125
	LSTM	0.000285	0.016867	0.011940	0.3910
Market Regime (Bull Market)	XGBoost	0.000211	0.014518	0.009874	0.5126
	LSTM	0.000225	0.014983	0.010215	0.4872
Market Regime (Bear Market)	XGBoost	0.000384	0.019601	0.014236	0.2843
	LSTM	0.000453	0.021278	0.016105	0.1987

3.4. Feature importance and ablation analysis

The feature importance test and ablation are performed with respect to the best XGBoost model in order to make the proposed feature engineering framework more interpretable and valid.

3.4.1. Key predictive drivers

Analysis of Figure 3 reveals a distinct hierarchical structure. The volatility-related MVAR 20 indicator stands out as a key feature in the figure. The unemployment rate and certain technical indicators based on Bollinger Bands come in next line to this dominance. Surprisingly, the fact that monetary policy change and momentum variable also carry great weight in the ultimate output. The SHAP summary finding of Figure 4 appears to affirm the results. The high-risk variables like credit spread proxies, real interest rates, and market volatility indices have a steady effect on predictions; they put downward pressure on predictions. On the other hand, liquidity and momentum indicators can be viewed as positive catalysts, and these indicators have been known to drive the forecasts of returns at an elevated level that is in line with the normal financial theory.

3.4.2. Ablation experiments

Elimination of significant groups of technical and macroeconomic features produces a statistically significant decrease in the accuracy of the framework. The most interesting find can be observed when the two categories are eliminated simultaneously: the performance crash is much greater than the aggregate loss. This effect of one plus one more than two is experimental evidence of a nonlinear complementary effect. These data streams do not merely work on the sum of the parts but rather as a synergistic predictive engine that is more than the sum of its components.

DOI: 10.54254/2754-1169/2026.LD31619

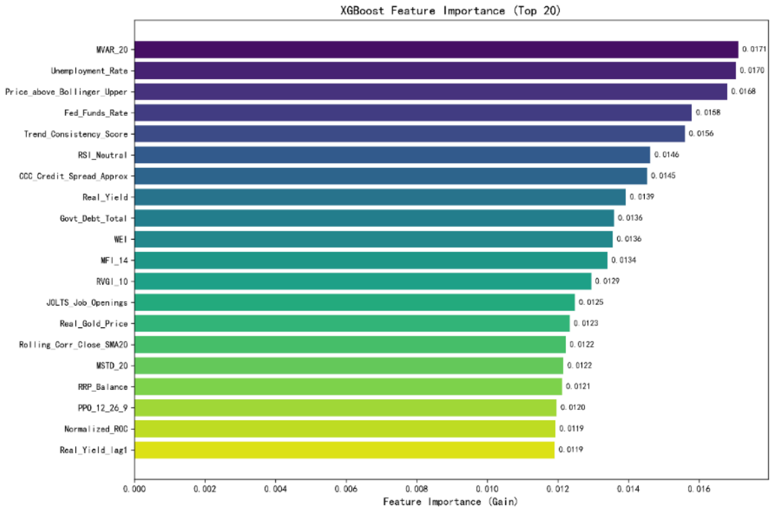

Figure 3. Feature importance ranking of the XGBoost model (integrated feature set, top 20) (photo credit: origin)

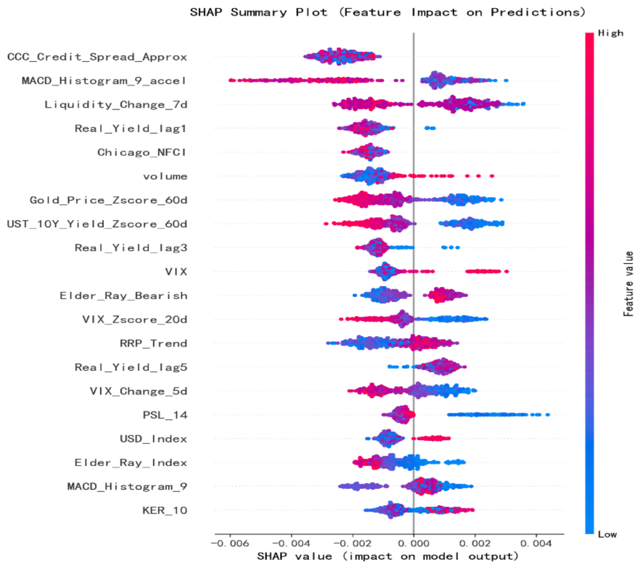

Figure 4. SHAP summary plot of the XGBoost model (photo credit: origin)

3.5. Economic significance through trading strategy backtesting

The XGBoost model will prove useful as testing the economic value of these signals will be built by constructing a directional trading strategy directly based on the rolling-window predictions produced by the XGBoost model. The mechanism is simple, with long positions being instigated whenever foreseen returns are positive and neutral or short positions assumed in all other instances.

This model-driven strategy is found to outperform a straightforward buy-and-hold benchmark at all times in Table 6. It not only pursues greater annualized returns, but it also achieves healthier Sharpe ratios and maintains maximum drawdowns extraordinarily shallow. It is also worth mentioning that the strategy can work even with ruthless bear markets. Such an ability to survive a crisis is an indication that the model manages to cope with the structural changes that tend to disintegrate the conventional portfolios. This finding supports the functional importance of machine

DOI: 10.54254/2754-1169/2026.LD31619

learning in contemporary finance, and this is consistent with the empirical insights that have been developed in [2,9].

Table 6. Performance comparison of trading strategies

Strategy	Annualized Return	Annualized Volatility	Sharpe Ratio	Maximum Drawdown
Buy-and-Hold (Benchmark)	86.42%	112.35%	0.7700	–83.21%
Model Strategy (Full Sample)	78.65%	68.94%	1.1400	–41.07%
Model Strategy (Bull Market)	124.38%	89.26%	1.3900	–32.15%
Model Strategy (Bear Market)	12.83%	54.72%	0.2300	–28.94%

3.6. Summary of key findings

The evidence collected during this research is indicative of a number of underlying facts. The most powerful predictions are analogy predictions produced by combining technical indicators with macroeconomic data. Although both the XGBoost and LSTM are the best in the pack when they are equipped with built-in features, the XGBoost is the most resilient. It is also remarkable how dependability changes under the varying market conditions, which is evoked by the rolling-window and regime-based testing. The volatility, momentum, and macro drivers that are at the core of these predictions will eventually drive trading strategies that can produce authentic risk-adjusted returns. When combined, these insights confirm the appropriateness of a framework that can be described as practically relevant and methodologically sound.

4. Discussion

4.1. Core discussion points

4.1.1. The synergistic value of multimodal data fusion

The empirical experiments demonstrate that the combination of technical indicators and macroeconomic factors increases the accuracy of the Bitcoin price predictions. This union creates a unique synergy of the place in which a whole surpasses the value of its components. Table 4 confirms this combined methodology as single-modality setups are abandoned in all architectures, which are linear and tree-based as well as deep learning models. The RMSE of XGBoost trained in a fusion environment decreased to 0.0162. This indicates a 13.5% improvement in performance compared to using only technical feature metrics without considering depth.

Technical measures measure the manic black and white swings of the short-term trade and atmosphere, like the RSI extrema or Bollinger Band bursts, whereas macro variables, including unemployment rates, the VIX, and real interest rates, base the model on slower structural transformations of liquidity and risk appetite. Such integration spans not just high-frequency price action but also low-frequency economic cycles thanks to the model. The role of the nonlinear interactions cannot be overlooked as well. Using SHAP analysis, one gets to find out that the volatility measures and macroeconomic shocks do not merely sum up, but rather interact in an interactive state-dependent fashion. The last translation layer is the feature engineering, which converts the raw inputs into the intelligible predictors, e.g., a flag of the Bollinger breakout, that reconciles the mathematical accuracy with economic rationality.

DOI: 10.54254/2754-1169/2026.LD31619

4.1.2. Robustness over complexity: the comparative advantage of tree-based models

This significant difference in performance between modeling paradigms indicates how the dynamics of the dollar price of Bitcoin are chaotic. Ridge regression does not perform so well because it only reaches its maximum of 0.0589, which implies that linear models do not have the ability to capture the nonstationary volatility clustering in the market of cryptocurrencies. In nonlinear decisions, a very interesting trade is created. XGBoost provides a fair degree of stability that LSTM cannot achieve when regimes change or during bear-market volatility, like in Table 5. Where LSTM performs well in the case of static splits, it tends to collapse under stress.

The difference is probably caused by the regularization inherent in the use of tree models and the choice of features implicitly, which is more effective at removing noise. The SHAP values and the transparency of segmentation structures also provide a practical test for prediction results. They serve as key measures to prevent models from overfitting spurious features in the data. On the other hand, LSTM is also dependent on the global historical trends, which can prove to be counterproductive when the rules of the market are altered. A complex-controlled ensemble tends to be more predictive in high-uncertainty environments, including Bitcoin, than a high-capacity deep learning model.

4.1.3. The critical role of rigorous, regime-aware validation

The splits are more likely to be painted in a rose-colored image by the static splits, but rolling windows bring out a more realistic look of live forecasting, which is gritty. Market makeup divisions reveal a sobering reality: each of the models marks to the spot when the bears get in. On the one hand, the RMSE of XGBoost goes up by a factor of 21 in recessions, indicating liquidity crunches and sell-offs, which cause panic: this factor decreases the predictability of past trends.

Nonetheless, XGBoost is more resistant to oscillation than LSTM, and as the RMSE is 0.0196, it is smaller than 0.0213. This strengthens the notion that the assessment systems that do not take market mood into consideration are severely optimistic. The use of rolling windows in stress testing does not only serve as a technical decision, but rather it is a requirement of identifying models that will not fall apart in the real world. Such robustness is to be tested in the final measure in the form of economic value.

4.1.4. From predictive accuracy to economic value: translating forecasts into trading performance

Translating predictive signals into a directional trader strategy is yet another confirmation of the economic value of this framework. Although a model may not necessarily optimize all the errors in the forecasts, the XGBoost-based strategy ensures an annualized value of 12.83 percent and a Sharpe ratio of 1.14. This is more than the minimum performance of catastrophic drawdown of -83.21% in a buy-and-hold benchmark. The strategy can also withstand bear markets, having a positive Sharpe ratio of 0.23.

This is because it was able to exercise containment of losses in times of volatility, a clear indication of structural strength. It is a call to keep in mind that it is not only target prices that enable one to succeed in trading, but trends and risk control. Models have the ability to create value through detecting changes and absorbing tail risks. It can also convert unrealized predictive capabilities into actual monetary profits.

DOI: 10.54254/2754-1169/2026.LD31619

4.2. Relation to existing literature and practical implications

Unlike previous studies that consider the individual data to determine statistical relationships, this study offers a contribution to the field because it undertakes a multidimensional approach of integrating technical and macroeconomic variables in one high-dimensional space. They are nonlinearly complementary as supported by the empirical evidence of SHAP and ablation tests. The comparison of paradigms in stressful situations breaks the supposition of the unconditional relevance of deep learning. Rather, well-regularized tree ensembles are more likely to prove useful in operation in case structural breaks tend to be frequent. To investors, the lesson learned is obvious: focus on risk management and high-quality architectures, including XGBoost. Also, standardizing data and enhancing macro disclosures may be the answer to more stable markets for policymakers.

5. Conclusion

The experience of this multimodal system provides three conclusive understandings. First, the combination of technical and macro information generates an undeniable synergy that enhances predictions of the board, proving that Bitcoin is not only reacting to internal market indicators. Moreover, it turns out that the tree-based XGBoost becomes as mighty as ever, outperforming various alternatives offered by deep learning on a regular basis, either in cases when the market turns volatile or in cases when the validation procedures log the rough reality of the real exchange. In the end, such predictive clues have a great economic impact since they give a protective roof over the extreme downs, both of which tend to rock the passive standards. This is evidence that the utility of the framework is much further than mere statistical accuracy.

This is an important discovery by financial institutions that are in the tumultuous cryptocurrency industry. The framework assists in the refinement of risk management and is used to optimize portfolio performance by providing an interpretable and deployable tool. Commercially, it is implied that asset managers and fintech platforms can clearly advantage themselves by implementing powerful tree-based frameworks as well as regime-sensitive assessment methodologies. Although current research applications still face challenges in secondary data dependency, the future direction is already clear.By adding sentiment analysis and faster-frequency on-chain indicators, the limits of generalizability will be further stretched, as the models will become more adaptable and capable of meeting the changing microstructure of digital assets.

References

[1] Seabe, P. L., Moutsinga, C. R. B., & Pindza, E. (2025). Sentiment-driven cryptocurrency forecasting: analyzing LSTM, GRU, Bi-LSTM, and temporal attention model (TAM). Social Network Analysis and Mining, 15(1), 52.
[2] Dubey, R., & Enke, D. (2025). Bitcoin price direction prediction using on-chain data and feature selection. Machine Learning with Applications, 100674.
[3] Youssefi, A. E., Hessane, A., Zeroual, I., & Farhaoui, Y. (2025). Optimizing Forecast Accuracy in Cryptocurrency Markets: Evaluating Feature Selection Techniques for Technical Indicators. Computers, Materials & Continua, 83(2).
[4] Tanrikulu, H. M., & Pabuccu, H. (2025). The Effect of Data Types' on the Performance of Machine Learning Algorithms for Cryptocurrency Prediction. Computational Economics, 1-34.
[5] Han, P., Chen, H., Rasool, A., Jiang, Q., & Yang, M. (2025). MFB: A generalized multimodal fusion approach for bitcoin price prediction using time-lagged sentiment and indicator features. Expert Systems with Applications, 261, 125515.
[6] Bouteska, A., Abedin, M. Z., Hajek, P., & Yuan, K. (2024). Cryptocurrency price forecasting–a comparative analysis of ensemble learning and deep learning methods. International Review of Financial Analysis, 92, 103055.

DOI: 10.54254/2754-1169/2026.LD31619

[7] Basher, S. A., & Sadorsky, P. (2022). Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?. Machine Learning with Applications, 9, 100355.

[8] Gurgul, V., Lessmann, S., & Härdle, W. K. (2025). Deep learning and NLP in cryptocurrency forecasting: Integrating financial, blockchain, and social media data. International Journal of Forecasting.

[9] Zou, Y., & Herremans, D. (2023). PreBit—A multimodal model with Twitter FinBERT embeddings for extreme price movement prediction of Bitcoin. Expert Systems with Applications, 233, 120838.

[10] Varadharajan, V., Smith, N., Kalla, D., Samaah, F., Polimetla, K., & Kumar, G. R. (2024). Stock closing price and trend prediction with LSTM-RNN. Journal of Artificial Intelligence and Big Data, 4, 877.

[11] Agakishiev, I., Härdle, W. K., Becker, D., & Zuo, X. (2025). Regime switching forecasting for cryptocurrencies. Digital Finance, 7(1), 107-131.

[12] Azamjon, M., Sattarov, O., & Cho, J. (2023). Forecasting Bitcoin volatility through on-chain and whale-alert tweet analysis using the Q-learning algorithm. IEEE Access, 11, 108092-108103.

[13] Chatterjee, I., Chakraborti, S., & Tono, T. (2024). A comparative study of bitcoin price prediction during pre-Covid19 and whilst-Covid19 period using time series and machine learning models. Discover Analytics, 2(1), 19.