Social Network Analysis and Mining (2025) 15:93
https://doi.org/10.1007/s13278-025-01520-0

ORIGINAL ARTICLE

Benchmarking modeling architectures for cryptocurrency price prediction using financial and social media data

Tulika Shrivastava[1] · Basem Suleiman[1,2] · Sachit A. J. Desa[1,3] · Muhammad Johan Alibasa[4,5] · Andrey Tregubov[1] · Jiewen Zhang[1]

Received: 23 March 2025 / Revised: 4 August 2025 / Accepted: 6 August 2025

Abstract

The volatility of cryptocurrencies necessitates reliable short-term price prediction models for informed investment decisions. This work presents two benchmarking studies that predict cryptocurrency price over hourly and daily time horizons using market indicators and social media data. Study 1 used BERT-based sentiment analysis of hourly Twitter data combined with financial indicators, while Study 2 applied VADER sentiment analysis to daily Twitter and Google Trends data alongside financial indicators. Both studies systematically evaluated statistical models (ARIMA, ARIMAX), machine learning approaches (SVR), and deep learning architectures (1D-CNN, LSTM) including ensemble, multi-modal, and hybrid configurations. Particular attention was given to the influence of lag periods, data aggregation, and sentiment analysis nuances on cryptocurrency price. Empirical results identify LSTM as the best-performing singular prediction model, achieving a 64.5% reduction in RMSE (4.56e−03) compared with the SVR baseline in Study 1. In Study 2, the hybrid LSTM + ARIMA model delivered the strongest performance, reducing RMSE by 32.5% (RMSE=2.55e+02) relative to the best performing singular baseline. Hybrid architectures combining LSTM with ARIMA or ARIMAX consistently achieved the lowest RMSE values, outperforming all other configurations and proving especially effective at capturing price movements and turning points. These findings demonstrate how combining statistical methods with deep learning can address non-stationarity, improve sentiment preprocessing, and enhance model interpretability.

Keywords Cryptocurrency price prediction · Hybrid models · Sentiment analysis · Financial data · Social media data

1 Introduction

Recent years have seen a boom in cryptocurrency investment. Generally, a cryptocurrency may be defined as a digital peer-to-peer electronic payment system, in which transaction records are maintained by a decentralized system via cryptography. The popularity of this currency form is driven by: (i) the security of cryptographic transactions, which make fraud and hacking difficult; (ii) absence of a controlling entity, such as a bank or financial institution; and (iii) global accessibility, which ensures that these assets are traded globally without incurring fees or levies. However,

✉ Sachit A. J. Desa
sachit.desa@manchester.ac.uk

Tulika Shrivastava
tshr3215@uni.sydney.edu.au

Basem Suleiman
b.suleiman@unsw.edu.au

Muhammad Johan Alibasa
johan.alibasa@monash.edu

Andrey Tregubov
andrey.tregubov@sydney.edu.au

Jiewen Zhang
jiewen.zhang@sydney.edu.au

1 School of Computer Science, The University of Sydney, Sydney, Australia

2 School of Computer Science, The University of New South Wales, Sydney, Australia

3 School of Environment, Education and Development, The University of Manchester, M13 9PL Manchester, United Kingdom

4 Faculty of Information Technology, Monash University Indonesia, Banten, Indonesia

5 Faculty of Engineering and Technology, Sampoerna University, Jakarta, Indonesia

Published online: 30 August 2025

despite their appeal, cryptocurrencies are still viewed as a speculative investment. This is in-part a result of their round-the-clock trading, which adds to their price volatility. Due to this unpredictable pricing nature, a strong forecasting model is a valuable tool for investors considering cryptocurrency trading, especially over the short-term.

Kristoufek (2015) argues that the fundamental price of cryptocurrency is driven by their *'usage in trade, money supply, and price level'*. Other studies have extended this perspective by leveraging market metrics for forecasting price fluctuations (Khedr et al. 2021). A separate stream of work, often interlinked with price forecasting, has investigated the role of social media sentiment as a price driver of cryptocurrencies. However, existing studies often limit their scope to selective algorithms, feature sets, or simplified sentiment processing approaches, leaving critical gaps in understanding the interplay between market dynamics and sentiment-driven behavior.

This work bridges these gaps by presenting the results of two short-term cryptocurrency price prediction studies that base prediction on market indicators and social media data. Both studies adopt common predictive baseline models such as ARIMA, CNN, and LSTM. The studies explore advanced hybrid architectures and diverse feature selection techniques, offering valuable perspectives on the application of machine learning and deep learning for financial time-series forecasting. The specific research contributions of this work are as follows:

- Identifying singular and hybrid model architectures that optimize forecasting of financial time-series data.
- Evaluating the effects of social media data preprocessing and feature selection on cryptocurrency price prediction accuracy.
- Benchmarking the performance of combined model architectures on multi-format datasets.

2 Related work

Numerous studies have sought to predict cryptocurrency price using market indicators and social media data. This section reviews the state of research on singular prediction models, network architectures, and the role of social media sentiment in cryptocurrency price forecasting. Also highlighted are the key knowledge gaps across these areas and in our understanding of the factors driving short-term cryptocurrency price movements.

2.1 Singular price prediction models

Statistical models like ARIMA have been widely used for short-term cryptocurrency price forecasting. Hua (2020) highlights ARIMA's effectiveness in short-term predictions but notes its limitations for long-term horizons. Conversely, Support Vector Regression (SVR), a non-parametric model, is adept at capturing non-linear, non-stationary processes, producing competitive results against deep learning methods (Khedr et al. 2021).

Deep learning architectures such as 1D-CNN and LSTM have demonstrated superior predictive performance in this domain (Khedr et al. 2021; Mohanty et al. 2018; Chen et al. 2020; Ji et al. 2019; Huang et al. 2021). LSTM models are particularly effective at learning long-term dependencies due to their gated architecture, enabling them to incorporate the influence of historical events on current price movements (Seabe et al. 2025; Tiwari et al. 2025). CNNs, on the other hand, have been successfully employed for automatic feature extraction and for capturing complex non-linear relationships (Fawaz et al. 2018; Ji et al. 2019). Together, these models define a baseline for forecasting accuracy, which more advanced or hybrid architectures seek to improve upon.

2.2 Network architectures

Several network architectures show strong potential for short-term price prediction. Ensemble models that leverage LSTM or CNN as base learners often outperform singular models by mitigating the biases of individual classifiers. Ji et al. (2019) highlight the consistent superiority of ensemble architectures over traditional methods. Building on this, multi-modal models, where feature subsets are assigned to specialized base learners, further enhance forecasting accuracy (Khedr et al. 2021; Valencia et al. 2019; Abraham et al. 2018). Similarly, hybrid architectures that integrate statistical methods (e.g., ARIMA) with neural networks (e.g., RNN) combine the strengths of both approaches: statistical models capture linear trends and seasonality, while neural networks are well-suited for modeling non-linear dynamics. When carefully configured, such hybrid models exhibit improved predictive performance and greater robustness (Fathi 2019).

Transformers have recently gained attention for time-series forecasting, with studies showing that coupling them with LSTM can enhance prediction accuracy (Tanwar and Kumar 2022; Murray et al. 2023; Wu et al. 2024). Despite these benefits, their high computational demands and inconsistent performance at short-term horizons limit their practicality for real-time cryptocurrency prediction. In particular, the intensive resource requirements of Transformers create bottlenecks for applications that demand rapid processing (Nauen et al. 2024). Although combined architectures often achieve superior accuracy, significant challenges remain in configuring these models, processing data efficiently, and adapting them to multi-modal contexts for short-term price forecasting (Wu et al. 2024).

2.3 Social media sentiment

Market indicators such as price highs, lows, and trade volume are established drivers of cryptocurrency price (Kristoufek 2015; Khedr et al. 2021). The influence of social media sentiment, however, remains less definitive. While Bitcoin prices have been shown to correlate with sentiment extremes and tweet volume, the overall impact of sentiment is still debated. Abraham et al. (2018) argue that tweet volume is a stronger predictor than sentiment itself, advocating for more sophisticated models to explore this relationship in volatile markets. Expanding on this, Wolk (2019) and Youssfi et al. (2023) incorporate Google Trends into their analyses, demonstrating the predictive utility of search interest and negative sentiment in short-term forecasting. Recent approaches have refined sentiment modeling using transformer-based embeddings and context-aware labeling, where sentiment is trained on market-derived outcomes rather than human annotation, yielding stronger predictive power (Dashtaki et al. 2024).

Although prior studies suggest that social media sentiment can affect short-term cryptocurrency prices, there is limited consensus on the precise temporal dynamics, particularly at the 1-hour and 1-day intervals. Frohmann et al. (2023) and Wolk (2019) report that sentiment from high-frequency Twitter activity can influence intraday price fluctuations, with Wolk (2019) specifically identifying a 1-hour lag. Similarly, Li et al. (2018) find strong correlations between hourly sentiment scores and short-term price movements. In contrast, Bhatt et al. (2023) highlights a relationship between daily sentiment and daily price changes, suggesting that the influence of sentiment may vary depending on the temporal aggregation window.

2.4 Motivations and research gaps

LSTM consistently demonstrates superior performance in cryptocurrency price prediction compared to other architectures. However, optimal configurations for deploying LSTM and CNN remain unclear, particularly regarding model parameterization and prediction horizons. The integration of social media data into these architectures presents additional challenges, specifically in distinguishing significant market events from general sentiment fluctuations. A more robust assessment of temporal factors, particularly how prediction horizons influence data aggregation methods, is needed to better understand their effects on model performance and feature importance.

This work advances the understanding of multi-modal architectures for short-term cryptocurrency price prediction through empirical evaluation of both market indicators and social media features. The comprehensive benchmarking of model performance provides evidence-based insights into the effective integration of these data. In view of the limited understanding of the temporal dynamics of social media sentiment, particularly at hourly and daily lags, this study proposes a comparative assessment of how sentiment signals interact with price movements across different temporal windows, and how predictive performance varies with respect to lag granularity.

3 Data collection and processing

Study 1 was set-up to investigate the influence of high-frequency sentiment signals on short-term price movements, specifically over an hourly prediction horizon. The use of BERT for sentiment inference reflects the need for a context-aware, fine-grained model capable of capturing cryptocurrency-specific language and nuances common in social media discussions (Moradi-Kamali et al. 2025; Singh and Bhat 2024). Twitter was selected as the primary sentiment source due to its prominence as a real-time platform for market commentary and news dissemination. Hourly aggregation of both sentiment and market data allows for a detailed examination of intraday price dynamics, aligning with prior studies that emphasize rapid market responses to online sentiment shifts.

Study 2 extends the investigation to a daily prediction horizon, enabling the inclusion of broader sentiment indicators and reduced noise from short-term fluctuations. In this context, VADER is adopted as a lightweight yet effective sentiment tool capable of processing large volumes of data efficiently while still capturing general sentiment trends Hutto and Gilbert (2014); Chadha and Chaudhary (2023). In addition to Twitter, the study incorporates Google Trends data to account for latent investor interest beyond social media, offering a more comprehensive view of public sentiment. Daily aggregation is aligned with the longer lag structures observed in previous research and is better suited for detecting sustained market trends across cryptocurrencies such as Bitcoin and Ethereum.

3.1 Study 1: Sentiment inference model and market indicators

Twitter data used for study 1 was sourced from Kaggle and covered a time period of 3 years, from January 2017 to December 2019 (Dutta 2021). The data was made available in its raw format with information relating usernames, urls, timestamps, likes, retweets, etc...The dataset consisted of a total of 16 M tweets.

In the interest of selecting an appropriate sentiment analysis model, three inference models were evaluated for their accuracy: VADER (rule-based), TextBlob (machine learning), and BERT (deep learning). The models were evaluated over labeled twitter data, with BERT performing notably better across accuracy, precision, recall, and F1-scores. In a held-out validation set of manually labelled tweets, BERT achieved higher overall accuracy than VADER, a margin large enough to guide our model choice. The detailed confusion matrices were discarded once aggregated sentiment scores had been generated in accordance with our data-retention protocol. BERT was selected as the preferred sentiment analysis model and generated a sentiment label (positive, neutral, or negative) for each tweet in the dataset. The labels were categorically aggregated on an hourly basis and matched with the time increment and cryptocurrency price (hourly representation of market data). The number of likes for each tweet was counted toward the classified sentiment label of that tweet. For instance, if a tweet had 100 likes, and it was classified positive, then a count of 101 was added to the total number of positive opinions. A preliminary manual comparison of sentiment scores before and after applying the weighting in Equation 1 revealed a noticeable reduction in sentiment-score skewness, suggesting that normalisation helps mitigate influencer bias. The final twitter dataset consisted of: timestamp (hourly), number of positive opinions, negative opinions, and neutral opinions.

Cryptocurrency market data, recorded at 1-minute intervals, was sourced from Kaggle for the same period as the Twitter data. This dataset included details such as date, opening price, closing price, high price, low price, and transaction volume. Data was aggregated to hourly intervals, retaining the timestamp, highest price, lowest price, transaction volume, and closing price for the next time step. Figure 1 presents a candlestick chart showing Bitcoin's opening price and its correlation with Twitter sentiment counts. The chart illustrates how peaks and troughs in price often align with fluctuations in sentiment volume, highlighting the potential influence of high-frequency social media activity on short-term market dynamics. The trends observed in the figure underscore the motivation for incorporating sentiment analysis for price prediction, particularly for capturing intraday volatility.

3.2 Study 2: Sentiment weighting, Google Trends, and market indicators

In Study 2, Twitter data related to Bitcoin and Ethereum was collected over three years via API. Data preprocessing involved removing extraneous elements such as hashtags, URLs, and mentions while retaining text-based information. Automated accounts (algorithmically-driven profiles or"bots") were also excluded from the dataset, consistent with study 1. VADER was chosen for sentiment analysis in this study due to its lightweight architecture and ability to generate both sentiment labels and compound scores, representing overall sentiment. VADER assigns a positive, neutral, or negative label to each tweet, as well as calculating a compound score. The score represents the overall sentiment of a string of text (i.e. sum of individual word scores).

To derive the daily sentiment score, a weighted average method was employed. This approach is supported by studies that demonstrate how the number of retweets or followers can influence engagement, implicitly addressing 'influencer bias' Rodriguez-Ibanez et al. (2023); Xu et al. (2022). The weighted average took into account the metadata for each tweet: number of replies, likes, quotes, followers, accounts followed, and total number of tweets. To avoid the skew introduced by influential accounts, the metadata was normalised before the weighted average was calculated. Eq (1) presents the calculation of daily sentiment score.

$$\text{sentiment score} \rightarrow \frac{\sum(\text{compound score}_{\text{norm}} \times \text{weight}_{\text{norm}})}{\sum(\text{weight}_{\text{norm}})} \tag{1}$$

Google Trends data (January 2019-March 2022) was included to capture a broader spectrum of investor sentiment, accounting for individuals not actively engaged on Twitter. Features such as search interest and related topics

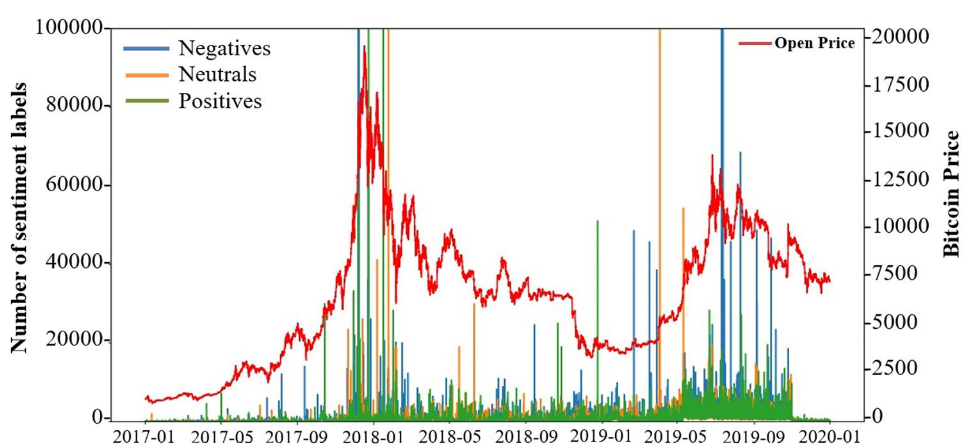


Fig. 1 Candlestick chart for Bitcoin price from January 2017 to December 2019

were transformed into count data and added to the social media dataset. Specifically, Google Trends data was transformed into count data using a threshold-based conversion method. Search interest values (0-100) were mapped to count values based on intensity levels, with higher search interest corresponding to higher count values. This transformation preserved the relative importance of search trends while providing a more concrete representation of public interest. The resulting count data was then normalized to account for overall search volume variations across different time periods.

Financial market data related to Bitcoin and Ethereum were sourced from Yahoo Finance (2022). The data consisted of opening price, closing price, high, and low market values, and did not require any preprocessing.

3.3 Evaluating lag-times

A comprehensive analysis of lag structures for Bitcoin and Ethereum was conducted using three complementary statistical approaches: Granger causality testing to assess directionality, cross-correlation analysis to identify temporal dependencies, and Autoregressive Distributed Lag (ARDL) modeling to capture dynamic relationships. To identify the most appropriate lag for each cryptocurrency, we tested lag values ranging from 1 to 10 time steps and selected those with the strongest and most consistent signals across all three methods.

For Bitcoin, Granger causality tests indicated that tweet volume significantly influenced closing price at a 1-lag structure ($p \leq 0.05$), with no consistent significance at higher lags. This was further supported by cross-correlation analysis, which revealed a clear peak at lag 1. In contrast, Ethereum displayed a delayed reaction to sentiment signals: Granger causality results were statistically significant only at 5 lags, and cross-correlation showed the strongest association at this interval. ARDL models further confirmed that these lag structures minimized residual error and improved model fit. To validate these findings, we tested alternative lag structures (e.g., 2-4 lags for Ethereum), but found that they either reduced predictive performance or failed to produce statistically significant relationships. Figure 2 visualizes the Granger causality relationship between tweet volume and Bitcoin price at lag 1. The Granger causality relationship highlights that tweet volume is a statistically significant predictor of subsequent price movements, as evidenced by the alignment of peaks at a one-hour lag. This finding reinforced the hypothesis that high-frequency sentiment signals can exert an immediate influence on cryptocurrency markets.

Table 1 summarizes the experimental configurations for both studies, highlighting differences in temporal scale and sentiment analysis methods. Study 1 focuses on hourly predictions using BERT-based Twitter sentiment, while Study 2 adopts a daily horizon with VADER sentiment, Google Trends data, and longer lag structures (particularly for Ethereum) to capture broader and more sustained market movements.

4 Methodology

In both studies, 1D-CNN and LSTM models were employed as individual predictors to establish baseline accuracies for cryptocurrency price prediction. Building upon these baselines, more sophisticated modeling architectures were

Table 1 Experimental configuration across studies

Feature	Study 1	Study 2
Sentiment analysis model	BERT	VADER
Social media data	twitter	twitter & google trends
Market data	bitcoin	bitcoin & Ethereum
Lag period	1 lag	1 lag (Bitcoin) 5 lags (Ethereum)
Prediction horizon	1-hour	1-day

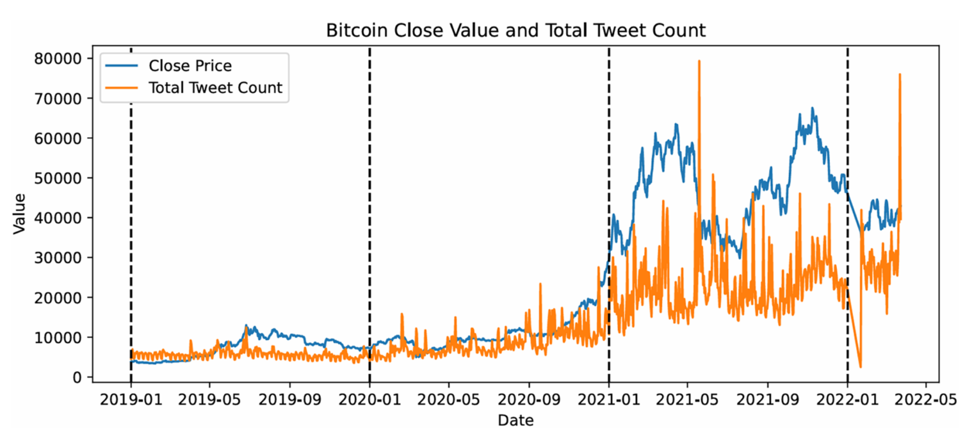


Fig. 2 Granger causality for Bitcoin close price and tweet volume

developed to enhance predictive performance and surpass the benchmarks set by individual models. The experimental framework, outlined in Tables 2 & 3, systematically examines price prediction across model architectures, feature combinations, and feature processing techniques.

The decision to predict cryptocurrency prices rather than *returns* was driven by the specific goals of this research. While financial modeling often favors *returns* due to their stationarity, price prediction aligns more closely with the practical needs of traders and market participants, who often prioritize forecasting absolute price levels for decision-making. Returns, defined as percentage changes in price over time, remove trends and seasonality, making them ideal for econometric models. However, the current approach in both studies addresses non-stationarity in price prediction through established preprocessing techniques. In doing so, both studies maintain interpretability of direct price predictions. Furthermore, by leveraging deep learning models (1D-CNN and LSTM), which are more flexible in handling varied data characteristics, both studies were able to model complex temporal dependencies directly from price data.

4.1 Baseline models

The **A**uto**R**egressive, **I**ntegrated, **M**oving **A**verage (ARIMA) is a widely used statistical time series prediction model. ARIMA operates under the assumption of stationarity, where statistical properties like mean and variance remain constant over time. To satisfy this assumption, the input time series were tested for stationarity using the Augmented Dickey-Fuller (ADF) test and transformed via log and first-order differencing. For cases involving additional independent variables, ARIMAX was used, allowing the inclusion of external factors such as sentiment or trend data. This extension is particularly beneficial for capturing relationships between market dynamics and external influences. The parameters, q, representing current values based on past residual errors, and, p, representing lagged observations, were tuned via a parameter grid search using the Akaike Information Criterion (AIC).

The choice of Support Vector Machine (SVR) in study 1 was inspired by its effectiveness in capturing non-linear relationships to successfully predict Bitcoin price (Altan et al. 2019). The parameters tuned for SVR were: a regularization parameter (C), margin separation parameter (ε), and width of the radial basis function (RBF) kernel (γ). Optimal values for these parameters were identified through k-fold cross-validation to prevent over-fitting and ensure robust performance.

The LSTM models used in this study required the following hyperparameters to be tuned: prediction window size, number of hidden units in the LSTM block, and batch size. For the 1D-CNN model, the parameters to be tuned included

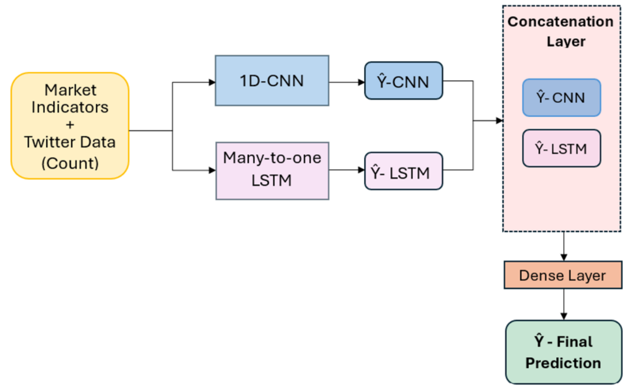


Fig. 3 Ensemble model architecture

Table 2 Experimental setup for study 1

Method	Algorithm	Dataset Features
Singular	ARIMA	Market indicators
	SVR	Market indicators
		Market indicators+Twitter
	CNN	Market indicators
		Market indicators+Twitter
	LSTM	Market indicators
		Market indicators+Twitter
Ensemble	CNN+LSTM	Market indicators
		Market indicators + Twitter
Multimodal	CNN+LSTM	Market(LSTM)+Twitter (CNN)
		Market(CNN)+Twitter (LSTM)

window size, number of conv-1d blocks, and batch size. Hyperparameter tuning was performed through randomized search on a predefined parameter grid, with evaluation based on R^2, RMSE, and MAPE metrics. For robust training, and in order to prevent information leakage, the following measures were implemented: a. Testing and training data were split chronologically to ensure future data did not influence past predictions; b. Transformations were performed on the training data and these same transformations replicated on the validation and test sets using parameters derived from training sets; and c. Temporal cross-validation was used to ensure that validation folds respected time-series structure.

4.2 Study 1: Ensemble and multi-modal models

Bespoke architectures built in study 1 include an ensemble model and multi-modal model. Both modeling architectures employed 1D-CNNs and LSTMs as base learners. Since the hyperparameters of 1D CNN and the LSTM model were tuned separately, only parameters related to optimal window size and batch size were tuned for the ensemble and multi-modal models. Tuning the window and batch size allowed for the optimization of model performance by controlling the trade-off between computational time and accuracy.

The ensemble model consists of two base-learners, CNN and LSTM, which accept the same input data. Each base learner generates its own output, that is concatenated and fed into a dense layer. This layer optimizes the final prediction by calculating a weighted combination of individual predictions. Figure 3 illustrates the architecture. The multi-modal model, differing only in input modalities, separates data streams such that one base learner receives market indicators while the other processes sentiment data. The experimental setup for study 1 is presented in Table 2. The setup highlights how different combinations of market indicators and Twitter sentiment were used to assess the added value of high-frequency sentiment signals in hourly price prediction.

4.3 Study 2: Hybrid models

Study 2 employs a hybrid modeling approach that combines deep learning models (LSTM or 1D-CNN) with statistical methods (ARIMA or ARIMAX). For financial data, LSTM and 1D-CNN were integrated with ARIMA, while analysis incorporating both financial and non-financial data used ARIMAX. The hybrid architecture operates in two stages: first, the statistical model generates residuals, which are then used as inputs for the deep learning model to predict the final residual values. These predicted residuals are transformed back to the original scale for comparison with the test data. By incorporating residuals as an additional input feature, the deep learning component captures patterns and nuances not fully explained by the statistical model. This two-stage process leverages the complementary strengths of both paradigms–statistical methods account for underlying data structures, while deep learning models detect complex temporal dependencies. Figure 4 illustrates this workflow.

Study 2 evaluated four primary feature categories: market indicators (MI), Twitter sentiment (TS), Tweet count (TC), and Google trends (GT). These features were tested in multiple combinations to assess their relative contributions to prediction accuracy and to understand feature interactions. Table 3 presents the experimental configurations for Study 2, which systematically tested different feature combinations of market indicators, Twitter sentiment, tweet counts, and Google Trends data. Table 3 illustrates how the study evaluated the incremental contribution of social media and search interest signals in daily cryptocurrency price prediction.

Table 3 Experimental setup for study 2

Method	Algorithm	Dataset Features
Singular	CNN/LSTM	Market Indicators
		Market Indicators, TC
		Market Indicators, TS
		Market Indicators, GT
		Market Indicators, TS, TC & GT
Hybrid	CNN + ARIMA	Market Indicators
	LSTM + ARIMA	Market Indicators
	CNN + ARIMAX	Market Indicators, TC
		Market Indicators, TS
		Market Indicators, GT
		Market Indicators, TS, TC & GT
	LSTM + ARIMAX	Market Indicators, TC
		Market Indicators, TS
		Market Indicators, GT
		Market Indicators, TS, TC & GT

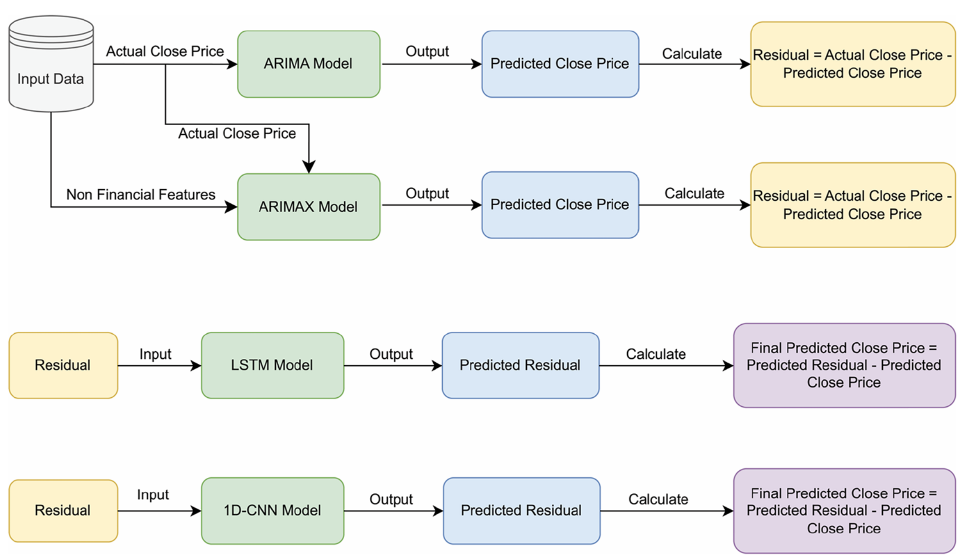


Fig. 4 Hybrid model workflow

4.4 Model training, tuning, and ablation analysis

4.4.1 Sentiment analysis

In Study 1, three sentiment classifiers were evaluated: VADER (rule-based), TextBlob (lexicon-based machine learning), and BERT (deep learning). Each was tested on two labeled Twitter datasets: one containing general-topic tweets and the other finance-related content. BERT, fine-tuned on both datasets, achieved the highest accuracy, precision, recall, and F1-scores, and was therefore selected. Its ability to capture subtle, context-specific language made it particularly suited for real-time cryptocurrency commentary. In Study 2, sentiment was assessed at a daily horizon using VADER, a lightweight and efficient model capable of processing large volumes of data. Its use is consistent with prior research validating its effectiveness for longer prediction horizons than hourly intervals Hutto and Gilbert (2014); Chadha and Chaudhary (2023); Doan (2025). Specifically, (Lupu and Donoiu 2025) demonstrate VADER's rule-based simplicity to be effective in capturing daily market sentiment, where it outperforms more complex models.

4.4.2 Temporal sensitivity analysis

Both studies explored a range of prediction horizons and data aggregation frequency, offering insights into temporal sensitivity. In Study 1, features were aggregated hourly to align with a 1 h prediction horizon, enabling evaluation of high-frequency sentiment signals and their impact on rapid intra-day price movements. In Study 2, the horizon was extended to daily forecasts, reducing short-term noise and allowing for the inclusion of broader sentiment indicators such as Google Trends. Rolling forecast experiments with ARIMA and ARIMAX tested horizons of 1, 7, 14, 30, 60, and 90 days, with training sets iteratively updated after each forecast window. As shown in Table 4 and Figure 5, accuracy was highest at shorter horizons and declined with longer ones. External features (Twitter sentiment, Tweet volume, Google Trends) yielded modest gains in the short term but did not prevent error accumulation over extended horizons.

Table 4 Rolling forecast performance for Ethereum using ARIMA

Rolling forecast period (days)	RMSE	MAPE
1	139.36	0.03
7	204.43	0.04
14	248.36	0.05
30	440.32	0.11
60	563.99	0.14
90	513.33	0.14

4.4.3 Hyperparameter selection

All models were trained with a fixed random seed to ensure reproducibility and reduce variance from random weight initialization and data shuffling. To prevent data leakage, a forward-chaining time-series cross-validation strategy was used, preserving chronological order in each fold. Data transformations, including normalization, were computed solely on the training fold and then applied to validation and test sets, simulating live trading conditions.

Hyperparameter tuning for deep learning models–LSTM and 1D-CNN–was performed using randomized search over defined parameter ranges. For LSTM, parameters included hidden units (64–512), input sequence length (20–100 time steps), and batch size (64–512). For 1D-CNN, the search space included the number of Conv1D layers (1–5), kernel sizes (3–7), filters per layer (16–128), and batch sizes (64–512). All deep learning models used the Adam optimizer, with learning rates sampled logarithmically between $1e^{-4}$ and $1e^{-2}$. Training was monitored with early stopping based on validation loss (patience=5 epochs, max=50 epochs).

For both statistical models, hyperparameter tuning was conducted using grid search. ARIMA and ARIMAX parameters (p, d, q) were explored in ranges of p = 0–5, d = 0–2, and q = 0–5, with final models selected using the lowest Akaike Information Criterion (AIC). Support Vector Regression (SVR) parameters were tuned across C (0.1–100), ϵ (0.001–10), and γ (0.001–1), minimizing RMSE under time-series cross-validation. Figure 5 present the hyperparameter ranges and selections for the models used across studies 1 & 2.

5 Results

Prediction results were evaluated over metrics of MAPE and RMSE, these values have been standardized via Z-scores for comparison. The models for study 1 were trained on Bitcoin data from January, 2017 to August, 2019 and Bitcoin price prediction was obtained for the period of September to December of 2019 (4 months). The models for study 2 were trained on cryptocurrency data from January, 2019 to February, 2021 and prediction was obtained between February, 2021 to August, 2022. A comparison of actual and predicted prices for both studies are presented in Tables 6 to 8. Table 6 reports Study 1 outcomes using singular, ensemble, and multi modal models, while Tables 7 and 8 summarize Study 2 results for Bitcoin and Ethereum, respectively, evaluating singular

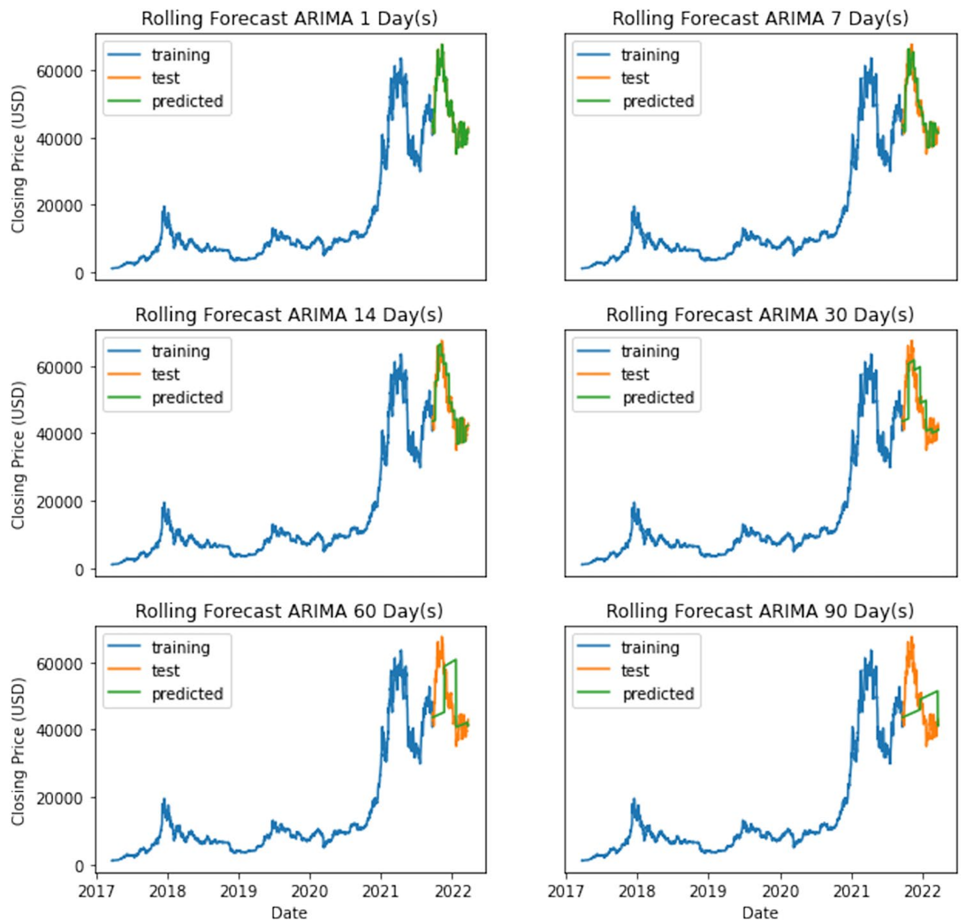


Fig. 5 Price Prediction using Rolling Forecast

Table 5 Model training ranges and selected parameters

Method	Algorithm	Range	Selected parameters
Statistical	ARIMA/ ARIMAX	p=0–5	p=0
		d=0–2	d=1
		q=0–5	q=2
Machine learning	SVR	C=0.1–100	C=10
		ϵ=0.001–10	ϵ=5.86
		γ=0.001–1	γ=0.02
Deep learning	CNN	Window=20–100	Window=48
		ConvBlock=1–5	ConvBlock=1
		Batch=64–512	Batch=256
	LSTM	Window=20–100	Window=48
		Hidden=64–512	Hidden=256
		Batch=64–512	Batch=256

and hybrid configurations with different feature combinations. All tables provide RMSE, MAPE, and standardized Z scores to enable consistent comparison of model performance. It is noted that in Study 1, the errors are based on adjusted data, so the numbers are small. In Study 2, the errors are based on real price values.

Key observations from both studies include:

- Multi-modal, ensemble, and hybrid architecture models demonstrate robust performance, generally outperforming singular models.
- LSTM is the leading time-series prediction model amongst singular models.

Table 6 Bitcoin price prediction for study 1

Method	Algorithm	Data Features	RMSE	MAPE	Zscore RMSE	Zscore MAPE
Singular Models	LSTM	MI	4.56e-03	2.10e-05	-1.09	-1.39
Multi-modal Models	CNN + LSTM	MI (LSTM) + TC (CNN)	4.63e-03	2.10e-05	-1.08	-1.39
Singular Models	LSTM	MI + TC	4.68e-03	2.20e-05	-1.07	-1.38
Multi-modal Models	CNN + LSTM	MI (CNN) + TC (LSTM)	1.28e-02	1.69e-04	-0.06	0.30
Singular Models	CNN	MI + TC	1.30e-02	1.69e-04	-0.05	0.30
Ensemble Models	CNN + LSTM	MI + TC	1.30e-02	1.69e-04	-0.05	0.30
Ensemble Models	CNN + LSTM	MI	1.36e-02	1.86e-04	0.03	0.49
Singular Models	CNN	MI	1.39e-02	1.93e-04	0.07	0.57
Singular Models	SVR	MI	2.53e-02	2.31e-04	1.47	1.01
Singular Models	SVR	MI + TC	2.81e-02	2.46e-04	1.82	1.18
Singular Models	ARIMA	MI	4.00e+03	4.68e+01	–	–

Table 7 Bitcoin price prediction for study 2

Method	Algorithm	Data Features	RMSE	MAPE	Zscore RMSE	Zscore MAPE
Hybrid	LSTM + ARIMA	MI	1725.3	0.027	-1.08	-1.11
	LSTM + ARIMAX	MI, TS, TC & GT	1866.9	0.028	-0.96	-1.07
	LSTM + ARIMAX	MI, TC	1901.5	0.028	-0.93	-1.07
	LSTM + ARIMAX	MI, GT	1903.3	0.028	-0.93	-1.07
	LSTM + ARIMAX	MI, TS	1921.1	0.058	-0.92	0.26
	CNN + ARIMAX	MI, TS, TC & GT	2026.2	0.033	-0.83	-0.85
	CNN + ARIMAX	MI, TS	2066.4	0.033	-0.80	-0.85
	CNN + ARIMAX	MI, TC	2090.6	0.033	-0.78	-0.85
	CNN + ARIMAX	MI, GT	2112.4	0.034	-0.76	-0.80
Singular Models	LSTM	MI	2557.9	0.044	-0.38	-0.36
Singular Models	LSTM	MI, GT	2926.6	0.052	-0.07	-0.01
Hybrid	CNN + ARIMA	MI	2983.2	0.051	-0.02	-0.05
Singular Models	LSTM	MI, TS	3387.5	0.058	0.32	0.26
Singular Models	CNN	MI	3679.9	0.057	0.57	0.21
Singular Models	LSTM	MI, TS, TC & GT	3801.8	0.066	0.67	0.61
Singular Models	LSTM	MI, TC	4075.7	0.075	0.90	1.00
Singular Models	CNN	MI, TC	4151.6	0.070	0.96	0.78
Singular Models	CNN	MI, TS	4520.3	0.077	1.27	1.09
Singular Models	CNN	MI, TS, TC & GT	4742.0	0.083	1.46	1.36
Singular Models	CNN	MI, GT	5743.1	0.109	2.31	2.50

Table 8 Ethereum price prediction for study 2

Method	Algorithm	Data Features	RMSE	MAPE	Zscore RMSE	Zscore MAPE
Hybrid	LSTM + ARIMAX	MI, TS, TC & GT	153.2	0.033	-0.61	-0.61
	LSTM + ARIMAX	MI, GT	153.6	0.033	-0.60	-0.61
	LSTM + ARIMAX	MI, TC	153.7	0.033	-0.60	-0.61
	LSTM + ARIMAX	MI, TS	154.4	0.033	-0.60	-0.61
Hybrid	CNN + ARIMAX	MI, TS, TC & GT	156.5	0.034	-0.60	-0.60
	CNN + ARIMAX	MI, GT	157.1	0.034	-0.60	-0.60
	CNN + ARIMAX	MI, TS	158.7	0.034	-0.60	-0.60
	CNN + ARIMAX	MI, TC	161.4	0.035	-0.59	-0.59
Hybrid	LSTM + ARIMA	MI	177.6	0.043	-0.56	-0.53
Singular Models	LSTM	MI	212.4	0.051	-0.49	-0.47
	CNN	MI, TS	236.3	0.055	-0.44	-0.43
	CNN	MI	262.3	0.065	-0.39	-0.35
	CNN	MI, TC	284.9	0.066	-0.34	-0.35
	CNN	MI, TS, TC & GT	314.6	0.073	-0.28	-0.29
Singular Models	LSTM	MI, TS, TC & GT	1293.1	0.306	1.66	1.59
	LSTM	MI, TS	1402.9	0.342	1.88	1.88
	LSTM	MI, TC	1495.9	0.370	2.07	2.10
	LSTM	MI, GT	1504.1	0.373	2.08	2.12

- Combined model architectures, with LSTM as a base-learner, demonstrate superior results to networks without LSTM as a base-learner.
- Market indicators, alone, are strong price predictors, even without social media data.

Figures 6 to 8 present a visual comparison of the actual price versus the predicted price for the two best performing models in studies 1 & 2. Figure 6 presents Bitcoin predictions from Study 1, comparing an LSTM model using market indicators with a multi-modal LSTM–CNN configuration. Study 2 results for Bitcoin are presented in Figure 7, featuring hybrid LSTM+ARIMA and LSTM+ARIMAX models, while Figure 8 presents Ethereum predictions generated by hybrid LSTM+ARIMAX models with varying feature sets. Collectively, these figures highlight how the best-performing models align with actual price movements across different time horizons and data configurations.

The range of Bitcoin price during testing in study 1 ranged from 6,600 to 10,400 USD across 4 months. In contrast, for study 2, the Bitcoin price fluctuated between 39,600 and 67,000 USD, whilst the Ethereum price ranged between 2,300 to 4,900 USD across 16 months. From figure 6 to 8 we infer:

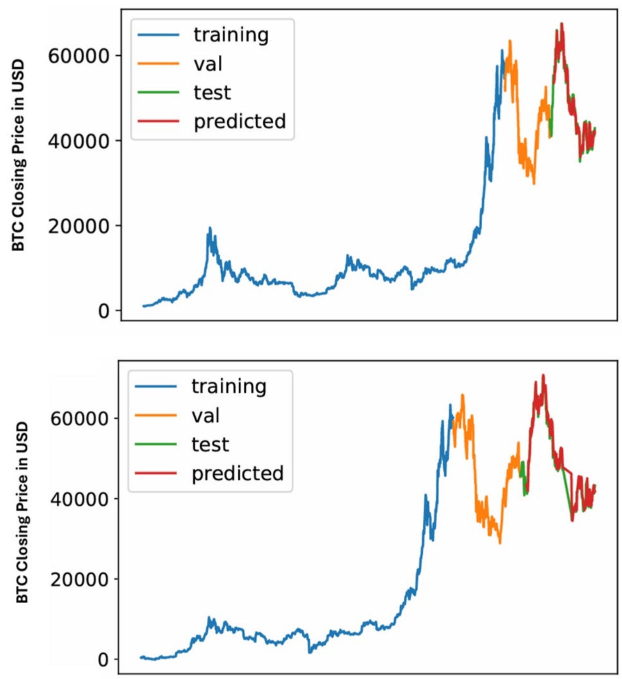


Fig. 7 Two best performing models for Bitcoin in study 2: (Top) Hybrid model of LSTM+ARIMA (market indicators), (Bottom) Hybrid model of LSTM+ARIMAX (market indicators, twitter count, sentiment, and google trends)

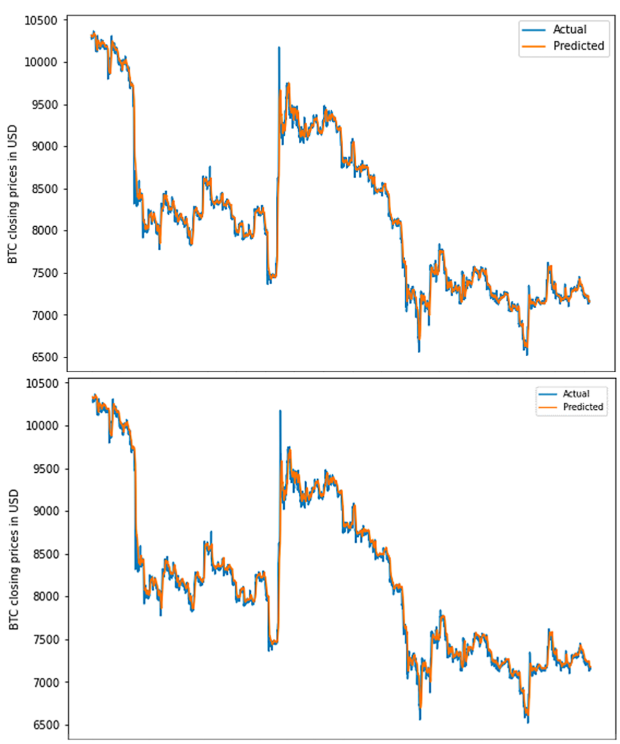


Fig. 6 Two best performing models of study 1: (Top) LSTM with market indicators, (Bottom) multi-modal model of LSTM (market indicators) and CNN (twitter count)

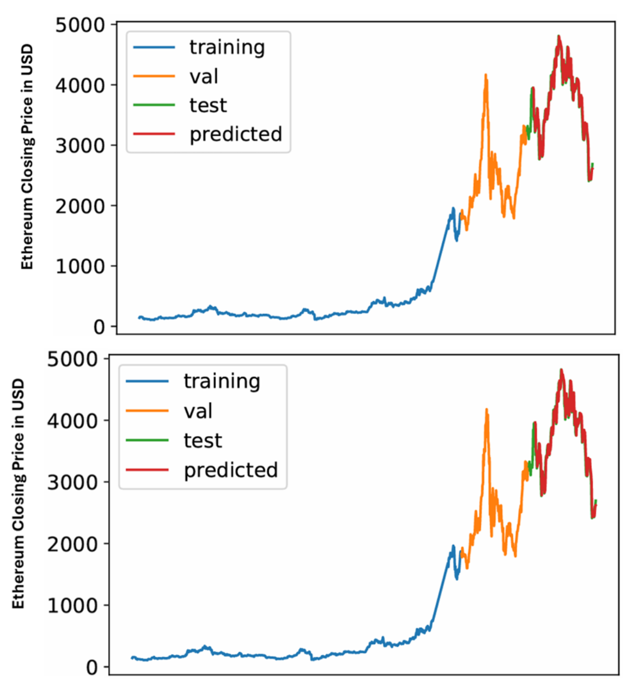


Fig. 8 Two best performing models for Ethereum in study 2: (Top) Hybrid model of LSTM+ARIMAX (market indicators, twitter count, sentiment, and google trends), (Bottom) Hybrid model of LSTM+ARIMAX (market indicators and google trends)

- While Study 2 is able to predict upturns and downturns, as shown by the predicted line (red) aligning with the actual price (green) at peaks and troughs, latent patterns and/or other factors still influence these inflections.
- Hybrid architectures in study 2 are able to adequately capture steep increases, decreases, and sudden price inflections, reinforcing their robustness.
- Generally, the architectures explored in both studies are able to maintain strong predictive power across a broad pricing range.
- Micro and macros price fluctuations are generally well-captured across the models in Study 1 & 2.

6 Discussion

6.1 Efficiency of network architectures

Network architectures with two base learners generally outperform singular models, with the notable exception of LSTM in Study 1. Among such configurations, multi-modal models typically achieve higher accuracy than ensemble models, while hybrid models consistently emerge as the best predictors.

In Study 1, the multi-modal and ensemble architectures combined LSTM and CNN. By contrast, hybrid architectures integrated statistical models (ARIMA or ARIMAX) with deep learning models (LSTM or CNN). In this design, residuals from the statistical model were used as inputs for the deep learning component, enabling the hybrid models to capture patterns and nuances not fully explained by the statistical approach.

For Bitcoin and Ethereum price prediction in Study 2, incorporating LSTM as a base learner markedly improved accuracy. The results confirm that LSTMs hold a clear advantage because they are well suited to modeling the temporal and long-term dependencies of non-stationary data (Seabe et al. 2023; Patel et al. 2023). This study extends prior work by showing that LSTM-based hybrids outperform alternative architectures even when input data are limited. Notably, CNN-based hybrid models (CNN+ARIMA/ARIMAX) did not reach the same level of accuracy as their LSTM-based counterparts, even when provided with the full feature set (MI, TS, TC, and GT).

Another important finding is the stability of hybrid models relative to both ensemble and multi-modal configurations. The range of Z-scores for multi-modal models corresponds to approximately one standard deviation, whereas for hybrid models it narrows to 0.32 for Bitcoin and 0.05 for Ethereum. The large performance gap between the two multi-modal models suggests that their accuracy is highly sensitive to the type of data processed by each base learner, highlighting the robustness of hybrid approaches.

6.2 Feature importance and temporal dynamics

The empirical results offer clear insights into feature importance for cryptocurrency price prediction. Across both studies, market indicators consistently emerged as the strongest predictors. This was particularly evident in Study 1, where the LSTM model trained solely on market indicators achieved the lowest RMSE (4.56e−03) and MAPE (2.10e−05) scores (Table 6), outperforming more complex models that incorporated social media features. A similar pattern was observed in Study 2: the LSTM+ARIMA hybrid using only market indicators achieved an RMSE of 1725.3, while adding social media features (Twitter, sentiment, Google Trends) resulted in slightly higher RMSE values (1866.9-−1921.1). These findings indicate that although social media signals contain predictive value, their contribution is often diminished by noise, particularly at the daily horizon.

Temporal aggregation of social media data proved to be another critical factor. Study 1 employed hourly aggregation of Twitter data, while Study 2 used daily aggregation of Twitter, sentiment, and Google Trends data. Neither approach produced notable improvements in accuracy, despite prior research linking Twitter activity with cryptocurrency price fluctuations. The influence of lag structures was especially noteworthy. While Critien et al. (2022) found that a 3-day lag outperformed a 7-day lag and suggested that a 1-day lag might be optimal, our results show that even a 1-day lag was insufficient to overcome the inherent noise in social media signals.

Further evidence of these complex temporal dynamics is provided by related studies. Said et al. (2023) observed increased volatility on the second day at a 5% significance level, while Kjaerland et al. (2018) found Google search impacts more than twice as strong with two lags (0.42% price increase) compared with a single lag (0.16%). Together, these studies suggest that the relationship between social media signals and price movements may operate on longer time scales than commonly assumed.

Finally, the restriction to Twitter and Google Trends data may have influenced our findings. Mai et al. (2018) report that cryptocurrency prices are significantly shaped by a“silent majority”of less active users, with internet forum discussions showing stronger predictive power than tweets at the inter-day level. This points to a valuable direction for future research: incorporating alternative data sources such as forums to mitigate the noise observed in Twitter and Google-based signals.

6.3 Dynamics between market indicators and social media data

The comparative performance of market-based and sentiment-based predictions reveals several important insights.

Market indicators demonstrate remarkable predictive power, particularly for short-term forecasting. This is evidenced by the consistent performance of market indicator-only models across both Bitcoin and Ethereum predictions (Tables 6-8). Two of the three experimental setups demonstrate that models trained exclusively on market indicators outperform those using the full feature set, suggesting that market variables are more reliable predictors of price movements.

Figure 9 presents a probability density plot estimating the distribution of RMSE Z-scores across the evaluated models for Bitcoin and Ethereum. The figure highlights the narrower spread of Ethereum model performances compared to Bitcoin, indicating greater stability and robustness in Ethereum predictions across different feature combinations. This aligns with Said et al. (2023), who reported that Ethereum is less sensitive to sentiment-driven shocks than Bitcoin. The visualization therefore underscores a key finding of this study: while hybrid models improve accuracy for both cryptocurrencies, the predictive stability of Ethereum models suggests that social media signals exert a weaker and less volatile influence compared to market indicators. By comparing the full distribution of model errors rather than isolated metrics, Figure 9 provides stronger evidence of model consistency and reliability.

Sentiment analysis, while intuitively appealing, showed limited effectiveness in improving short-term predictions. This is most clearly demonstrated in Study 2's results (Table 7), where models incorporating sentiment features often performed worse than their market-only counterparts. For instance, the addition of sentiment features to the CNN model increased the RMSE from 3679.9 to 4520.3. At the 1-hour prediction horizon, Twitter sentiment appeared to have an almost negligible influence on price, while market indicators maintained strong predictive power. The temporal horizon of predictions emerged as a crucial factor in model performance. While the 1-day prediction horizon showed improved accuracy compared to hourly predictions, the current approach of aggregating data into single records per time period (hourly or daily) may limit model generalizability. This limitation is particularly evident in the performance degradation when adding social media features, suggesting that the models struggle to effectively incorporate this additional information.

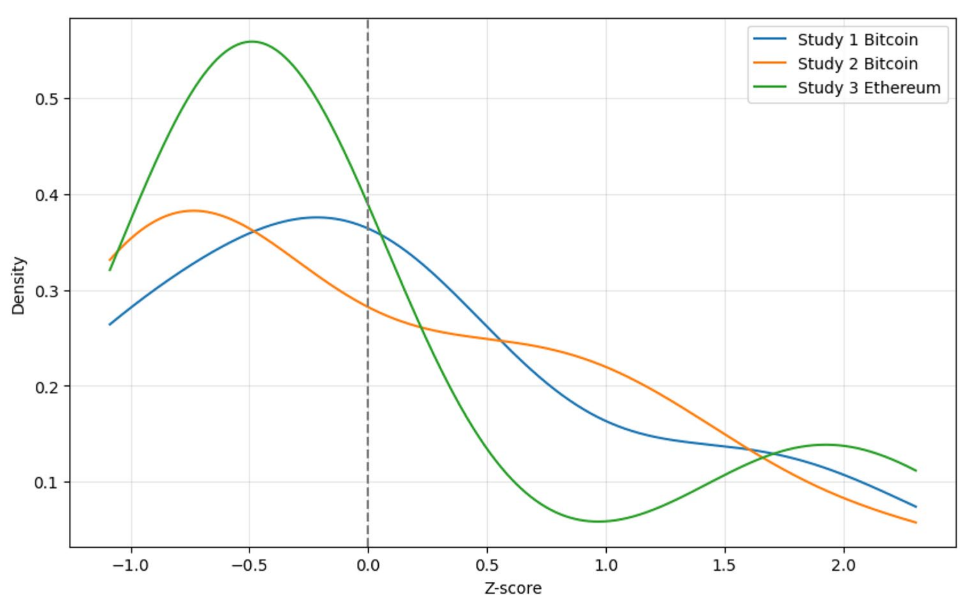
6.4 Comparison with emerging model architectures

Current research in cryptocurrency price prediction establishes Long Short-Term Memory (LSTM) networks as foundational. Their inherent ability to capture long-range dependencies in complex, volatile financial time series makes them a cornerstone for this task (Kanaparthi 2024). Variations like Gated Recurrent Units (GRU) and Bi-directional LSTMs further demonstrate how the core LSTM architecture can be refined for enhanced performance Ashok (2024).

Beyond singular models, hybrid architectures are increasingly recognized as a robust approach that excel by combining the strengths of different deep learning components. Studies show that models integrating LSTM with other techniques often outperform standalone LSTMs in price prediction (Ladhari and Boubaker 2024; Tiwari et al. 2025; Lupu and Donoiu 2025).

An emerging trend in cryptocurrency price prediction is the use of Transformer-based models, often in combination with LSTMs or GRUs. This hybrid approach leverages the Transformer's advanced attention mechanisms to capture long-range patterns and contextual relationships, while

Fig. 9 Probability distribution of Z-scores



drawing on the proven ability of LSTMs to model sequential and short-term dependencies (Khaniki and Manthouri 2024).

6.5 Summary of key findings

The two studies make several notable contributions to the literature on cryptocurrency price prediction. First, they confirm the superior effectiveness of hybrid architectures, particularly those combining statistical methods (ARIMA/ARIMAX) with deep learning (LSTM). Across both studies, hybrid models consistently outperformed singular models, achieving lower error rates and better Z-scores, as shown in Tables 6 and 7.

Second, the findings challenge the assumed importance of social media sentiment in short-term price prediction. While prior research has emphasized the influence of social media signals, our results suggest a more nuanced relationship, especially at shorter time horizons. As illustrated in Figures 6–8, models relying primarily on market indicators achieved prediction accuracies comparable to, or better than, those incorporating social media features.

Third, the studies advance understanding of model stability and reliability. Comparative analysis of Z-scores across different architectures shows that hybrid models not only deliver higher accuracy but also maintain consistent performance under varying market conditions. This is particularly evident in the Ethereum predictions, where hybrid models sustained stable performance across diverse feature combinations (Table 8).

Overall, the findings indicate that while social media sentiment may influence cryptocurrency prices, its effect is complex and less direct than previously assumed. The strong and consistent performance of market indicators, especially when integrated into sophisticated hybrid models, underscores the continued reliability of technical analysis for short-term price prediction.

7 Limitations & future work

This study provides a comparative analysis of two experimental setups for cryptocurrency price prediction, covering Bitcoin and Ethereum under different temporal and modeling conditions. While this design enables broader benchmarking, it also limits direct comparability, as the studies differ in prediction horizons (1 h vs. 1 day), sentiment sources, and inference tools. While these differences allow for broader benchmarking across real-world contexts, they limit the direct comparability of results and preclude a unified experimental framework. The main limitations of this approach, and of our work more broadly, are outlined below.

The studies do not test a singular theory of price formation; instead, they assess how predictive performance varies across feature sets, lag periods, and model types. Although lag structures were selected based on prior empirical evidence and validated through statistical analysis, a more systematic exploration of lag sensitivity could yield stronger theoretical insights. Similarly, the use of Twitter and Google Trends data constrains sentiment analysis to specific platforms and time periods, which may limit generalizability to newer or alternative data sources (e.g., forums, news media).

In terms of evaluation, the study reports standardized performance metrics (RMSE, MAPE, and Z-scores) but does not include formal statistical testing (e.g., Diebold-Mariano tests) or confidence intervals. As a result, minor differences between similarly ranked models should be interpreted cautiously.

Finally, while price prediction was chosen for its interpretability and relevance to trading applications, this decision entails trade-offs. RMSE and MAPE capture magnitude-based error, but they overlook directional accuracy and investment utility. Returns, in contrast, are often more statistically stable and better suited for risk-adjusted evaluation. The reliance on price prediction and error magnitude metrics therefore limits the ability to fully assess model effectiveness in real-world trading scenarios.

Building on the findings of this study, and broader literature gaps, some promising directions for future research include:

- Lag sensitivity and residual analysis: Incorporating time-step-level residuals could support hypothesis testing and provide deeper insights into error distributions and lag structure effects.
- Return-based forecasting: Moving beyond price prediction to return-based forecasting would allow evaluation using risk-adjusted financial indicators such as directional accuracy, Sharpe ratio, and volatility-adjusted returns, offering a more realistic measure of trading utility.
- Broader sentiment and data retention: Retaining raw tweet content and intermediate outputs in newly collected datasets would enable richer diagnostic analyses and allow for the inclusion of additional sentiment sources (e.g., forums, news media) to improve generalizability.

8 Conclusion

This work demonstrates the value of hybrid modeling approaches in advancing short-term cryptocurrency price prediction. By integrating statistical and deep learning methods, hybrid models deliver both improved accuracy and stable performance across volatile market conditions. The key findings, across both studies, can be summarized as follows:

- Hybrid architectures are superior to ensemble models and singular models: Models combining ARIMA/ARIMAX with LSTM consistently outperformed singular approaches, achieving lower error rates and better Z-scores across both studies.
- Market indicators remain more reliable than social media sentiment: While sentiment plays a role, its influence is more nuanced than previously assumed, with market-driven models often surpassing those incorporating social features.
- Hybrid models have stable performance: Beyond accuracy, hybrid approaches demonstrated more consistent performance across varying market conditions, particularly in Ethereum predictions.

The implications of these findings extend beyond academic benchmarking: these models can inform risk management strategies, algorithmic trading systems, and fintech applications that depend on rapid, data-driven decision-making. As cryptocurrency markets continue to mature and diversify, adaptable hybrid architectures offer a promising foundation for building intelligent decision-support tools. Future research should explore longer forecasting horizons, alternative sentiment data sources, and model adaptability across asset classes. Such efforts will help clarify the nuanced role of sentiment while further refining predictive frameworks that bridge technical analysis with modern AI-driven methods.

Acknowledgements The authors would like to acknowledge several masters students that helped in the implementation and running of the models presented within this study.

Author contributions T.S., A.T., and J.Z. undertook modeling for the project and generated results M.J.A. and B.S. provided high-level guidance and advice over the course of the project S.D. helped in the preparation of the figures, tables, and manuscript

Funding The authors declare that there was no funding received for the research associated with this study.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Conflict of interest The authors declare no Conflict of interest.

References

Abraham J, Higdon DW, Nelson J, Ibarra J (2018) In Cryptocurrency price prediction using tweet volumes and sentiment analysis. [Online]. Available: https://api.semanticscholar.org/CorpusID:52950647

Altan A, Karasu S, Bekiros S (2019) Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. Chaos, Solitons Fractals 126:325–336

Ashok P (2024) Cryptocurrency price prediction using deep learning algorithms: a comparative study. Commun Appl Nonlinear Anal 32:537–551

Bhatt S, Ghazanfar M, Amirhosseini M (2023) Sentiment-driven cryptocurrency price prediction: A machine learning approach utilizing historical data and social media sentiment analysis. Mach Learn Appl An Int J (MLAIJ) 10(2/3):1–15

Chadha R, Chaudhary A (2023) A study analyzing an innovative approach to sentiment analysis with vader. J Eng Des Anal 6:23–27

Chen Z, Li C, Sun W (2020) Bitcoin price prediction using machine learning: An approach to sample dimension engineering. J Comput Appl Math 365:112395

Critien JV, Gatt A, Ellul J (2022) Bitcoin price change and trend prediction through twitter sentiment and data volume. J Financ Innov 8:05

Dashtaki SM, Chagahi MH, Moshiri B, Piran MJ (2024) Multi-source hard and soft information fusion approach for accurate cryptocurrency price movement prediction. [Online]. Available: https://arxiv.org/abs/2409.18895

Doan M (2025) Sentiment classification of bitcoin-related tweets using vader: Analyzing temporal sentiment trends in cryptocurrency markets. J Curr Res Blockchain 2:88–102

Dutta G (2021) Bitcoin tweets - 16m tweets with sentiment tagged,. [Online]. Available: https://www.kaggle.com/datasets/gauravduttakiit/bitcoin-tweets-16m-tweets-with-sentiment-tagged

Fathi O (2019) Time series forecasting using a hybrid ARIMA and LSTM model. Velvet consulting, pp.1-7

Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller P (2018) Deep learning for time series classification: a review, CoRR, vol. abs/1809.04356. [Online]. Available: http://arxiv.org/abs/1809.04356

Frohmann M, Karner M, Khudoyan S, Wagner R, Schedl M (2023) Predicting the price of Bitcoin using sentiment-enriched time series forecasting. Big Data Cogn Comput 7(3):137

Hua Y (2020) Bitcoin price prediction using arima and lstm, E3S Web of Conf

Huang X, Zhang W, Huang Y, Tang X, Zhang M, Surbiryala J, Iosifidis V, Liu Z, Zhang J (2021) Lstm based sentiment analysis for cryptocurrency prediction. International conference on database systems for advanced applications. Springer International Publishing, Cham, pp 617–621

Hutto CJ, Gilbert E (2014) Vader: a parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the international AAAI conference on web and social media (ICWSM)

8(1):216–225. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Ji S, Kim J, Im H (2019) Comparative study of bitcoin price prediction using deep learning. Mathematics 7(10):898

Kanaparthi V (2024) Robustness evaluation of lstm-based deep learning models for bitcoin price prediction in the presence of random disturbances. Int J Innove Sci Mod Eng 12:14–23

Khaniki MAL, Manthouri M (2024) Enhancing price prediction in cryptocurrency using transformer neural network and technical indicators, [Online]. Available: https://arxiv.org/abs/2403.03606

Khedr A, Arif I, P V P, El-Bannany M, Alhashmi M, Sreedharan M, (2021) Cryptocurrency price prediction using traditional statistical and machine-learning techniques: a survey. Intell Syst Account Financ Manag 28(1), 3, 34

Kjærland F, Meland M, Oust A, Øyen V (2018) How can bitcoin price fluctuations be explained? Int J Econ Financ Issues 8(3):323–332

Kristoufek L (2015) What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis. PLoS ONE 10:04

Ladhari A, Boubaker H (2024) Deep learning models for bitcoin prediction using hybrid approaches with gradient-specific optimization. Forecasting 6(2):279–295

Li TR, Chamrajnagar AS, Fong XR, Rizik NR, Fu F (2018) Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. Front Phys 7:98

Lupu R, Donoiu PC (2025) Sentiment matters for cryptocurrencies: evidence from tweets. Data 10(4):50

Mai F, Shan J, Bai Q, Wang S, Chiang R (2018) How does social media impact bitcoin value? A test of the silent majority hypothesis. J Manag Inf Syst 35:19–52

Mohanty P, Patel D, Patel P, Roy S (2018) Predicting fluctuations in cryptocurrencies' price using users' comments and real-time prices, in 2018 7th international conference on reliability, infocom technologies and optimization (trends and future directions) (ICRITO), pp. 477–482

Moradi-Kamali H, Rajabi-Ghozlou MH, Ghazavi M, Soltani A, Sattarzadeh A, Entezari-Maleki R (2025) Market-derived financial sentiment analysis: context-aware language models for crypto forecasting. [Online]. Available: https://arxiv.org/abs/2502.14897

Murray K, Rossi A, Carraro D, Visentin A (2023) On forecasting cryptocurrency prices: a comparison of machine learning, deep learning, and ensembles. Forecasting 5(1):196–209

Nauen TC, Palacio S, Raue F, Dengel A (2024) Which transformer to favor: a comparative analysis of efficiency in vision transformers . [Online]. Available: https://arxiv.org/abs/2308.09372

Patel N, Vasani N, Jadav NK, Gupta R, Tanwar S, Polkowski Z, Alqahtani F, Gafar A (2023) F-LSTM: Federated learning-based LSTM framework for cryptocurrency price prediction. Electron Res Arch 31(10):6525–6551

Rodríguez-Ibánez M, Casánez-Ventura A, Castejón-Mateos F, Cuenca-Jiménez PM (2023) A review on sentiment analysis from social media platforms. Expert Syst Appl 223:119862

Said FF, Somasuntharam RS, Yaakub MR, Sarmidi T (2023) Impact of google searches and social media on digital assets' volatility. Humanit Soc Sci Commun 10(1):1–17

Seabe PL, Moutsinga CRB, Pindza E (2023) Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach. Fractal Fract 7(2):203

Seabe P, Moutsinga C, Pindza E (2025) Sentiment-driven cryptocurrency forecasting: analyzing lstm, gru, bi-lstm, and temporal attention model (tam). Soc Netw Anal Min 15:05

Singh S,Bhat M (2024) Transformer-based approach for ethereum price prediction using crosscurrency correlation and sentiment analysis. [Online]. Available: https://arxiv.org/abs/2401.08077

Tanwar A, Kumar V (2022) Prediction of cryptocurrency prices using transformers and long short term neural networks, In. international conference on intelligent controller and computing for smart power (ICICCSP) 2022:1–4

Tiwari D, Bhati B, Nagpal B, Al-Rasheed A, Getahun M, Othman S (2025) A swarm-optimization based fusion model of sentiment analysis for cryptocurrency price prediction. Sci Rep 15:03

Valencia F, Gómez-Espinosa A, Valdés-Aguirre B (2019) Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. Entropy 21(6):589

Wołk K (2019) Advanced social media sentiment analysis for short-term cryptocurrency price prediction. Expert Syst 37:11

Wu J, Zhang X, Huang F, Zhou H, Chandra R (2024) Review of deep learning models for crypto price prediction: implementation and evaluation,

Xu QA, Chang V, Jayne C (2022) A systematic review of social media-based sentiment analysis: Emerging trends and challenges. Decis Anal J 3:100073

Yahoo!Finance, Crypto real time prices and latest news, 2022, updated Daily. [Online]. Available: https://finance.yahoo.com/crypto/

Youssfi AN, Bouchakwa M, Jamoussi Y (2023) Bitcoin price prediction considering sentiment analysis on twitter and google news, In proceedings of the 27th international database engineered applications symposium, ser. IDEAS '23.New York, NY, USA: Association for Computing Machinery, p. 71–78. [Online]. Available: https://doi.org/10.1145/3589462.3589494