Research and Analysis of Cryptocurrency Price Prediction Based on Deep Learning

Zhaoliang Xu

{mcdxcmd@gmail.com}

Faculty of Business and Economics, Monash University, Melbourne, 3168, Australia

Abstract. Cryptocurrency markets trade 24/7 and show extreme volatility, heavy-tailed returns, and regime shifts, making forecasting difficult yet valuable for risk management and trading. This review synthesizes recent AI research on cryptocurrency prediction with a focus on deep learning. The paper classifies studies into CNN-based models, RNN and hybrid architectures, attention/Transformer methods, graph neural networks capturing cross-asset effects, and decomposition or multimodal pipelines that integrate signals such as sentiment. The paper compares work by forecasting targets (price, return, direction), data modalities (OHLCV, indicators, macro stress, text), and evaluation protocols (time-aware splits and walk-forward validation). Reported improvements are often regime-dependent and highly sensitive to leakage-prone design choices, complicating fair comparison and deployment. The paper concludes with recommendations for standardized benchmarks, uncertainty-aware modeling, and strictly time-aligned multimodal inputs.

Keywords: Cryptocurrency forecasting; Deep learning; Transformer; Graph neural network; Sentiment analysis

1 Introduction

Cryptocurrencies have become a prominent segment of global financial markets. It is drawing attention from investors, researchers, and policymakers because of rapid innovation and extreme price fluctuations. Unlike traditional assets traded within limited market hours. Cryptocurrency markets operate continuously and can respond quickly to various information flows such as macro shocks, platform incidents, regulatory announcements, and sentiment-driven narratives. These features often translate into pronounced volatility, heavy-tailed returns, and frequent regime shifts. It makes forecasting difficult while increasing the practical value of reliable predictive tools for risk management, portfolio allocation, market making, and trading strategy design.

Classical statistical forecasting methods and pure linear time-series models typically assume relatively stable dynamics. However, cryptocurrency price processes are widely considered as nonlinear and nonstationary. So model performance may vary sharply across markets and data frequencies.

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365165

This has motivated a shift toward AI methods, especially deep learning, because neural models can learn nonlinear mappings from historical prices and engineered features, and can be extended to incorporate additional signals.

Several deep learning families are commonly used for cryptocurrency forecasting. CNN-based approaches are often applied to extract local temporal patterns and exploit cross-asset information; Zhang et al. propose a CNN framework with weighted channels and attentive memory to capture interdependencies among major cryptocurrencies [1]. RNN variants (e.g., LSTM/GRU) remain popular for sequential dependence, and hybrid designs are frequently adopted in noisy, high-frequency settings; Peng et al. develop an attention-based CNN–LSTM for multiple-cryptocurrency trend prediction [2]. Attention and Transformer-style architectures have also gained popularity because self-attention can represent longer-range dependencies with efficient computation; empirical studies report competitive performance but note sensitivity to data scale and evaluation choices [3][4].

Beyond purely temporal forecasting, researchers have increasingly turned to richer data sources and network-based representations. GNN methods encode relationships among cryptocurrencies so that spillover dynamics and market-wide shocks can be captured alongside each asset's own price history. Yin et al. extend this by incorporating a financial stress index into a GNN-based framework, reflecting the view that macro-level stress propagates through inter-asset connections before appearing in individual prices [5]. Twitter and other social platforms have separately motivated work treating sentiment as a direct forecasting input. Koltun and Yamshchikov find that including Twitter sentiment improves both prediction performance and outcomes in a simple trading strategy [6]. Han et al. take a broader multimodal approach for Bitcoin, fusing time-lagged sentiment from tweets and news with technical indicators; their emphasis on temporal alignment reflects the risk that carelessly constructed sentiment features can introduce look-ahead bias [7]. Jin and Li pursue yet another angle, first decomposing the raw price series into trend and high-frequency components and then applying deep learning to each before reconstructing a combined forecast—a pipeline they find improves stability in nonstationary conditions [8].

These advances notwithstanding, several difficulties continue to limit cross-study comparison and real-world applicability. Model performance can deteriorate sharply when market regimes shift, because patterns learned under one regime may not persist into new conditions. On the evaluation side, non-temporal data splits, preprocessing steps that inadvertently incorporate future data, and inconsistent forecast horizons can all push reported accuracy well above what would be achievable in live deployment. Comparability suffers further because studies measure performance differently: some prioritize MAE or RMSE, others track directional accuracy, and still others report trading simulation returns. Bouteska et al. argue that standardized, temporally rigorous evaluation is a prerequisite for identifying approaches that genuinely generalize, a view reinforced by a recent survey calling for clearer reproducibility standards [9][10].

This review accordingly takes stock of deep learning research on cryptocurrency price forecasting. The paper organizes the literature by architectural family—CNN-based models, RNN and hybrid variants, attention and Transformer designs, GNN approaches, and decomposition or multimodal pipelines—and characterize the data inputs, prediction targets, and evaluation practices common across these studies. The paper then draw conditional conclusions, flag methodological pitfalls including data leakage, and identify the challenges and research directions most relevant to building forecasting systems that are both accurate and deployable.

The paper proceeds as follows. Section 2 surveys the principal deep learning families and their trade-offs in the cryptocurrency context. Section 3 synthesizes empirical evidence by task, input modality, and evaluation design. Section 4 addresses open challenges and future directions before Section 5 concludes.

2 Main deep learning techniques for cryptocurrency forecasting

The deep learning architectures applied to cryptocurrency forecasting address distinct structural properties of financial time-series data, from local temporal patterns and sequential dependence to long-range dependencies, cross-asset network effects, and multi-scale dynamics. The discussion below is organized by architectural family and focuses on modeling intuition and practical trade-offs rather than asserting a universally superior approach, since the most appropriate choice typically depends on data characteristics, the forecasting objective, and evaluation design.

2.1 CNN-family and temporal feature extraction

Convolutional neural networks (CNNs) are widely used to extract local temporal patterns from sliding windows of OHLCV data or technical indicators. With 1D convolutions, models can learn short term patterns, such as brief momentum moves or local reversals. Pooling or attention layers then aggregate these local feature maps into predictions. A widely adopted extension is multi-channel modeling, in which individual cryptocurrencies or distinct feature groups occupy separate input channels, enabling the network to detect co-movement patterns that would be invisible from any single asset's history in isolation. Zhang et al. build on this structure with a CNN framework that combines weighted channel aggregation and an attentive memory mechanism to represent interdependencies among major cryptocurrencies [1].

CNNs are computationally efficient and generally straightforward to optimize, making them well-suited to settings that require rapid inference or large-scale feature screening. Their principal limitation is a bounded receptive field: without augmentation by recurrent or attention components, purely convolutional models may miss dependencies that span time horizons longer than the filter length, and the local statistical regularities they learn can degrade when underlying market regimes shift.

2.2 RNN-family (LSTM/GRU) and hybrid architectures

Recurrent neural networks, and LSTM and GRU variants in particular, have long been a staple of financial time-series forecasting because their gating mechanisms selectively retain relevant history across time steps, making them naturally suited to multivariate windows of price and indicator features. The inherent noisiness of cryptocurrency data—especially pronounced at higher sampling frequencies—has motivated hybrid designs that prepend convolutional layers for initial feature extraction before passing the result to a recurrent module for sequential modeling. Peng et al. develop an attentive CNN–LSTM model for high-frequency trend prediction across multiple cryptocurrency pairs, in which the convolutional layers extract short-term pattern features, the LSTM layers capture temporal dynamics, and an attention mechanism re-weights the most informative segments of each input window [2]. Compared with purely convolutional designs, recurrent architectures typically require longer training times and can be sensitive to abrupt regime transitions, which may reduce generalization on data drawn from market conditions not well represented in the training period.

2.3 Transformer and attention-based models

Transformer architectures replace sequential recurrence with self-attention, allowing the model to directly relate any two time points in an input window and thereby capture longer-range dependencies—such as delayed market reactions to news or multi-day recurring patterns—without the distance-dependent attenuation that affects recurrent networks. Self-attention also permits parallel computation over the time dimension, which can accelerate training on larger datasets. Dinçer and Kilimci compare recurrent and Transformer-based approaches for cryptocurrency prediction, finding that attention-based designs offer competitive performance in multivariate settings [3]. Kehinde et al. develop the Helformer, an attention-centric model adapted specifically to the characteristics of cryptocurrency price series [4]. That said, Transformer models are prone to overfitting when training data are limited or market regimes shift rapidly, and their performance is notably sensitive to choices of input window length, forecast horizon, and feature composition, underscoring the importance of rigorous evaluation when benchmarking against recurrent baselines.

2.4 Graph neural networks and cross-asset / network effects

Graph neural networks (GNNs) treat cryptocurrencies as nodes. Edges connect nodes based on relationships like correlation or co-movement. This is motivated by frequent spillovers and market-wide shocks in crypto markets. Through message passing, GNNs learn features that combine an asset's own history with signals from related assets. They can also be combined with temporal modules to support dynamic forecasting. Yin et al. integrate a financial stress index into a GNN-based prediction strategy. It reflects the view that macro stress and cross-asset connections can transmit into crypto prices [5]. In practice, results can be sensitive to how the graph is built. Key choices include the similarity measure. The estimation window also matters. People also need to decide whether the graph is static or time-varying. Transparent reporting is therefore important so results remain comparable across studies.

2.5 Decomposition-enhanced and multimodal pipelines

To handle noise and multi-scale behavior, some studies perform signal decomposition before applying deep learning. The goal is to separate high-frequency fluctuations from smoother components. Jin and Li combine frequency decomposition with deep learning in a "decompose–learn–reconstruct" pipeline to improve forecasting under nonstationarity [8]. Another enhancement is multimodal forecasting, where price-based inputs are augmented with external signals such as social-media sentiment or news. Koltun and Yamshchikov incorporate Twitter sentiment for cryptocurrency prediction [6]. And Han et al. propose a multimodal fusion approach using time-lagged sentiment together with indicator features [7]. These enhancements can add explanatory power, but they also bring new risks. Common issues include time misalignment and noisy text signals. Leakage can also occur when inputs are not aligned in a strictly time-causal way. These issues motivate the evidence synthesis and protocol discussion in the next section.

3 Evidence synthesis and comparative insights

This section summarizes evidence from prior studies. It organizes findings by forecasting tasks, data inputs, and evaluation protocols. Cryptocurrency markets are highly nonstationary. As a result, reported performance can change a lot. It depends on the time period. It also depends on the sampling frequency. Label design matters. The split strategy matters as well. The goal is to summarize conditional findings from the literature. And focusing on what tends to work in specific settings rather than claiming universal superiority.

3.1 Forecasting targets and experimental settings

Prior work covers multiple forecasting targets. Some studies predict the next-step price. Others target returns or directional movement, such as up or down. This choice is common when the objective is closer to trading decisions than to minimizing raw price error. Target selection also varies with data frequency. For daily datasets, predicting price levels or returns is often relatively stable, whereas intraday studies frequently adopt direction or trend labels in order to reduce short-term market noise. Peng et al., for instance, study short-interval trend prediction using an attention-based CNN–LSTM architecture [2]. In contrast, research comparing multiple model families typically relies on daily data so that broader conclusions can be drawn, although results still depend heavily on both the prediction horizon and prevailing market regimes [3][9].

Forecast horizon is another important design choice. Multi-step forecasting tends to accumulate errors and is particularly sensitive to regime changes, which may reduce the apparent advantage of more complex models when the evaluation procedure does not strictly follow time order. For this reason, one-step-ahead prediction remains the dominant evaluation target, with multi-step horizons appearing mainly as supplementary robustness checks rather than primary benchmarks [3][4].

3.2 Input features and data modalities

OHLCV variables together with standard technical indicators serve as the baseline input set in nearly all reviewed studies. CNN-based models apply sliding-window convolutions to identify localized temporal patterns and can treat separate assets or feature groups as independent input channels, capturing co-movement that would be invisible from a single-asset perspective [1]. Sequential architectures—LSTM-based or Transformer-based—work with multivariate time windows that may incorporate a wider range of engineered variables; performance in these settings is noticeably sensitive to feature scaling choices and window length decisions [2][3][4].

Beyond baseline market variables, two further data sources appear with notable frequency. Network-based approaches model inter-asset relationships as a graph, with edges derived from return correlations or price similarities, and propagate spillover signals through message passing; some implementations also incorporate macro-level features such as financial stress indexes [5]. Sentiment indicators extracted from social media or financial news are a second common augmentation, typically combined with price-based inputs rather than used alone. When lagged appropriately to precede the forecasting window, they can improve accuracy for coins with high social-media visibility [6][7], though their benefit is uneven—text-derived signals are noisy and topic-sensitive, and their contribution varies considerably across assets and time periods. A third approach decomposes the raw price series into trend and higher-frequency residual components before training, on the premise that isolating different timescales may simplify the learning problem, though how much this helps depends on the decomposition technique used and how the component forecasts are recombined [8].

3.3 Evaluation protocols and common pitfalls

Differences in evaluation design are among the most consequential sources of incomparability across the literature. Financial time series exhibit autocorrelation and regime-dependent dynamics that make random train-test splits inappropriate: when future observations are sampled into the training set, information leaks across time, inflating reported accuracy in ways that do not reflect genuine out-of-sample performance. Walk-forward validation addresses this directly: the model is retrained on data available before each test window and evaluated on the observations immediately ahead, preserving the temporal ordering of financial data.

Leakage risk extends beyond the train-test split to every stage of feature construction. Technical indicators must use only information available at each forecasting timestamp; normalization statistics must be estimated from training data and applied without refitting to the test set; sentiment aggregates must be lagged far enough back to rule out any overlap with the forecast target. Studies further complicate comparison by measuring performance along quite different dimensions—MAE, RMSE, and MAPE each weight errors differently; directional accuracy captures something else again; and back testing outcomes are highly sensitive to assumptions about transaction costs and position sizing. Comparative analyses have repeatedly stressed that consistent experimental design and transparent documentation are prerequisites for drawing reliable conclusions [9], and survey work reinforces this with calls for reproducible benchmarks that include full specifications of data splits, feature pipelines, and hyperparameter search procedures [10].

3.4 Conditional performance patterns across methods and settings

The literature does not point to a single dominant architecture; instead it reveals a set of conditional patterns linking architectural choices to data characteristics. Hybrid CNN-RNN designs are most common in high-frequency or noisy settings, where an initial convolutional stage filters short-term fluctuations before the recurrent component handles sequential dynamics [2]. Transformer and attention models tend to perform well in multivariate settings where longer historical context matters, though this advantage is fragile and often disappears when training data are limited, regularization is insufficient, or evaluation conditions inadvertently favour the model [3][4]. GNN approaches add the most value when cross-asset spillovers strongly drive price movements, but performance depends heavily on graph construction choices and whether the topology is periodically updated [5]. Decomposition pipelines and multimodal inputs can add useful information during volatile or news-driven episodes, yet they also create new failure modes—misaligned signals, unstable text features, subtle leakage—that require careful design to avoid [6][7][8]. Overall, reported gains tend to reflect improvements in input quality rather than architectural novelty alone, and their durability depends on whether evaluation was genuinely time-ordered and free of leakage [9][10].

4 Discussion on challenges and future directions

4.1 Key challenges

Nonstationarity is the most persistent obstacle. Regime changes triggered by regulatory announcements, macroeconomic events, or shifts in the investor base can alter the statistical relationships a model learned from history, sometimes abruptly, leaving it without useful predictive power soon after deployment.

Evaluation leakage is a second major concern. The most visible form is applying random splits to time-ordered data, but subtler forms arise from fitting normalization statistics on the full dataset before splitting, or from failing to lag sentiment inputs far enough back in time. Each produces accuracy estimates that overstate real deployability.

Poor reproducibility compounds both problems. Many studies omit key details—the exact data collection window, how missing values were handled, the scope of hyperparameter search—making it very difficult to tell whether a reported gain reflects a genuine methodological contribution or simply a favourable experimental setup [9][10]. Complex models may also require frequent retraining to stay current, and their opacity creates difficulties for risk management and regulatory review.

4.2 Future research directions

Progress on comparability will require shared datasets and standardized evaluation procedures: walk-forward validation with predetermined retraining schedules, publicly available baseline implementations, and detailed reporting of feature construction and tuning choices [9][10]. Better uncertainty quantification is a clear parallel need; models that output calibrated prediction intervals would serve risk managers far better than point forecasts, and evaluating probabilistic calibration should become routine practice.

Multimodal forecasting involving sentiment or macro signals remains promising but will require more rigorous causal alignment between text collection and trade timestamps, alongside more robust signal extraction from noisy social media data [6][7]. Finally, advances in interpretability and computational efficiency would help close the gap between research performance and deployment, making it practical for practitioners to trust and maintain models in fast-moving markets.

5 Conclusion

This review summarizes recent work on AI-based cryptocurrency forecasting. It organizes the literature by major deep learning approaches and compares studies across tasks, inputs, and evaluation settings. CNN-family models are often used to capture short-term patterns and multi-asset signals. Hybrid CNN–RNN models are common when the data are noisy, especially at high frequency. Attention and Transformer models can capture longer-range patterns when there is enough data and the testing setup is sound. GNN-based methods model how coins move together and how broader stress signals can spill over into prices. Decomposition and multimodal pipelines try to make forecasts more stable by separating different time scales and by adding signals such as sentiment from news or social media.

Across these approaches, reported gains are highly sensitive to regime shifts and evaluation choices. This sensitivity shows why tests should follow the time sequence. Leakage must be avoided. Studies should clearly describe their data and methods. Future work should focus on shared benchmarks. Models should handle uncertainty better. Multi-source inputs should be aligned in time. These changes can make results easier to reproduce and more useful in real-world crypto markets.

References

[1] Zhang, Z., Dai, H.N., Zhou, J., Mondal, S.K., García, M.M., Wang, H.: Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels. Expert Systems with Applications. 183:115378. (2021) https://doi.org/10.1016/j.eswa.2021.115378

[2] Peng, P., Chen, Y., Lin, W., Wang, J.Z.: Attention-based CNN–LSTM for high-frequency multiple cryptocurrency trend prediction. Expert Systems with Applications. 237(Part B):121520. (2024) https://doi.org/10.1016/j.eswa.2023.121520

[3] Dinçer, N.G., Kilimci, Z.H.: From LSTM to GPT-2: Recurrent and transformer-based approaches to cryptocurrency prediction. Symmetry. 18(1):32. (2025) https://doi.org/10.3390/sym18010032

[4] Kehinde, T.O., Adedokun, O.J., Joseph, A., Kabirat, K.M., Akano, H.A., Olanrewaju, O.A.: Helformer: An attention-based deep learning model for cryptocurrency price forecasting. Journal of Big Data. 12:81. (2025) https://doi.org/10.1186/s40537-025-01135-4

[5] Yin, W., Chen, Z., Luo, X., Kirkulak-Uludag, B.: Forecasting cryptocurrencies' price with the financial stress index: A graph neural network prediction strategy. Applied Economics Letters. 31(7):630–639. (2024) https://doi.org/10.1080/13504851.2022.2141436

[6] Koltun, V., Yamshchikov, I.P.: Pump It: Twitter sentiment analysis for cryptocurrency price prediction. Risks. 11(9):159. (2023) https://doi.org/10.3390/risks11090159

[7] Han, P., Chen, H., Rasool, A., Jiang, Q., Yang, M.: MFB: A generalized multimodal fusion approach for Bitcoin price prediction using time-lagged sentiment and indicator features. Expert Systems with Applications. 261:125515. (2025) https://doi.org/10.1016/j.eswa.2024.125515

[8] Jin, C., Li, Y.: Cryptocurrency price prediction using frequency decomposition and deep learning. Fractal and Fractional. 7(10):708. (2023) https://doi.org/10.3390/fractalfract7100708

[9] Bouteska, A., Abedin, M.Z., Hajek, P., Yuan, K.: Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods. International Review of Financial Analysis. 92:103055. (2024) https://doi.org/10.1016/j.irfa.2023.103055

[10] John, D.L., Binnewies, S., Stantic, B.: Cryptocurrency price prediction algorithms: A survey and future directions. Forecasting. 6(3):637–671. (2024) https://doi.org/10.3390/forecast6030034