Independent Study Final Report

Level 3

Pattern Recognition in Cryptocurrency Markets Using Machine Learning

by

Malinga K.A. (214130R)

Supervisor: Ms. Manawathilake K.C.T.

Faculty of Information Technology

University of Moratuwa

2025

Independent Study Final Report

Level 3

Pattern Recognition in Cryptocurrency Markets Using Machine Learning

by

Malinga K.A. (214130R)

Student	Signature	Date
Malinga K.A.		**23/11/2025**
Supervisor	**Signature**	**Date**
Ms. Manawathilake K.C.T.		24.11.2025

Faculty of Information Technology

University of Moratuwa

2025

Pattern Recognition in Cryptocurrency Markets Using Machine Learning

Mr. Malinga K.A.
Department of Information Technology
University of Moratuwa
Moratuwa, Sri Lanka
malingaka.21@uom.lk

***Abstract*— This review of the literature is a critical analysis of current developments in deep learning and machine learning methods to predict cryptocurrency prices and identify patterns. We critically evaluate 16 articles covering both methodological frameworks of classical statistical models to cutting-edge deep learning architecture, sentiment analysis models, and automated trading solutions. The review finds the major trends in the methodology, compares the performance in various methods, and critically examines limitations such as data requirements, the risk of overfitting, and limitations on practical implementation. We emphasize that although LSTM networks are proven to be effective in the temporal pattern recognition (with 52–99 percent accuracy across studies), much work needs to be done to ensure the reliability of long-term prediction, model interpretability, and the use of multi-modal data streams with it. The review can form a background of the cryptocurrency market prediction study because it outlines the gaps that need to be addressed and outlines the future promising directions.**

Keywords—Cryptocurrency; Machine Learning; Pattern Recognition; Deep Learning; Forecasting

I. INTRODUCTION

Cryptocurrency markets have grown and become volatile as never before since the introduction of the Bitcoin in 2008 by Nakamoto as a decentralized peer-to-peer electronic cash system [15]. Bitcoin market capitalization in itself increased to more than 1 trillion in 2021 since 2016, 9 billion, and the crypto market is greater than 2 trillion [12]. This virulent expansion, on the one hand, coupled with such a high price volatility, standard deviations of 4–8 percent a day on the major cryptocurrencies versus 1–2 percent on traditional stocks [7] on the other, opens both opportunities and challenges to price prediction.

Conventional statistical methods like ARIMA do not have the non-linear, non-stationary properties of cryptocurrency markets [14]. With the appearance of machine learning (ML) and deep learning (DL), more opportunities became available to model these complicated dynamics. Nevertheless, the high rate of methodology development—starting with the classical support vector machines and the more sophisticated recurrent neural networks—requires the critical assessment of their relative advantages, shortcomings and feasibility.

The current literature review critically analyzes 16 studies in order to answer the following three basic questions: (1) What ML/DL models show the best performance in cryptocurrency price prediction? (2) What can be considered the critical limitations and failure modes of existing approaches? (3) What are the methodological gaps that need to be filled by the future research? In contrast to earlier surveys that focus largely on enumerating methods of measuring it, this review takes a critical approach to the matter, considering the problem of reproducibility, bias in data sources, and the gap between reported accuracy and actual trading profitability that is usually ignored.

II. METHODOLOGICAL EVOLUTION

A. From Classical Statistics to Deep Learning

The historical development of deep learning architectures based on traditional time-series techniques can be seen as an improvement as well as a continuity of challenges. Initial prediction studies of cryptocurrencies used ARIMA models, which, according to McNally et al., provide only directional forecasting of 50.05 percent with 53.74 percent RMSE of Bitcoin, basically pseudo-random behavior. This underperformance is due to the inherent assumptions of ARIMA of stationarity and linearity being systematically breached in the cryptocurrency markets whose nature involves regime changes, fat-tailed distributions and values of kurtosis of 10–56, versus 4–11 with traditional stocks.

Support Vector Machines (SVM) presented a significant shift, and various reports of accuracy 55–95 percent are presented in the literature [10], [3]. Interestingly, Hitam and Ismail have 95.5 percent accuracy with Ethereum with SVM which only uses four input features (open, high, low, close), which is much higher than deep learning methods in their comparative analysis paper [3]. This paradoxic finding, that simpler models are better than complex ones, deserves critical analysis. The dataset used by the authors was only 526 Ethereum observations, which makes overfitting a question since the model seems relatively simple. In addition, they had a one-year test period (January 2017–2018) which was in a mostly bull market which would likely overstate accuracy measures that would not necessarily be accurate in bear markets.

B. Recurrent Neural Networks and LSTM Dominance

Long Short-Term Memory (LSTM) networks have become the leading structure used in cryptocurrency prediction, as they are used in 14 of the 16 articles reviewed. The conceptual attractiveness is strong: the gated architecture of LSTM directly deals with the vanishing gradient problem that afflicts conventional RNNs, allowing it to learn long-term temporal dependencies that are necessary to learn market trends [11].

Empirical evidence seems to be impressive. According to Ho et al., the accuracy of Bitcoin prediction was 99.87 percent using Linear Regression and 0.08 percent using LSTM [1]. Accuracy of 52.78 percent and RMSE of 6.87 percent were recorded by McNally et al., much higher in comparison to ARIMA [14]. Kumar and Rath have shown that LSTM performs better than Multi-Layer Perceptron (MLP) on daily, hourly, and minute-wise Ethereum data with MAPE scores of 3.67 percent, 1.38 percent, and 2.21 percent respectively [2].

Nevertheless, there are alarming trends to be observed. To begin with, reported accuracies range all over the place, 52 percent to 99 percent, even though the methodology is analogous, indicating that they are highly sensitive to the choice of dataset, preprocessing options, and even evaluation procedures. McNally et al. fully disclosed training error of less than 1 percent but validation error of 7–8 percent, which revealed significant overfitting in their study [14]. This is an open reporting, which is not common in the literature; most studies indicate the performance in test sets but nothing more.

Second, the tradeoff between the cost and benefit of computation is not given due attention. McNally et al. discovered that training LSTM took 3.1x longer than training RNN to improve by 2.5 percent (52.78 vs. 50.25) accuracy [14]. Although 67.7 percent speedup was obtained with GPU acceleration, the difference in accuracy over random guessing (50 percent) is quite low and this questions the usefulness of acceleration, especially when transaction costs are taken into account which most research does not.

Third, one of the hyperparameters is the temporal window, which is treated inconsistently. McNally took 24 days for RNN and 100 days for LSTM [14], whereas Patel et al. observed that the LSTM models paradoxically provide short-term patterns of history in spite of their long-term memory capability [11]. This difference between theoretical ability and the actual practice has not been fully answered.

C. Gated Recurrent Units and Architectural Variants

The extensive comparison of GRU, LSTM, and bidirectional LSTM (bi-LSTM) by Hamayel and Owda offers indispensable details about the trade-offs of the architecture [6]. Surprisingly, the simpler GRU model had better performance: 0.2454 percent MAPE in Bitcoin compared with 1.1234 percent in LSTM and 5.990 percent in bi-LSTM respectively. This is in contrast to the general belief that the more the model complexity, the better it performs.

Of particular interest is the poor performance of the bi-LSTM (it is always poorest in all three cryptocurrencies being tested). Bidirectional processing in theory allows learning into the past and future context, but this benefit did not become a reality. The authors do not provide any explanation but this might refer to the fact that crypto markets are inherently future-oriented-there is no information about the future in the training data as it is in the case of such tasks as machine translation. This points to a very serious methodological issue, namely the transfer of architectures that have been proven to be efficient in other areas without testing their suitability to financial time series.

D. Hybrid and Ensemble Approaches

The acknowledgment of the inability of a single model to describe all the facets of cryptocurrency processes has contributed to the development of hybrid methodologies. Cheng et al. used SARIMA (seasonal trends), Facebook Prophet (holiday effects), and models based on machine learning [7]. The SDAE-B (Stacked Denoising Autoencoder with Bagging) of Zhang et al. obtained an impressive list of results: MAPE of 0.016, RMSE of 131.643, and directional accuracy of 0.817 when predicting Bitcoin with blockchain metrics and public attention, and macroeconomic variables included into the price predictor [9].

The systemic accuracy of the deep learning architecture by Liu et al. based on Stacked Denoising Autoencoders (SDAE) with 40 determinants has beat the standard methods by a lower MAPE of 72 percent versus reductions of 40 percent and 43 percent for BPNN and SVR respectively [16]. The benefit of the study is that it has a wholesome feature engineering in terms of cryptocurrency markets, public attention (Baidu Index, Google Trends) and macroeconomic indicators (stock indices, commodities, exchange rates).

Nevertheless, the ensemble methods add new problems. The ensemble approach used by Chowdhury et al. of GBT, Neural Network and Linear Regression attained 92.4 percent accuracy on IOTA, and 51.6 percent on Bitcoin only [13]. This type of cryptocurrency specificity implies that models are potentially identifying idiosyncratic trends, as opposed to market dynamics that are universalizable. Moreover, the complexity of the ensemble methods makes interpretation more difficult-a limitation that is essential in the context of financial uses of machine learning where explainability is necessary to comply with regulation and mitigate risk.

III. CRITICAL EVALUATION OF PERFORMANCE CLAIMS

A. The Accuracy–Profitability Gap

At the point where reported prediction accuracy has been compared to practical trading profitability, there arises a central criticism to consider. The performance in the study presented by McNally et al. was at the stage of 52.78 percent directional accuracy-almost as good as the flip of the coin-but at least, it was the best result of the research [14]. The extensive comparison of 12 different cryptocurrencies at different times (daily, 60-min, 30-min, 15-min) by Akyildirim et al. showed that SVM only reached 54–59% accuracy [10]. These small numbers are in sharp contrast to some other studies where 95–99% accuracy has been claimed [7].

This problem is infrequently revealed by Singh in his thesis, which applied LSTM-based algorithmic trading which, on Bitcoin test data, made a profit of 21,436 USD, and performed 3.4 times better than ARIMA, which made 6,326 USD [7]. Nonetheless, this backtest took zero transaction costs and did not consider slippage and used past data where the model was aware ex post that particular patterns were predictive. The costs of market impact, latency, and adaptive behavior of other market participants would be encountered during real-world implementation.

This is inherently a flaw because the accuracy of the classification is confusing two different questions: "Is it possible to predict direction of prices? And can we make money off those prophesies?" The 55 percent accuracy model is marginal, but when it is wrong in times of high volatility and right in times of stable volatility, it could be highly profitable. On the other hand, a 70 percent accuracy will not be of use when transaction costs will eat up marginal gains. This was only reported by Mudassir et al., where directional accuracy was

65 percent and daily forecast error at 1.44 percent on average, yet the profitability was not analyzed with transaction costs [16].

B. Dataset Limitations and Reproducibility Concerns

Systematic analysis demonstrates worrying dataset habits. Kumar and Rath made 1,200 daily observations of Ethereum (daily observations of about 3.3 years) only [2], whereas McNally et al. made 1,066 days of observation of Bitcoin [14]. Models will be susceptible to regime-specific overfitting because of these small samples. The thesis by Singh presents minute-level data that present 512,640 observations annually [7], whereas most of the research works presented daily data-a strange, somewhat puzzling decision to make because deep learning ought to be enhanced by higher-frequency data.

There are disturbing variability in training-test splits. Ho et al. trained between August 2017–August 2020 and tested at an unspecified later time [1]. Kumar and Rath used a 60/40 split, and McNally used 80/20. This discrepancy does not allow cross-study comparison. Worse still, some studies do not give enough detail such that one can replicate their data preprocessing, normalization preferences, or their treatment of missing data and outliers.

Chowdhury et al. eliminated outliers in Bitcoin, Ethereum and Litecoin as 0.9–2.2% of the data points [13]. Although such removal appears minor, extreme events in cryptocurrency markets hold important information about market dynamics. The candlestick pattern recognition discussion by Uzun et al. is an exemplary best practice, as it offers an outlier analysis and its justification in detail, though such rigor is not the rule of thumb.

C. Evaluation Metric Selection and Reporting Bias

Research also has varied and unsuitable measures of evaluation. Most commonly used are Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and classification accuracy, but they are used differently. MAPE is proportional and is effective in price level prediction. RMSE highlights bigger errors, which is appropriate in volatility forecasting. Directional prediction is best captured by classification accuracy.

Khan et al. compared volatility forecasting with NNETAR, CSS, and GMDH-NN, and they did not identify a universally best model. CSS was optimal for Bitcoin and XRP, while GMDH-NN was optimal for Tether [8]. Such cryptocurrency-induced variation in performance indicates that reported accuracies could be due to lucky modelasset matching, not true methodological superiority.

Reporting bias is evident. Studies with extraordinarily high findings (95–99 percent accuracy) are less methodologically detailed and transparent on limitations compared to those with modest findings (52–65 percent accuracy). The open discussion of overfitting (1 percent training error vs. 7–8 percent validation error) by McNally et al. [14] and the extensive cross-asset and cross-timeframe testing by Akyildirim et al. [10] are best practices but rare.

IV. BEYOND PRICE DATA: MULTI-MODAL APPROACHES

A. Social Media Sentiment Integration

At the stage of the prediction accuracy reported being set at par with practical trading profitability there comes a pivotal critique to be taken into account. The study by McNally et al. had the performance of the process at 52.78 percent directional accuracy-nearly as good as the flip of the coin-but at least it was the best result of the research (ref14). The comprehensive comparison of the 12 various cryptocurrencies at various points (daily, 60-min, 30-min, 15-min) by Akyildirim et al. revealed that SVM only achieved 54-59 percent accuracy [10]. These low figures are in drastic opposition to other studies where 95-99

This issue is rarely disclosed by Singh in his thesis, which used the algorithmic trading based on LSTM, which on the test data of Bitcoin was able to bring profit in 21, 436 USD, and performed 3.4 times more effectively than ARIMA that earned 6, 326 USD USD 7-6326. However, this backtest used zero transaction costs, and did not model slippage or used past data in which the model was aware that a specific pattern at ex-post was predictive. Market impact costs, latency costs and adaptive behavior costs of other market participants would be experienced in real life implementation.

This is a weakness in itself due to the fact that the correctness of the classification is bewildering two different questions: "Is it possible to predict direction of prices? And shall we make money out of them prophesies? The 55 percent accuracy model is already being marginal, but when it proves to be wrong during the period of high volatility and is correct during the period of stable volatility, the business would have been a very profitable one. On one hand, 70percent accuracy will not come in handy when marginal gains will be devoured by transaction costs. This was reported by Mudassir et al. only with a directional accuracy of 65 percent and a daily forecast error of 1.44 percent on average but the profitability was not considered with transaction costs as in the case of the former 16.

B. Blockchain and Network Metrics

Some of the studies included blockchain-related parameters, such as hash rate, mining difficulty, volume of transactions, number of active addresses, which it was hypothesized that they would reflect basic network health. According to McNally et al., these features do not provide much predictive power [14] whereas Liu et al. used them in their large 40 feature system that performed better. This mismatch can be due to various feature engineering strategies or effects of interaction that arise at the point of having an adequate level of model complexity.

The wavelet coherence analysis of Singh showed that the closing prices of Bitcoin are highly correlated with network hash rate in the long-term scale but isolated temporal correlations were observed at 34, 45 and 47-day lags. These unconventional lag structures break the traditional time-series modeling assumptions and propose that cryptocurrency dynamics might be periodic based on changes in mining difficulty and not on calendar effects.

C. Macroeconomic and Cross-Asset Relationships

The most comprehensive feature engineering in the literature under review is the fact that Liu et al. use 23 macroeconomic factors (stock indices, commodities, exchange rates, VIX) in their SDAE model, which is the most complex feature engineering that literature is reviewed so far [9]. This is in recognition of the fact that cryptocurrency prices do not develop as autonomous entities but are subject to the world financial circumstances.

But Akyildirim et al. observe that Bitcoin demonstrates time-varying relationships with conventional assets, which are relatively short-run but have longer-term relationships [10]. Such non-stationarity in cross-assets relationships makes it difficult to model. What is predictive in one market regime can be unpredictive or counterproductive in another, although most studies suppose that relationships are expected to remain constant.

V. ALGORITHMIC TRADING AND PRACTICAL IMPLEMENTATION

A. Reinforcement Learning Approaches

The thesis by Singh is the most extensive coverage of cryptocurrency algorithmic trading of the reviewed papers [7]. He had an LSTM strategy with tuned purchase/sell thresholds ($T_b = 152$, $T_s = -115$ in case of Bitcoin) that (average) made a superset of 21,436 as compared to 6326 with ARIMA. Deep Reinforcement Learning models with Deep temporal difference (TD3, PPO, DDPG) demonstrated annualized returns of 23-25 percent, which was significantly higher than buy-and-hold (1015 percent).

Importantly, the fuzzy risk analysis by Singh showed that the algorithmic trading caused smaller volatility of returns than buy-and-hold strategies in question, which indicated that such strategies had risk-adjusted advantages over simple returns. His data-driven volatility forecasting (DDVF) was more computationally efficient (0.17-0.67 seconds compared to 140-1,535 seconds) and generated more consistent volatility and Value-at-Risk forecasts on cryptocurrencies due to its data-driven forecasting (0.17-0.67 seconds vs. neural network based 140-1,535 seconds).

These promising results are however curbed by a number of limitations. First, backtests assume that the execution is done at closing price without slippage and market impact. Second, the best thresholds were identified post-using test data—forward testing on data that is actually not seen would probably display degraded performance.

Third, the transaction costs borne on the trades of 0.1-0.5 percent was not included but they might be eating huge amounts of reported returns due to the high-frequency trading.

B. DeFi and Alternative Applications

Singh did not only analyze spot trading but also considered 10 Uniswap V3 pools as being decentralized finance (DeFi) liquidities providers [7]. The results that the gross returns (with fees) were constantly lower Value-at-Risk than plain token holding indicate that liquidity provision might be a risk-adjusted benefit. Nevertheless, this was analyzed when the DeFi was expanding considerably; the bear market can radically change risk profiles, or some large-scale protocol hacks.

The rule-based candlestick pattern recognition by Uzun et al. was 80-89 percent accurate with all three types of Bitcoin, Ethereum, and Litecoin [4]. Their interpretability property in comparison with black-box ML models deals with the interpretability gap, yet the patterns they analyze (Advance Block, Doji Star, Evening Star) are a small part of the technical patterns that traders use. Moreover, they were validated by manual expert annotation (Kappa of Cohen 0.78- 0.85), which implies expert bias.

VI. CRITICAL RESEARCH GAPS AND LIMITATIONS

A. Long-term Prediction Failures

A systematic trend can be observed: models are not bad in 1-day forecasts but they get worse and worse in longer forecasts. Patel et al. observed that LSTM models concentrate on short-term trends even though they can be theoretically long-term memory capacity 4. Kumar and Rath demonstrated that Ethereum prediction errors have been rising at a rate of 3.67

Such a universal constraint implies that there is a fundamental issue: cryptocurrency markets can be unpredictable in the long-term (except at the short level) because they are sensitive to unpredictable externalities (regulatory announcements, giant hacks, powerful tweets). Short-term momentum and mean-reversion models do not work when they are subject to regime changes. However, this limitation is rarely addressed in most studies, and when it is part of a brief mention.

B. Model Interpretability and Explainability

To a surprisingly lesser degree, the black box nature of deep learning models is addressed. Only Singh directly covers interpretability with a suggestion of the integration of fuzzy set theory to measure uncertainty in forecasts [7]. Explainability is needed in financial applications to provide regulatory compliance, risk management, and build stakeholder trust, although most of the research focuses on predictive accuracy instead of interpretability.

The rule-based methodology developed by Uzun et al. provides full transparency- all the patterns are identified with good reasons behind it [4]. This is slightly less accurate than deep learning (80-89 vs. possibly 90-95) but this price could be justified by its practical use. There would be advantages to the field by systematically comparing the accuracy-interpretability tradeoffs instead of addressing them as independent issues.

C. Overfitting and Generalization Failures

The fact that McNally et al. reported their 1 per cent training error and 7-8 per cent validation error honestly illustrates the issue of overfitting concern. This issue of cryptocurrency research is made worse by several factors:

Ideas: Most studies have limited data (1,000-2,000 observations), which is not enough to train deep networks with thousands of parameters. In 4 cryptocurrencies, 1,296 observations were used by Khan et al. 1,277 records were used in 1 study by Hamayel and Owda 1,296 records were used in 1 study by reference 8.

Proliferation of parameters: Multilayer LSTM networks with hundreds of neurons have thousands of parameters, which are larger than sample sizes. Unless aggressive regularization (dropout, early stopping) is used, overfitting is unavoidable.

Hyperparameter optimization on test data: A number of studies seem to have hyperparameter-tuned to maximize test set performance, which is against the correct protocols of validation. Application of Bayesian optimization (Hyperas) on validation data held out by Singh is best practice 6, although uncommon.

Regime-specific patterns: Bull market models will not work in bear markets. The majority of the studies take a continuous time interval instead of a stratified sampling under various market conditions, which restricts the generalization.

D. Transaction Costs and Market Microstructure

Studies do not consider transaction costs although they are of critical significance to the profitability of trading with rare exceptions 7. Cryptocurrency exchanges are charged 0.1-0.5 percent per trade and slippage is 0.1-0.3 percent on market orders. A 55

Almost no attention is paid to market microstructure effects. Bf exchange Akyildirim et al. applied the Bitfinex data only at any rate, whereas other sources utilize aggregate indices. Executions quality is also significantly affected by exchange-specific effects (liquidity, order book depth, trading fees, maker-taker incentives) and the models that were trained on a particular exchange might not carry over to other places.

E. External Validity and Market Adaptation

The question is whether predictive models will be effective when they are adopted widely, and it is still a basic issue that has not been fully addressed. The market efficiency theory posits that abusive patterns must be eliminated because they are being known. The studies that have been reviewed assume that the market is a static environment, yet effective prediction strategies would initiate adaptive behavior on the part of other players.

DRL agents developed by Singh demonstrated that market conditions influence the effectiveness of strategies 7 but they did not test how a high adoption of models could change the patterns that models utilize. This is a problem of reflexivity peculiar to financial applications and it is worth treating explicitly.

VII. FUTURE RESEARCH DIRECTIONS

A. Methodological Improvements

Some of the promising directions come out of this literature review:

Interpretable ensemble methods: Ensembles can be even stronger when they are composed of multiple interpretable models, yet ensembles today have become uninterpretable. Another significant direction is the development of ensemble methods that will ensure transparency and enhance generalization.

Method: Train models to tell the difference between training and test distributions: This would help detect overfitting prior to deployment, which would enhance generalization.

Causal inference: It would be better to go beyond correlation and achieve causality between features and price changes to be more robust to regime changes.

idea: Transfer learning: Instead of training models on a case-by-case basis, training foundation models that learn across assets can make data more efficient.

B. Data and Feature Engineering

High-frequency data: The majority of research works are conducted on a daily basis even though tick-level data is available. According to Singh, minute-level data gives 512,640 observations per year [7] but it is not used fully.

Alternative data sources: In addition to Twitter sentiment, introducing on-chain metrics (exchange inflows/outflows, whale transactions), derivatives data (funding rates, options implied volatility), and decentralized exchange activity may be the new signals.

Bot filtering: According to Kraaijeveld and De Smedt, 1-14% of cryptocurrency tweets contain bot presence 10. Sentiment analysis reliability would be enhanced by the development of strong bot detection and filtering techniques.

C. Practical Implementation Research

Entry: Integrating transaction costs, slippage, and market effects Explicitly modeled costs, slippery and market effects would put a study in real world reality.

Risk management: More complete trading solutions would be to go further than prediction of returns and to volatility prediction and tail risk estimation and optimization of portfolio.

Real-world validation Live trading experiments, albeit with small capital, would be invaluable in giving insights on deployment challenges that backtests fail to capture.

D. Regulatory and Ethical Considerations

With the evolution of cryptocurrency markets, there are heightened regulatory oversight. Research should address:

Market manipulation detection: With the detection model, instead of taking advantage of the patterns, the creation of the market manipulation detecting model would help enhance the integrity of the market.

Compliance with explainable AI: Financial regulators are demanding more transparency of algorithmic trading systems. This requirement is met by developing interpretable models that can sustain the competitive performance.

Access equity: The benefits of high-frequency trading go to well-endowed traders. Democratize benefits Research that guarantees wider access to prediction tools would be possible.

VIII. CONCLUSION

This literature review of 16 papers reveals substantial progress in applying machine learning to cryptocurrency price prediction, particularly through LSTM architectures and hybrid ensemble approaches. However, several concerning patterns emerge: inflated accuracy claims disconnected from trading profitability, insufficient attention to overfitting and generalization, limited reproducibility due to incomplete methodological reporting, and neglect of transaction costs and market microstructure effects.

The most robust findings suggest: (1) Deep learning approaches, particularly LSTM networks, genuinely outperform traditional statistical methods for short-term prediction (1-3 days); (2) Hybrid models integrating multiple data sources (price, sentiment, blockchain metrics) provide incremental improvements over single-source approaches; (3) Simpler models like SVM sometimes outperform complex deep learning, particularly with limited data; (4) Prediction accuracy deteriorates rapidly for longer time horizons, suggesting fundamental limits to cryptocurrency predictability.

Critical research gaps include the absence of real-world trading validation, limited understanding of model failure modes, insufficient attention to interpretability, and the unexplored reflexivity problem-how widespread model adoption might eliminate the patterns models exploit.

Future research should prioritize reproducibility through detailed methodological reporting, comprehensive evaluation including transaction costs and cross-regime validation, development of interpretable ensemble methods, and explicit treatment of practical deployment challenges. Only by addressing these limitations can the field move from promising academic results to reliable, deployed systems that add value in real-world cryptocurrency trading and risk management.

REFERENCES

[1] A. Ho, R. Vatambeti, and S. K. Ravichandran, “Bitcoin price prediction using machine learning and artificial neural network model,” *Indian Journal of Science and Technology*, vol. 14, no. 27, pp. 2300–2308, 2021.

[2] D. Kumar and S. K. Rath, “Predicting the trends of price for Ethereum using deep learning techniques,” *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, pp. 103–114, 2020.

[3] N. A. Hitam and A. R. Ismail, “Comparative performance of ML algorithms for cryptocurrency forecasting,” *Indonesian J. Elect. Eng. Comput. Sci*, 2018.

[4] I. Uzun et al., “Candlestick pattern recognition in cryptocurrency price time-series data,” *Computation*, vol. 12, no. 7, 2024.

[5] G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[6] M. J. Hamayel and A. Y. Owda, “A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM algorithms,” *AI*, vol. 2, no. 4, pp. 477–496, 2021.

[7] J. Singh, “Data-Driven Risk Forecasting and Algorithmic Trading Models for Cryptocurrencies,” 2022.

[8] F. U. Khan et al., “Forecasting returns volatility of cryptocurrency by applying various deep learning algorithms,” *Future Business Journal*, vol. 9, no. 1, 2023.

[9] M. Liu et al., “Forecasting the price of Bitcoin using deep learning,” *Finance Research Letters*, vol. 40, p. 101755, 2021.

[10] E. Akyildirim et al., “Prediction of cryptocurrency returns using machine learning,” *Annals of Operations Research*, 2020.

[11] M. Patel, S. Tanwar, and R. Gupta, “A deep learning-based cryptocurrency price prediction scheme,” *Journal of Information Security and Applications*, vol. 55, 2020.

[12] R. Amirzadeh et al., “Applying artificial intelligence in cryptocurrency markets: A survey,” *Algorithms*, vol. 15, no. 11, 2022.

[13] R. Chowdhury et al., “An approach to predict and forecast cryptocurrency index prices using machine learning,” *Physica A*, vol. 551, 2020.

[14] S. Mcnally et al., “Predicting the price of bitcoin using machine learning,” *PDP*, 2018.

[15] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008.

[16] S. Otabek and J. Choi, “A comprehensive review of cryptocurrency trading strategies,” *IEEE Access*, 2024.