iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VI **Month of publication:** June 2025

DOI: https://doi.org/10.22214/ijraset.2025.72248

Cryptocurrency Price Forecasting using LSTM: A Review

Assistant Professor Neha Verma, Rittik Kumbhbakar, Navneet Narayan, Sachin Dongre, Ajit Jadhav,
Department of Artificial Intelligence and Data Science, Ajeenkya DY Patil School of Engineering, Lohegaon, Pune, India

Abstract: The rapid growth and volatility of cryptocurrency markets have driven significant interest in accurate price forecasting models. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), have emerged as a powerful tool for time-series prediction due to their ability to capture long-term dependencies in sequential data. This review explores recent advancements in cryptocurrency price forecasting using LSTM models, comparing methodologies, datasets, preprocessing techniques, and performance metrics. Key challenges such as overfitting, data noise, and market unpredictability are also discussed. The paper concludes by highlighting research gaps and proposing directions for future development in LSTM-based crypto forecasting.
Keywords: Cryptocurrency forecasting, LSTM, GRU, Bi-LSTM, Deep Learning, Time Series Forecasting, Financial Markets, ARIMA

I. INTRODUCTION

Bitcoin and Ethereum which have revolutionized the financial sector by providing decentralized, borderless and transparent substitutes to fiat currencies. But the extremely volatile and speculative environment of crypto markets complicates things when it comes to predicting, with any degree of certainty, the direction of asset prices. Conventional statistical techniques frequently fail to account for these complex, nonlinear and time-varying structures in cryptocurrency price times series.

Background with the development of artificial intelligence and deep learning, the Long Short-Term Memory (LSTM) networks as a specific type of Recurrent Neural Network (RNN) models have been showing promising results for time series forecasting applications. LSTM models are made for processing sequential data where long, short term dependencies are relevant, hence are ideal for modeling the very dynamic nature of crypto assets.

The goal of this review paper is to present an overview and a comparison of relevant works, and recent works using LSTM models for cryptocurrency price prediction. It emphasizes architectural diversities, data processing methodologies, and evaluation metrics and comparative performance. The paper also details constraints faced by current methods and offers directions for future work. The aim is to present researchers, as well as practitioners, with a brief overview of the state of the art and the main problems posed by LSTM-based forecasting in the context of cryptocurrencies.

II. LITERATURE REVIEW

The literature emphasizes the complexities of cryptocurrency price prediction based on high volatility, heavy-tailed distribution, and market inefficiencies. Early literature concentrated on volatility modeling, whereas current research utilizes machine learning and deep learning models such as ARIMA, MLP, and LSTM to enhance prediction accuracy. Comparative studies identify that LSTM and hybrid models perform better than naive approaches. Nevertheless, they have limitations such as limited data scope, overfitting risks, non-reality of data in real time, and challenges in interpreting complex models. They imply that richer datasets and more interpretable forecasting methods are essential. [1]

Researchers conducted a study using Long Short-Term Memory (LSTM) networks to predict short-term prices of the top 100 cryptocurrencies based on market capitalization. Researchers gathered hourly information for Bitcoin, Ethereum, and Ripple from CoinMarketCap and normalized it with the tanh function. The research team trained a two-layer LSTM neural network consisting of 1000 neurons per layer with dropout using the Adam optimizer and mean absolute error as the loss function. The training loss results demonstrated a downward trend which confirmed the effectiveness of the learning process. The research demonstrated that LSTM networks efficiently capture short-term market movements in the volatile cryptocurrency space while offering a wider analytical range than earlier studies. [2]

The study showed machine learning methods produce more accurate cryptocurrency price predictions compared to traditional approaches by matching actual results more closely. The study highlighted decision trees, support vector machines (SVM), and neural networks (NN) as effective machine learning approaches.

The research indicated that adding cryptocurrencies to multi-asset portfolios provides improved diversification and better risk management. The study began to examine deep learning techniques including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Bi-Directional LSTM (Bi-LSTM) as sophisticated methods for cryptocurrency prediction. Global interest in cryptocurrencies expanded as observers noted their decentralized structure together with cryptographic security mechanisms and blockchain technology which facilitates secure yet pseudo-anonymous transactions. The groundwork established here opened opportunities for more research into deep learning models' predictive potential in cryptocurrency markets. [3]

Recent researches examined the application of machine learning for cryptocurrency price prediction, with an emphasis on Bitcoin. The process entailed dividing the dataset between 80% training and 20% test sets, and the closing price was chosen as the prominent feature to predict. The new target column was formed by moving closing prices five days forward, and the final five entries were designated as missing. Different machine learning models, such as decision trees, linear regression, and support vector machines (SVM), were utilized, some reaching up to 98.79% accuracy. The research noted that machine learning models, and more so hybrid models incorporating ML and deep learning, outperformed standard statistical models, which are poor at handling heterogeneous data. It found that machine learning provides a strong foundation for price prediction in the extremely volatile cryptocurrency market, supporting its increasing use over traditional methods. [4]

A comprehensive study of Bitcoin price prediction with LSTM networks outlined a systematic approach with a strong focus on time series data and training of models. The dataset was preprocessed by dropping extraneous features, and lag plots were utilized to examine correlations on different time scales with higher correlations observed within shorter time periods. Following the division of the data into training and test sets (reserving 60 samples for testing), scaling was done after splitting to avoid data leakage. A data generator function with a five-day lookback time was used to stage input for the LSTM model. The performance of the model was measured with RMSE, and predictions were plotted against actual values. Although the LSTM model was effective, the study pointed out that additional features could enhance accuracy. It suggested the use of high-quality datasets from sites such as Kaggle and called for further examination of deep learning methodologies in cryptocurrency forecasting. [5]

An empirical analysis compared the accuracy of ARIMA, Random Forest, and LSTM models for predicting the prices of Bitcoin, Ethereum, and Dogecoin using 1,500 data points divided into 80% training and 20% testing sets. Twelve characteristic variables, such as closing price, volume, and technical indicators such as CMO, RSI, and MACD, were applied to predict price movements. Accuracy metrics like MSE, RMSE, MAE, and R^2 were used to measure model accuracy. Results indicated that ARIMA failed miserably with negative values of R^2, whereas Random Forest exhibited fair performance ($R^2 \approx 0.73$) for Bitcoin and Ethereum but performed less than optimally for Dogecoin. LSTM surpassed both models, recording values of R^2 well above 0.85, with an all-time high of 0.94243 for Bitcoin, proving its better ability to handle intricate time-series patterns. The research concluded that LSTM is the best-performing model in the task of predicting cryptocurrency prices, reiterating the usefulness of deep learning for modeling highly volatile financial markets. [6]

III. METHODOLOGY

This paper discusses different methodologies adopted for the prediction of cryptocurrency prices based on deep learning and machine learning approaches. The research works reviewed here are methodical in nature with steps involving data collection, preprocessing, development of models, training, and evaluation. The objective is to unify these methodologies into a common framework for easier understanding and comparison.

A. Data Collection

The majority of the studies reviewed employed historical prices collected from publicly available databases like Yahoo Finance, Kaggle, and CoinMarketCap. The datasets mostly contain series of daily or hourly open, high, low, close, and volume values for popular cryptocurrencies like Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), Dogecoin, Bitcoin Cash, Solana, and Binance Coin (BNB). The data covers several years, often between 2015 and 2023.

B. Data Preprocessing

Before model creation, raw data went through various preprocessing steps. The missing values were treated by imputation methods like forward fill. The columns that were redundant were dropped, keeping mostly closing price data only. In a few instances, the data was assigned as a "price" column for ease. The time series data was next normalized through methods like MinMax scaling or hyperbolic tangent (tanh) transformation to make all features fall in similar ranges, which is necessary for enhanced model convergence and precision. Furthermore, label shifting was performed by adding a "prediction" column through which future values were shifted upward to be used as targets.

C. Data Splitting
To measure forecasting performance, the datasets were divided into training and test datasets. One typical split ratio was 80:20, with the training set consisting of the early part of the data and testing set for validation. A few studies also applied windowing methods, wherein a sliding window was employed to generate time-sequenced input-output pairs appropriate for time series modeling. The information was usually reformatted into three dimensions (samples, time steps, features) to be compatible with LSTM-based models.

D. Forecasting Models
Different models of forecasting were employed in the literature. Deep models like Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Gated Recurrent Unit (GRU), and hybrid models like ARIMA-LSTM and ARIMA-MLP were employed. Traditional machine learning models like Support Vector Machine (SVM), Linear Regression, and Decision Tree were also utilized. Ensemble methods like AdaBoost and LightGBM, as well as sophisticated systems like Adaptive Neuro-Fuzzy Inference System (ANFIS), were employed in some studies.

E. Training Configuration
Optimizers such as Adam were utilized to train models with parameters including 100 neurons per layer, 2–3 hidden layers, batch sizes from 4 to 32, and up to 200 epochs. Time series inputs were handled by generator functions or explicit data windowing. Hyperparameters were either manually adjusted or tuned through trial-and-error to achieve optimal performance.

F. Evaluation Metrics
Model performance was validated by statistical error measures. The most commonly utilized measures were Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). In some instances, graphical representations of actual versus predicted prices were employed to qualitatively evaluate model performance. Naive models such as random walk, white noise, and simple trading strategies were also instantiated as baselines for comparison

IV. CHALLENGES AND RESEARCH GAPS

A. Neglect of External and Behavioural Factors
Most current research draws mostly from historical price series without considering the effect of external factors like regulation change, microeconomic signals, and environmental issues. In addition, psychological variables such as investor opinion, cognitive biases, and market psychology are really considered, even though they have considerable effects on cryptocurrency price fluctuations.

B. Lack of Real-Time and High-Frequency Forecasting Models
Most models that have been proposed do not find applications in real-time systems in highly volatile markets. Very little has been explored in terms of techniques that can handle streaming data, learn from new patterns in real-time, and accommodate high-frequency trading environments.

C. Model Robustness and Overfitting Challenges
Deep learning architectures like LSTM, GRU, and Bi-LSTM have a tendency to be overfitted when trained on small or short time-series data. Future work is required to enhance the generalizability and resilience of models on different sets of data and market scenarios.

D. Limited Comparative and Hybrid Modeling Approaches
There is limited thorough comparative research that analyses the performance of various machine learning models, such as ARIMA, Random Forest, and complex natural structures. There is also an untapped potential in hybrid model that merge classical statistical methods and deep learning techniques.

E. Underexplored Multivariate and Cross-Market Analysis
Most studies use single cryptocurrency univariate models. There is a gap in research on developing multivariate models that include variables such as trading volumes, sentiment scores, macroeconomic variables, and inter-cryptocurrency and inter-financial asset correlations.

F. Lack of Long-Term, Cultural, and Ethical Perspectives
Current studies tend to concentrate on short-term prediction with the lack of emphasis on long-term trend analysis. Cultural and geographical differences affecting cryptocurrency adoption are unresearched. Algorithmic trading ethical issues of fairness, interpretability, and market manipulation are largely untackled.

V. FUTURE SCOPE

A. Incorporation of External and Behavioural Variables
Future models can incorporate the impact of external drivers like government policies, overall macroeconomic directions, and environmental influences. Moreover, adding behavioural finance drivers like social media sentiment from stock investors, search patterns, and psychological trends can dramatically improve the effectiveness of forecasting models.

B. Development of Real-Time and Adaptive Forecasting Systems
The research must concentrate on creating models with real-time price forecasting potential by dealing with high-frequency streaming data. Online learning and adaptive algorithms will play a key role in implementing these models in real-time trading platforms in order to make correct and timely decisions.

C. Enhancement of Model Robustness and Generalization
Future research should tackle overfitting issues by examining regularisation methods, dropout layers, transfer learning and ensemble techniques. More importance should be given to the verification of model performance on various datasets and market environments to make it more applicable.

D. Exploration of Hybrid and Comparative Modeling Frameworks
Comparative studies are needed in order to assess both traditional statistical model and more complex machine learning architectures in different settings. In addition, researchers need to create hybrid frameworks that blend the strengths of several algorithms, i.e., ARIMA-LSTM or GRU-RF models, to attend greater accuracy and interpretability.

E. Design of Multivariate and Cross-Domain Models
For more comprehensive insight into market trends, the next generation of work should design multivariate models that take into account multiple cryptocurrencies and their associated assets at the same time. Cross-market correlation with stocks, commodities, and foreign exchange might reveal latent patterns and correlation to help optimise portfolios.

F. Long-Term Analysis and Ethical Implementation
Longitudinal analyses of sustainability and long-term patterns of cryptocurrency markets are needed to determine their maturity and resilience. Moreover, studies need to consider ethical implications, such as model decision-making transparency, protection of market manipulation, and development of equitable and fair algorithmic trading platforms.

VI. CONCLUSION
The review paper examined the existing state of cryptocurrency price prediction with deep learning models including LSTM, GRU, Bi-LSTM, and hybrid strategies. Although such models demonstrate encouraging performance in dealing with the non-linear and volatile nature of cryptocurrency prices, major challenges still persist, such as inadequate incorporation of external variables (such as sentiment and regulations), real-time prediction limitations, model interpretability, and inadequate attention to ethical aspects. The survey indicates that there is a call for more potent, interpretable, and hybrid models, real-time and multivariate data sources, and the analysis of behavior and cross-market effects. Fulfilling these deficits can result in more accurate and trustworthy forecasting systems that facilitate well-informed decision-making in the dynamic and fast-changing realm of digital finance.

REFERENCES

[1] M. Z. A. P. H. K. Y. Ahmed Bouteska, "Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods," International Review of Financial Analysis, vol. 92, 2024.

[2] O. T. A. Ernes Korkmaz, "Cryptocurrency Price Forecasting Using LSTM with Short Time Series Data," International Journal of Advanced Natural Sciences and Engineering Researches, vol. 8, no. 2, pp. 367-376, 2024.

[3] C. R. B. M. E. P. Phumudzo Lloyd Seabe, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach," MDPI, 2023.

[4] D. C. a. S. K. Saroj, "Cryptocurrency Price Prediction Using Supervised Machine Learning Algorithms," ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, vol. 12, no. 1, 2023.

[5] S. B. N. K. S. G. M. S. S. Tushar Maheshwari, "Bitcoin Price Prediction Using LSTM," in ICCS: International Conference on Computing Sciences, 2024, 2023.

[6] L. Pan, "Cryptocurrency Price Prediction Based on ARIMA, Random Forest and LSTM Algorithm," BCP Business & Management,, vol. 38, 2023.