Cryptocurrency Price Prediction with LSTM and Transformer Models Leveraging Momentum and Volatility Technical Indicators

Siddharth Penmetsa
NCSSM department of math and computer science
North Carolina School of Science and Mathematics
Durham, USA
penmetsa150@gmail.com

Maruthi Vemula
NCSSM department of math and computer science
North Carolina School of Science and Mathematics
Durham, USA
vemula.maruthimukesh@gmail.com

***Abstract*— Accurate cryptocurrency price prediction is essential to investors and researchers for analyzing trends and advising financial decisions, as price prediction is fundamental to making beneficial investment decisions. Due to the high volatility and unpredictability of the cryptocurrency market, it is difficult to predict these prices based on cryptocurrency time series data accurately. This research paper presents a two-fold analysis of the effectiveness of neural networks and deep learning to predict cryptocurrency prices and proposes a novel approach to cryptocurrency price prediction. This is done by considering Long-Short Term Memory (LSTM) and Transformer neural networks that use historical price features in addition to volatility and momentum technical indicators, along with historical price features, and testing these models on Bitcoin (BTC), Ethereum (ETH) and Litecoin (LTC). Momentum and volatility technical indicators such as Relative Strength Index (RSI), Bollinger Bands %B and Moving Average Convergence/Divergence (MACD) are not commonly used in cryptocurrency machine learning models. Still, the addition of these features can give better insight into the general trend of the price. By adding volatility and momentum features to our LSTM and Transformer models, we see a significant increase in price prediction accuracy, and we also find that Transformers tend to outperform LSTM models in price prediction and trends of cryptocurrency data.**

Keywords—Cryptocurrency, LSTM, Transformer, Volatility and Momentum Technical Indicators, Price Prediction

I. Introduction

Cryptocurrencies are digital assets used as forms of exchange that exploit cryptography for verifying and recording transactions in a decentralized system, offering an exchange of value independent of state control. The scarcity of the asset is created by the complexity of equations used to validate transactions, thus making it extremely resilient to fraudulent transactions [1]. The original cryptocurrency, Bitcoin, peaked in 2021 with a market cap of $1 trillion [2]. The sharp increase in value for this cryptocurrency, prompted by its decentralized nature, has provoked institutions to pursue stakeholdership in this venture through creating platforms of trade or coordinating operations with the cryptocurrency sector. This trend of investing in cryptocurrencies has matriculated largely into the world and has caused the need for tools to analyze these transactions at very high efficiencies.

Since cryptocurrencies are decentralized through the blockchain and free from control and intervention, they are found to have high volatility as shown by their high beta values [3]. Certain cryptocurrencies (such as Bitcoin and Ethereum) are also found to not possess a high causal relationship with other cryptocurrencies, further reinforcing the unpredictable nature of cryptocurrency prices [3]. Bitcoin, a popular cryptocurrency, has been shown to have a correlation with numerous different variables, among those being investor sentiment and the volatility index (VIX) of Bitcoin's closing price [4]. The correlations that cryptocurrencies have with these different variables along with the general volatility of the market pose risks for investors trying to gain profit.

Due to the volatile nature of cryptocurrencies, it has become increasingly important for investors to be able to accurately predict cryptocurrency prices in order to manage risks, diversify their cryptocurrency portfolio, and ultimately gain profit off the market. Strategies and algorithms for cryptocurrency price prediction can greatly advise investors in making short and long-term investment decisions.

Traditional analysis of Bitcoin and other cryptocurrencies involves utilizing financial and economic data through various numerical methods in order to predict trends. However, along with these, there are several other internal and external components that motivate the flux in the price of cryptocurrencies [5]. Some of these factors include technical indicators including market beta, trading volume, and volatility [6]. Due to the inability of traditional financial models to estimate Bitcoin prices, machine learning models, particularly those based on artificial neural networks, have increased in prominence [7]. Although many models have been developed to predict cryptocurrency prices with great accuracy, many of these models fail to utilize the technical indicators mentioned above [8],[9].

We propose a novel approach to predicting cryptocurrencies with LSTM and Transformer architectures with additional technical indicators that can express the momentum and volatility of the cryptocurrency market. We chose specifically to focus on LSTM because existing research shows that LSTM is very commonly used in price prediction [8],[9],[10]. Transformers, on the other hand, have been used [11], but not nearly as often as other methods, but we find that it is very accurate in sequence prediction, which is why we have chosen it for our study. We found 6 specific indicators to possibly be fruitful for price prediction: Trading volume, Relative Strength Index (RSI), Moving Average Convergence/Divergence (MACD), Signal Line, Histogram, and Bollinger Bands %B. Trading volume is often considered to be the most important technical indicator because it tells us whether the market is bullish or bearish, RSI is a great indicator of volatility and momentum and MACD (along with the Signal Line and Histogram) simplifies price movement in order to better show large trends and shows bearish and bullish movement[12]. Bollinger Bands %B is a great indicator to show the strength of a trend, overbought/oversold conditions, and the general volatility of the market. We know that LSTM is particularly effective at time series forecasting due to its ability to capture short-term and long-term dependencies in time series data and so we propose we use these technical indicators in conjunction with basic cryptocurrency price data (opening price, closing price, high price, and low price) to feed into an LSTM for predictions of the market. We also find Transformers effective in time-series forecasting due to their self-attention mechanism, so we also propose feeding this data into the transformer to predict cryptocurrency prices.

By analyzing the effectiveness of these models through comparison of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) with LSTM and Transformer models that don't use the technical indicators mentioned above, we hope to understand the improvement that these models could possibly have.

II. Related Work

Our investigation starts by looking into the models used to predict the prices of cryptocurrencies. According to [13], several distinct, modern deep-learning models are utilized in order to predict the price of Bitcoin. In this study, the authors checked the Spearman correlation coefficient amongst various Bitcoin prices and features. In doing so, they found that market price, followed by market cap, were the most important features for predicting the price of Bitcoin. Using these as factors for predicting the price of Bitcoin, they went on to compare different deep-learning models and found that an LSTM model performed the best over the other models used in the paper when using an input sequence size of 10.

For other statistical analysis methods for predicting the prices of Bitcoin, we reference the work of the authors of [14] and the references within. McNally et al. [15] compared 2 prediction models, recurrent neural networks (RNN's) and LSTM models, to the autoregressive integrated moving average (ARIMA) [16] model for price direction prediction. In their results, [15] showed that both RNN and LSTM models were better for predicting cryptocurrency price direction predictions than the ARIMA model. Additionally, Saad and Mohaisen [17] conducted an in-depth examination of various Bitcoin blockchain metrics, including the count of Bitcoin wallets, unique addresses, block mining complexity, hash rate, and more. They identified features with a strong correlation to Bitcoin prices and utilized them to develop forecasting models. Their research encompassed a range of regression techniques, encompassing linear regression, random forests [18], gradient boosting [19], and neural networks. Similarly, Jang and Lee [20] expanded their research beyond blockchain data by incorporating macroeconomic indicators like the S&P 500, Euro stoxx 50, DOW 30, NASDAQ, and exchange rates of major fiat currencies. They explored three forecasting methodologies: a Bayesian neural network (BNN) ([21], Chapter 5.7), linear regression, and support vector regression (SVR) [22]. Their findings indicated the superior performance of the BNN model in comparison to the other two. In a subsequent study, Jang et al. [23] introduced a rolling window LSTM approach, demonstrating its enhanced predictive accuracy over models based on linear regression, SVR, neural networks, and LSTM. In addition to these technical factors, Kim et al. [24], Li et al. [25], and Kanji et al. [26] examined the role of social data in predicting Bitcoin price fluctuations.

In addition to the models mentioned above, a new dominant sequence transduction model has been popularized for cryptocurrency price forecasting. [27] first introduced a transformer model as a faster deep learning model relative to LSTM models and RNNs. This transformer model has been adapted for time-series forecasting and utilized in various forms such as a linear transformer [28] or a temporal fusion transformer [29]. Given their efficiency, our paper utilizes an architecture similar to a temporal fusion transformer, based on the architecture by New Wu et al. [30] in addition to an LSTM model in order to compare their performances.

However, none of the works mentioned above considered technical indicators to be included in their models for price prediction. Moreover, unlike most of the previous studies, our paper provides the performance of our models on multiple cryptocurrencies, in addition to Bitcoin, to show the effect of the inclusion of additional technical indicators in cryptocurrency price prediction across multiple currencies in the market.

III. Data Used in The Study

Our dataset consists of 1463 days of data from June 16, 2019, to June 16, 2023. The price data was taken from Trading View [31].

Before using this data in our models, we must preprocess the data to ensure that the model uses quality and accurate data and will be able to make accurate predictions. There are multiple steps of preprocessing that are highlighted in our pipeline that we put our data through.

- Data Cleaning-Removing duplicate and null values that might be in our data.
- Data Integration-Where we convert all values of our data into a mutual data type for feeding into our machine learning models. (for our research, we put it into a Pandas data frame and then into a NumPy array).
- Data Transformation-Using feature scaling to scale all values into a similar range which makes gradient descent (a common optimization algorithm used in machine learning programs) faster and more efficient. After testing MinMaxScaler, StandardScaler and MaxAbsScaler normalization techniques from the Sci-kit Learn library, we found StandardScaler to be most effective, and thus what we chose to feature scale.
- Data Splitting-Next we split our datasets into training data, cross-validation data and test data. We train the data using the training data, we keep track of how well our model generalizes while training with our cross-validation data and we use our test data to see how well our model works. For all of our models, we decided to use a 85-5-10 split where our training data takes up the first 85% of our data, the test data takes up the last 10% of our data and the cross-validation takes up the rest of the data.

TABLE I Definition of variables and data source

Feature	Description	Type of Feature
Closing Price	Price of the cryptocurrency when the market opens on a particular day.	Basic feature
Opening Price	Price of the cryptocurrency when the market closes on a particular day.	Basic feature
High Price	Highest price of the cryptocurrency on a particular day.	Basic feature
Low Price	Lowest Price of the cryptocurrency on a particular day.	Basic feature
Trading Volume	Number of units of the cryptocurrency that were traded on a particular day.	Technical Indicator
Relative Strength Index (RSI)	Momentum oscillator that measures the speed and change of price movements.	Technical Indicator
Moving Average Convergence/Divergence (MACD)	Trend-following momentum indicator based on the Exponential Moving Average (EMA). Used in conjunction with the Signal line and Histogram to provide information about times to sell or buy.	Technical Indicator
Bollinger Bands %B	Indicator derived from the Bollinger Bands that can be used to identify overbought or oversold conditions.	Technical Indicator
Signal Line	Acts as a trigger to buy and sell, used in conjunction with the MACD.	Technical Indicator
Histogram	Represents the difference between the MACD and Signal line, sign of the Histogram indicates Bullish or Bearish market movement.	Technical Indicator

IV. METHODOLOGY

A. *Pipeline and Process*

The pipeline for the study is shown in figure 1. First, we preprocess the historical price and technical indicator data for Bitcoin, Ethereum and Litecoin. Then, we apply our LSTM and Transformer models, and hyperparameter tune accordingly. We then can get our prediction results and interpret the performance of our models.

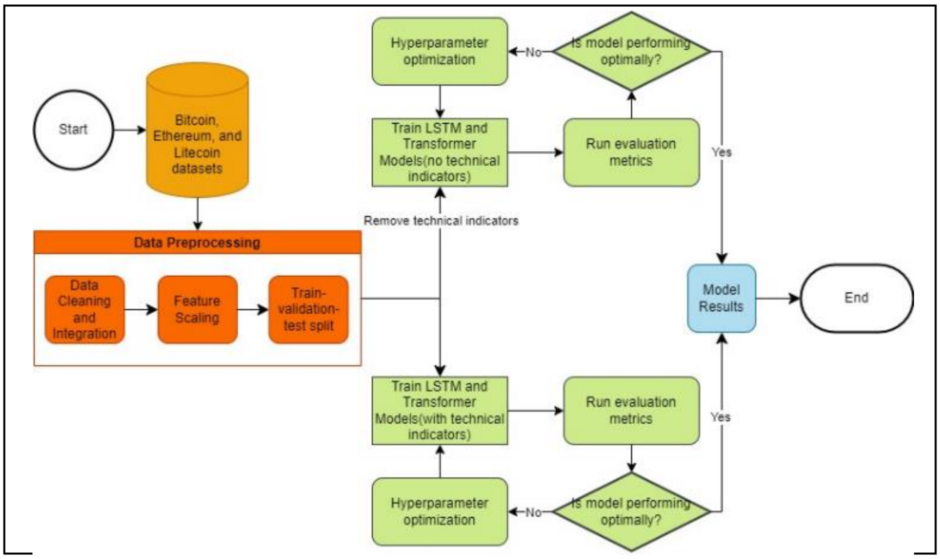

Fig. 1: Schematic pipeline diagram showing our process to results.

B. *LSTM Model*

The LSTM neural network was proposed by [31] which is a type of Recurrent Neural Network (RNN) that can process sequential data for prediction. LSTM *models* have a particular advantage over RNN's due to their ability to learn long-term dependencies by fixing the vanishing/exploding gradient problem (that traditional RNN's have) with structures in repeating cells called gates. LSTM *models* inhibit a similar chain-like structure like ordinary RNN's but, in an LSTM cell there are four neural layers that combine in

the input, forget and output gates. The input gate updates new information, the forget gate gets rid of unimportant information and the output gate passes information. This makes LSTM *models* particularly fruitful for time series forecasting and our use case. Figure 2 shows a diagram for a single cell of an LSTM *model*.

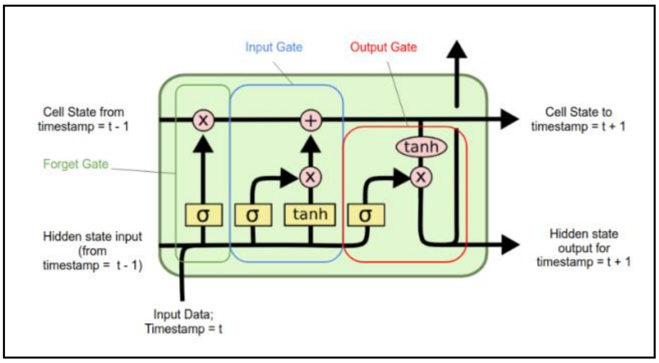

Fig. 2 Structure of an LSTM cell showcasing its 3 gates.

LSTM models have been used successfully in cryptocurrency price prediction in the past [10],[29] due to their effectiveness in time series forecasting. In addition to LSTM layers that find patterns in our time series data, we use dropout layers to prevent overfitting as well as dense layers to get a result. Figure 3 shows a table with specifications regarding our model. The input for our LSTM without indicators is (10,146,4) and input for LSTM with indicators is (10,146,10), where the first number (10) is the length of the past data used to forecast prices, the second number (146) is the length of the test data, and the last number (4 or 10) is the number of features in the input data.

TABLE II LSTM PARAMETERS

Layer	**Parameters**
LSTM	100 units, return sequences on
Dropout	Rate of 0.2
LSTM	200 units, return sequences on
Dropout	Rate of 0.2
LSTM	100 units
Dropout	Rate of 0.2
Dense	50 units
Dropout	Rate of 0.2
Dense	(10 units for w/ indicators and 4 units for w/o indicators)

Every LSTM and Dense layer had the Xavier initialization for weights and biases, each LSTM layer had Tanh activations and sigmoid recurrent activations. Every LSTM used a mean absolute error loss function in training and 2000 epochs were used for every model. Each model used the Adam optimizer for learning rate adjustment, although initial learning rates varied for each cryptocurrency.

C. *Transformer Model*

The Transformer neural network was first proposed in [27] as an architecture for natural language processing (NLP) and machine translation. This model works very well with sequential data in NLP and machine translation, so sequential time series data is a natural extension of its uses. Rather than relying on recurrence for processing sequential data, the transformer is based off a mechanism called "attention" which allows the model to focus on different parts of the input sequence simultaneously, enabling parallel processing and the handling of long-term dependencies. The architecture of a transformer is quite complicated, but it is very effective. Figure 3 shows a diagram for a basic encoder and decoder transformer layer.

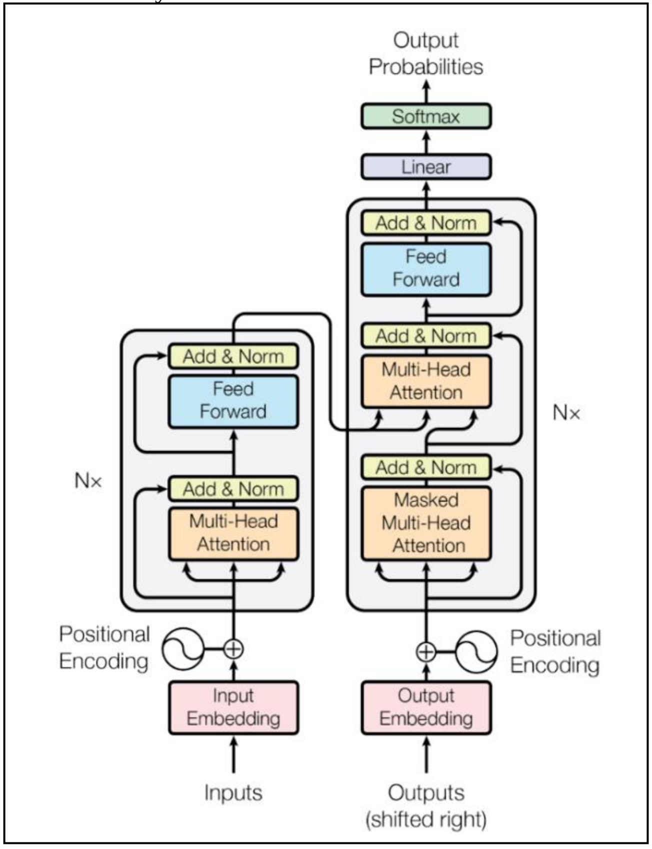

Fig 3. Structure of a simple transformer model with N encoder and decoder layers.

The parameters used in our models are shown below.

TABLE III TRANSFORMER PARAMETERS

Layers	**Parameters**
Positional encoding for encoder	Sub-layer output:512, dropout rate:0.2
Number of encoder layers	3
Number of attention heads	4
Encoder dropout	Rate of 0.2
Feedforward encoder dimension	2048
Positional encoding for decoder	Sub-layer output:512, dropout rate:0.2
Number of decoder layers	3
Decoder dropout	Rate of 0.2
Feedforward decoder dimension	2048

For this model, the Mean Absolute Error (MAE) and the Adam optimizer to optimize gradient descent during training. We also used the ReLU activation function in our feedforward neural networks and the Xavier initialization for weights and biases. We used 75 epochs for training each model, but initial learning rates varied across cryptocurrencies.

IV. Results And Model Performance

Using the proposed models, experimentation was done on three of the most popular cryptocurrencies: Bitcoin, Ethereum, and Litecoin [31]. A 10-day window was utilized for predicting the price of the cryptocurrency the next day. A dataset split of 85, 5, and 10 (in percentages) for training, validation and test respectively were used for training and testing the models. To assess the performance of the models, we used the following evaluation metrics: Mean Absolute Error (MAE) loss [Figure 4], Mean Absolute Percentage Error (MAPE) [Figure 5], and Root Mean Squared Error (RMSE) [Figure 6]. We show a graph of the 2 Transformer prediction values vs. true prices in figure 7.

$$\frac{1}{n}\sum_{n=1}^{n}\left|A_i - P_i\right| \qquad \frac{100}{n}\sum_{n=1}^{n}\frac{\left|A_i - P_i\right|}{A_i} \qquad \sqrt{\sum_{n=1}^{n}\frac{(P_i - A_i)^2}{n}}$$

Fig 4. MAE formula Fig 5. MAPE formula Fig 6. RMSE formula

Fig 7. Graphs of transformer predicted prices and true prices.

We found that our models showed this behavior across all three cryptocurrencies, Bitcoin, Ethereum and Litecoin. The following tables show the evaluation metrics on all of our LSTM and Transformer models.

TABLE IV LSTM EVALUATION METRICS

Model	MAE	RMSE	MAPE
BTC w/o technical indicators	844.57	1012.76	3.18%
BTC w/ technical indicators	697.87	893.55	2.75%
ETH w/o technical indicators	54.13	70.16	3.10%
ETH w/ technical indicators	54.08	64.70	3.08%
LTC w/o technical indicators	2.98	3.71	3.27%
LTC w/ technical indicators	2.65	3.44	2.96%

TABLE V TRANSFORMER EVALUATION METRICS

Model	MAE	RMSE	MAPE
BTC w/o technical indicators	540.65	724.65	2.08%
BTC w/ technical indicators	506.17	704.57	1.96%
ETH w/o technical indicators	45.15	59.17	2.55%
ETH w/ technical indicators	41.60	54.78	2.38%
LTC w/o technical indicators	2.50	3.36	2.82%
LTC w/ technical indicators	2.26	3.28	2.56%

We see that in the case of all 3 cryptocurrencies from Table IV and Table V, the transformer with technical indicators performs the best relative to all the other models. In addition, it can generally be seen that transformers outperform LSTM models in their predictive capabilities and augmenting technical indicators improves the model, regardless of the architecture. One interesting observation we found in our research is that using more complicated Transformer architectures, such as increasing the number of self-attention heads, led to less accurate results, so not only are these complex transformers more computationally expensive, but they are also less effective when dealing with sequence-to-sequence models. LSTM models do not show this same effect, as increasing complexity in LSTM models would almost always lead to a better prediction, but there are diminishing returns after a certain level of complexity, so there must be a tradeoff between computational cost and efficiency.

References

[1] Tredinnick, L. (2019). Cryptocurrencies and the blockchain. *Business Information Review*, *36*(1), 39-44. https://doi.org/10.1177/0266382119836314

[2] Pero, M. C. (2022). Understanding bitcoin and its utility for special operations forces (Ph.D. dissertation). Naval Postgraduate School, Monterey, CA.

[3] Gupta, H., & Chaudhary, R. (2022). An Empirical Study of Volatility in Cryptocurrency Market. *Journal of Risk and Financial Management*, *15*(11), 513. https://doi.org/10.3390/jrfm15110513

[4] López-Cabarcos, M. Á., Pérez-Pico, A. M., Piñeiro-Chousa, J., & Šević, A. (2021). Bitcoin volatility, stock market and investor sentiment. Are they connected? *Finance Research Letters*, *38*, 101399. https://doi.org/10.1016/j.frl.2019.101399

[5] Bouri, E., Jalkh, N., Molnár, P., & Roubaud, D. (2017). Bitcoin for energy commodities before and after the December 2013 crash: diversifier, hedge or safe haven? *Applied Economics, 49*(50), 5063–5073.

[6] Sovbetov, Y. (2018). Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litcoin, and monero. *Journal of Economics and Financial Analysis*, *2*(2), 1–27.

[7] Wang, J., & Wang, J. (2015). Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks. *Neurocomputing*, *156*, 68–78.

[8] Hamayel, M. J., & Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. *AI*, *2*(4), 477-496. https://doi.org/10.3390/ai2040030

[9] Jay, P., Kalariya, V., Parmar, P., Tanwar, S., Kumar, N., & Alazab, M. (2020). Stochastic Neural Networks for Cryptocurrency Price Prediction. *IEEE Access*, *8*, 82804-82818. https://doi.org/10.1109/ACCESS.2020.2990659

[10] Seabe, P. L., Moutsinga, C. R. B., & Pindza, E. (2023). Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach. *Fractal and Fractional*, *7*(2), 203. https://doi.org/10.3390/fractalfract7020203

[11] Tanwar, A., & Kumar, V. (2022). Prediction of Cryptocurrency prices using Transformers and Long Short term Neural Networks. In *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICICCSP53532.2022.9862436

[12] Asaad, R. (2022). Technical Analysis of Digital Currencies and the Most Important Indicators for Trading. *Academic Journal of Nawroz University*, *11*, 337-344. https://doi.org/10.25007/ajnu.v11n4a1481

[13] Ji, S., Kim, J., & Im, H. (2019). A comparative study of bitcoin price prediction using deep learning. Mathematics, 7(10), 898.

[14] Alessandretti, L., ElBahrawy, A., Aiello, L.M., & Baronchelli, A. (2018). Anticipating Cryptocurrency Prices Using Machine Learning. *Complexity, 2018*, 8983590:1–8983590:16.

[15] McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. In *Proceedings of the 2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)* (pp. 339–343), Cambridge, UK.

[16] Box, G., & Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.

[17] Saad, M., & Mohaisen, A. (2018). Towards characterizing blockchain-based cryptocurrencies for highly-accurate predictions. In *Proceedings of the IEEE INFOCOM 2018—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 704–709), Honolulu, HI, USA.

[18] Ho, T.K. (1995). Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition* (Vol. 1, pp. 278–282), Montreal, QC, Canada.

[19] Freund, Y., & Schapire, R.E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci., 55*, 119–139.

[20] Jang, H., & Lee, J. (2018). An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. *IEEE Access, 6*, 5427–5437.

[21] Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[22] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Mach. Learn., 20*, 273–297.

[23] Jang, H., Lee, J., Ko, H., & Lee, W. (2018). Predicting Bitcoin Prices by Using Rolling Window LSTM model. In *Proceedings of the KDD Data Science in Fintech Workshop*, London, UK.

[24] Kim, Y.B., Kim, J.G., Kim, W., Im, J.H., Kim, T.H., Kang, S.J., & Kim, C.H. (2016). Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. *PLoS ONE, 11*.

[25] Li, T.R., Chamrajnagar, A.S., Fong, X.R., Rizik, N.R., & Fu, F. (2018). Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model. arXiv preprint arXiv:1805.00558.

[26] Kanji, A. M., Chaudhary, I., Shankar, R. L., & Srinivasa, G. (2022). Predicting the Price Direction of Bitcoin Using Twitter Data and Machine Learning. In *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)* (pp. 46-52). IEEE.

[27] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*.

[28] Xu, C., Li, J., Feng, B., & Lu, B. (2023). A Financial Time-Series Prediction Model Based on Multiplex Attention and Linear Transformer Structure. *Applied Sciences, 13*(8).

[29] Murray, K., Rossi, A., Carraro, D., & Visentin, A. (2023). On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles. *Forecasting, 5*, 196-209.

[30] Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. arXiv preprint arXiv:2001.08317.

[31] **TradingView.** (n.d.). Bitcoin to USD exchange rate; Ethereum to USD exchange rate; Litecoin to USD exchange rate. Retrieved September 15, 2023, from https://www.tradingview.com/symbols/BTCUSD/, https://www.tradingview.com/symbols/ETHUSD/, https://www.tradingview.com/symbols/LTCUSD/.

[32] **Hochreiter, S., & Schmidhuber, J.** (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735-1780.

[33] **Ryan, T.J.J.** (2020). A single LSTM cell [Image]. Medium.com. https://medium.com/analytics-vidhya/lstms-explained-a-complete-technically-accurate-conceptual-guide-with-keras-2a650327e8f2