Exploring the Predictive Power of On-chain Metrics in Bitcoin Price Forecasting

Graham Comerford
MSc in Blockchain and Distributed Ledger Technologies
Dublin City University
Dublin, Ireland
graham.comerford4@mail.dcu.ie

***Abstract*— Given the volatility of the cryptocurrency market and the significant financial implications of Bitcoin's price movements, accurate forecasts are indispensable tools for investment and risk mitigation strategies. Despite extensive research, the prediction of Bitcoin's price trajectory remains a formidable challenge due to the market's inherent volatility and complex nature. This study investigates the potential benefits of integrating on-chain metrics into models designed for Bitcoin price prediction. To the best of our knowledge this is the first such study. Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Recurrent Neural Network (BiRNN) are employed, two machine learning models noted for their efficacy in handling time-series data. Our findings indicate that while BiRNN falls short of BiLSTM in terms of performance, the inclusion of on-chain metrics noticeably enhances the predictive capability of both models.**

I. Introduction

Cryptocurrency markets have witnessed substantial growth and evolution in the past decade, with the total market capitalisation hitting a peak value of $2.8 trillion in 2021 [28]. Bitcoin has been at the forefront of this market, currently representing 50% of the total market cap [29]. With such growth comes periods of high volatility. Bitcoins average volatility in 2021 was 4.56% which is high when compared to gold's average of 1.2% or major currencies which average between 0.5% and 1% [36]. In spite of this volatility, a significant portion of the financial sector has shown interest in cryptocurrency investments. In fact, 38% of traditional hedge funds are now allocating investments in this domain, and over 300 specialised cryptocurrency hedge funds have emerged in response to the rising demand [30].

In this rapidly evolving landscape, the value of accurate Bitcoin price predictions is substantial. Such forecasts act as vital instruments in financial market participation, directing decisions on when to buy, sell, or hold an asset. Beyond informing trading decisions, price predictions serve a crucial role in risk management, empowering investors to evaluate potential risks associated with their investments and optimise their portfolios accordingly. Further, these predictions enable strategic hedging, where anticipated market trajectories inform investments in assets that can counterbalance potential losses, enhancing the resilience of investment portfolios.

Despite the best efforts of traders and researchers, predicting Bitcoins price movements remains a challenging task due to the inherent volatility and complexity of the market. Researchers have utilised a range of methodologies to forecast Bitcoin prices, such as traditional statistical techniques and machine learning algorithms. Notably, many of these studies have concentrated primarily on using price and volume data, often neglecting other potentially significant data sources.

On-chain metrics, which are derived from the analysis of transactions recorded on the Bitcoin blockchain, provide a rich source of information about the network's activity. They encompass various factors, such as transaction volumes and flows, user behaviour, and mining information. These metrics offer insights that extend beyond traditional price and volume data, potentially aiding in the prediction of Bitcoin price movements.

However, the potential of on-chain metrics in Bitcoin price prediction has been largely underexplored in the existing literature. This paper aims to fill this gap by investigating the value of on-chain metrics, in combination with technical indicators, in forecasting Bitcoin prices. Two machine learning models, namely the Bidirectional Long Short-Term Memory (BiLSTM) [38], and Bidirectional Recurrent Neural Network (BiRNN) [39][40] which are known for their efficacy in handling time-series data, are employed.

The paper is structured as follows: Section II provides a technical background on the Bitcoin indicators, both technical and on-chain, and the machine learning techniques used in this study. Section III reviews the related literature, focusing on statistical forecasting, machine learning, feature selection, and deep learning in the context of Bitcoin price prediction. Section IV outlines our methodology. Section V presents the results of our study, followed by a discussion in Section VI. Finally, Section VII concludes the paper and suggests potential directions for future research.

II. TECHNICAL BACKGROUND

A. Technical Indicators

Technical Indicators are statistical calculations based on historical trading activity such as price and volume. These indicators are used to forecast potential price changes and trends in financial markets. They have been shown to offer significant predictive power and profitability in Bitcoin trading [31]. They provide a heuristic approach to understand market psychology and sentiment, making them valuable for both investors and traders. They typically fall into four categories: trend, momentum, volatility, and volume indicators. Trend indicators evaluate the direction of market movements. Momentum indicators assess the speed of price changes. Volatility indicators quantify the extent of price fluctuations, while volume indicators measure the intensity of investor participation. Notably, these indicators do not provide absolute predictions but rather provide a probabilistic view of potential market scenarios.

B. On-chain Metrics

On-chain metrics provide a novel source of data in the cryptocurrency market, focusing on transaction activities within the blockchain. These metrics yield insights into investor behaviour and blockchain network health. By analysing transaction volumes, transaction fees, active addresses, and other on-chain data, one can obtain a holistic view of the underlying 'fundamentals' of a cryptocurrency, beyond what is available from price and volume data alone. In this study, a comprehensive set of on-chain metrics is utilised to train the predictive models.

C. Machine Learning Price Prediction

Machine learning techniques are increasingly being used in financial markets to predict asset prices [32]. These techniques include both supervised learning, where a model is trained on historical data to predict future prices, and unsupervised learning, where patterns are identified without prior training. Specifically, in this research, BiRNN and BiLSTM networks are employed. Both BiRNNs and BiLSTMs are types of artificial neural networks designed to recognize patterns over time, making them particularly well-suited for time-series data. BiRNNs can model complex relationships between inputs and outputs with time lags, while BiLSTMs, a special kind of RNN, can capture long-term dependencies in time-series data, which is particularly relevant in volatile financial markets. By combining technical indicators and on-chain metrics with these advanced machine learning techniques, we aim to enhance the accuracy of Bitcoin price prediction models.

The combination of these three areas - technical indicators, on-chain metrics, and machine learning techniques for price prediction - forms the basis of this study. In particular, the study seeks to evaluate the value of integrating on-chain metrics with technical indicators for Bitcoin price prediction, applying the RNN and LSTM models on the compiled dataset.

III. RELATED LITERATURE

A. Statistical Forecasting

Statistical forecasting involves building mathematical models that allow the prediction of future values based on historical data. Statistical forecasting is well established and has been used in the financial industry for almost a century. As highlighted in [23] some corporate investors have obligations to disclose how investment decisions are made to their customers. This excludes them from using certain machine learning models as the decision making process within the model can not be explained easily by the institution. For this reason statistical forecasting is an active area of research in bitcoin price prediction.

In [1], Bitcoin volatility forecasting is examined using historical volatility and exchange order book data, with features including volume, spread, depth, and slope for both bid and ask sides. A temporal mixture model, a weighted sum of specialised component models, outperforms standard models that focus on a single data set, suggesting that integrating multiple data sets can enhance prediction accuracy. However, certain models, when using both data sets, do not outperform their counterparts, emphasising the importance of not only feature selection but also model selection for effective Bitcoin forecasts. The study recommends exploring additional data sources like social media, blockchain data, and exchange data for future analysis.

In [2], the correlation between Bitcoin returns and transaction activity is analysed using bivariate Vector Autoregressive (VAR) models, revealing strong links and suggesting a focus on Bitcoin's microstructure for future work. [3] also explores transaction activity's predictive power on Bitcoin returns, finding that increased transaction activity correlates with higher returns in a bull market and lower returns in a bear market. Similarly, [4] emphasises the Bitcoin microstructure and introduces a framework for analysing the Bitcoin transaction graph's local topography through chainlet motifs, which are recurring topographical structures. Utilising chainlets resulted in more competitive pricing models compared to traditional time series models. It is clear from [2], [3] and [4] that analysing the information from the blockchain itself will form an important part of building a pricing model.

B. Machine learning-based Forecasting

1) Machine Learning

Research in the area of bitcoin has focused heavily on training models based on historical price data. Linear regression, decision trees, random forests and neural network techniques have been studied as means to forecast bitcoins price. An important area of research is identifying which machine learning techniques offer the best results for forecasting bitcoin price. In [5] a review of the machine learning algorithms used in price prediction is carried out. These include ARIMA, Linear Regression, LSM, BGLM, GARCH, SVM, LSTM, NARX, and MLP. The NARX model was found to have the most accurate predictions achieving 62% accuracy. Unfortunately it is difficult to further interpret these results as the testing and evaluating process is not documented. In [6] a more thorough examination of the various models is documented. A variety of nonlinear models are trained using a large range of technical indicators representing market momentum, trend, sentiment, and volume. A nonlinear Random Forest model is found to be the best performing model. This seems to go against the logical assumption that a time series style model would perform better than a classification style model. However there is literature to support this finding. In [7] Random Forests were found to outperform existing models when predicting stock prices. In [9] Random Forest models are created using a range of technical indicators and are found to have accuracy of 92%, 85%, 87% for 3 day, 5 day, 10 day directional predictions. In [10] SVM, ANN, NB, and Random Forest models are compared and the Random Forest is found to have the highest forecasting performance in the continuous dataset. Reference [8] gives a possible explanation for the advantage that Random Forests have. It attributes the advantage to the fact that they allow nonlinear predictor interactions that are missed by other style models. The research in [6] is very useful, not only in the way that it highlights the Random Forest model as one suitable for further examination, but also in that it provides a clear overview of how the effectiveness of models and features can be tested.

2) Feature selection

The importance of feature selection has been covered widely in the literature. In [6] a quantitative analysis is carried out with regards to the impact the individual features have on the model's accuracy. Through this analysis a shift is identified in the bitcoin market from being dominated by trend following and momentum in earlier years to one driven by sentiment: fear, greed, and google search activity in more recent years. This finding strengthens the position that including features beyond the typical price and volume information in the training of models may lead to enhanced accuracy in forecasts. A number of researchers have included non-typical features in their training data. Reference [11] documents an LSTM model trained using bitcoin price and social sentiment gathered from twitter and analysed using Apache Flume. Results were not satisfactory at 50% accuracy and 60% precision. An explanation for these unsatisfactory results may be found in [15] where it is found that it is not the sentiment that is a highly predictive feature for price but rather the overall volume of tweets or google searches. This highlights the importance of examining the data and trying different variations of the same feature in order to arrive at the optimal combination for training the model.

In [12] an attempt is made to quantify the importance of business cycle variables like interest rates, inflation, and market volatility for predicting bitcoin price. To do so Random Forest models are trained using bitcoin price, a variety of technical indicators, and macroeconomic variables. It is found that yield on the US ten-year bond and the oil volatility index are the most important when predicting price direction. These results highlight the impact that the broader macroeconomic climate has on bitcoin price. However it raises a question. Is it worthwhile including macro variables in the training data if the prediction horizon is short? In [13] this problem is addressed and it is suggested that for low frequency daily price prediction, less complex algorithms should be used but with several higher dimension features. For high frequency price prediction a low amount of features should be used but with a complex algorithm. These findings are echoed in [14] where it is found that as the prediction horizon increases the importance of features other than the bitcoin price increase. This paper is a valuable contribution to the literature as it documents a process for quantifying the impact of individual features in a bitcoin prediction model. Impact is quantified by randomly permuting every feature vector with a standard normalised vector and then measuring the decrease in prediction accuracy. If there is a large drop in accuracy then it means that the prediction model relies heavily on that feature for predictions. This process could be implemented when evaluating the impact that the on-chain indicators have on the accuracy of the proposed models predictions.

3) Deep learning

Recently several papers have focused on using deep learning techniques to forecast bitcoin price. In [18] several models were created using different techniques and several features. The techniques used were Artificial neural network, Stacked artificial neural network, Support vector machines, and LSTM. The features included Simple Moving Average, Exponential Moving Average, Relative Strength Index, Weighted Moving Average, Standard Deviation, Variance, Triple Moving Exponential and Rate of Change. The importance of feature selection and its relation to data frequency was highlighted. Several techniques were used to identify the important features and these were carried out separately for each prediction horizon. Overall the LSTM was found to have the best performance with next day prediction accuracy of up to 65%. Reference [17] studied a hybrid approach of LSTM and Bayesian Optimisation. They made the novel choice to use weekly data in order to try to minimise the detrimental effect of bitcoins volatility on the quality of their models predictions. Most research up to this point had used minute range or daily range data. In [19] LSTM and MLP models were trained with exchange activity and social features. It was found that the accuracy of the models increases with the additional features and that the LSTM is the best model of the two for price prediction. Similarly [20] compared LSTM with a Gradient Boosting Model and found that the LSTM performs 7% better. The research presented in [17], [18], [19], and [20] all point to LSTM models as being worthy of future examination with regard to bitcoin price prediction.

A novel approach was taken in [21] where the most frequent edges of the bitcoin transaction network were used to make predictions about bitcoins price. The process of creating the transaction network for this study involved creating a node for every address in the bitcoin blockchain and a directed link between two nodes for each transaction as outlined in [22]. The frequency of use of each edge was quantified daily and then the most frequently used ones were taken for further examination. This group of the most frequently used edges was further narrowed down using a Random Forest classifier which eliminated the edges that had a weak impact on price. The remaining edges were used to train a single-hidden layer feedforward neural network (SLFN) alongside the log of bitcoin returns. The model achieved 60.05% accuracy. An interesting finding is that the model achieved this result using only the most active 0.45% of edges in the transaction network. These findings show that the bitcoin transaction set holds valuable predictive power over price. As most of the on-chain indicators are drawn from the transaction set these findings justify their further examination in relation to bitcoin price prediction.

The existing literature on Bitcoin price prediction primarily explores the use of statistical forecasting, machine learning, and deep learning, with features including historical price data, technical indicators, social sentiment data, and macroeconomic variables. Notably, some studies have started considering transactional activity data from the Bitcoin blockchain. Despite this diversity in approach, there is a lack of emphasis on on-chain metrics in Bitcoin price prediction models. The inclusion of the metrics, as proposed in this study, could potentially enhance the effectiveness of these models and provide a deeper understanding of the fundamental factors driving Bitcoin's price dynamics.

IV. Methodology

The methodology presented here is rooted in the 7 steps of machine learning [33], a well-structured, reliable, and replicable process that lays the groundwork for any machine learning project. In this study the 7 step methodology allows us to carefully work through the machine learning process, ensuring the reliability of the results and setting the stage for future research and practical applications in the field. The research methodology followed in this work includes the following steps: data collection, data preparation, model selection, model training, model evaluation, hyperparameter tuning, prediction.

A. Data Collection

The machine learning process begins with the collection of two separate datasets, 'ta_df'[1] and 'oc_ta_df'[2], spanning from January 2nd, 2015 to December 30th, 2022. The first dataset, 'ta_df', contains Bitcoin's price and volume data collected from Kaggle [34]. This dataset is then enriched with 86 unique technical indicators calculated using the TA library [35] in Python. These indicators encapsulate various aspects such as volume, volatility, trend, and momentum. Also, additional metrics like the daily return, daily log return, and cumulative return are incorporated[3]. The 'oc_ta_df' dataset extends 'ta_df' by supplementing it with 123 Bitcoin on-chain metrics[4] fetched from Glassode, a prominent provider of on-chain analysis [27]. Each metric is downloaded individually from Glassnode as a csv and then manually merged.

B. Data Preparation

After collecting the datasets, the next step is to prepare them for training models. The datasets are inspected for missing values and the data is visualised to grasp its nature. A correlation analysis is also performed to identify potentially significant features. This is followed by a scaling operation, employing MinMaxScaler from the sklearn.preprocessing module, which will standardise the features within a range of 0 and 1. This step diminishes the effect of the variability of high magnitude features on the model.

Once scaled, the datasets are partitioned into training and testing sets in a 9:1 ratio. A 60-day look-back period and a 1-day forecast period are used, following the standard practice in time series modelling and forecasting. Consequently, the model utilises the past 60 days of data to predict the price for the forthcoming day.

C. Model Selection

For this study, the selection gravitated towards two related models, BiRNN and BiLSTM. The preference for these models arises from key model attributes that align well with the nature of our datasets. The strength of BiRNNs and BiLSTMs in time series analysis is their capability to handle temporal dependencies. These models, in contrast to traditional ones that presuppose independence of observations, acknowledge the correlation between current and preceding observations. This recognition of dependence is a critical aspect of financial time series data.

Furthermore, BiLSTMs demonstrate superior ability in retaining past information over extended sequences. In financial forecasting, this attribute proves highly beneficial as long-standing trends and cyclical patterns can significantly influence future prices. Importantly, both BiRNN and BiLSTM models can adeptly handle multivariate time series data, which is essential for this study due to the extensive number of features in both datasets.

In the preliminary exploration phase, bidirectional models of RNN and LSTM outperformed their unidirectional counterparts in predictive performance. This superior performance may be attributed to their enhanced ability to comprehend data patterns during the training phase when they have access to complete sequences. As per [37],

[1] ta_df.csv
[2] oc_ta_df.csv
[3] description-of-technical.pdf
[4] description-of-on-chain.pdf

BiLSTMs, with their extra reverse layer, mine more data than unidirectional LSTMs, leading to a deeper data understanding. By leveraging future information for a specific time step, these models create a robust internal representation that facilitates effective generalisation to unseen data. Consequently, despite functioning unidirectionally during prediction, their bidirectional learning in training may equip them to handle price prediction tasks more proficiently.

In this study, the specific architectures utilised include a BiRNN model, which is composed of two primary layers of Bidirectional SimpleRNNs. Each of these is followed by a Dropout layer to prevent overfitting, and the network concludes with two Dense layers designed for regression predictions. Likewise, the BiLSTM model is structured with two main layers of Bidirectional LSTMs, each succeeded by a Dropout layer to mitigate overfitting. This model also culminates with two Dense layers intended for regression predictions. The detailed hyperparameters, including the number of units in each layer for the two models, are elaborated in Tables I and II.

D. Model Training

For each model type, training is performed on each dataset for a total of 50 epochs. The learning rate is set at 0.001, which determines the step size at each iteration while moving towards a minimum loss. The training process employs the Adam optimizer, a popular choice for training deep learning models due to its efficient computation and little memory requirement.

During the training phase 10% of the training set is set aside for validation. The validation set provides a reliable estimate of the model's performance on unseen data during training. The loss function utilised is the Mean Squared Error, a common choice for regression problems as it punishes large errors more due to its squared nature.

To ensure the robustness of the results and to account for the inherent randomness of machine learning training, five models are trained for each dataset. This allows for the performance to be averaged over multiple runs, leading to more robust results.

E.Model Evaluation

The models are evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the R-squared (R^2) statistic. These metrics allow for the robust evaluation of model performance by offering different perspectives on the model's error characteristics.

The Mean Absolute Error (MAE) provides a direct measure of the average magnitude of the forecast errors, expressing the typical prediction error in the same units as the data. The Mean Squared Error (MSE) squares each error before averaging them, thereby giving more weight to larger errors. The Root Mean Squared Error (RMSE), as the square root of the MSE, also emphasises larger errors but reverts to the original units of the output. The Mean Absolute Percentage Error (MAPE), on the other hand, expresses the forecast error as a percentage, providing a relative measure of error and thus a perspective on the cost of the error.

TABLE I. BiRNN MODEL ARCHITECTURE

Layer Type	**Exp1 Units**	**Exp2 Units**	**Other Parameters**
Bidirectional SimpleRNN	384	448	return_sequences=True
Dropout	-	-	rate=0.2
Bidirectional SimpleRNN	128	256	return_sequences=False
Dense	25	25	-
Dense	1	1	-

TABLE II. BiLSTM MODEL ARCHITECTURE

Layer Type	**Exp1 Units**	**Exp2 Units**	**Other Parameters**
Bidirectional LSTM	256	288	return_sequences=True
Dropout	-	-	rate=0.2
Bidirectional LSTM	128	128	return_sequences=False
Dense	25	25	-
Dense	1	1	-

Finally, the R-squared (R^2) statistic quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables. An R^2 score closer to 1 indicates that the model can explain a large portion of the variance in the output variable, thus demonstrating a strong predictive power.

Taken together, these metrics provide a comprehensive evaluation of the models' performance, offering a detailed understanding of their strengths and weaknesses, and paving the way for future improvements.

F. Hyperparameter Tuning

The performance of a machine learning model can be significantly influenced by its hyperparameters, necessitating careful tuning. This study employs two approaches, both manual experimentation, and automated tuning via Keras Tuner, a Python library dedicated to hyperparameter optimization. In the manual tuning process, potential hyperparameters are iteratively tested and refined based on the model's performance on the test set. In contrast, Keras Tuner leverages random search to efficiently and effectively traverse the model's hyperparameter space. The tuning process is conducted over ten trials, with each trial consisting of three executions. This approach helps mitigate the impact of the inherent stochasticity of neural networks. The optimal set of hyperparameters is then identified based on the model's performance on the validation set, paving the way for robust and accurate predictions.

G. Prediction

Upon finalising the model training and hyperparameter tuning phases, the models are put to the test with a rigorous evaluation process. This process involves assessing the models' performance on the test set, a data subset not encountered during training, providing an unbiased estimate of real-world performance. This evaluation can expose any disparities between training performance and performance on unseen data, highlighting potential overfitting or underfitting issues. The predictions are also plotted against actual values, as visual comparisons can illuminate any patterns in prediction errors, providing insights for potential model improvement.

V. Results

The study was divided into two distinct experimental stages to ensure a thorough evaluation of the models. In the initial stage, the models were manually hyperparameter tuned to achieve optimal performance, followed by an evaluation phase. The second stage involved automatic hyperparameter tuning using Keras Tuner's RandomSearch on the 'ta_df' dataset, with the goal of optimising the number of units in all layers. Subsequently, the models were evaluated again. The purpose of conducting these distinct experiments was to confirm that the observed improvements due to the inclusion of on-chain metrics were not merely a result of hyperparameter tuning. If the models, even when optimised on the 'ta_df' dataset, performed better on the 'oc_ta_df' dataset, this would provide a strong testament to the value of the on-chain metrics. The full set of results is available in Table III.

A. Experiment 1

When trained on the ta_df dataset, the BiRNN model had a high RMSE varying from 1496.01 to 4015.49 and R2 scores from -0.081 to 0.397. The wide range of these metrics, especially the negative R2 score, indicates that the RNN model had difficulty accurately predicting the Bitcoin price using only technical indicators.

When trained on the ta_df dataset, the BiLSTM model showed a varied performance. The RMSE fluctuated from 1031.21 to 2289.43, indicating a significant fluctuation in the model's ability to accurately predict the Bitcoin price. The R2 scores varied from 0.673 to 0.934.

Training the BiRNN model on the oc_ta_df dataset, however, led to an improvement in performance. The MAE range was reduced to 984.6 to 3499.9, and the R2 scores ranged from 0.133 to 0.897. This suggests that the inclusion of on-chain metrics could enhance the BiRNN model's performance, similar to the effect observed with the BiLSTM model.

On the other hand, when the BiLSTM model was trained on the dataset incorporating both technical indicators and on-chain metrics (oc_ta_df), the performance showed an overall improvement. The MAE ranged from 962.6 to 1291.6, which is notably lower than the range observed with the ta_df dataset. This suggests that including on-chain metrics in the feature set can potentially enhance the accuracy of the BiLSTM model in predicting Bitcoin prices. The R2 scores, ranging from 0.854 to 0.910, also improved, further corroborating this observation.

The variability in the models' performance across different runs highlights the inherent complexity and uncertainty involved in predicting Bitcoin prices. Table IV summarises the averaged results obtained from both the BiLSTM and BiRNN models and for both datasets (ta_df and oc_ta_df). The MAE, MSE, RMSE, MAPE, and R2 scores for each run are presented, The enhanced performance of the models when trained on the dataset including on-chain metrics (oc_ta_df) suggests that these metrics provide valuable information for predicting Bitcoin prices.

Table III. Experiment results

Dataset	Model	Run	MAE Exp1	MSE Exp1	RMSE Exp1	MAPE Exp1	R2 Exp1	MAE Exp2	MSE Exp2	RMSE Exp2	MAPE Exp2	R2 Exp2
ta_df	BiRNN	1	763.3	1.063e6	1031.2	3.64	0.93	1120.998	1.897e6	1377.353	5.358	0.882
ta_df	BiRNN	2	1243.2	2.132e6	1460.1	6.06	0.87	15732.875	3.424e8	18505.276	75.257	-20.339
ta_df	BiRNN	3	1592.1	3.365e6	1834.3	7.78	0.79	9963.441	1.040e8	10196.329	48.179	-5.478
ta_df	BiRNN	4	2017.8	5.242e6	2289.4	9.55	0.67	1878.252	4.911e6	2216.09	9.665	0.694
ta_df	BiRNN	5	986.5	1.778e6	1333.4	4.54	0.89	2220.149	6.006e6	2450.808	11.078	0.626
ta_df	BiLSTM	1	3893.6	16.124e6	4015.5	18.95	-0.005	2187.17	5.806e6	2409.461	10.721	0.638
ta_df	BiLSTM	2	4222.1	19.271e6	4389.8	20.26	-0.20	2045.03	4.939e6	2222.375	10.055	0.692
ta_df	BiLSTM	3	2924.0	9.679e6	3111.2	14.58	0.40	1340.657	2.538e6	1593.074	6.563	0.842
ta_df	BiLSTM	4	1100.0	2.238e6	1496.0	5.07	0.86	1522.456	3.088e6	1757.285	7.359	0.808
ta_df	BiLSTM	5	3991.0	17.353e6	4165.7	19.28	-0.08	3174.295	1.099e7	3314.553	15.41	0.315
oc_ta_df	BiRNN	1	1291.6	2.349e6	1532.7	6.25	0.85	5526.305	3.173e7	5633.066	27.239	-0.977
oc_ta_df	BiRNN	2	1054.4	1.701e6	1304.3	5.15	0.89	3515.056	1.424e7	3773.448	16.802	0.113
oc_ta_df	BiRNN	3	1178.6	2.237e6	1495.8	5.71	0.86	2607.563	1.031e7	3211.324	13.261	0.357
oc_ta_df	BiRNN	4	982.6	1.611e6	1269.3	4.54	0.90	1698.837	4.055e6	2013.785	8.461	0.747
oc_ta_df	BiRNN	5	962.6	1.440e6	1200.0	4.69	0.91	3967.338	1.804e7	4247.511	19.914	-0.124
oc_ta_df	BiLSTM	1	1477.0	3.691e6	1921.1	7.46	0.77	2633.771	8.076e6	2841.762	13.405	0.497
oc_ta_df	BiLSTM	2	1414.0	3.178e6	1782.6	6.51	0.80	848.6	1.276e6	1129.78	4.151	0.92
oc_ta_df	BiLSTM	3	1747.6	3.821e6	1954.7	8.42	0.76	3316.774	1.187e7	3445.472	16.371	0.26
oc_ta_df	BiLSTM	4	984.6	1.661e6	1288.8	4.55	0.90	1497.96	3.141e6	1772.206	7.533	0.804
oc_ta_df	BiLSTM	5	3500.0	13.915e60	3730.3	17.47	0.13	985.204	1.513e6	1230.113	4.808	0.906

B. Experiment 2

The performance of the BiRNN and BiLSTM models trained on the ta_df dataset demonstrated a significant degree of variation. For the BiRNN model, the RMSE spanned from 1377.35 to an exceptional 18505.28, indicating a substantial inconsistency in accurately predicting Bitcoin prices. The R2 scores also showed wide-ranging results, from a high of 0.881 to an extreme low of -20.33, underlining the model's instability when utilising only technical indicators.

Conversely, when trained on the same ta_df dataset, the BiLSTM model displayed more reliable performance with an RMSE ranging from 1593.07 to 3314.55 and R2 scores from 0.315 to 0.841, reflecting a lesser degree of variability compared to the BiRNN model.When the BiRNN model was trained on the dataset incorporating both technical indicators and on-chain metrics (oc_ta_df), an overall enhancement in performance was observed. The RMSE spanned from 2013.78 to 5633.07, showing a significant reduction compared to the results obtained with the ta_df dataset. R2 scores also improved, ranging from -0.977 to 0.747, highlighting the positive effect of including on-chain metrics on model accuracy.

A similar improvement was observed with the BiLSTM model trained on the oc_ta_df dataset. The RMSE values were notably lower, ranging from 1129.78 to 3445.47, with R2 scores consistently positive, ranging from 0.26 to 0.92.

Table iv. Averaged results of experiment 1

DF	Model	MAE	MSE	RMSE	MAPE	R2
oc_ta_df	bilstm	1093.935	1.87e6	1360.423	5.267	0.884
oc_ta_df	birnn	1824.639	5.25e6	2275.504	8.882	0.673
ta_df	bilstm	1320.597	2.72e6	1589.691	6.313	0.831
ta_df	birnn	3226.132	1.29e7	3435.640	15.631	0.194

Table v. Averaged results of experiment 2

DF	Model	MAE	MSE	RMSE	MAPE	R2
oc_ta_df	bilstm	1856.462	5.18e6	2273.866	9.253	0.678
oc_ta_df	birnn	3463.02	1.57e7	3775.827	17.136	0.023
ta_df	bilstm	2053.921	5.47e6	2339.35	10.022	0.659
ta_df	birnn	6183.143	9.19e7	6949.171	29.907	-4.723

Table V summarises the averaged results obtained from both the BiLSTM and BiRNN models and for both datasets (ta_df and oc_ta_df). Although the models' hyperparameters were optimised for the ta_df dataset, the oc_ta_df dataset outperformed the ta_df dataset, suggesting that the on-chain metrics provide valuable information for predicting Bitcoin prices. This experiment provides compelling evidence of the value of incorporating on-chain metrics in models for predicting Bitcoin prices.

VI. Discussion

Predicting Bitcoin prices is complex due to the high volatility of cryptocurrencies [1]. This complexity arises from numerous internal and external factors like regulatory news, technological advancements, macroeconomic environment, and social media sentiment. Although on-chain metrics offer insights into transactional and usage elements of a blockchain, they fail to capture the full range of these external influences.

In addition, the inherent characteristics of financial time series, which often exhibit non-stationarity and a high noise-to-signal ratio, further compound the complexity of prediction tasks. Recurrent Neural Networks, like other machine learning models, can struggle to discern patterns within noisy data. This struggle can lead to inconsistencies and variability across different training runs, as the model may interpret noise as meaningful information, thereby affecting its learning process and subsequent performance.

In our study, we observed that BiRNNs consistently underperform relative to the BiLSTM models, generating higher prediction errors and demonstrating a poorer fit to the data. This could be attributed to inherent weaknesses in RNNs such as the challenges in capturing long-term dependencies and the issues related to vanishing and exploding gradients. These limitations provide opportunities for future research to explore and mitigate these challenges.

Furthermore, this study demonstrates the value of including on-chain metrics in enhancing the predictive performance of machine learning models for Bitcoin prices. Nevertheless, this enhancement needs to be interpreted with a degree of caution. Although on-chain metrics boost predictive abilities, their incorporation does not inherently guarantee the successful execution of trading strategies.

In financial forecasting, an important consideration is a model's directional accuracy – its ability to predict the correct direction of price changes. Interestingly, a model with lower error, as measured by standard metrics like Mean Squared Error (MSE) or Mean Absolute Error (MAE), isn't always the most effective. A model might provide precise price predictions yet fail to correctly anticipate the direction of price changes. Such a model could lead to incorrect trading decisions and consequent financial losses.

Traditional evaluation metrics such as MSE, MAE, and R-squared primarily measure the deviation between predicted and actual values, assuming a linear relationship between input features and the target variable. However, these metrics might overlook the importance of correctly predicting the direction of price changes – a crucial aspect of financial forecasting.

Therefore, the use of alternate evaluation metrics that capture a model's success in predicting price direction or its profitability, is proposed. These could include mean directional accuracy, profitability metrics, or risk-adjusted return metrics like the Sharpe ratio.

Looking ahead, one promising research avenue could involve the development of ensemble techniques that leverage both classification models (for direction prediction) and regression models (for estimating the magnitude of price movements). This combined approach might help harness the strengths of both methodologies, potentially leading to more robust and profitable prediction models

While academic exploration of price prediction is an interesting field contributing to our understanding of market dynamics, translating these models into real-world trading applications remains a challenge. The accuracy levels needed for a model to be reliably profitable in real-world scenarios often significantly surpass those achieved in academic studies. Even minor prediction errors can culminate in substantial financial losses over time.

Moreover, real-world trading factors such as transaction costs, risk management, timing of trades, often overlooked in academic studies, can critically influence the profitability of a trading strategy. Therefore, while our study enhances the literature by exploring the potential of on-chain metrics in price prediction, the leap from a theoretically sound prediction model to a practically profitable trading strategy remains a significant challenge that deserves further research.

VII. Conclusion

This paper aimed to evaluate the value of incorporating on-chain metrics into models for bitcoin price prediction. To the best of our knowledge this is the first such study. This study found that on-chain metrics, when combined with technical indicators, enhance the performance of both BiLSTM and BiRNN models in predicting Bitcoin prices. This suggests that on-chain analysis can provide valuable insights into Bitcoin's valuation. However, the models' improved performance does not guarantee the profitability of a trading strategy based on these predictions. Cryptocurrency markets are highly volatile and influenced by a wide array of factors, many of which are difficult to quantify or predict.

Furthermore, the inherent complexity and uncertainty involved in predicting Bitcoin prices were highlighted by the variability in the models' performance across different runs. This observation underscores the necessity of robust testing methodologies and the importance of cautious interpretation of the results.

The research identified several areas for future investigation. One is the exploration of other on-chain metrics not covered in this study. The new field of on-chain analysis is continually evolving, with researchers and practitioners developing new metrics to glean insights from the blockchain. Incorporating these novel metrics into machine learning models may further improve their performance.

Another potential area for future research is the development of models that can account for the non-stationarity and high noise-to-signal ratio typically observed in financial time series. Such models could potentially offer more accurate predictions of Bitcoin prices.

Finally, the transition from academically interesting prediction models to practically profitable trading strategies remains a substantial challenge. Future research could focus on addressing this challenge by incorporating real-world considerations such as transaction costs, risk management, and the timing of trades into the models.

The findings of this paper underscore the compelling potential of on-chain analysis as a tool for predicting Bitcoin prices, while also highlighting the challenges inherent to this task. As this new field matures, it is reasonable to anticipate enhancements in the accuracy and dependability of Bitcoin price prediction models in the foreseeable future.

References

[1] T. Guo, A. Bifet, and N. Antulov-Fantulin, "Bitcoin volatility forecasting with a glimpse into buy and sell orders," in 2018 IEEE International Conference on Data Mining (ICDM), pp. 989-994, IEEE, 2018. [Online]. Available: https://doi.org/10.1109/ICDM.2018.00123

[2] D. Koutmos, "Bitcoin returns and transaction activity," Economics Letters, vol. 167, pp. 81–85, 2018. [Online]. Available: https://doi.org/10.1016/j.econlet.2018.03.021

[3] L. Hau, H. Zhu, M. Shahbaz, and W. Sun, "Does transaction activity predict Bitcoin returns? Evidence from quantile-on-quantile analysis," The North American Journal of Economics and Finance, vol. 55, p. 101297, 2021. [Online]. Available: https://doi.org/10.1016/j.najef.2020.101297

[4] C. G. Akcora, A. K. Dey, Y. R. Gel, and M. Kantarcioglu, "Forecasting bitcoin price with graph chainlets", Advances in Knowledge Discovery and Data Mining, Cham: Springer International Publishing, pp. 765–776. 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-93040-4_60

[5] P. V. Rane and S. N. Dhage, "Systematic erudition of bitcoin price prediction using machine learning techniques," in 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), pp. 594-598. IEEE, 2019. [Online]. Available: https://doi.org/10.1109/ICACCS.2019.8728424

[6] N. Gradojevic, D. Kukolj, R. Adcock, and V. Djakovic, "Forecasting Bitcoin with technical analysis: A not-so-random forest?," International Journal of Forecasting 39, no. 1, pp. 1-17, 2023. [Online]. Available: https://doi.org/10.1016/j.ijforecast.2021.08.001

[7] L. Khaidem, S. Saha, and S. R. Dey, "Predicting the direction of stock market prices using random forest," arXiv preprint arXiv:1605.00003, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1605.00003
[8] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," The Review of Financial Studies 33, no. 5, pp. 2223-2273, 2020. [Online]. Available: https://doi.org/10.1093/rfs/hhaa009
[9] S. Inder and S. Sharma, "Predicting the Movement of Cryptocurrency 'Bitcoin' Using Random Forest," in Data Science and Computational Intelligence: Sixteenth International Conference on Information Processing (ICInPro), Bengaluru, India, Oct. 22–24, 2021, pp. 166-180. [Online]. Available: https://doi.org/10.1007/978-3-030-91244-4_14
[10] H. Pabuçcu, S. Ongan, and A. Ongan, "Forecasting the movements of Bitcoin prices: an application of machine learning algorithms," Quantitative Finance and Economics 4, no. 4, pp. 679-692, 2020. [Online]. Available: https://doi.org/10.3934/QFE.2020031
[11] P. Mohanty, D. Patel, P. Patel, and S. Roy, "Predicting fluctuations in cryptocurrencies' price using users' comments and real-time prices," in 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), pp. 477-482. IEEE, 2018. [Online]. Available: https://doi.org/10.1109/ICRITO.2018.8748792
[12] S. A. Basher, and P. Sadorsky, "Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?," Machine Learning with Applications 9, 100355, 2022. [Online]. Available: https://doi.org/10.1016/j.mlwa.2022.100355
[13] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," Journal of Computational and Applied Mathematics 365, 112395, 2020. [Online]. Available: https://doi.org/10.1016/j.cam.2019.112395
[14] P. Jaquart, D. Dann, and C. Weinhardt, "Short-term bitcoin market prediction via machine learning," The journal of finance and data science 7, 45-66, 2021. [Online]. Available: https://doi.org/10.1016/j.jfds.2021.03.001
[15] A. Mittal, V. Dhiman, A. Singh and C. Prakash, "Short-Term Bitcoin Price Fluctuation Prediction Using Social Media and Web Search Data", 2019 Twelfth International Conference on Contemporary Computing (IC3), pp. 1-6, 2019. [Online]. Available: https://doi.org/10.1109/IC3.2019.8844899
[16] "Glassnode Academy," Glassnode. [Online]. Available: https://academy.glassnode.com/. Accessed: Jan 30, 2023
[17] E. S. Pour, H. Jafari, A. Lashgari, E. Rabiee, and A. Ahmadisharaf, "Cryptocurrency price prediction with neural networks of LSTM and Bayesian optimization," European Journal of Business and Management Research 7, no. 2, pp. 20-27, 2022. [Online]. Available: https://doi.org/10.24018/ejbmr.2022.7.2.1307
[18] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach," Neural computing and applications, pp. 1-15, 2020. [Online]. Available: https://doi.org/10.1007/s00521-020-05129-6
[19] A. Misnik, S. Krutalevich, S. Prakapenka, P. Borovykh, and M. Vasiliev, "Neural Network Approximation Precision Change Analysis on Cryptocurrency Price Prediction," in Proceedings of the II International Scientific and Practical Conference, pp. 96–101, 2018
[20] D.-H. Kwon, J.-B. Kim, J.-S. Heo, C.-M. Kim, and Y.-H. Han, "Time series classification of cryptocurrency price trend based on a recurrent LSTM neural network," Journal of Information Processing Systems 15, no. 3, pp. 694-706, 2019. [Online]. Available: https://doi.org/10.3745/JIPS.03.0120
[21] M. T. Kurbucz, "Predicting the price of bitcoin by the most frequent edges of its transaction network," Economics letters 184 p. 108655, 2019. [Online]. Available: https://doi.org/10.1016/j.econlet.2019.108655
[22] D. Kondor, M. Po´sfai, I. Csabai, G. Vattay, "Do the Rich Get Richer? An Empirical Analysis of the Bitcoin Transaction Network", PLoS ONE 9(2): e86197, 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0086197
[23] H. A. Dolatsara, E. Kibis, M. Caglar, S. Simsek, A. Dag, G. A. Dolatsara, and D. Delen, "An interpretable decision-support systems for daily cryptocurrency trading," Expert Systems with Applications 203, p. 117409, 2022. [Online]. Available: https://doi.org/10.1016/j.eswa.2022.117409
[24] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, pp. 281-305, Feb. 2012
[25] J. Bartos, "Does Bitcoin follow the hypothesis of efficient market?," International Journal of Economic Sciences, vol. IV, no. 2, pp. 10-23, 2015. [Online]. Available: 10.20472/ES.2015.4.2.002
[26] S. Yan, "Understanding LSTM and Its Diagrams," ML Review. [Online]. Available: https://blog.mlreview.com/understanding-lstm-and-its-diagrams. March 13, 2016
[27] "Glassnode," Glassnode. [Online]. Available: https://glassnode.com/. Accessed: May 10, 2023
[28] TradingView, "Crypto Total Market Cap, $," TradingView. [Online]. Available: https://www.tradingview.com/symbols/TOTAL/. Accessed: Jul. 24, 2023
[29] TradingView, "Market Cap BTC Dominance, %," TradingView. [Online]. Available: https://www.tradingview.com/symbols/BTC.D/. Accessed: Jul. 24, 2023
[30] PwC, "4th Annual Global Crypto Hedge Fund Report," PricewaterhouseCoopers International Limited, June 2022. [Online]. Available: https://www.pwc.com/gx/en/financial-services/pdf/4th-annual-global-crypto-hedge-fund-report-june-2022.pdf. Accessed: Jul. 24, 2023
[31] R. Hudson and A. Urquhart, "Technical trading and cryptocurrencies," Ann. Oper. Res., vol. 297, pp. 191-220, Feb. 2021. [Online]. Available: https://doi.org/10.1007/s10479-019-03357-1
[32] Z. Hu, Y. Zhao, and M. Khushi, “A Survey of Forex and Stock Price Prediction Using Deep Learning,” Applied System Innovation, vol. 4, no. 1, p. 9, Feb. 2021. [Online]. Available: https://doi.org/10.3390/asi4010009
[33] Y. Guo, "The 7 Steps of Machine Learning," Towards Data Science, 23 November 2017. [Online]. Available: https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e. Accessed: July 24, 2023
[34] S. Hussain, "Bitcoin USD (BTC-USD)(Daily Records of Bitcoin)," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/shaheerhussain/bitcoin-usd-btc-usd. Accessed: May 30, 2023
[35] D. López Padial, "Technical Analysis Library in Python," GitHub, 2023. [Online]. Available: https://github.com/bukosabino/ta. Accessed: May 30, 2023
[36] "The Bitcoin Volatility Index," Buy Bitcoin Worldwide. [Online]. Available: https://buybitcoinworldwide.com/volatility-index/. Accessed: 25-July-2023
[37] M. Yang and J. Wang, "Adaptability of financial time series prediction based on BiLSTM," in Procedia Computer Science, vol. 199, pp. 18-25, 2022. [Online]. Available: https://doi.org/10.1016/j.procs.2022.01.003
[38] F. A. Gers, J. Schmidhuber and F. Cummins, "Learning to forget: Continual prediction with LSTM," Neural Computation, vol. 12, no. 10, pp. 2451-2471, Oct. 2000. [Online]. Available: https://doi.org/10.1162/089976600300015015
[39] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," Neural Computation, vol. 1, no. 2, pp. 270-280, 1989. [Online]. Available: https://doi.org/10.1162/neco.1989.1.2.270
[40] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," in IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673-2681, Nov. 1997. [Online]. Available: https://doi.org/10.1109/78.650093