Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, No. 4, October : 2023

HYBRID TRANSFER LEARNING FOR BITCOIN PRICE PREDICTION: ALEXNET AND LSTM IN HARMONY

Juttu Suresh[1], Hima Bindu Kunchanapalli[2], Ch Swapna[3] Assistant Professor, Department of Information Technology, Malla Reddy Engineering College and Management Sciences, Hyderabad, Telangana.

Abstract
Bitcoin, one of the major cryptocurrencies, presents great opportunities and challenges with its tremendous potential returns accompanying high risks. The high volatility of Bitcoin and the complex factors affecting them make the study of effective price forecasting methods of great practical importance to financial investors and researchers worldwide. In this paper, we propose a novel approach called AlexNet-LSTM, which combines a AlexNet and a Long Short-Term Memory (LSTM) to implement Bitcoin closing price prediction. Specifically, the Multi-scale residual module is based on one-dimensional convolution, which is not only capable of adaptive detecting features of different time scales in multivariate time series, but also enables the fusion of these features. LSTM has the ability to learn long-term dependencies in series, which is widely used in financial time series forecasting. By mixing these two methods, the model is able to obtain highly expressive features and efficiently learn trends and interactions of multivariate time series. In the study, the impact of external factors such as macroeconomic variables and investor attention on the Bitcoin price is considered in addition to the trading information of the Bitcoin market. We performed experiments to predict the daily closing price of Bitcoin (USD), and the experimental results show that AlexNet-LSTM significantly outperforms a variety of other network structures. Furthermore, we conduct additional experiments on two other cryptocurrencies, Ethereum and Litecoin, to further confirm the effectiveness of the AlexNet-LSTM in short-term forecasting for multivariate time series of cryptocurrencies.

Keywords: AlexNet, Long Short-Term Memory, Bitcoin Price

I. Introduction
Bitcoin is the world's first distributed super sovereign digital currency, proposed and established by Satoshi Nakamoto [1] in 2009. It relies on an electronic payment system based on cryptography and P2P (Point to Point) technology. By using encryption algorithms and automatic authentication mechanisms, it is difficult to crack or forge, making its transactions more secure and transparent. Since the birth of Bitcoin, it has quickly gained widespread attention. As a new type of cryptocurrency, Bitcoin conducts 24/7 transactions and is able to exchange for many major currencies at a low cost of foreign exchange. Compared to other traditional financial assets, Bitcoin provides investors a new type of portfolio management. Existing research [2], [3] indicated that Bitcoin has an apparent role in the portfolio management market. Empirical analysis by Anne H et al. [4] also corroborates the investability of Bitcoin. In recent years, its rich portfolio and the potential for high returns have attracted the attention of an increasing number of financial investors. However, the Bitcoin market is highly volatile and subject to frequent speculative bubbles [5], [6], thus its trading risks are enormous. Consequently, as discussed above, the study of effective price forecasting methods is of great practical importance to investors, researchers, and policymakers around the world due to the tremendous opportunities and challenges posed by Bitcoin. In recent years increasingly machine learning methods, especially deep neural networks(DNNs), have been applied to market forecasting in cryptocurrencies. In terms of the Bitcoin market, it is highly volatile and generates a large amount of highly non-linear transaction data. In order to

effectively explore these dynamic data, we need models which are able to analyze the internal interactions and hidden patterns in the data. Several researchers [7], [8] have shown that DNN models are well suited for learning lagged correlations between stepwise trends in large financial time series.
In a literature review [9] of comparative studies between artificial neural networks and traditional statistical models, it was found that in 72% of 96 cases, artificial neural network models were shown to have better predictive performance. Consequently, there is no doubt that due to the highly nonlinear and volatile nature of financial markets, DNN models are increasingly applied in the field of finance, especially for financial time series forecasting. In this paper, a novel DNN model consisting of a Multiscale Residual Convolutional Neural Network with LSTM is proposed for Bitcoin price prediction. In the study, not only the influence of transaction information such as Bitcoin historical price is considered on the closing price of Bitcoin, but also external influences such as macroeconomic factors and investors' attentions are introduced. The proposed model consists of two main parts. First, is the proposed multi-scale residual module based on one-dimensional convolution. It contains an identity mapping and three branching networks, where the size of the convolutional kernel is different for each branching network.
Then, the second part is the LSTM network, which is able to further learn the trends of multivariate time series and the interactions between the series, and output the final predicted values. Different from previous financial multivariate time series forecasting, we construct a multi-scale residual module, which can also be called a three-bypass residual module, in which information from these bypasses can be shared with each other. Moreover, this structure can extract potential features that have a high impact on bitcoin price in multiple time scales and integrate them into highly expressive feature vectors after the concatenate operation. In addition to the Bitcoin price prediction experiments, we also conducted extensive experiments on datasets of two different cryptocurrencies (Litecoin and Ethereum) to confirm the strong ability of the proposed AlexNet-LSTM model for short-term price prediction of cryptocurrencies.
The remainder of the paper is organized as follows. Section II reviews the relevant literature. Section III introduces the methodology, including the proposed model. Section IV presents our experiments, while our empirical results and discussion are shown in Section V. Finally, Section VI is the conclusion of this paper.

II. RELATED LITERATURE
In this section, the related work about price prediction of the cryptocurrency and the basis knowledge to the AlexNet-LSTM are provided.
A. Neural Network Approaches
Salim Lahmiri et al. [10] first applied the deep neural networks (DNNs) to cryptocurrency price prediction, elucidated the short-term predictability of cryptocurrencies, and found that the predictive accuracy of Long Short-term Memory (LSTM) neural networks is higher than that of Generalized Regression Neural Networks (GRNN) [11]. Subsequently, a growing body of research has emerged in this area [12], [13]. LSTM neural networks [14] have been shown to be significantly effective in forecasting bitcoin prices due to their ability to identify long-term dependencies and store both long-term and short-term temporal information. Researchers have obtained better results with the LSTM network, in their works to predict bitcoin prices and compare them with the performance of different models [15], [16]. Further, Convolutional Neural Networks (CNNs) have also been applied to cryptocurrency market forecasting. In a study [17], researchers combined CNN with LSTM neural network for high-frequency market trend prediction for a variety of cryptocurrencies. Their empirical analysis shows that the addition of convolutional layers improves the prediction performance and that the hybrid network structure provides the best prediction results in the experiment. Yan Li et al. [18] propose a hybrid neural network model

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, No. 4, October : 2023

based on CNN and LSTM neural networks, and experimental results show that CNNLSTM hybrid neural network can effectively improve the accuracy of value prediction and direction prediction compared to single-structure neural network. The hybrid neural network model combining CNN with LSTM has also been used by many researchers for time series prediction of financial data such as Gold volatility prediction [19], stock prices, etc. Its excellent performance has also been demonstrated in many other fields [20]–[22].

B. Time Series Forecasting

Time series research has always been an important field of machine learning. By building up neural networks based on deep learning, we are able to extract and exploit the hidden information represented by digital currency raw data of digital currencies in order to make accurate and efficient price predictions [10]. In recent years, researchers have continuously made progress in the field of financial time series forecasting [23]–[25]. Many traditional research approaches focus on learning internal patterns, such as autocorrelation, in a particular time series. However, for reality scenarios, especially in cryptocurrencies, stock markets and other financial domains, in most cases we are dealing with multivariate time series. Changwei Hu et al. [26] developed a deep learning structural time series model to handle correlated multivariate time series inputs. Their model is able to leverage dependencies among multiple correlated time series and extract weighted differencing features for better trend learning. The existence of close correlations among many multivariate time series motivates us to consider not only intra-series pattern learning but also inter-series pattern learning when dealing with such tasks.

C. Transfer learning approaches

As one of the milestones in the evolution of CNN, AlexNet [27] has achieved impressive, recordbreaking performance on many challenging tasks. AlexNet share some similarities with Highway networks, such as residual blocks and shortcut connections. AlexNet simplifies the training of very deep networks by bypassing signals from one layer to the next through identity connections. The basic idea underlying residual learning is the branching of gradient propagation paths. For CNNs, this idea was first introduced in the form of parallel paths in the inception models of [28]. Many research works [29], [30] have exploited the multilevel features in CNNs by skip-connections and found them to be effective for a variety of visual tasks. GoogLeNet [28], [31] proposed an “Inception module” that connects feature mappings generated by filters of different sizes to increase the diversity of feature extraction. Meanwhile, GoogLeNet increased the network width by skipping connections to improve the robustness and expressiveness of the network. Gao Huang et al. [32] proposed the Dense Convolutional Network (DenseNet) to exploit the potential of the network through feature reuse. In contrast to AlexNet, it introduces a direct connection from any layer to all subsequent layers. In addition, DenseNet combines features by concatenating them, rather than combining features through summation. These works show us the utility of skip connections, and DenseNet also shows us how to connect feature maps via concatenation. Our work is partly inspired by these two ideas and explores their application to model building for short-term price forecasting of cryptocurrencies.

III. METHODOLOGIES

In this section, the details of our proposed method AlexNet-LSTM and basic blocks are described.

A. AlexaNet

AlexNet, which was first proposed by Alex Krizhevsky et al. in the 2012 ImageNet Large Scale Visual Recognition Challenge, is a fundamental, simple, and effective CNN architecture, which is mainly composed of cascaded stages, namely, convolution layers, pooling layers, rectified linear unit (ReLU) layers and fully connected layers. Specifically, AlexNet is composed of five convolutional layers, the first layer, the second layer, the third layer and the fourth layer followed by the pooling layer, and the fifth layer followed by three fully-connected layers. For the AlexNet architecture, the convolutional kernels are extracted during the back-propagation optimization procedure by optimizing the whole cost function with the stochastic gradient descent (SGD) algorithm. Generally, the convolutional layers act upon the input feature maps with the sliding convolutional kernels to generate the convolved feature maps, and the pooling layers operate on the convolved feature maps to aggregate the information within the given neighborhood window with a max pooling operation or average pooling operation. The reason why AlexNet is successful can be attributed to some of the practical strategies, for instance, the ReLU non-linearity layer and the dropout regularization technique. The ReLU, as shown in Equation (1), is a half-wave rectifier function, which can significantly accelerate the training phase and prevent overfitting. The dropout technique can be regarded as a kind of regularization by stochastically setting a number of the input neurons or hidden neurons to be zero to reduce the co-adaptations of the neurons, which is usually utilized in the fully connected layers in the AlexNet architecture.

$$f(x) = max(x, 0) \quad (1)$$

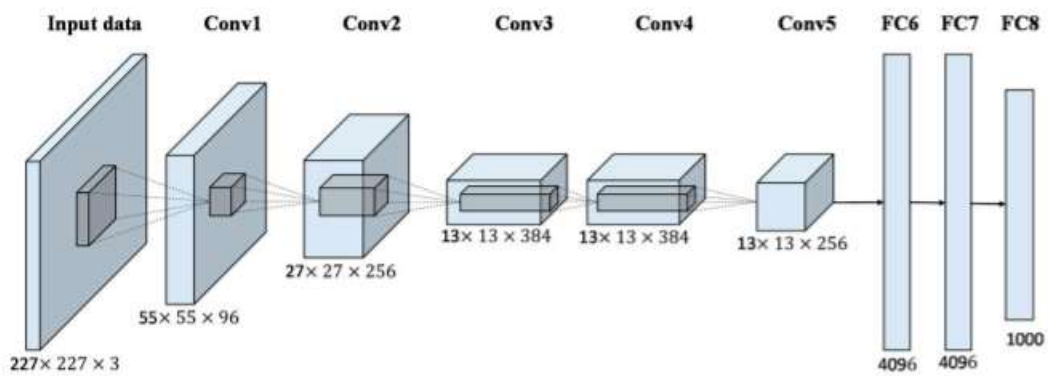


Figure 1. The AlexNet architecture.

The transfer mechanism and the pre-training mechanism allow the CNN network parameters to be transferred from natural data datasets to Bitcoin price data datasets. The reason why this can succeed can be explained, to some extent, by the similarities between natural data datasets and price data datasets, and the category compatibility. It can also be easily understood that the large and complicated ImageNet datasets can help to obtain a well-trained AlexNet architecture, and well-trained network parameters are important for initializing the subsequent classification framework. Therefore, the pre-training mechanism helps the AlexNet architecture to perform the Bitcoin price data scene classification task. Based on the introduction of the convenient and comprehensive representation ability of the pre-trained AlexNet architecture in dealing with Bitcoin price data scene classification, the pre-training mechanism also makes the AlexNet architecture an end-to-end classification pipeline. The pre-trained AlexNet network architecture is shown in Figure 1and each residual unit can be represented in the following Equations 1 and 2.

$$yt = h(xt) + F(xt, wt) \quad (1)$$

$$xt+1 = f(yi) \quad (2)$$

Where F is a residual function, f is a ReLU function, wt is the weight matrix, and xt and yt are the inputs and outputs of the t-th layer. The function h is an identity mapping given by Equation 3:

$$h(xt) = xt \quad (3)$$

In a residual block, skip connections can effectively aggregate historical information, reduce the loss of features and information to some extent, and enable the network to learn richer content. Motivated by the idea of Residual units, our proposed model applies the method of skip connections.

B. LSTM

LSTM is based on the Recurrent Neural Network (RNN) model, which can effectively solve the gradient disappearance and gradient explosion problems in the RNN model. The addition of a special gate control mechanism makes it possible to solve the problem of long-term dependence, and it is suitable for time series processing and prediction, natural language generation [33]–[35]. Fig. 2 gives the basic unit of the LSTM neural network. Its basic unit is the memory block, which contains the memory cell and three gates that control the memory cell, namely, the input gate, the output gate, and the forget gate. Forget Gate is mainly used to calculate the degree of information retention and discarding. The output ft represents the probability of forgetting the state of the underlying cell layer, and is calculated as follows:

$$ft = \sigma(Wfxt + Ufht-1 + bf) \quad (4)$$

In the Equation (4), the xt indicates the input of the current cell, ht−1 presents the output of the previous cell, σ is the sigmoid function. In the input gate, the sigmoid activation function is used to calculate which information needs to be updated. Then, use the tanh activation function to get a vector of candidate values C˜ t, after which updates the previous state Ct−1 to Ct. The formulas are Equations 5, 6 and 7.

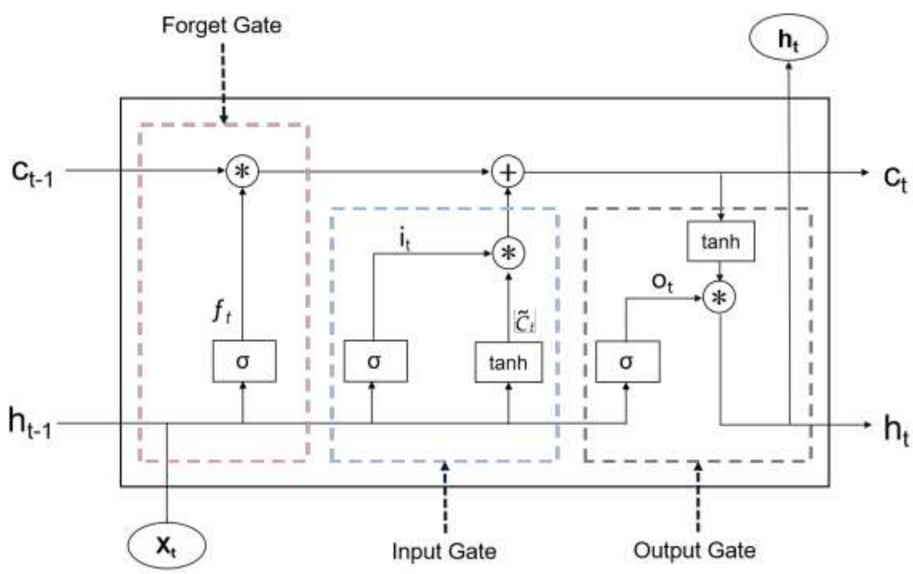


Fig. 2. Basic unit of LSTM network

$$i_t = (W_i x_t + U_i h_{t-1} + b_i) \quad (5)$$

$$\tilde{C}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (6)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (7)$$

The output gate is used to calculate the extent of the information output at the current moment. The information is filtered by the sigmoid activation function to obtain ot. Then the tanh activation function is used to obtain the desired information.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (8)$$

$$h_t = o_t \cdot tanh(C_t) \quad (9)$$

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, No. 4, October : 2023

C. The Proposed Network Architecture.

Inspired by AlexNet, and based on the need to deal with time series problems, we developed a novel building block, Multi-scale Residual Convolutional block (AlexNet) based on one-dimensional time convolution. It is combined with an LSTM neural network to form a new hybrid network architecture (AlexNet-LSTM) to perform cryptocurrency time series prediction. In the following, we will first introduce how to design the multi-scale residual block, followed by the proposed AlexNet-LSTM model, i.e., how the new building block can be combined with LSTM to predict the bitcoin price. The Multi-scale Residual Block. This is the first part of AlexNet-LSTM which is utilized to extract potential features with high expressiveness at different time scales from the dataset. We constructed a three-bypass convolutional layer with convolutional kernels of different sizes, and joined a jump connection to further aggregate historical information efficiently. Due to the high volatility of Bitcoin and the short-term predictability of cryptocurrencies [10], the Bitcoin market is suitable for short-term forecasting, and the longer the forecast period, the worse the forecast performance.

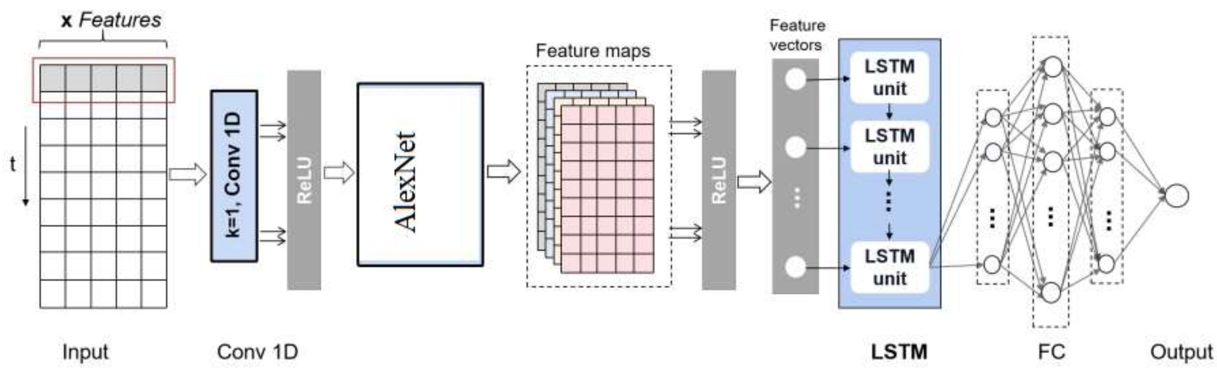


Figure 3. The overall architecture of AlexNet-LSTM model

Therefore, we select 1D convolutional kernels of size 1, 2, and 3 to slide the sequences in the temporal domain, which means that such multi-scale residual module can simultaneously extract the trends and the hidden interactions of the data within 1, 2, and 3 adjacent days in the sequences. In addition, it is known that the window size of the kernel in a one-dimensional convolution will affect the network learning effect. It is likely to miss local feature information when the window size is too large; When the window size is small, local features are easy to extract, but may reduce the correlation of local features. Hence, the "Multi-scale" design of the proposed model can also take into account the advantages of both large and small windows. When it comes to the way to combine the extracted features of three bypasses, here we draw on the idea of DenseNet, which differs from AlexNet in that we never combine features by summing them before they are passed to the next layer; instead, we combine features by concatenating them. We know that the concatenation operation is not viable when the featuremap size changes. Therefore, in order to keep the featuremaps size uniform, we need to perform a zero-padding when performing 1D convolution. Then, the identity mapping is concatenated in the depth direction with the feature maps of the three paths and fed into the subsequent layer. At the end of the module, the use of 1 × 1 2D convolutional kernels enables cross-channel information interaction and expansion. In this way, the feature-maps extracted by the three prediction cycles will be fused and the network will adaptive extract useful information from these hierarchical features. Meanwhile, non-linearity can be added while keeping the feature map scale constant.

Hybrid AlexNet-LSTM model: The network consists of two main parts, the first is the multi-scale residual module for extracting features in the multivariate time series, and the second is the LSTM layer

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, No. 4, October : 2023
for learning pattern changes and predicting prices. Fig. 3 shows the structure of the AlexNet-LSTM neural network. The network contains an input layer, a 1D convolutional layer, the multi-scale residual module, an LSTM layer, a fully connected layer, and an output layer. During the construction of the AlexNet-LSTM, we first use a set of 1D convolutional kernels with a window size of 1 to slide in the time direction for initial feature extraction and feature augmentation of the sequence. Then, the multiscale residual module is utilized to extract and integrate local features across multiple time scales. Finally, an LSTM network is used to learn relationships in time steps for price forecasting.

$$f(x) = \begin{cases} x, (x > 0) \\ 0, (x \leq 0) \end{cases} \qquad (10)$$

We introduce activation functions to incorporate nonlinearities in the network. The common activation functions include Sigmoid, Tanh, ReLU, and SELU. In this paper, the rectified linear unit (ReLU) is selected as the nonlinear activation function, which is expressed as the Equation (10).

IV. EXPERIMENTS

A. Data Collection and pre-processing

The dataset used in this experiment includes the daily closing price of Bitcoin and 10 types of internal (Bitcoin trading datas) and external (macroeconomic variables and investor attention) information that have an impact on the price of Bitcoin from October 25, 2015 to October 17, 2020, for a total of 1820 records. The dataset is divided into 80% as training set and 20% as test set. The training set is further divided into a training set (80%) and a validation set (20%) for evaluating performance and avoiding overfitting.

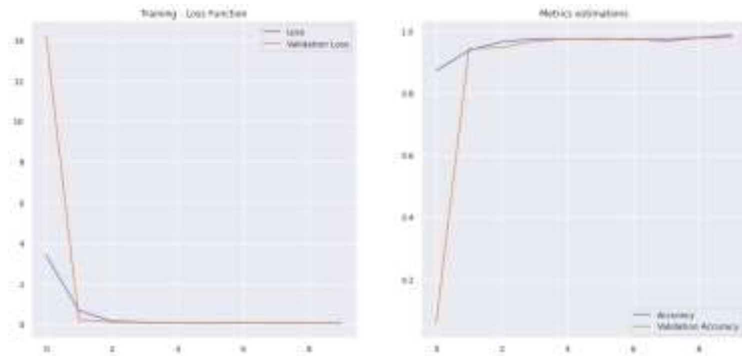
Table I shows all the attributes that compose the dataset. The internal information in the dataset refers to the daily transaction datas of Bitcoin, including the open price, the closing price, the highest price, the lowest price, the weighted price, the trade volume in Bitcoins and the market's currency. They are collected from the offical website 1. As for external factors, we take the macroeconomic factors and investor attention into consideration. As a new type of investment instrument, the price of bitcoin is considered to be related to several macroeconomic variables.

Table 1. List of input attributes

Transaction Information	Macroeconomic Variables	Investor Attention
the Open price	S&P 500 Index	Google Trends
the Close price	GVZ	-
the Highest price	VIX	-
the Lowest price	-	-
the Weighted price	-	-
Volume(BTC)	-	-
Volume(Currency)	-	-

The study by Leon Li considered four representative volatility indices published by the Chicago Board Options Exchange(CBOE), including the Volatility Index(VIX) and the Gold Price Volatility Index(GVZ). He believed that volatility indices can be used to speculate and trade on market sentiment regarding future volatility [36]. Some researchers have found that there is a relationship between the price of bitcoin and gold, crude oil, and stock market indices [37]. Therefore, this paper introduces macroeconomic variables that have been proven to have predictive power for the bitcoin market to predict bitcoin prices. Ultimately, the following macroeconomic factors were chosen for the experiment: S&P 500 Index, GVZ, VIX. They are all daily data from the Wind Database.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, No. 4, October : 2023



(a) (b)

Figure 4: Loss Function and Accuracies of Training set and validation set

The loss function for both learning and validation datasets are represented in graph in Figure 4(a). Here the loss values for learning and validation datasets are obtained as 0.051 and 0.044 respectively. The prediction rate of the CNN classifier is 98.68% for the train set and 98% for the validation set. Whereas the Figure 4(b) presents the training and validation datasets accuracies in which number of epochs are considered in X-axis and percentage of accuracy is represented on Y-axis.

Table 2. Performance comparison of various approaches.

Method	Accuracy	F1-score	Precision	Specificity	Sensitivity
SVM [14]	85.633	91.695	85.145	85.870	88.844
PNN [17]	92.437	91.968	86.606	88.622	89.866
LSTM [18]	92.878	95.548	89.174	93.606	94.027
RNN [20]	96.466	96.173	93.935	96.607	97.217
Proposed AlexNet-LSTM	98.584	97.368	99.674	97.380	99.875

Table 1 shows that the proposed AlexNet-LSTM method resulted in superior performance than the conventional methods like SVM [14], PNN [17], LSTM [18], and RNN [20]. For all the performance metrics, the proposed approach resulted in superior performance.

V. CONCLUSION

In this study, a hybrid method consisting of a multi-scale residual block and an LSTM network is proposed to predict the bitcoin price. Specifically, multi-scale residual block in this hybrid model is able to extract rich features at different time scales and also strengthen the representational ability of these 4https://www.quandl.com/ features. Besides, the utilization of local residual learning leads to a reduction in computational complexity as well as improves the performance of the DNN. Different from some traditional studies, we introduce external influences such as macroeconomic variables and investor attention when performing price forecasting to combine various factors that may affect bitcoin prices. Furthermore, sufficient experimental results confirm that our proposed model has better prediction results than other single-structured models. The efficient feature extraction and integration capability of the proposed residual block is also demonstrated in the experimental comparison. In addition, we perform price prediction with the AlexNet-LSTM for two other major cryptocurrencies, ETH and LTC, further demonstrating the effectiveness of our model for multivariate time series forecasting of cryptocurrencies. In summary, compared with several state-of-art works, the proposed DNN model achieves better prediction performance, but it also has some limitations, since the Bitcoin market is particularly sensitive to some national policies, regulatory and market events, etc. More future work could be focused on comprehensive metrics which measure the investor's attention to more timely detection of bitcoin market volatility and thus more accurate price prediction.

REFERENCES

[1] S. Nakamoto and A. Bitcoin, "A peer-to-peer electronic cash system," Bitcoin.–URL: https://bitcoin. org/bitcoin. pdf, vol. 4, 2008.
[2] A. Dyhrberg, "Bitcoin, gold and the dollar – a garch volatility analysis," Finance Research Letters, vol. 16, 11 2015.
[3] A. H. Dyhrberg, "Hedging capabilities of bitcoin. is it the virtual gold?" Finance Research Letters, vol. 16, pp. 139 – 144, 2016.
[4] A. H. Dyhrberg, S. Foley, and J. Svec, "How investible is bitcoin? analyzing the liquidity and transaction costs of bitcoin markets," Economics Letters, vol. 171, pp. 140 – 143, 2018.
[5] J. E.-T. Cheah and J. Fry, "Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin," Economics Letters, vol. 130, 02 2015.
[6] S. Corbet, B. Lucey, and L. Yarovaya, "Datestamping the bitcoin and ethereum bubbles," Finance Research Letters, vol. 26, pp. 81 – 88, 2018.
[7] R. Adcock and N. Gradojevic, "Non-fundamental, non-parametric bitcoin forecasting," Physica A: Statistical Mechanics and its Applications, vol. 531, p. 121727, 2019.
[8] M. Nakano, A. Takahashi, and S. Takahashi, "Bitcoin technical trading with artificial neural network," Physica A: Statistical Mechanics and its Applications, vol. 510, pp. 587 – 609, 2018.
[9] M. Khashei and M. Bijari, "An artificial neural network (p,d,q) model for timeseries forecasting," Expert Systems with Applications, vol. 37, no. 1, pp. 479 – 489, 2010.
[10] S. Lahmiri and S. Bekiros, "Cryptocurrency forecasting with deep learning chaotic neural networks," Chaos, Solitons & Fractals, vol. 118, pp. 35–40, 2019.
[11] D. F. Specht et al., "A general regression neural network," IEEE transactions on neural networks, vol. 2, no. 6, pp. 568–576, 1991.
[12] H. Liang, W. Lei, P. Y. Chan, Z. Yang, M. Sun, and T.- S. Chua, "Pirhdy: Learning pitch-, rhythm-, and dynamics-aware embeddings for symbolic music," Proceedings of the 28th ACM International Conference on Multimedia, Oct 2020. [Online]. Available: http://dx.doi.org/10.1145/3394171.3414032
[13] F. Zhu, W. Lei, C. Wang, J. Zheng, S. Poria, and T.-S. Chua, "Retrieving and reading: A comprehensive survey on open-domain question answering," 2021.
[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[15] N. Uras, L. Marchesi, M. Marchesi, and R. Tonelli, "Forecasting bitcoin closing price series using linear regression and neural networks models," arXiv preprint arXiv:2001.01127, 2020.
[16] P. Linardatos and S. Kotsiantis, "Bitcoin price prediction combining data and text mining," in Advances in Integrations of Intelligent Methods. Springer, 2020, pp. 49–63.
[17] S. Alonso-Monsalve, A. L. Suarez-Cetrulo, A. Cervantes, and D. Quin- ´ tana, "Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators," Expert Systems with Applications, vol. 149, p. 113250, 2020.
[18] Y. Li and W. Dai, "Bitcoin price forecasting method based on cnn-lstm hybrid neural network model," The Journal of Engineering, vol. 2020, 01 2020.
[19] A. Vidal and W. Kristjanpoller, "Gold volatility prediction using a cnnlstm approach," Expert Systems with Applications, vol. 157, p. 113481, 2020.
[20] C. Tian, J. Ma, C. Zhang, and P. Zhan, "A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network," Energies, vol. 11, no. 12, p. 3493, 2018.
[21] W. Lei, X. Wang, M. Liu, I. Ilievski, X. He, and M.-Y. Kan, "Swim: A simple word interaction model for implicit discourse relation recognition," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 4026–4032. [Online]. Available: https://doi.org/10.24963/ijcai.2017/562

[22] W. Lei, X. Jin, M.-Y. Kan, Z. Ren, X. He, and D. Yin, "Sequicity: Simplifying task-oriented dialogue systems with single sequence-to-sequence architectures," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 1437–1447. [Online]. Available: https://www.aclweb.org/anthology/P18-1133
[23] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the arima model," in 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. IEEE, 2014, pp. 106–112.
[24] R. Adhikari and R. Agrawal, "A combination of artificial neural network and random walk models for financial time series forecasting," Neural Computing and Applications, vol. 24, no. 6, pp. 1441–1449, 2014.
[25] J. Cao, Z. Li, and J. Li, "Financial time series forecasting model based on ceemdan and lstm," Physica A: Statistical Mechanics and its Applications, vol. 519, pp. 127–139, 2019.
[26] C. Hu, Y. Hu, and S. Seo, "A deep structural model for analyzing correlated multivariate time series," in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). IEEE, 2019, pp. 69–74.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
[29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[30] S. Yang and D. Ramanan, "Multi-scale recognition with dag-cnns," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1215–1223.
[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.
[32] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
[33] W. Lei, X. Jin, M.-Y. Kan, Z. Ren, X. He, and D. Yin, "Sequicity: Simplifying task-oriented dialogue systems with single sequence-tosequence architectures," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 1437–1447.
[34] X. Jin, W. Lei, Z. Ren, H. Chen, S. Liang, Y. Zhao, and D. Yin, "Explicit state tracking with semi-supervisionfor neural dialogue generation," in Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 1403–1412.
[35] L. Pan, W. Lei, T.-S. Chua, and M.-Y. Kan, "Recent advances in neural question generation," arXiv preprint arXiv:1905.08949, 2019.
[36] L. Li, "Risk of investing in volatility products: A regime-switching approach," Investment Analysts Journal, pp. 1–16, 2020.
[37] S. Vassiliadis, P. Papadopoulos, M. Rangoussi, T. Konieczny, and J. Gralewski, "Bitcoin value analysis based on cross-correlations," Journal of Internet Banking and Commerce, vol. 22, no. S7, p. 1, 2017.
[38] A. Urquhart, "What causes the attention of bitcoin?" Economics Letters, vol. 166, pp. 40–44, 2018.
[39] A. Yelowitz and M. Wilson, "Characteristics of bitcoin users: an analysis of google search data," Applied Economics Letters, vol. 22, no. 13, pp. 1030–1036, 2015.
[40] Z. Da, J. Engelberg, and P. Gao, "In search of attention," The Journal of Finance, vol. 66, no. 5, pp. 1461–1499, 2011.