A Data Driven Machine Learning Model for Predicting Crypto Currency Prices

ANKITA SONI[1], DR. SACHIN PATEL[2]
[1, 2]SAGE University, Indore

***Abstract*— Data driven machine learning models have been extensively used for forecasting crypto prices, which is typically challenging due to the random and volatile nature of crypto prices. In addition, businesses that rely on crypto price predictions for financial research and investments stand to lose a lot of money if even a little estimate is off. In recent times, methods based on artificial intelligence and machine learning have gained popularity for predicting cryptocurrency prices. This is because these methods outperform traditional statistical methods. To forecast cryptocurrency prices, the suggested study uses a scaled-back propagation technique based on steepest descent in conjunction with data pre-processing utilizing the discrete wavelet transform (DWT). It has been demonstrated that the suggested method achieves a lower mean square percentage error than the previously used method.**

***Index Terms*- Data Analytics, Crypto Currencies, Bitcoin Prices, Deep Neural Networks, Forecasting Error.**

I. INTRODUCTION

Cryptos' prominence has grown in tandem with the digitization and dissemination of resources. Cryptocurrencies like Bitcoin, Ethereum, etc., have seen massive investment as a result [1]. Nevertheless, investments in cryptocurrency are fraught with risk due to their highly unpredictable, fluctuating, and volatile pricing. In addition, historical crypto data frequently displays noise in the form of random swings, volatility, and trend deviations [2]. Due to the difficulty in pattern detection caused by this chaotic activity, inaccurate forecasts are produced. Therefore, before submitting the time series crypto data to any machine learning or deep learning model for pattern detection, it is required to filter off the baseline noise [3]. Crypto trend analysis is a time series regression problem, but it's made even more complicated by the fact that it depends on a lot of non-numerical factors, like a country's political climate, the stability of its government, the severity of financial crises and trade wars, the rate of global economic downturn, public opinion about a company, and so on. This causes stock trends to fluctuate, which can show patterns that don't match up with past data [4].

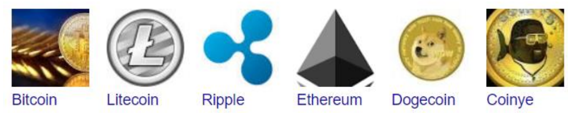

Fig.1 Common Crypto Currencies

Global events, such as natural disasters, geopolitical tensions, or pandemics, can also impact crypto trends [5]. These events can create uncertainty in the markets and cause investors to become more risk-averse and hence it is essential to note that data trends are inherently variable and can be influenced by a wide range of factors [6]. It's also worth mentioning that past performance does not guarantee future results, so investors should exercise caution when making investment decisions based on historical trends [7].. Cryptocurrency prediction is basically a time series prediction problem. Mathematically:

$$P = f(t, v) \qquad (1)$$

Here,
P represents crypto price
f represents a function of
t is the time variable
v are other influencing global variables

The dependence of crypto process over time makes it somewhat predictable under similar other conditions of global influencing variables. However, even the slightest of changes can derail the prediction completely [8].

Statistical techniques are not found to be as accurate as the contemporary artificial intelligence and machine learning based approaches [9]. In this paper, a back propagation based scaled conjugate gradient algorithm is used in conjugation with the discrete wavelet transform (DWT) for forecasting crypto price trends. The evaluation of the proposed approach has been done based on the mean absolute percentage error (MAPE). A comparative MAPE analysis has also been done w.r.t. previously existing techniques [10].

II. DEEPNETS

Deep learning has evolved as one of the most effective machine learning techniques which has the capability to handle extremely large and complex datasets [11]. It is training neural networks which have multiple hidden layers as compared to the single hidden layer neural network architectures [12].

The architectural view of a deep neural network is shown in figure 1. In this case, the outputs of each individual hidden layer is fed as the input to the subsequent hidden layer. The weight adaptation however can follow the training rule decided for the neural architecture. There are various configurations of hidden layers which can be the feed forward, recurrent or back propagation etc [13].

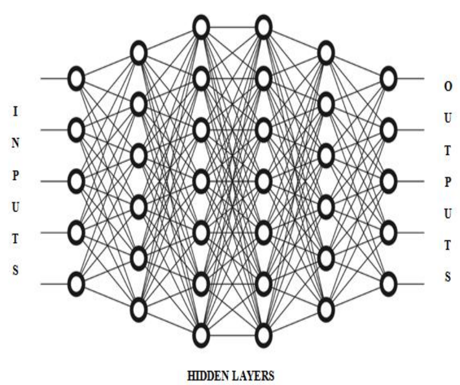

Fig.2 The Deep Neural Network Architecture

The figure above depicts the deep neural network architecture with multiple hidden layers. The output of the neural network however follows the following ANN rule:

$$Y = f(\sum_{i=1}^{n} X_i . W_i + \theta_i) \qquad (2)$$

Where,
X are the inputs
Y is the output
W are the weights
ϴ is the bias.

Training of ANN is of major importance before it can be used to predict the outcome of the data inputs.

III. BACK PROPAGATION

Back propagation is one of the most effective ways to implement the deep neural networks with the following conditions [14]:

1) Time series behavior of the data
2) Multi-variate data sets
3) Highly uncorrelated nature of input vectors

The essence of the back propagation based approach is the fact that the errors of each iteration is fed as the input to the next iteration. [15]. The error feedback mechanism generally is well suited to time series problems in which the dependent variable is primarily a function of time along with associated variables. Mathematically,

$$Y = f(t, V_1 \ldots . V_n) \qquad (3)$$

Here,
Y is the dependent variable
f stands for a function of
t is the time metric
V are the associated variables
n is the number of variables

The back propagation based approach can be illustrated graphically in figure 2.

In case of back propagation, the weights of a subsequent iteration doesn't only depend on the conditions of that iteration but also on the weights and errors of the previous iteration mathematically given by [16]:

$$W_{k+1} = f(W_k, e_k, V) \qquad (4)$$

Here,
W_{k+1} are the weights of a subsequent iteration
W_k are the weights of the present iteration
e_k is the present iteration error
V is the set of associated variables

In general, back propagation is able to minimize errors faster than feed forward networks, however at the cost of computational complexity at times. However, the trade off between the computational complexity and the performance can be clearly justified for large, complex and uncorrelated datasets for cloud data sets [17].

IV. GRADIENT DESCENT BASED TRAINING

The gradient descent algorithms (GDAs) generally exhibit:

1) Relatively lesser memory requirement
2) Relatively faster convergence rate

The essence of this approach is the updating of the gradient vector g, in such as way that it reduces the errors with respect to weights in the fastest manner. Mathematically, let the gradient be represented by g and the descent search vector by p, then [18]:

$$p_0 = -g_0 \qquad (5)$$

Where,

g_0 denotes the gradient given by $\frac{\partial e}{\partial w}$

The sub-script 0 represents the starting iteration

The negative sign indicates a reduction in the errors w.r.t. weights [15].

The tradeoff between the speed and accuracy is clearly given by the following relations [16]:

$$W_{k+1} = W_k - \alpha g_x\ , \ \ \alpha = \frac{1}{\mu} \qquad (6)$$

Here,

w_{k+1} is the weight of the next iteration

w_k is the weight of the present iteration

g_x is the gradient vector

μ is the step size for weight adjustment in each iteration.

There are several ways to implement the back propagation technique in the neural networks [19]. One consideration however always remains that of the least time and space complexity so as to reduce the amount of computational cost that is associated with the training algorithm. The essence of the scaled conjugate gradient algorithm is the fact that it has very low space and time complexity making it ideally suited to large data sets to be analyzed in real time applications where the time is a constraint. The training rule for the algorithm is given by:

$$A_0 = -g_0 \qquad (7)$$

A is the initial search vector for steepest gradient search

g is the actual gradient

$$w_{k+1} = w_k + \mu_k g_k \qquad (8)$$

Here,

w_{k+1} is the weight of the next iteration

w_k is the weight of the present iteration

μ_k is the combination co-efficient

V. THE DISCRETE WAVELET TRANSFORM

The wavelet transform is an effective tool for removal of local disturbances. Crypto prices show extremely random behavior and local disturbances. Hence conventional Fourier methods do not render good results for highly fluctuating data sets. Mathematically, the wavelet transform can be given as [20]

$$Z\,(S, P) = \int_{-\infty}^{\infty} z(\mathrm{t})\ ((S, P, t))dt \qquad (9)$$

Here,

S denotes the scaling operation

P denotes the shifting operation

t denotes the time variable

Z is the image in transform domain

z is the image in the spatial domain

The major advantage of the wavelet transform is the fact that it is capable of handling fluctuating natured data and also local disturbances. The DWT can be defined as [21]:

$$W\Phi\ (\mathrm{Jo, k}) = \frac{1}{\sqrt{M}} \sum_n S(n).\,\Phi(n)_{jo\prime k} \qquad (10)$$

The data is divided in the ration of 70:30 for training and testing data set bifurcation.

The final performance metrics computed for system evaluation are:

1) Mean Absolute Percentage Error (MAPE)

$$MAPE\ = \frac{100}{M} \sum_{t=1}^{N} \frac{E - E_t|}{E_t} \qquad (11)$$

Here E_t and $E_t^{\sim}$ stand for the predicted and actual values respectively.

The number of predicted samples is indicated by M.

2) Regression

The extent of similarity between two variables is given by the regression.

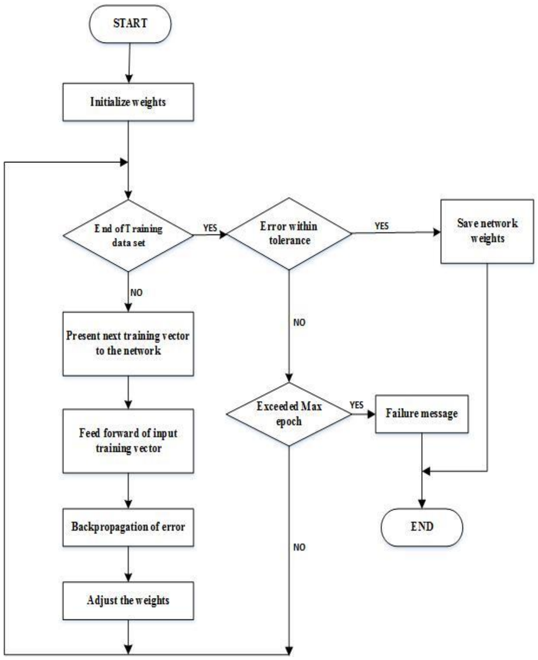

Fig.3 Flowchart of Back Propagation

VI. RESULTS

The simulations results have been presented next. The dataset has beem obtained from Kaggle for the Bitcoin prices.

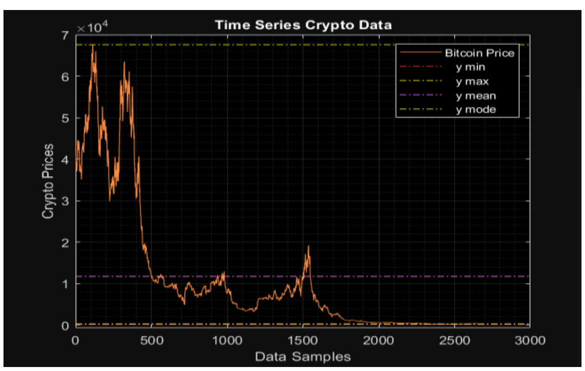

Fig.4 Original Bitcoin Price

The figure above depicts variation in Bitcoin price.

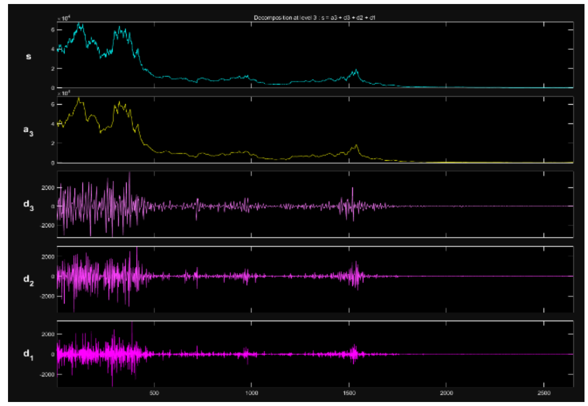

Fig.5 Symlet Decomposition at Level 3

The figure above depicts the Symlet decomposition in terms of approximate and detailed co-efficients for the data.

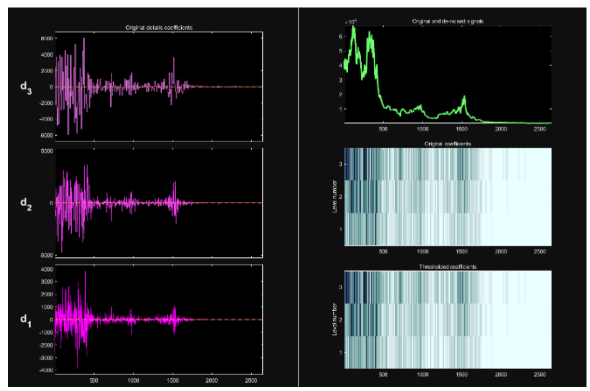

Fig.6 Denoising data using Symlet

The figure above depicts the denoising process using Symlet.

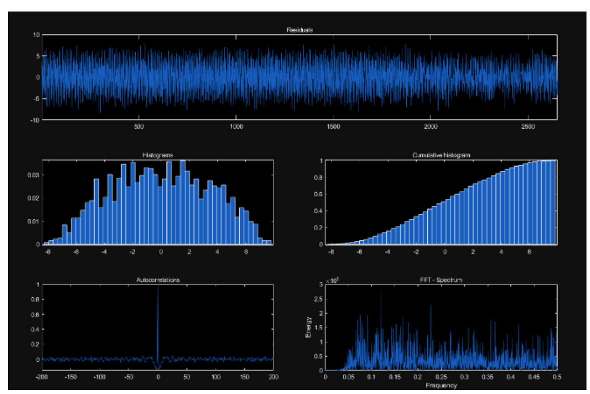

Fig.7 Multi-Resolution Analysis

The figure above depicts the multi resolution analysis of noise baseline.

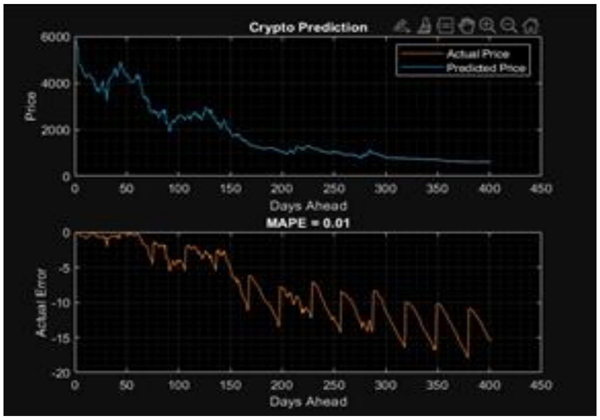

Fig.8 Predicted and Actual Crypto Behavior

The figure above depicts the predicted and actual crypto behavior.

From the above figures, it can be concluded that the proposed system attains the following results:
1) Iterations:1000
2) MSE=0.114
3) MAPE of Proposed work=0.01%
4) MAPE of Previous work [1]=0.06%

CONCLUSION

Previous observations indicate that crypto price prediction falls under the category of time series prediction that is highly dependent and sensitive to outside forces. As a result, getting really accurate predictions is generally a challenge. The proposed model employs a model of a deep neural network that makes use of back propagation. To train the network how to function, we employ the adaptive gradient descent algorithm, or GDA. In order to prepare the data for analysis, the discrete wavelet transform is used. Regarding the accuracy for the benchmark datasets utilized, the suggested method also surpasses prior systems, as the proposed model attains an MAPE of just 0.01%.

REFERENCES

[1] M. Rafi, Q. A. K. Mirza, M. I. Sohail, M. Aliasghar, A. Aziz and S. Hameed, "Enhancing Cryptocurrency Price Forecasting Accuracy: A Feature Selection and Weighting Approach With Bi-Directional LSTM and Trend-Preserving Model Bias Correction," in IEEE Access, vol. 11, pp. 65700-65710, 2023

[2] G. Kim, D. -H. Shin, J. G. Choi and S. Lim, "A Deep Learning-Based Cryptocurrency Price Prediction Model That Uses On-Chain Data," in IEEE Access, vol. 10, pp. 56232-56248, 2022.

[3] Z. Shahbazi and Y. -C. Byun, "Improving the Cryptocurrency Price Prediction Performance Based on Reinforcement Learning," in IEEE Access, vol. 9, pp. 162651-162659, 2021

[4] M. Ertz and É. Boily, ‘‘The rise of the digital economy: Thoughts on blockchain technology and cryptocurrencies for the collaborative economy,’’ Int. J. Innov. Stud., vol. 3, no. 4, pp. 84–93, Dec. 2019.

[5] V. Buterin. (2013). Ethereum White Paper: A Next Generation Smart Contract & Decentralized Application Platform. [Online]. Available: https://github.com/ethereum/wiki/wiki/White-Paper

[6] D. Vujicic, D. Jagodic, and S. Randic, ‘‘Blockchain technology, bitcoin, and ethereum: A brief overview,’’ in Proc. 17th Int. Symp. INFOTEHJAHORINA (INFOTEH), Mar. 2018, pp. 1–6.

[7] M. Mudassir et al., ‘‘Time-series forecasting of Bitcoin prices using highdimensional features: A machine learning approach,’’ Neural Comput. Appl., 2020.

[8] S. Asante Gyamerah, ‘‘Are bitcoins price predictable? Evidence from machine learning techniques using technical indicators,’’ 2019, arXiv:1909.01268.

[9] J.-Z. Huang, W. Huang, and J. Ni, ‘‘Predicting bitcoin returns using highdimensional technical indicators,’’ J. Finance Data Sci., vol. 5, no. 3, pp. 140–155, Sep. 2019.

[10] R. Adcock and N. Gradojevic, ‘‘Non-fundamental, non-parametric bitcoin forecasting,’’ Phys. A, Stat. Mech. Appl., vol. 531, Oct. 2019, Art. no. 121727.

[11] D. Philippas, H. Rjiba, K. Guesmi, and S. Goutte, ‘‘Media attention and bitcoin prices,’’ Finance Res. Lett., vol. 30, pp. 37–43, Sep. 2019.

[12] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, ‘‘Cryptocurrency price prediction using tweet volumes and sentiment analysis,’’ SMU Data Sci. Rev., vol. 1, no. 3, p. 1, 2018.

[13] D. Shen, A. Urquhart, and P. Wang, ‘‘Does Twitter predict bitcoin?’’ Econ. Lett., vol. 174, pp. 118–122, Jan. 2019.

[14] N. Aslanidis, A. F. Bariviera, and Ó. G. López, ‘‘The link between cryptocurrencies and Google trends attention,’’ Finance Res. Lett., vol. 47, Jun. 2022, Art. no. 102654, doi: 10.1016/j.frl.2021.102654.

[15] Y. B. Kim, J. G. Kim, W. Kim, J. H. Im, T. H. Kim, S. J. Kang, and C. H. Kim, ‘‘Predicting fluctuations in cryptocurrency transactions based on user comments and replies,’’ PLoS ONE, vol. 11, no. 8, Aug. 2016, Art. no. e0161197.

[16] D. Liang, F. Ma and W. Li, "New Gradient-Weighted Adaptive Gradient Methods With Dynamic Constraints," in IEEE Access, 2020, vol. 8, pp. 110929-110942.

[17] X. Huang, J. Guan, B. Zhang, S. Qi, X. Wang and Q. Liao, "Differentially Private Convolutional Neural Networks with Adaptive Gradient Descent," 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC), Hangzhou, China, 2019, pp. 642-648.

[18] AC Wilson, R Roelofs, M Stern, “The marginal value of adaptive gradient methods in machine learning”, Proceedings in Advances in Neural Information Processing Systems 30 (NIPS 2017).

[19] Y Xue, Y Wang, J Liang, “A self-adaptive gradient descent search algorithm for fully-connected neural networks”, Neurocomputing, Elsevier 2022. Vol.478, pp. 70-80.

[20] I Khandelwal, R Adhikari, G Verma, “Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition”, Procedia in Computer Science, vol.48, pp. 173-179.

[21] A Butorova, E Baglaeva, I Subbotina, “Application of the Wavelet Data Transformation for the Time Series Forecasting by the Artificial Neural Network”, Proceedings in New Trends in the Applications of Differential Equations in Sciences, Springer 2022, pp 365–370.