

Evaluation of Weight Decay Regularization Techniques for Stock Price Prediction using Gated Recurrent Unit Network

Arjun Singh Saud^{1*} and Subarna Shakya²

¹Central Department of Computer Science and IT, Tribhuvan University, Kirtipur, Kathmandu, Nepal

²Department of Electronics and Computer Engineering, Pulchowk Campus, Tribhuvan University, Lalitpur, Nepal

*CORRESPONDENCE:

Arjun Singh Saud

Central Department of Computer Science and IT, Tribhuvan University, Kirtipur, Kathmandu, Nepal
Email: arjunsaud@cdcsit.edu.np

ISSN : 2382-5359(Online),
1994-1412(Print)

DOI:

<https://doi.org/10.3126/njst.v20i1.39379>

ACCESS THE ARTICLE ONLINE



CONFLICT OF INTEREST: None

Copyright: The Author(s) 2020. This is an open access article under the [CC BY](#) license.



ABSTRACT

Stock price forecasting in the field of interest for many stock investors to earn more profit from stock trading. Nowadays, machine learning researchers are also involved in this research field so that fast, accurate and automatic stock price forecasting can be achieved. This research paper evaluated GRU network's performance with weight decay regularization techniques for predicting price of stocks listed NEPSE. Three weight decay regularization technique analyzed in this research work were (1) L1 regularization (2) L2 regularization and (3) L1_L2 regularization. In this research work, six randomly selected stocks from NEPSE were experimented. From the experimental results, we observed that L2 regularization could outperform L1 and L1_L2 regularization techniques for all six stocks. The average MSE obtained with L2 regularization was 4.12% to 33.52% lower than the average MSE obtained with L1 regularization, and it was 10.92% to 37.1% lower than the average MSE obtained with L1_L2 regularization. Thus, we concluded that the L2 regularization is best choice among weight regularization for stock price prediction.

Keywords: L1 and L2 regularization, stock market forecasting, weight regularization.

1. INTRODUCTION

Forecasting of stock price trends is one of the hot topics of research specifically among researchers from finance, statistics, and computer science (Yu and Yan 2020). Stock traders have a keen interest in this type of research and research results because they want to buy stocks that tend to increment shortly and want to sell stocks that don't tend to increment (Murkute & Sarode 2015). Fundamental analysis and technical analysis are two widely used approaches by researchers from the financial domain (Vijh *et al.* 2020). On the other hand, time series prediction tools and technique are popular among researchers from statistics and computer science (J *et al.* 2018).

As concluded by the random walk hypothesis (Fama 1965) and the efficient market hypothesis (Fama 1970), it is impossible to accurately predict stock prices. It is because stock market trading data is noisy, and it is a function comprising of many known and unknown parameters. Known parameters include past financial performance, past dividend history, growth, different financial ratio etc. and unknown parameters include political changes, rumours, future financial performance, growth potential etc. However, many research works have successfully predicted stock prices with adequate performance measures. Research in stock market prediction is mainly focused in following three dimensions: (1) prediction of stock prices (Nikou *et al.* 2019; Shen & Shafiq 2020) (2) prediction of the index (Moghaddam *et al.* 2016; Sakarya *et al.* 2015) and (3) and prediction of stock trading signals (Arevalo *et al.* 2016; Zhang *et al.* 2017).

Although technical analysis is a popular tool among researchers from the financial domain, it is challenging to analyze the large volume of stock trading data and identify patterns. Therefore, machine learning models became popular among researchers for predicting trends of the stock price. Artificial neural network (ANN) and deep learning models are recently widely used by researchers to predict stock price trends (Lv *et al.* 2019). These models have shown improved performance in stock price prediction. One of the main problems associated with ANN and machine learning model is over fitting. Regularization is one of the techniques to cope with the over fitting problem. Weight decay regularization is one of the widely used classes of regularization techniques (Good fellow *et al.* 2016). L1 and L2 regularization techniques are the two regularization techniques that belong to this class. This research paper analyzed gated recurrent unit (GRU) network performance with different weight decay regularization techniques for predicting stock prices and identified the best weight decay regularization technique.

This paper is organized as follows. Related theories and models used for performing this research work and methodology adapted to carry out this research work are described in Section 2. Experimental results and its analysis are presented in Section 3. Finally, Section 4 presents the conclusion of the research work and the recommendations.

2. MATERIALS AND METHODS

Related theories and models and methodology adapted to carry out this research work are described in this section.

2.1. Related Theories and Models

Two major theories and models used in this research work are gated recurrent unit (GRU) network, weigh decay regularization techniques.

2.1.1 Gated Recurrent Unit (GRU) Network

Recurrent neural networks (RNNs) suffer from the vanishing/exploding gradient descent problem. Long short-term memory (LSTM) network and GRU network are variants of RNN that can cope with this type of problem. The concept behind GRU and LSTM network is similar. However, the structure of the GRU network is simple (Cho *et al.* 2014). It uses only two gates and fewer parameters than LSTM. Therefore, it can be trained faster than LSTM as well as it can be generalized with fewer data. Weiss *et al.* (2018) argued that LSTM has higher expressive power than GRU and yields better results when working with vast amounts of data. But GRU performs well for a moderate amount of data. The network’s mathematical formulation is provided in Equations 1 to 4, and its architecture is provided in Figure 1 (Cho *et al.* 2014).

$$Z_t = \sigma(W_z x_t + U_z H_{t-1}) \tag{1}$$

where, z is update gate

$$R_t = \sigma(W_r x_t + U_r H_{t-1}) \tag{2}$$

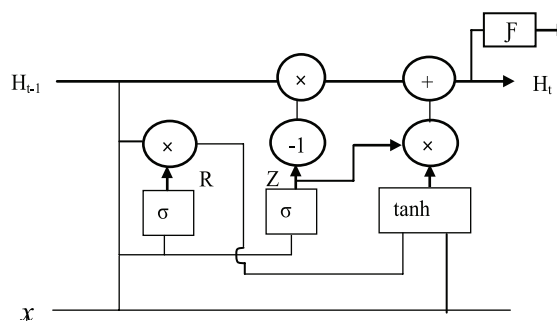
where, R is reset gates

$$H'_t = \tanh(W_h x_t + (R_t \times H_{t-1}) U_h) \tag{3}$$

where, H'_t ia candiadte hidden state

$$H_t = (Z_t \times H'_t) + ((1 - Z_t) \times H_{t-1}) \tag{4}$$

where, H_t is and hidden state



2.1.2 Weight Decay Regularization

Generalization is the primary objective of any machine learning model. Over fitting is the problem that prevents the

model from generalization. Regularization is the technique that helps models to generalize well by preventing over fitting. Its main objective is to reduce the variance of the model without substantial increase of bias. Weight decay regularization is the technique that penalizes the neural network weights by adding a term to its loss function. The main objective of weight decay regularization is to encourage small weights because large weights are a symbol of over fitting and make neural network unstable (Good fellow *et al.* 2016). Three approaches used for weight decay regularization are (1) L1 regularization (2) L2 regularization and (3) L1_L2 regularization.

L1 regularization adds the sum of the absolute value of weights as a penalty term to the loss function, and L2 regularization adds the sum of squares of weights as a penalty term to the loss function. L1_L2 regularization combines both regularization techniques and adds both terms as penalty term to the loss function. Loss functions used for L1 and L2 regularization is shown in Equations 5 and 6, respectively.

$$L(x) = L(x) + \lambda \sum |w| \tag{5}$$

$$L(x) = L(x) + \lambda \sum w^2 \tag{6}$$

where λ is the hyperparameter whose value ranges between 0 to 1

2.2 Methodology

This sub-section describes historical stock trading data, data preprocessing, performance measures, data preparation, and the GRU network configuration used in the research work.

2.2.1 Stock Data

In this study, historical trading data of stocks listed in Nepal Stock Exchange (NEPSE) were used. The trading data of NEPSE stocks were collected from Nepal Stock Exchange (Nepal Stock Exchange Ltd *n.d.*). Six NEPSE stocks were selected randomly. Stocks representing different sectors were selected for the experiment. The data for NEPSE stocks was daily data from March 2010 to October 2020. The six NEPSE stocks used in this research work were Sanima Bank Limited (SANIMA), Citizens Bank Limited (CZBIL), Asian Life Insurance (ALICL), Shikhar Insurance Limited (SICL), Chilime Hydropower (CHCL), and Butwal Power Company (BPCL).

2.2.2 Data Preprocessing

First, the data was arranged in the order of oldest to newest

date and the unnecessary features like stock symbol, amount, number of transactions etc. were dropped. The target attribute ‘Next Day’s Close Price’ was then generated, which is merely close price shifted back by 1 position. Finally, the input features were normalized using standard scalar.

2.2.3 Performance Measure

Mean squared error (MSE) was used for measuring the performance of stock price prediction. MSE is the average of the squares of errors. It can be calculated using Equation (7).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2 \tag{7}$$

where y is vector of observed values and p is the vector of predicted values

2.2.4 Data Preparation and Output

The data were divided into training, validation, and test sets in 8:1:1 ratio. This study aimed to predict stocks close price for the day $t+1$ using input features from $(t-N+1)^{th}$ day to t^{th} day, where t is the current trading day, and N is the window size. Window size 5 was used in this research work. Therefore, the input data used in this research work was a combination of N independent variables $d_{t-N+1} \dots d_{t-1}, d_t$ and a dependent variable c_{t+1} , where d_t is a tuple (v_t, h_t, l_t, c_t) . The symbols $v_t, h_t, l_t,$ and $c_t,$ denotes volume, high price, low price, and close price respectively for the i^{th} trading day. Finally, the predicted price was reverse scaled, the MSE between observed values and predicted price was calculated, and the result was plotted.

2.2.5 GRU Configuration

Configuration of the GRU network used in this research work was $4 \times 64 \times 64 \times 64 \times 64 \times 1$. Hyperbolic tangent activation function was used in each hidden layer, and the linear activation function was used in the output layer of the GRU network. The network was trained using a mini-batch gradient descent algorithm with batch size 32 and Adam optimizer.

3. RESULTS AND DISCUSSION

In this research, experiments were conducted for each of the six stocks, using L1, L2, and L1_L2 regularization. Each experiment was executed 10 times, and MSE was captured for each execution. Further, the average MSE was calculated for each regularization technique.

3.1 RESULTS

Table 1: MSE's Obtained for SANIMA, CZBIL, and ALICL

Execution#	SANIMA-MSE			CZBIL-MSE			ALICL-MSE		
	L1	L2	L1-L2	L1	L2	L1-L2	L1	L2	L1-L2
1	147.64	126.08	152.22	39.24	29.27	38.89	903.46	956.93	1261.86
2	141.78	100.85	161.20	38.42	30.71	45.87	1075.87	892.65	974.94
3	146.29	117.64	161.45	40.48	29.66	43.88	997.36	948.78	1054.83
4	143.26	115.69	165.19	37.44	30.80	43.74	929.36	974.31	959.10
5	144.51	125.47	158.00	38.68	32.74	40.97	1172.01	985.39	1027.72
6	154.31	110.85	166.07	38.62	32.04	42.24	1090.89	1067.29	1061.39
7	157.19	107.43	150.64	45.24	27.78	42.58	932.18	877.44	1209.22
8	148.95	106.43	153.32	39.56	27.79	43.27	901.39	863.71	1138.37
9	151.80	114.07	156.53	40.41	31.07	43.40	980.12	949.13	964.87
10	142.79	119.00	163.12	41.00	30.96	41.03	932.07	989.79	1019.11

Table 2: MSE's Obtained for SICL, CHCL, and BPCL

Execution#	SICL-MSE			CHCL-MSE			BPCL-MSE		
	L1	L2	L1-L2	L1	L2	L1-L2	L1	L2	L1-L2
1	4338.44	3778.21	4405.85	429.97	367.59	423.38	146.73	93.08	154.47
2	4266.57	3692.90	4777.00	462.15	324.43	460.14	146.89	96.56	156.98
3	4259.59	3496.01	4369.73	422.55	356.02	424.98	142.88	95.25	160.73
4	4517.52	3845.10	4707.12	403.95	372.76	437.40	141.51	107.40	156.37
5	4759.49	3509.73	4489.09	433.45	330.09	453.72	143.76	95.68	160.51
6	4188.14	3696.49	4400.53	434.97	344.00	460.41	143.74	103.29	164.15
7	4851.22	3499.54	4688.48	405.65	320.43	421.75	165.79	92.67	162.41
8	4732.71	4414.97	4959.40	446.90	317.47	445.40	150.03	107.56	161.72
9	4542.39	3607.00	4663.82	420.84	327.70	447.02	161.22	110.83	161.58
10	4108.35	3819.72	5246.92	414.09	317.49	521.37	159.66	96.32	148.75

Table 3: Average MSE for Each Stock

Stock Symbol	MSE		
	L1	L2	L1_L2
SANIMA	147.85	114.35	158.77
CZBIL	39.91	30.28	42.59
SICL	4456.44	3735.96	4671.79
ALICL	991.47	950.54	1067.14
CHCL	427.45	337.8	449.56
BPCL	150.22	99.86	158.77

3.2 Result Analysis and Discussion

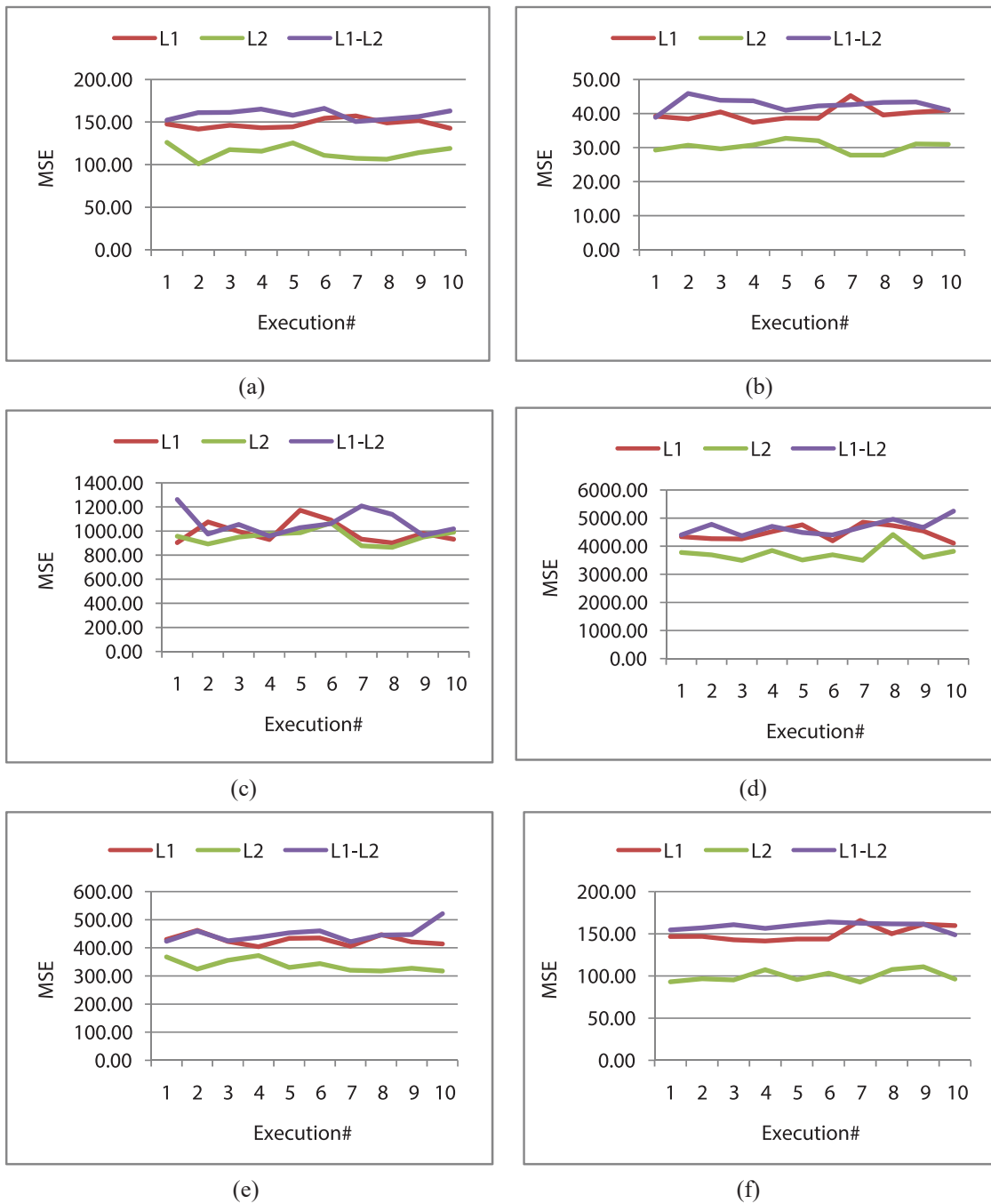


Fig.2. MSE's of (a) SANIMA (b) CZBIL (c) ALICL (d) SICL (e) CHCL (f) BPCL

If we look at the graphs of Figure 2, we can see that, for all stocks, the lowest MSE is obtained for stock price predicted using GRU network with L2 regularization. Whereas the highest MSE is obtained for stock price predicted using GRU network with L1_L2 regularization. If we look at the graphs of Figures 2(a), 2(b), 2(e), and 2(f), we can see that L1_l2 regularization performed worst for SANIMA, CZBIL, CHCL, and BPCL. However, if we observe the graphs of Figure 2(c) and 2(d), the worst performer is not clear for ALICL and SICL. Therefore, to draw a concise picture of the MSE, the average of MSE's was computed, and the result is plotted in Figure 3.

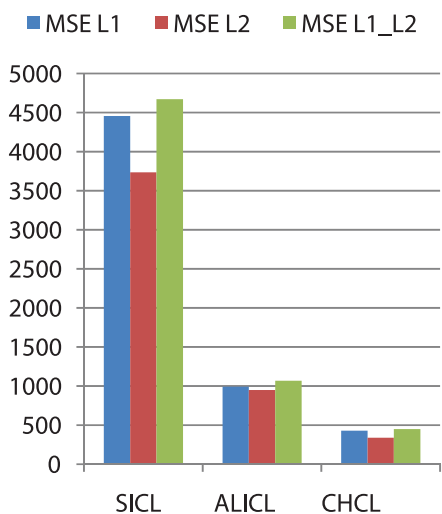
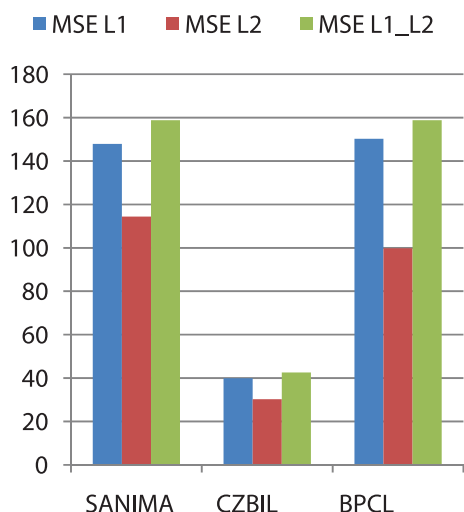


Fig.3. Average MSE of Stocks

If we look at Figure 3, we can observe that average MSE obtained with L2 regularization is lowest for all stocks whereas MSE obtained with L1_L2 regularization is highest for all stocks. The average MSE obtained with L2 regularization was 4.12% to 33.52% lower than the average MSE obtained with L1 regularization, and it was 10.92% to 37.1% lower than the average MSE obtained with L1_L2 regularization. Further, the average MSE obtained with L1 regularization was 4.83% to 7.38% lower than the average MSE obtained with L1_L2 regularization.

4. CONCLUSION

This research work evaluated the impact of weight decay regularization techniques in predicting the next day’s stock price. We randomly selected six stocks listed in NEPSE and predicted their price using GRU network with various weight decay regularization techniques.

From the experimental results, we observed that the L2 regularization technique outperformed L1 and L1_L2 regularization in terms of MSE of predicted stock prices. The average MSE obtained with L2 regularization was 4.12% to 33.52% lower than the average MSE obtained with L1 regularization, and it was 10.92% to 37.1% lower than the average MSE obtained with L1_L2 regularization. Thus, we concluded that L2 regularization is the best weight decay regularization technique used with machine learning models for predicting the next day’s stock price.

There are many ways in which this research can be extended. This research work has only experimented stocks listed in NESPE. Therefore, replicating the research work with stocks listed in other exchanges will be an excellent idea to enhance the reliability of the results obtained in this research work. Further, regularization hyper parameter of L2 regularization can be analyzed for finding its best value so that stock price can be predicted with lower MSE.

ACKNOWLEDGEMENT

The authors are incredibly thankful to the University Grant Commission (UGC), Nepal and Central Department of Computer Science and Information Technology, TU, Nepal for their support to this research work. Besides, we are thankful to the Nepal Stock Exchange (NEPSE) and Central Depository Service and Clearing (CDSC) Ltd for providing support and cooperation during the data collection required for this research work.

REFERENCES

1. Arevalo, A., J. Nini, G. Hernandez and J. Sandoval. 2016. *High-frequency trading strategy based on deep neural networks. Lecture Notes in Computer Science* **9773**:424–436.
2. Cho, K., B.V. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Pages 1724-1734 in *A. Moschitti, B. Pang, W. Daelemans*, editors. *Proceedings, Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 25-29 October 2010, Qatar. ACL, [Pennsylvania, United States](#).
3. Fama, E. F. 1965. Random Walks in Stock Market Prices. *Financial Analysts Journal* **51**:75–80.
4. Fama, E. F. 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* **25**:383–417.
5. Goodfellow, I., Y. Bengio and A. Courville. 2016.

- Deep Learning* (Adaptive Computation and Machine Learning series). First edition. The MIT Press, Massachusetts, USA.
6. J, K., I. Sengupta and S. Chaudhury. 2018. Stock Market Prediction Using Time Series Analysis. Id 3168423 in P. Venkata Krishna, and M. Usha Rani, editors. Proceedings, *IADS International Conference on Computing, Communications & Data Engineering (CCODE)*, 7-8 February 2018, Tirupati, India. SSRN: Elsevier, Amsterdam, Netherland.
 7. Lv, D., S. Yuan, M. Li and Y. Xiang. 2019. An Empirical Study of Machine Learning Algorithms for Stock Daily Trading Strategy. *Mathematical Problems in Engineering*, 7816154.
 8. Moghaddam, A.H., M.H. Moghaddam and M. Esfandyari. 2016. Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, **21**: 89–93.
 9. Murkute, A. and T. Sarode. 2015. Forecasting Market Price of Stock using Artificial Neural Network. *International Journal of Computer Applications*, **124**:11–15.
 10. Nepal Stock Exchange Ltd. (n.d.). Retrieved June 12, 2020, from <http://nepalstock.com.np/stockWisePrices>
 11. Nikou, M., G. Mansourfar and J. Bagherzadeh. 2019. Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, **26**: 164–174.
 12. Sakarya, Á., M. Yavuz, A.D. Karaoazlan and N. Ozdemir. 2015. Stock Market Index Prediction with Neural Network during Financial Crises: A Review on Best-100. *Financial Risk and Management Reviews*, **1**:53–67.
 13. Saud, A.S. and S. Shakya. 2020. Analysis of lookback period for stock price prediction with RNN variants: A case study on the banking sector of NEPSE. *Procedia Computer Science*, **167**: 788–798.
 14. Shen, J. and M.O. Shafiq. 2020. Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*, **7**: 66.
 15. Vijh, M., D. Chandola, V.A. Tikkiwal and A. Kumar. 2020. Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*, **167**: 599–606.
 16. Weiss, G., Y. Goldberg and E. Yahav. 2018. On the practical computational power of finite precision RNNs for language recognition. Pages 740-745 in [I. Gurevych](#), and [Y. Miyao](#), editors. Proceedings, 56th Annual Meeting of the Association for Computational Linguistics, 20-25 July 2020, Melbourne, Australia. ACL, [Pennsylvania, United States](#).
 17. Yu, P. and X. Yan. 2020. Stock price prediction based on deep neural networks. *Neural Computing and Applications*, **32**:1609–1628.
 18. Zhang, L., C. Aggarwal and G.J. Qi. 2017. Stock price prediction via discovering multi-frequency trading patterns. Pages 2141-2149 in V. Dhar, J. Tang, and J. Furnkranz, editors. Proceeding, *23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 15-17 August 2017, Halifax, Canada. ACM, New York, United States.