HED › Hradec Economic Days

Forecasting of FOREX Price Trend Using Recurrent Neural Network - Long Short-term Memory

Michal DOBROVOLNY [1],*, Ivan SOUKAL [1], Kok Cheng LIM [2] , Ali SELAMAT [3] and Ondrej KREJCAR [1]

[1] University of Hradec Kralove, Hradec Kralove, Czech Republic; michal.dobrovolny@uhk.cz; ivan.soukal@uhk.cz; ondrej.krejcar@uhk.cz

[2] Universiti Teknologi Malaysia, Johor Bahru, Malaysia; kokcheng@uniten.edu.my

[3] Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia; aselamat@utm.my

* Correspondence: michal.dobrovolny@uhk.cz

Abstract: Algorithms of neural networks (NN) can search and represent both structured and not structured data, we employ then on financial time-series. This paper describes the use of Long short-term memory (LSTM) for FOREX pair EUR/USD price prediction. Aim of the paper is to test and proposes the best time block to predict based on a daily FOREX data. We employ the mean of absolute errors and the least mean squared errors to assess prediction results in order to find the time block. We tested time blocks from ten to fifty-eight days and 100 or 300 epochs. Training dataset contained daily exchange rate data from 1.4.1971 until 9.5.2019. The best performing network has been trained for 30-day period and 100 epochs. This paper also describes the effect of training for a high number of epochs.

Keywords: neural networks; FOREX; forecasting; long short-term memory; time block

JEL Classification: C45; C53; F31

1. Introduction

Forecasting is one of the essential tasks that humans are trying to achieve. That is why economic time-series were the aim of prediction for a very long time. In the late 90' financial time series, such as stock market and Foreign Exchange (FOREX), were described as a phenomenon with nearly like a random-walk process behavior, with statistical properties different at different points in time as the process is time-varying, making the prediction almost impossible (Hellstrom and Holmstrom 1998). Twenty-one years later is our set of tools wider and computational power incomparable. With their use we try algorithms to learn from repetitions and patterns to predict the next period. The methods are moving forward very quickly yet long-term or extremely short-term remains as a significant challenge and mainly intraday trading needs to be explored (Pradeepkumar and Ravi 2018).

Algorithms of neural networks (NN) can search and represent both structured and not structured data — for instance, natural language processing, time series or image data (Abdel-Nasser and Mahmoud 2019; Pena-Barragan et al. 2011). In image data processing can be found examples about fixing an image (Wolterink et al. 2017; Yang et al. 2018), compression (Sun et al. 2020), super-resolution (Ledig et al. 2017) or image classification (Krizhevsky et al. 2012). Algorithms of NN more often occur in the first position in many competitions. Image processing and time series are a very developed field. Some of the examples are image understanding, fixing old image damage, styling image and fake image generating.

Researchers try to improve their models with many variations. For instance, Francesco Rudo added a new layer of reinforcement learning to his model. With this new model, he got Return of Investment 98.23 % and with a reduction of drawdown 15.97%. (Rundo 2019)

One of the main problems is the time block used for a prediction. This is the aim of our research. We perform a mid-term FOREX analysis and test a different time block in order to find the least mean of absolute errors (MAE) and the least mean squared errors (MSE). Time block of a prediction is crucial to prediction accuracy. With an extended period, we get too far into the future and lost accuracy. With

a short time block, we also get very inaccurate results. It is caused by missing information about the previous trend.

The paper is divided into several parts. In the first chapter, we describe the connection between FOREX and neural networks. In the second chapter, we explain RNN and LSTM network architecture. Also, we describe the data we used and the methodology of processing. The third chapter describes the result of time series prediction we have achieved.

2. Methodology

2.1. Topic Overview

Regarding the Web of Science database, the topic of FOREX and forecast a query "FOREX AND forecasting" shows an increasing trend for the last decade. Since 2015 there is almost three times the number of published articles, see chart below. About 30% is related to computer science.
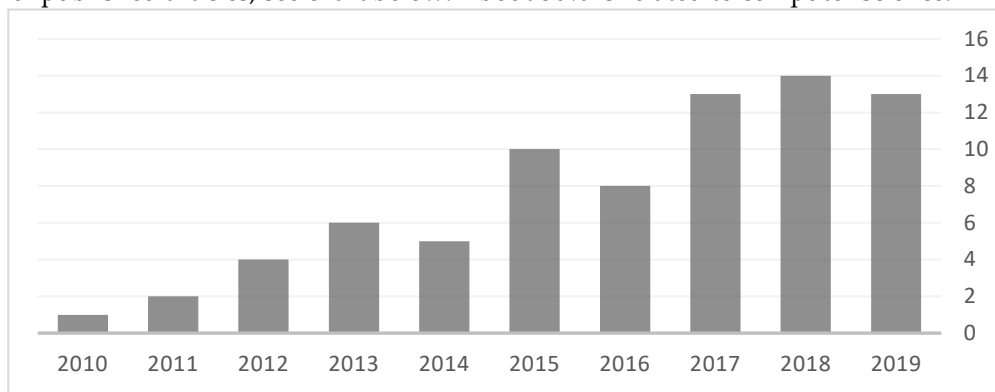

Figure 1. Yearly count of articles focused on FOREX.

Keyword analysis in **Figure 2** shows the connection between FOREX and keywords such as forecasting, machine learning, neural networks, prediction in a total of 76 keywords. The minimum occurrence of keyword was 5 times. The amount of all used keywords is 1,687 from 423 articles. This analysis suggested suitability of employing recurrent neural networks (RNN) for FOREX value prediction. We also found out there are not many research papers using long short-term memory (LSTM).
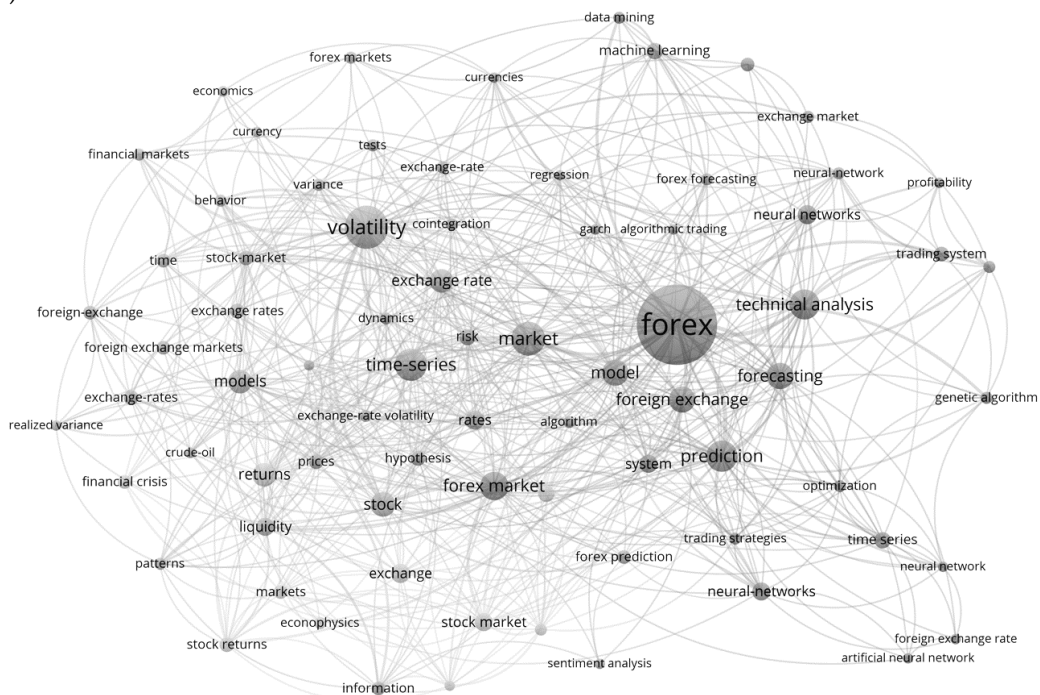

Figure 2. Keyword head map of a search for "FOREX" word.

2.2. Recurrent Neural Networks

Humans do not start their thinking from scratch every second. As we do something, we understand and we build our experience on previous knowledge. We do not throw everything away and start thinking from scratch again. During the reading, we understand every word in the context of other words.

Traditional neural networks don't have a memory. For instance, imagine your network is trying to solve a classification task. To be correct classify every frame of a movie. Every point of the timeline has to be described. It is impossible to understand a movie in context if the network does not have any information about previous events in the movie.

RNN is a network with architecture which address this issue. RNN contains a loop to allow the network to persist information. A loop allows information to be passed from one step to the next one. (Hochreiter and Schmidhuber 1997)

This loop makes a recurrent neural network less clear for understanding. For better understanding, we can imagine recurrent networks as multiple copies of the same network. Each new one is created from the previous network. Consider it as a loop.

This brings us to the general use of RNN. They are high with time series. The incredible success of RNN applied to a variety of problems: speech recognition, language modeling, translation, image captions, traffic predictions and our primary target price predictions. (Abdel-Nasser and Mahmoud 2019; Carapuço et al. 2018; Sidehabi et al. 2016; Zhao et al. 2017)

A special kind of RNN is LSTM. They are capable of learning long-term dependencies and were introduced in 1997. LSTM has been proven to be high performing on various problems. (Hochreiter and Schmidhuber 1997)

LSTM is designed to avoid the long-term dependency problem. They keep information for long periods. As we described in the previous part, all RNN have some form of a loop like repeating modules. LSTM also has a loop like a repeating module. The architecture of the LSTM cell is illustrated in Figure 3.
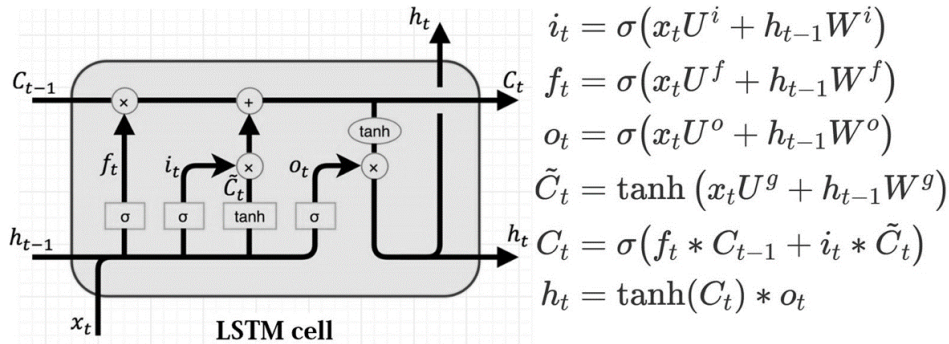

$$i_t = \sigma\big(x_t U^i + h_{t-1} W^i\big)$$
$$f_t = \sigma\big(x_t U^f + h_{t-1} W^f\big)$$
$$o_t = \sigma\big(x_t U^o + h_{t-1} W^o\big)$$
$$\tilde{C}_t = \tanh\big(x_t U^g + h_{t-1} W^g\big)$$
$$C_t = \sigma\big(f_t * C_{t-1} + i_t * \tilde{C}_t\big)$$
$$h_t = \tanh(C_t) * o_t$$

Figure 3. An illustration of LSTM cell with memory pipe. (Varsamopoulos et al., 2020)

The key part in a LSTM cell is C line. C represents a memory pipeline. LSTM has an ability to add or remove information to the memory regulated by gates. Gates x and + are a way to let information thought optionally. The σ layer outputs a real number between zero and one. This value represents the impact of information. Zero is for no impact one and one is for very important one.

2.3. Dataset, HW and Data Treatment

FOREX currency pairs are divided into three groups - major, minor and exotic. (Broto 2013; Laherrere and Sornette 1998) Major currencies are the most used. For instance, EUR/USD (our choice) or GBP/USD and others. Minor currencies are less traded and have lower liquidity - for instance, Norwegian krone. For different pairs a different behavior can be observed. Hence, similarly to a problem of various industry evaluation a different approach has to be taken in order to assess the development (Hedvicakova and Kral 2019), i.e. one strategy won't fit for all and for each pair the processing will results in a different NN settings. Regarding the settings, one of advantages is that FOREX is one of the most available sources of data which provides enough training data. It is one of

the reasons why we choose FOREX for our time series forecasting. Specifically, a public dataset from kaggle.com with the only date and close value of EUR/USD price. (Mahesh 2019)

For better results, we are using for pre-processing MinMaxScaler from library Sklearn. This feature scaler gets minimal value and maximal value. This value then recalculates all dataset values to fit the given range in the constructor.

Feature scaling is a technique to standardize the independent features in a fixed range value. It is done to reduce network size. For instance, with the adequately distributed value, we can use a small network. We used feature scaling for the price value. That means our results are also scaled values. As for post-processing, we also used the same instance of feature scaler on predicted values.

To reach our goal, we divided the input dataset into slices in a range from 10 to 60 days. This period is a shape for training and also for later predictions.

The trained model is pure LSTM architecture from 1997. (Hochreiter and Schmidhuber 1997) We are using the LSTM layer provided by Keras framework. Our architecture uses three output units connected into a fully connected Dense layer. We used mean squared error (MSE) loss function with RMSprop optimizer.

LMST input has as many dimensions as time block used for a particular model. Generally, we are using a structure with shape (N, 1).

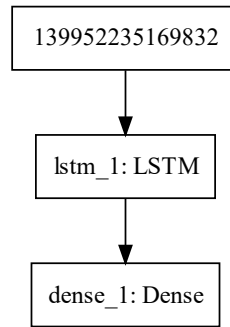Our model is drawn by Keras utility method mode_to_dot in Figure 4.

Figure 4. Model architecture with one LSTM and one Dense layer.

LSTM, RNN and NN generally are very performance demanding. Our computation computer has two dedicated cards with a total of 7,934 CUDA cores. These cards are one of the top-performing gaming cards of current NVIDIA cards. Because of frameworks support, we decided to use NVIDIA cards only. One of our cards is 1080TI with 11,176 MB graphic memory and 1,607 MHz max clock rate. Another one is 2080TI with 11,019 MB of graphics memory and 1,545 max clock rate. Used processor is i7-8700 with 3.20 GHz clock.

We used as a programming language Python in version 3.7.3. For programing in Python, we used a web-based interactive computational environment Jupyter Notebook. Keras is our main used framework. Keras is a top-level framework with high abstraction on top of Tensorflow. In new versions of Tensorflow (2.+) is Keras integrated.

Prediction validation was calculated by MSE and mean absolute error (MAE). Both methods are commonly used for neural network loss calculation. In our case, used loss function was MSE.

$$e_t = y_t - \hat{y}_t \tag{1}$$

where y is the actual value and $\hat{y}$ is predicted value.

For validation of results, we used MSE

$$MSE = {\sum_{t=1}^{N} e_t^2}/{N} \tag{2}$$

where e_t is the difference between the predicted and the the actual value. N is a number of observations.

Mean absolute error is a measure of the difference between two continuous variables.

$$MAE = \left.\sum_{t=1}^{N} |e_t|\right/N \tag{3}$$

where N is the number of observations.

3. Results

FOREX price forecasting is an exciting field of study. Our very simple model with only one LSTM layer shows excellent performance on this dataset. We proved the correlation between input time block and prediction accuracy. The best performing model has an MSE value 0.003052 and MAE 0.002390.
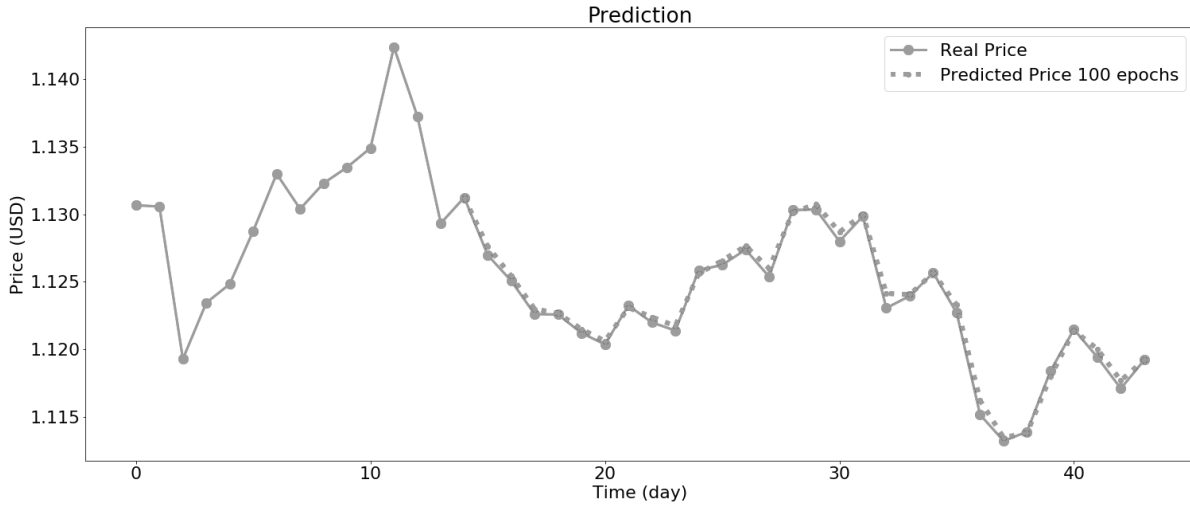

Figure 5. Best performing LSTM network on 30 days period trained for 100 epochs.

Impact of time block input on LSRM forecast accuracy can be seen in Table 1. Some time blocks are really close to the best performing 30 days period. An exciting result has a 26 days result. This block shows one of the worst accuracies. The worst block is 38 days, with 100 epochs.

We also test the impact of training length i.e. an impact of epochs counts on the accuracy. In the case of three times more epochs, we got very different results. For instance, the best performing 30-day period has 59 % worse accuracy. For a difference between 100 and 300 epoch training length see fig below.
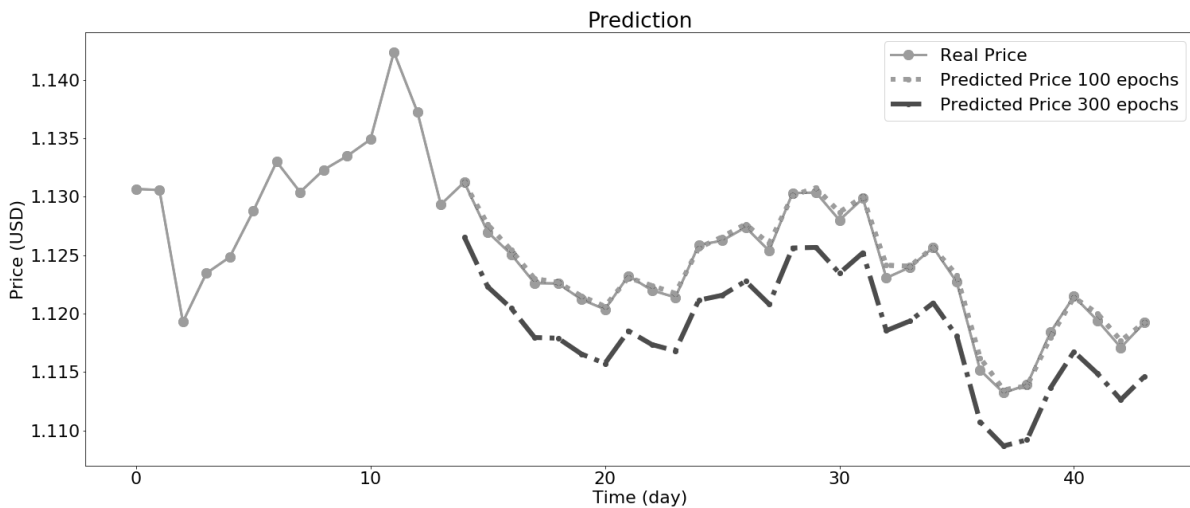

Figure 6. Comparison of 100 and 300 epoch training length.

Table 1. Table 1 represents results of calculated errors MSE and MAE. Best performing combination is bold and second is italic. The best performing network is a combination of 30 days

period and 100 epochs of training. It is essential to know that these results are directly dependent on our architecture.

Table 1. Results table of MSE and MAE value for every tested network.

Time period in days	Number of epochs	MSE	MAE
10	100	0.003427	0.002776
14	100	0.004097	0.003015
18	100	0.004671	0.004050
22	100	0.004666	0.004056
26	100	0.005178	0.004641
30	**100**	**0.003052**	**0.002390**
32	100	0.004152	0.003531
34	100	0.003548	0.002768
36	100	0.003377	0.002634
38	100	0.006203	0.005553
40	100	0.003491	0.002788
42	100	0.004744	0.003843
44	100	0.003873	0.003068
46	100	0.003696	0.002895
48	100	0.003987	0.002976
50	100	0.004229	0.003231
52	100	0.004821	0.003797
54	100	0.003890	0.002917
56	100	0.006394	0.005796
58	100	0.003421	0.002556
10	300	0.003536	0.002752
14	300	0.003961	0.002930
18	300	0.005398	0.004468
22	300	0.005036	0.004475
26	300	0.003479	0.002725
30	300	0.005140	0.004565
34	300	0.004205	0.003436
38	300	0.005318	0.004628
42	300	0.004561	0.003661
46	300	0.004879	0.003844
50	300	0.004976	0.004276
54	300	0.003485	0.002605
58	300	0.005875	0.004888

Full length of training our 33 variations was around 27 hours. We used the method of parallel training to decrease the amount of final time.

4. Discussion and Conclusion

Although not many papers described RNN and LSTM employment for the FOREX prediction, a heading of our research can be compared to results of e.g. (Sermpinis, Dunis et al. 2012) who chose the same pair EUR/USD pair and also a daily data. A larger set of methods was employed where RNN was compared to e.g. results of autocorrelation-based models but with mixed results. When RNN was compared to a hybrid model of adaptive particle swarm optimization and radial basis function in another study by (Sermpinis, Theofilatos et al. 2012), the hybrid model outperformed it by 9% in annualized return and needed only a half of positions taken. Although a trading pair was EUR/GBP. Daily time frame data were employed also by (Bagheri et al. 2014; Khashei et al. 2008) although they employed a hybrid model-based algorithms unlike us.

Study (Persio and Honchar 2016) employed the same approach as we when utilizing RNN, LSTM, and adding Multi-layer Perceptron, and the Convolutional Neural Networks (CNN) for a purpose of FOREX analysis (EUR/USD pair). They applied it also on the S&P500 index. Unlike our approach their focus was on intraday trading with a minute-by-minute setting regarding FOREX. The best results of accuracy after training for different architectures by their novel Wavelet + CNN algorithm which outperformed other NN approaches, including RNN in both stocks as well as forex market, with a very high 83% accuracy. However, the variation of results was not very high with maximum of 4%preditction accuracy showing RNN suitability for the task. Study (Maknickienė and Maknickas 2012) based EUR/USD trading strategy on LSTM resulting in 4% profit over the test period of three-day trading steps with an overall accuracy of 65% percent of Pearson's correlation between predicting and historical values. RNN and CNN were employed and compared to C-RNN algorithm by (Ni et al., 2019) for nine volatile currency pairs in the data set with 10 years timespan. Comparison of mean squared errors of predicted values by different forecasting algorithms where C-RNN performed better by 10% compared to CNN.

Employed RNN and LSTM methods in our study are suitable and they were successfully employed by other studies. Although a model specialization into hybrid models creation seems highly promising (Pradeepkumar and Ravi 2018) notes that more attention should be paid to image processing algorithms for the needs of FOREX prediction. This opinion is consistent with rather problematic but still practiced trading of chartism where a certain "image" formation types are being searched for in order to enter a trading position.

The best performing network is a combination of 30 days period and 100 epochs of training. It is essential to know that these results are directly dependent on our architecture. We understand that any algorithm can be, at least temporarily, unsuccessful in an environment with an unexpected fundament changes such as economic distortion caused by the recent coronavirus. After reaching a certain point the market became bearish in a three-day panic which affected both the stock market as well as FOREX. Although algorithms can adjust in the after-initial-panic period the moment of panic itself cannot be handled properly since there not enough adequate cases of such events. Hence one of the main challenges will likely remain even after further research, in this promising and organically fast-growing research area, will be conducted.

References

Abdel-Nasser Mohamed, and Mahmoud Karar. (2019). Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing & Applications*: 7, 2727–2740. https://doi.org/10.1007/s00521-017-3225-z.

Alfarraj Motas, and Al-Regib Ghassan. (2018). Petrophysical-property estimation from seismic data using recurrent neural networks. Paper presented at the Conference *SEG Technical Program Expanded Abstracts 2018*, Anaheim, USA, October 15–18, pp. 2141–2146. https://doi.org/10.1190/segam2018-2995752.1.

Bagheri Ahmad, Mohammadi Peyhani Hamed, Akbari Mohsen. (2014). Financial forecasting using ANFIS networks with Quantum-behaved Particle Swarm Optimization. *Expert Systems with Applications*: 14, 6235–6250. https://doi.org/10.1016/j.eswa.2014.04.003.

Broto Carmen. (2013). The effectiveness of forex interventions in four Latin American countries. *Emerging Markets Review*: 17, 224–240. https://doi.org/10.1016/j.ememar.2013.03.003.

Carapuço Joao, Neves Rui, and Horta Nuno. (2018). Reinforcement learning applied to Forex trading. *Applied Soft Computing*: 73, 783–794. https://doi.org/10.1016/j.asoc.2018.09.017.

Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Paper presented at the Conference *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, July 21–26, pp. 105–114. Available online: https://arxiv.org/abs/1609.04802 (accessed on 02 January 2020).

Hedvicakova Martina, and Kral Martin. (2019). Benefits of KPIs for Industry Sector Evaluation: The Case Study from the Czech Republic. *E&M Ekonomie a Management*: 22(2), 97–113. https://doi.org/10.15240/tul/001/2019-2-007.

Hellstrom Thomas, and Holmstrom Kenneth. (1998). *Predicting the Stock Market*. Technical Report Serie IMa-TOM. 1997-07(August), Vasteras: Malardalen University. Available online: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=1D16C20A0B4047D0D042CEA2F8BEE9C9?doi=10.1.1.9.4827&rep=rep1&type=pdf (accessed on 02 January 2020).

Hochreiter Sepp, and Schmidhuber Jurgen. (1997). Long short-term memory. *Neural Computation*: 8, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Khashei Mehdi, Hejazi Seyed Reza, and Bijari Mehdi. (2008). A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets and Systems*: 7, 769–786. https://doi.org/10.1016/j.fss.2007.10.011.

Krizhevsky Alex, Sutskever Ilja, and Hinton E. Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*: 25.

Laherrere Jean Henri, and Sornette Didier. (1998). Stretched exponential distributions in nature and economy: „fat tails" with characteristic scales. *European Physical Journal B*: 4, 525–539. https://doi.org/10.1007/s100510050276

Mahesh. (2019). EUR/USD daily—Dataset. Available online: https://kaggle.com/maheshtaank/eurusd-daily (accessed on 02 November 2019).

Nijolė Maknickienė, and Algirdas Maknickas. (2012). Application of Neural Network for Forecasting of Exchange Rates and Forex Trading. Paper presented at the Conference 7th International Scientific Conference „Business and Management 2012. Selected Papers, Vilnius, Lithuania, November 1–2, pp. 122–127. https://doi.org/10.3846/bm.2012.017.

Lina Ni, Yujie Li, Xiao Wang, Jinquan Zhang, Jiguo Yu, and Chengming Qi. (2019). Forecasting of Forex Time Series Data Based on Deep Learning. *Procedia Computer Science*: 147, 647–652. https://doi.org/10.1016/j.procs.2019.01.189.

Peña-Barragán M. José, Ngugi K. Moffatt, Plant E. Richard, and Six Johan. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*: 6, 1301–1316. https://doi.org/10.1016/j.rse.2011.01.009.

Luca Di Persio, and Oleksandr Honchar. (2016). Artificial Neural Networks Approach to the Forecast of Stock Market Price Movements. *International Journal of Economics and Management Systems*: 01, 158–162.

Pradeepkumar Dadabada., and Ravi Vadlamani. (2018). Soft computing hybrids for FOREX rate prediction: A comprehensive review. *Computers & Operations Research*: 99, 262–284. https://doi.org/10.1016/j.cor.2018.05.020.

Rundo Francesco. (2019). Deep LSTM with Reinforcement Learning Layer for Financial Trend Prediction in FX High Frequency Trading Systems. Applied Sciences: 20, 4460. https://doi.org/10.3390/app9204460.

Sermpinis Georgios, Dunis Christian, Laws Jason, and Stasinakis Charalampos. (2012). Forecasting and trading the EUR/USD exchange rate with stochastic Neural Network combination and time-varying leverage. *Decision Support Systems*: 1, 316–329. https://doi.org/10.1016/j.dss.2012.05.039.

Sermpinis Georgios, Theofilatos Konstantinos, Karathanasopoulos Andreas, Georgopoulos Evstratios, and Dunis Christian. (2012). A Hybrid Radial Basis Function and Particle Swarm Optimization Neural Network Approach in Forecasting the EUR/GBP Exchange Rates Returns. In *Engineering Applications of Neural Networks*. Berlin: Springer Berlin Heidelberg, Vol. 311, pp. 413–422. https://doi.org/10.1007/978-3-642-32909-8_42.

Sidehabi S. Wetenriajeng, Indrabayu, and Tandungan Sofyan. (2016). Statistical and Machine Learning approach in forex prediction based on empirical data. Paper presented at the Conference 2016 International Conference on Computational Intelligence and Cybernetics (CYBERNETICSCOM), Makassar, Indonesia, November 22–24, pp. 63–68. https://doi.org/10.1109/CyberneticsCom.2016.7892568.

Sun Yubao, Chen Jiwei, Liu Quingshan, and Liu Guangcan. (2020). Learning image compressed sensing with sub-pixel convolutional generative adversarial network. *Pattern Recognition*: 98, 107051. https://doi.org/10.1016/j.patcog.2019.107051.

Varsamopoulos Savvas, Bertels Koen, and Almudever C. Garcia. (2020). Comparing neural network based decoders for the surface code. *IEEE Transactions on Computers*: 2, 300–311. https://doi.org/10.1109/TC.2019.2948612.

Wolterink M. Jelmer, Leiner Tim, Viergever A. Max, and Isgum Ivana. (2017). Generative Adversarial Networks for Noise Reduction in Low-Dose CT. *Ieee Transactions on Medical Imaging*: 12, 2536–2545. https://doi.org/10.1109/TMI.2017.2708987.

Yang Qingsong, Yan Pingkun, Zhang Yanbo, Yu Hengoyng, Shi Yongyi, Mou Xuanqin, Kalra Mannudeep, Zhang, Yi, Sun Ling, and Wang Ge. (2018). Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss. *Ieee Transactions on Medical Imaging*: 6, 1348–1357. https://doi.org/10.1109/TMI.2018.2827462.

Zhao Zheng, Chen Weihai, Wu Xingming, Chen C. Y. Peter, and Liu Jingmeng. (2017). LSTM network: A deep learning approach for short-term traffic forecast. *Iet Intelligent Transport Systems*: 2, 68–75. https://doi.org/10.1049/iet-its.2016.0208.