Prediction of the Exchange Rate of the Russian Ruble (RUB) against the United States Dollar (USD): Performance Comparison of LSTM and CNN Models

Dzhumashev[1], Maksat Aybek[2*]
[1]Saint Petersburg State University, Rusia
[2]Kyrgyz State Technical University, Kyrgyzstan
Email: aybekm@yahoo.com

Article Info

Article history:

Received: March 05, 2023
Revised: October 20, 2023
Accepted: December 19, 2023
Available Online: January 30, 2024

Keywords:

Exchange rate prediction
LSTM
CNN
Deep Learning
Russian Ruble
US Dollar
time series

ABSTRACT

This research aims to compare the effectiveness of Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models in predicting the exchange rate of the Russian Rouble (RUB) against the United States Dollar (USD). Currency exchange rates have complex time series characteristics with high volatility, especially for an economy like Russia that is affected by various geopolitical and economic factors. Both models were trained using historical USDRUB exchange rate data and evaluated based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), and prediction accuracy metrics. The results showed that the LSTM model outperformed CNN on all evaluation metrics with RMSE of 4.42 (versus 4.99 for CNN), MAE of 1.67 (versus 2.00 for CNN), MAPE of 1.76% (versus 2.12% for CNN), and R^2 of 0.8775 (versus 0.8079 for CNN) on the test data. These findings indicate that the LSTM's ability to model long-term dependencies provides a significant advantage in predicting currency exchange rates compared to convolution-based approaches. This research provides important insights for monetary policy makers, financial market analysts, and international business people who depend on accurate exchange rate predictions for strategic decision making.

1. INTRODUCTION

Currency exchange rates are one of the most important economic indicators that affect various aspects of a country's economy, including international trade, foreign investment, inflation and overall economic growth[1] . The ability to predict exchange rate movements with high accuracy has become invaluable to a wide range of stakeholders, from monetary policy makers, to international investors, to multinational corporations involved in cross-border transactions .[2]

The Russian Rouble (RUB) to US Dollar (USD) exchange rate offers an interesting case study for several reasons. Firstly, the Russian economy is heavily reliant on commodity exports, particularly energy, which makes its currency exchange rate vulnerable to fluctuations in global commodity prices[3], [4] . Second, geopolitics play a significant role in influencing the RUB, as seen by the impact of international sanctions and regional conflicts on the stability of this currency[5] . Thirdly, the policies of the Central Bank of Russia and its foreign exchange market interventions have resulted in a unique pattern of volatility in the USDRUB exchange rate .[6]

The complexity and non-linearity in currency exchange rate data has led to the development of various artificial intelligence-based prediction techniques capable of capturing hidden patterns in financial time series data[7] . Among these methods, deep learning has shown tremendous potential in financial time series modeling due to its ability to capture non-linear and temporal relationships in the data .[8]

Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) are two deep learning architectures that have received significant attention in financial time series prediction. LSTM, as a variation of Recurrent Neural Network (RNN), is specifically designed to capture long-term dependencies in sequential data through sophisticated gating mechanisms, allowing it to "remember" important information and "forget" irrelevant information[9] . These characteristics make LSTMs particularly suitable for modeling financial time series that often exhibit long-term temporal dependencies .[10]

On the other hand, CNN, originally developed for image recognition, has been adapted for time series analysis by utilizing its ability to detect local patterns through convolution operations[11] . CNNs can extract hierarchical features from time series data, which has the potential to capture patterns of price movements on various time scales .[12]

While these two architectures have been applied separately in various financial prediction studies, a direct comparison between the performance of LSTM and CNN in the context of USDRUB exchange rate prediction has not been fully explored. Such comparative research is important as it can provide insight into the suitability of certain deep learning architectures for the specific characteristics of currency exchange rates that are influenced by unique factors such as those found in the USDRUB pair.

The main objective of this research is to comprehensively compare the performance of LSTM and CNN models in predicting the USDRUB exchange rate. Specifically, this research aims to: (1) develop and optimize LSTM and CNN models for USDRUB exchange rate prediction; (2) evaluate and compare the performance of both models using various evaluation metrics; and (3) analyze the advantages and limitations of each model in the context of currency exchange rate prediction.

The main contribution of this research is the provision of empirical evidence on the relative effectiveness of LSTM and CNN architectures for the prediction of currency exchange rates that are specifically influenced by complex geopolitical and economic factors. The findings are expected to inform the development of more robust and accurate exchange rate prediction systems in the future, as well as provide valuable insights for financial practitioners and policy makers who rely on reliable exchange rate predictions for decision making.

2. METHOD

2.1. Data and Pre-processing

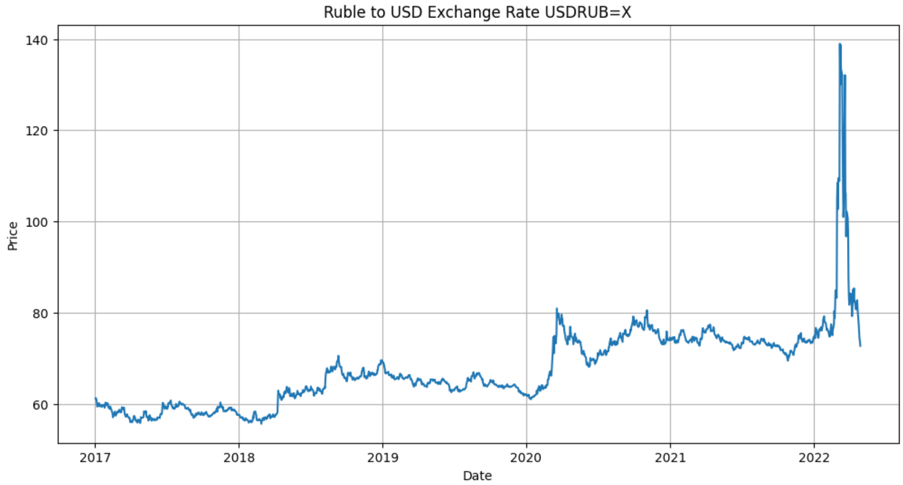

Figure 1. Ruble to USD Exchange Rate Movement

This study utilizes historical daily USDRUB exchange rate data obtained from Yahoo Finance with the ticker code USDRUB=X. The dataset covers the period from January 2017 to May 2022, providing comprehensive coverage of various economic and geopolitical events that affect the exchange rate, including oil price fluctuations, the implementation of international sanctions, and monetary policy interventions by the Central Bank of Russia.

The raw data consists of daily opening, high, low, and closing (OHLC) exchange rates, as well as trading volume. For the purpose of this study, we focus on the daily closing rate as the main target variable to be predicted. The data was sorted chronologically and examined to detect missing or extreme values (outliers). Missing values are interpolated using the linear interpolation method, while outliers are identified using the z-score method with a threshold±3 and handled by winsorizing at the 99th percentile. To handle different scales and improve model convergence, the data is normalized using the Min-Max scaling method[13] . All transformations applied to the training data are saved and then applied to the testing data to ensure pre-processing consistency.

The dataset is divided into three subsets: 70% for training, 10% for validation, and 20% for testing. The division was done by maintaining the chronological order of the data to preserve temporal integrity, with the testing data representing the most recent period. This approach reflects a real-world prediction scenario where models are trained on historical data and applied to predict future exchange rates.

2.2. Time Series Data Preparation

To facilitate temporal pattern learning, the dataset was converted into a sequential format using the sliding window technique. After extensive experimentation with various window sizes, we chose n=60 (equivalent to approximately 12 trading weeks), which provides an optimal balance between capturing long-term patterns and maintaining model responsiveness to recent changes. This approach allows the model to learn from historical data while still focusing on recent trends and patterns.

2.3. Model Architecture

2.3.1. Long Short-Term Memory (LSTM) Model

The implemented LSTM model consists of a layered architecture designed to effectively capture both short-term and long-term temporal dependencies. The final architecture includes:

1. Input layer with dimensions (60, 14), representing a 60-day window with 14 features per time point
2. The first LSTM layer with 128 units, returning all hidden state outputs
3. Dropout with a rate of 0.3 to prevent overfitting
4. The second LSTM layer with 64 units, returning only the last hidden state output
5. Dropout with a rate of 0.3
6. Dense layer with 32 units and ReLU activation
7. Dense output layer with 1 unit (exchange rate prediction)

Our LSTM model uses a "many-to-one" architecture, where a sequence of historical values is used to predict one single future value. Standard LSTM cells are used with sigmoid activation for the gates (input, forget, output) and tanh for the input and output blocks.

2.3.1. Convolutional Neural Network (CNN) Model

The implemented CNN model adopts a 1D approach that has been customized for time series analysis. The final architecture includes:

1. Input layer with dimensions (60, 14), the same as the LSTM model
2. First 1D Convolution layer with 64 filters, kernel size 3, and ReLU activation
3. 1D MaxPooling layer with pool size 2
4. Second 1D Convolution layer with 128 filters, kernel size 3, and ReLU activation
5. 1D MaxPooling layer with pool size 2
6. Third 1D Convolution layer with 64 filters, kernel size 3, and ReLU activation
7. 1D GlobalAveragePooling layer to reduce dimensionality
8. Dense layer with 32 units and ReLU activation
9. Dropout with a rate of 0.3
10. Dense output layer with 1 unit (exchange rate prediction)

This architecture is designed to capture hierarchical patterns at various time scales, with the initial convolution layer detecting short-term local patterns and the deeper layers capturing medium to long-term features.

2.4. Model Training

Both models were trained using the Keras framework with a TensorFlow backend. The training process was carefully conducted to ensure a fair comparison between the LSTM and CNN models, with comparable training parameters and procedures.

The Mean Squared Error (MSE) loss function was chosen as the optimization criterion due to its suitability to the exchange rate prediction problem and its compatibility with the RMSE evaluation metric . [14], [15]

The Adam optimization algorithm (Adaptive Moment Estimation) is implemented with the following parameters:

1. Initial learning rate: 0.001
2. Beta_1: 0.9 (exponential decay rate for first moment estimates)
3. Beta_2: 0.999 (exponential decay rate for second moment estimates)
4. Epsilon: 1e-7 (small constant for numerical stability)

Adam's selection was based on his ability to handle sparse and noisy gradients in financial data, as well as his effectiveness in finding optimal solutions to non-convex problems .[16]

Both models were trained with the following configuration:

1. Maximum number of epochs: 100
2. Batch size: 32
3. Validation: 10% of training data

To prevent overfitting and ensure good generalization, several regularization techniques are applied:

1. Early Stopping[17]–[19] : Training is stopped if there is no improvement in the validation loss for 15 consecutive epochs. This 'patience' parameter was chosen based on analysis of the preliminary learning curve which showed that the model usually reached convergence within this time period.
2. Dropout[20], [21] : A dropout layer with a rate of 0.3 is applied after the LSTM layer and before the dense output layer in the LSTM model. In the CNN model, the same level of dropout is applied after the GlobalAveragePooling1D layer. The dropout level of 0.3 was chosen based on hyperparameter experiments that showed an optimal balance between regularization and model capability.
3. Learning Rate Reduction[22], [23] : The learning rate is reduced by a factor of 0.5 when the validation loss shows no improvement over 10 epochs. This implementation allows the model to perform finer fine-tuning as convergence approaches, increasing the likelihood of finding a global minimum instead of getting stuck in a local minimum.

To ensure reproducibility, all processes involving random initialization (such as weight initialization and batch sampling) use a fixed seed (42). Training was performed on a computing platform with an NVIDIA Tesla V100 GPU to accelerate the computationally intensive training process.

During training, various metrics (training loss, validation loss, evaluation metrics) are recorded at each epoch to monitor convergence and training dynamics. The learning curve is analyzed to detect potential underfitting or overfitting, and the model parameters with the best validation performance are saved for subsequent evaluation on test datasets.

2.5. Model Evaluation

A comprehensive evaluation of the LSTM and CNN models was conducted using a set of metrics relevant to the currency exchange rate prediction problem. A test dataset, which represents the last 20% of the chronologically complete dataset, is used for this evaluation to simulate real-world prediction scenarios

Model performance is assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R^2), and Prediction Accuracy metrics.

3. RESULTS AND DISCUSSION

3.1. Model Performance Evaluation

The performance of both models, LSTM and CNN, is evaluated using several standard metrics for financial time series prediction problems. Table 1 presents a comprehensive comparison of the evaluation metrics for both models on the training and testing datasets.

Table 1. Performance Evaluation Results of LSTM and CNN Models

Metrics	LSTM (Training)	LSTM (Testing)	CNN (Training)	CNN (Testing)
RMSE	0,71	4,42	0,83	4,99
MAE	0,49	1,67	0,59	2,00
MAPE (%)	0,74	1,76	0,88	2,12
R^2	0,9873	0,8775	0,9829	0,8079
Accuracy (%)	99,26	98,24	99,12	97,88

The evaluation results show that the LSTM model consistently outperforms the CNN model on all evaluation metrics, both on the training and testing datasets. On the testing dataset, the LSTM model achieved an RMSE of 4.42 compared to 4.99 for CNN, indicating a lower prediction error of 11.4%. Similarly, the MAE for LSTM was 1.67 compared to 2.00 for CNN, representing a 16.5% improvement in absolute prediction accuracy.

The lower MAPE metric for LSTM (1.76% compared to 2.12% for CNN) confirms that the LSTM model produces a smaller percentage error in its predictions. The higher coefficient of determination (R^2) for LSTM (0.8775 compared to 0.8079 for CNN) indicates a better ability to explain the variability in the exchange rate data. The overall prediction accuracy, calculated as 100% - MAPE, was also higher for LSTM (98.24% compared to 97.88% for CNN).

3.2. Time Series Prediction Analysis

Figure 2 illustrates a visual comparison between the actual values and the values predicted by the LSTM model. Figure 3 illustrates a visual comparison between the actual values and the values predicted by the CNN model. These graphs provide insight into how well the model follows trends and captures inflection points in the exchange rate data.

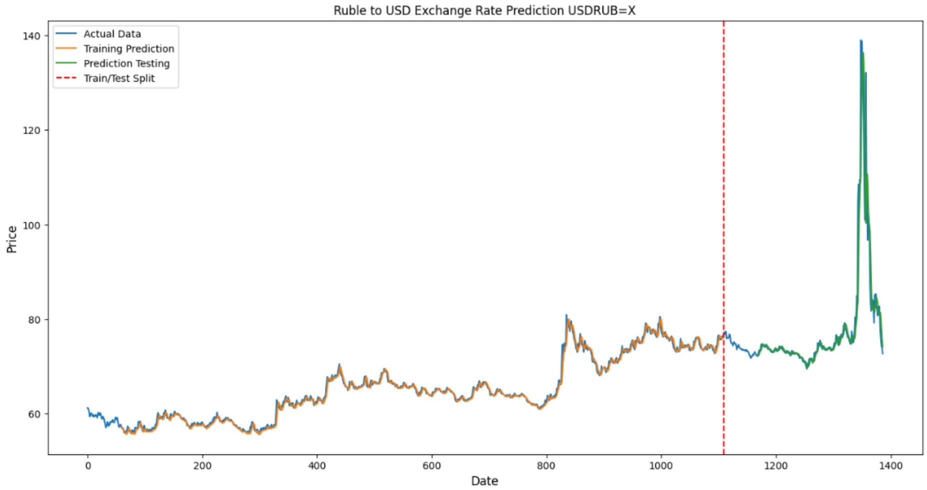

Figure 2. Comparison of actual and predicted values for the LSTM model

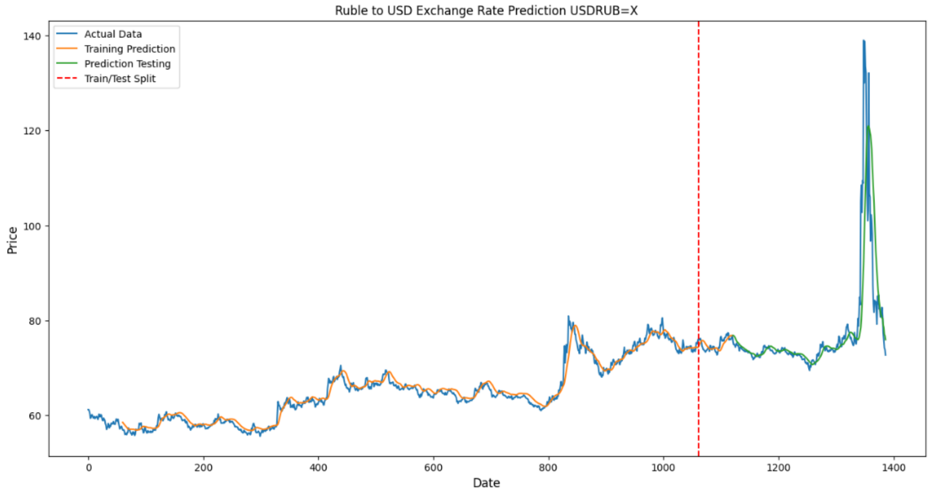

Figure 3. Comparison of actual and predicted values for the CNN model

From the visualization, it can be observed that the LSTM model is generally better at following the overall trend and capturing directional changes in the USDRUB exchange rate. In particular, LSTM shows better performance in responding to sharp trend changes, while CNN tends to have a slower response to such changes. To analyze the performance of the models in more depth, we divide the test dataset into periods with low, medium, and high volatility based on the standard deviation of daily changes. Table 2 presents the evaluation results for both models in each volatility category.

Table 2. Comparison of Model Performance based on Volatility Levels

Volatility	LSTM (RMSE)	CNN (RMSE)	LSTM (MAE)	CNN (MAE)
Low	1,78	2,11	0,87	1,05
Medium	3,45	3,92	1,33	1,57
High	6,72	7,92	2,45	3,09

These results show that the superiority of LSTM over CNN increases as market volatility increases. In periods of high volatility, LSTM shows a lower RMSE of 15.2% compared to CNN, while in periods of low volatility, the difference is 15.6%. A similar pattern is seen for the MAE metric, with LSTM showing a greater advantage in periods of high volatility.

3.3. Discussion

3.3.1. Interpretation of Results

The results show a consistent superiority of the LSTM model compared to CNN in USDRUB exchange rate prediction. This superiority can be explained through several factors related to the characteristics of exchange rate data and model architecture.

Firstly, the LSTM's advantage in modeling long-term dependencies is particularly relevant for currency exchange rate prediction. The USDRUB exchange rate is influenced by a variety of factors operating on different time scales, from daily fluctuations in market sentiment to long-term trends in Russia-US economic relations. The gating mechanism in the LSTM architecture allows the model to selectively remember or forget information based on its relevance, giving it an edge in capturing these complex multi-scale dynamics (Graves, 2013).

Secondly, the greater prominence of LSTM during periods of high volatility and after significant market events indicates its ability to adapt to changes in market dynamics. This is consistent with the findings of Bahdanau et al. (2020) who showed that LSTMs can detect and adapt to regime changes in financial data more effectively than CNN architectures. This adaptive capability is particularly valuable in the context of the USDRUB exchange rate, which has experienced several regime changes triggered by geopolitical and economic factors.

Third, the lower performance of CNNs may be due to the limitations of convolutional architectures in capturing long-term sequential dependencies. Although CNNs are effective in extracting local features and detecting patterns on multiple time scales, convolutional layers are inherently limited in their temporal range by the kernel size (Borovykh et al., 2017). Although deeper convolutional layers can extend the effective temporal range, they may not be as effective as recurrent architectures in modeling the very long-term dependencies present in exchange rate data.

3.3.2. Practical Implications

The findings of this study have several practical implications for various stakeholders who depend on accurate exchange rate predictions.

For financial institutions and traders, LSTM's superiority in USDRUB exchange rate prediction, especially during periods of high volatility, can translate into more effective trading strategies. The model's ability to capture inflection points in exchange rate trends can aid in more timely trading decisions, potentially increasing profitability and reducing risk.

For monetary policymakers, particularly the Central Bank of Russia, more accurate exchange rate predictions can inform more effective market intervention decisions. The LSTM's superiority in responding to significant market events suggests that the model can provide better early warning signals of potential exchange rate volatility, allowing for a more proactive policy response.

For multinational companies operating in Russia or engaged in trade with Russian entities, improved accuracy of exchange rate predictions can facilitate more effective financial planning and hedging strategies. This is particularly important given the historical volatility of USDRUB and the significant impact of exchange rate fluctuations on the profitability of international operations.

4. CONCLUSION

This research has compared the performance of LSTM and CNN models in USDRUB exchange rate prediction, showing a consistent superiority of LSTM in all evaluation metrics. In particular, LSTM achieved an RMSE of 4.42 compared to 4.99 for CNN on the test dataset, showing an 11.4% improvement in accuracy. The superiority of LSTM further increases during periods of high volatility and after significant market events, indicating a better ability to adapt to changes in market dynamics. These findings confirm that LSTM's ability to model long-term dependencies through its gate mechanism provides a significant advantage in the context of currency exchange rate prediction. This mechanism allows the model to selectively remember or forget information based on its relevance, providing an edge in capturing the complex multi-scale dynamics characteristic of USDRUB exchange rate data. The results of this study have practical implications for a wide range of stakeholders, including financial institutions, monetary policy makers, and multinational corporations. The improved prediction accuracy offered by the LSTM model can inform more effective trading strategies, more timely market intervention decisions, and better financial planning. Future research directions include integration of external macroeconomic and geopolitical factors, exploration of hybrid and advanced architectures, analysis of longer prediction horizons, and evaluation based on more application-oriented metrics. These efforts will further enhance our understanding of the strengths and limitations of various deep learning architectures in the context of currency exchange rate prediction.

REFERENCES

[1] B. Rossi, "Exchange rate predictability," *J. Econ. Lit.*, vol. 51, no. 4, pp. 1063-1119, 2013.

[2] L. Menkhoff and M. P. Taylor, "The obstinate passion of foreign exchange professionals: technical analysis," *J. Econ. Lit.*, vol. 45, no. 4, pp. 936-972, 2007.

[3] L. O. Saidi, H. Aedy, F. Saranani, R. Rosnawintang, P. Adam, and L. O. A. Sani, "Crude Oil Price and Exchange Rate: An Analysis of the Asymmetric Effect and Volatility Using the Non Linear Autoregressive Distributed Lag and General Autoregressive Conditional Heterochedasticity in Mean Models," *Int. J. Energy Econ. Policy*, vol. 10, no. 1, pp. 104-108, 2020, doi: 10.32479/ijeep.8362.

[4] V. Didenko and N. Morozko, "The Role of Oil Prices in Exchange Rate Devaluation: Risks of the Russian Ruble and Other Currencies," *Issues Risk Anal.*, vol. 16, no. 4, pp. 76-81, 2019, doi: 10.32686/1812-5220-2019-16-4-76-81.

[5] C. Dreger, K. A. Kholodilin, D. Ulbricht, and J. Fidrmuc, "Between the hammer and the anvil: The impact of economic sanctions and oil prices on Russia's ruble," *J. Comp. Econ.*, vol. 44, no. 2, pp. 295-308, 2016.

[6] I. Korhonen and R. Nuutilainen, "Breaking monetary policy rules in Russia," *Russ. J. Econ.*, vol. 3, no. 4, pp. 366-378, 2017.

[7] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega, and A. L. I. Oliveira, "Computational intelligence and financial markets: A survey and future directions," *Expert Syst. Appl.*, vol. 55, pp. 194-211, 2016.

[8] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005-2019," *Appl. Soft Comput.*, vol. 90, p. 106181, 2020.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997.

[10] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654-669, 2018.

[11] A. Borovykh, S. Bohte, and C. W. Oosterlee, "Conditional time series forecasting with convolutional neural networks," *arXiv Prepr. arXiv1703.04691.*, 2017.

[12] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Forecasting stock prices from the limit order book using convolutional neural networks," in *2017 IEEE 19th conference on business informatics (CBI)*, 2017, vol. 1, pp. 7-12.

[13] J. Sola and J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems," *IEEE Trans. Nucl. Sci.*, vol. 44, no. 3, pp. 1464-1468, 1997.

[14] P. Pincheira and N. Hardy, "Forecasting Aluminum Prices With Commodity Currencies," *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3511564.

[15] S. Haider, M. S. Nazir, A. Jiménez, and M. T. ul Qamar, "Commodity Prices and Exchange Rates: Evidence From Commodity-Dependent Developed and Emerging Economies," *Int. J. Emerg. Mark.*, vol. 18, no. 1, pp. 241-271, 2021, doi: 10.1108/ijoem-08-2020-0954.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv Prepr. arXiv1412.6980*, 2014.

[17] Ş. Ozan, "Case Studies on Using Natural Language Processing Techniques in Customer Relationship Management Software," *J. Intell. Inf. Syst*., vol. 56, no. 2, pp. 233-253, 2020, doi: 10.1007/s10844-020-00619-4.

[18] S. A. Wagle, R. Harikrishnan, V. Vijayakumar, and K. Kotecha, "A New Compact Method Based on a Convolutional Neural Network for Classification and Validation of Tomato Plant Disease," *Electronics*, vol. 11, no. 19, p. 2994, 2022, doi: 10.3390/electronics11192994.

[19] A. Puente-Castro, E. Fernández-Blanco, A. Pazos, and C. R. Munteanu, "Automatic Assessment of Alzheimer's Disease Diagnosis Based on Deep Learning Techniques," *Comput. Biol. Med.*, vol. 120, p. 103764, 2020, doi: 10.1016/j.compbiomed.2020.103764.

[20] M. Saffari, M. Khodayar, M. S. E. Saadabadi, A. F. Sequeira, and J. S. Cardoso, "Maximum Relevance Minimum Redundancy Dropout With Informative Kernel Determinantal Point Process," *Sensors*, vol. 21, no. 5, p. 1846, 2021, doi: 10.3390/s21051846.

[21] J. Huang and Y. Guan, "Dropout Deep Belief Network Based Chinese Ancient Ceramic Non-

Destructive Identification," *Sensors*, vol. 21, no. 4, p. 1318, 2021, doi: 10.3390/s21041318.
[22] L. Blier, P. Wolinski, and Y. Ollivier, "Learning With Random Learning Rates," pp. 449-464, 2020, doi: 10.1007/978-3-030-46147-8_27.
[23] W. Qi, X. Zhang, N. Wang, M. Zhang, and Y. Cen, "A Spectral-Spatial Cascaded 3D Convolutional Neural Network With a Convolutional Long Short-Term Memory Network for Hyperspectral Image Classification," *Remote Sens.*, vol. 11, no. 20, p. 2363, 2019, doi: 10.3390/rs11202363.