International Journal of Artificial Intelligent and Informatics
Vol. 3, No. 2, January 2025, pp. 60~ 66
ISSN: 2622-626X, DOI: http://doi.org/10.33292/ijarlit.v3i1.49

Comparison of CNN, CNN-GRU, and GRU Models for Prediction of Hryvnia (Ukraine) Exchange Rate against US Dollar

Sanoun Mostafa
[1]University of Doha for Science and Technology, Qatar
Email: sanounm@ymail.com

Article Info

Article history:

Received: March 20, 2025
Revised: May 12, 2025
Accepted: June 30, 2025
Available Online: July 30, 2025

Keywords:

exchange rate prediction
Ukrainian hryvnia
CNN
GRU
hybrid model
deep learning

ABSTRACT

This study aims to compare the performance of three neural network-based machine learning models, namely Convolutional Neural Network (CNN), hybrid CNN-Gated Recurrent Unit (CNN-GRU), and Gated Recurrent Unit (GRU) in predicting the exchange rate of the Ukrainian Hryvnia against the United States Dollar. The data used is sourced from Yahoo Finance in the range of 2018 to 2023. The evaluation results show that the CNN-GRU hybrid model provides the best performance with the highest test accuracy of 99.69% and an R^2 value of 0.6899. The CNN model achieved 98.99% test accuracy but with a negative R^2 (-1.0343), while the GRU model showed 97.94% test accuracy with a very low R^2 (-6.3755). This study reveals the advantages of the hybridization approach in modeling financial time series data by combining the feature extraction capabilities of CNN and the sequential modeling capabilities of GRU. These results provide important insights for the development of predictive models for volatile currency markets, especially for emerging economies such as Ukraine.

1. INTRODUCTION

Currency exchange rate volatility is one of the crucial economic parameters, especially for countries with unstable geopolitical conditions. Ukraine, as a country that has faced numerous political and economic challenges in recent years, has a complex and unpredictable exchange rate dynamics of the Hryvnia (UAH) against the United States Dollar (USD)[1], [2] . This exchange rate instability has a significant impact on economic stability, inflation, international trade and foreign investment in Ukraine .[3]

Currency exchange rate prediction is one of the challenging research topics in computational economics and finance. The complexity of the foreign exchange market, which is influenced by various factors such as macroeconomic conditions, geopolitical situations, monetary policy, and market sentiment, makes exchange rate prediction a difficult problem to solve with traditional approaches.[4] . In the era of industrial revolution 4.0, machine learning and deep learning-based approaches have opened up new opportunities in modeling and predicting complex and non-linear financial time series data .[5]

Several previous studies have explored the use of deep learning methods for currency exchange rate prediction. Convolutional Neural Network (CNN) has proven to be effective in identifying local patterns in time series data[6] . On the other hand, Recurrent Neural Network (RNN)-based architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have shown promise in modeling long-term sequential dependencies in financial data[7] . There is also a research trend that combines the advantages of various architectures, such as CNN-LSTM or CNN-GRU hybrid models, which aim to integrate the feature extraction capabilities of CNN with the sequential modeling capabilities of RNN .[8], [9]

Although there are many studies on exchange rate prediction using various deep learning techniques, there are limited studies that specifically compare the performance of CNN[10] , GRU, and hybrid CNN-GRU models for Ukrainian Hryvnia exchange rate prediction. Livieris et al. (2020)[6] conducted a comparison of several deep learning models for stock index prediction, but not specifically for currencies of countries with high volatility such as Ukraine. Meanwhile, Rehman et al. (2022)[11] proposed a hybrid model for exchange rate prediction but focused on major currencies such as EUR/USD or GBP/USD.

Predicting the Hryvnia exchange rate has its own complexities given Ukraine's economic and geopolitical dynamics. The prolonged conflict in eastern Ukraine and the annexation of Crimea by Russia since 2014 have had a significant impact on the volatility of the UAH/USD exchange rate (Plastun et al., 2019). In addition, economic reforms, changes in the monetary policy of the Central Bank of Ukraine, and fluctuations in foreign capital flows also affect the dynamics of the exchange rate[2] . These factors form a unique and challenging pattern of exchange rate movements to model.

The main motivation of this research is to explore and compare the capabilities of three deep learning architectures CNN, GRU, and CNN-GRU in predicting the UAH/USD exchange rate based on historical data. This research contributes in several aspects: first, it provides a comprehensive analysis of the effectiveness of deep learning models for currencies from emerging economies with high volatility; second, it evaluates the benefits of hybridization approaches in improving prediction accuracy; and third, it provides a methodological framework for the development of exchange rate prediction systems that can be adapted for other currencies.

Based on this background, this study aims to: (1) develop a UAH/USD exchange rate prediction model using CNN, GRU, and hybrid CNN-GRU algorithms; (2) compare the performance of the three models based on various evaluation metrics including RMSE, MAE, MAPE, R^2, and accuracy; and (3) identify the optimal model for Hryvnia exchange rate prediction that can support economic and financial decision making.

2. METHOD

2.1. Data Sources and Pre-processing

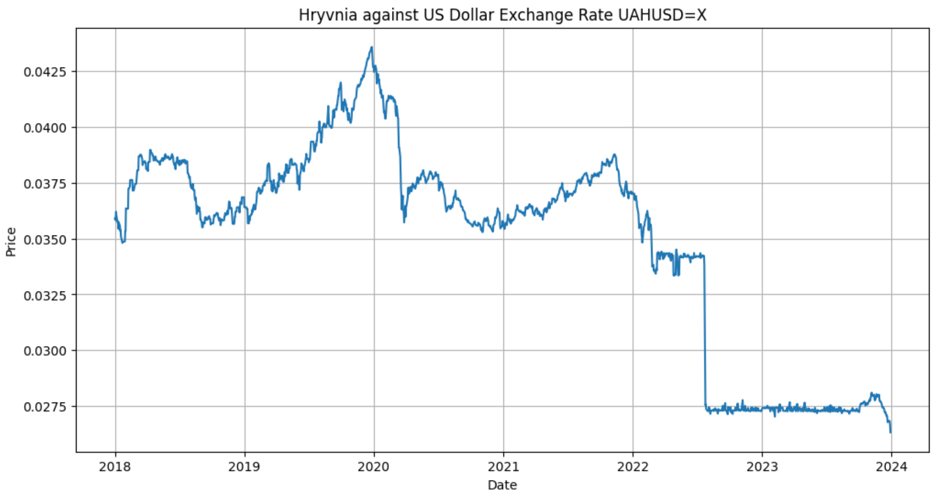

Figure 1. Historical data graph of Hryvnia to USD exchange rate movement

The data used in this research is sourced from Yahoo Finance, covering the daily exchange rate of the Ukrainian Hryvnia (UAH) against the United States Dollar (USD) in the period 2018 to 2023. The dataset provides information on the opening, highest, lowest, closing, and transaction volume exchange rates. In this research, the main focus is on the daily closing price.

Data pre-processing is an important step in data analysis, and consists of several crucial stages that aim to improve the quality and accuracy of machine learning models. In the first stage, data cleaning is done through handling missing values using linear interpolation. Linear interpolation is an effective method to fill in missing values by drawing straight lines between existing data points, resulting in more accurate estimates in the context of time or sequential data .[12]

Furthermore, detection and handling of outliers can be done using the Z-score method. Z-score calculates how far a value is from the mean in standard deviation units. This method is very useful for identifying inconsistent data that may affect the results of the analysis[13] . By identifying and removing outliers, we can ensure that the model built is more accurate and representative of the patterns in the data .[14]

In the third stage, data normalization is applied to scale the data into the same range. The technique used is Min-Max Scaler, which converts the values into a range between 0 and 1. This process is important to ensure all features contribute proportionally in the learning process, avoiding the dominance of features with a larger range of values[15], [16] . In addition, the use of Min-Max Scaler can accelerate the computational efficiency and convergence of the model, which makes the training process faster and more efficient .[17]

Finally, sequential feature formation is performed by creating a 60-day long sliding window as the model input to predict the next day's exchange rate. This technique allows the model to capture temporalities in the data and patterns that may not be visible in isolated observations. By preparing the data in this format, the model can learn to construct predictions based on the time sequence of the fed data[18], [19] . This design of sequential features is particularly important in the context of time modeling, such as exchange rate prediction, where relationships between times can greatly influence the results .[19]

In order to support the steps described, the application of normalization, outlier detection, and feature formation methods are fundamental aspects in ensuring the quality of the final model. In conclusion, good data pre-processing will significantly improve the performance and accuracy of machine learning models.

The data was divided into three subsets: training data (80%), and test data (20%). This division was done chronologically to maintain the temporal nature of the data, with training data coming from the earliest period, followed by validation data, and test data from the latest period.

2.2. Architecture

2.2.1. CNN Model

The CNN model developed consists of several layers: a one-dimensional convolution layer (Conv1D) with 128 filters and a kernel size of 2, followed by a max pooling layer (MaxPooling1D). The output of the max pooling layer is then flattened and passed to two dense layers with 50 and 1 neurons respectively. The ReLU activation function is used in the hidden layer, while the output layer uses a linear activation function to generate the exchange rate prediction. The complete architecture of the CNN model is shown in Table 1.

Table 1. CNN Model Architecture

Layer (type)	Output Shape	Parameters
conv1d_1 (Conv1D)	(None, 60, 128)	768
max_pooling1d_1(MaxPooling1D)	(None, 60, 128)	0
flatten_1 (Flatten)	(None, 7680)	0
dense_2 (Dense)	(None, 50)	384,050
dense_3 (Dense)	(None, 1)	51

The total parameters of the CNN model are 384,869, with all parameters trainable.

2.2.2. CNN-GRU Model

The CNN-GRU hybrid model integrates the advantages of CNN in local feature extraction with GRU's ability to capture long-term dependencies. The architecture starts with a convolution layer (Conv1D) with 128 filters, followed by max pooling. The output is then processed by a GRU layer with 50 units to capture temporal patterns. Finally, two dense layers with 25 and 1 neurons are used to generate predictions. The full architecture of the CNN-GRU model is shown in Table 2.

Table 2. CNN-GRU Model Architecture

Layer (type)	Output Shape	Parameters
conv1d (Conv1D)	(None, 60, 128)	768
max_pooling1d(MaxPooling1D)	(None, 60, 128)	0
gru (GRU)	(None, 50)	27,000
dense (Dense)	(None, 25)	1,275
dense_1 (Dense)	(None, 1)	26

The total parameters of the CNN-GRU model are 29,069, much more efficient than the CNN model.

2.2.3. GRU Model

The pure GRU model consists of two sequential GRU layers with 50 units each, separated by a dropout layer to prevent overfitting. The first GRU layer returns the full sequence to be processed by the second GRU layer, which only returns the final output. The final prediction is generated through two dense layers with 25 and 1 neurons. The full architecture of the GRU model is shown in Table 3.

Table 3. GRU Model Architecture

Layer (type)	Output Shape	Parameters
gru_2 (GRU)	(None, 60, 50)	7,950
dropout_2 (Dropout)	(None, 60, 50)	0
gru_3 (GRU)	(None, 50)	15,300
dropout_3 (Dropout)	(None, 50)	0
dense_2 (Dense)	(None, 25)	1,275
dense_3 (Dense)	(None, 1)	26

The total parameters of the GRU model are 24,551, making it the lightest model among the three models compared.

2.3. Model Training

All three models were trained using the Adam optimizer with an initial learning rate of 0.001. The loss function used was Mean Squared Error (MSE), which is commonly used for regression problems such as exchange rate prediction. To prevent overfitting, an early stopping technique with patience 10 was applied, which stops training if there is no improvement in the validation loss for 10 consecutive epochs. In addition, a learning rate reduction on plateau technique is also implemented, which reduces the learning rate by a factor of 0.5 if there is no improvement for 5 epochs.

2.4. Model Evaluation

The performance of both models was evaluated using several metrics to provide a comprehensive perspective:

1. Root Mean Square Error (RMSE): Measures the square root of the average square of the difference between the predicted value and the actual value .[20]
2. Mean Absolute Error (MAE): Measures the average absolute value of the difference between the predicted value and the true value .[21]
3. Mean Absolute Percentage Error (MAPE): Measures the average percentage absolute error relative to the true value.
4. Coefficient of Determination (R^2): Measures the proportion of variation in the dependent variable that can be explained by the independent variable.
5. Accuracy: Calculated as 100% - MAPE, shows the level of model accuracy in percentage form.

Evaluations were conducted on both training and test data to assess the generalization ability of the model and detect potential overfitting or underfitting.

3. RESULTS AND DISCUSSION

3.1. Model Performance

The evaluation results of the three models on the training and test data are presented in Table 4.

Table 4. Performance Comparison of CNN, CNN-GRU, and GRU Models

Model	Data	RMSE	MAE	MAPE (%)	R^2	Accuracy (%)
CNN	Train	0.00	0.00	0.74	0.9782	99.26
	Test	0.00	0.00	1.01	-1.0343	98.99
CNN-GRU	Train	0.00	0.00	0.53	0.9870	99.47
	Test	0.00	0.00	0.31	0.6899	99.69
GRU	Train	0.00	0.00	0.64	0.9852	99.36
	Test	0.00	0.00	2.06	-6.3755	97.94

The evaluation results show that all three models perform very well on the training data, with accuracy above 99% and R^2 values above 0.97. However, significant differences were observed in the models' performance on the test data.

The CNN-GRU model performed best on the test data with the highest accuracy of 99.69% and a positive R^2 value of 0.6899. The model shows minimal difference between performance on training and test data, indicating good generalization ability. The MAPE of the CNN-GRU model on the test data (0.31%) was even lower than on the training data (0.53%), indicating the adaptive ability of the model to new patterns in the data.

The CNN model came second in terms of accuracy on the test data with a value of 98.99%, however, it showed signs of overfitting with a negative R^2 value (-1.0343) on the test data. A negative R^2 value indicates that the model's prediction is less accurate than using the average value as prediction.

The pure GRU model showed the lowest performance on the test data with 97.94% accuracy and a strongly negative R^2 value (-6.3755). There was a significant drop in performance from the training data to the test data, indicating a serious problem in model generalization.

3.2. Model Performance Analysis

3.2.1. CNN Model

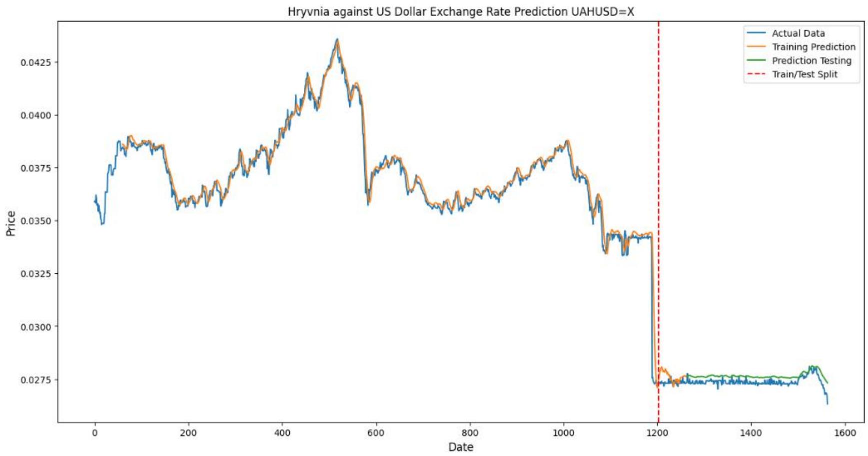

Figure 2. Comparison graph of train and test data of CNN model

The CNN model showed good ability in extracting local patterns from the UAH/USD exchange rate time series data. The large number of parameters (384,869) allows the model to capture the complexity of the data, but also increases the risk of overfitting. This can be seen from the difference in performance between the training and test data, especially in the R^2 value which decreased dramatically from 0.9782 to -1.0343.

Sezer et al. (2020) explain that CNN models are effective in detecting patterns in spatially structured data, but may be suboptimal for modeling complex long-term dependencies in financial data. The results of this study are in line with these findings, where the CNN model shows limitations in generalization to test data despite performing very well on training data.

3.2.2. CNN-GRU Model

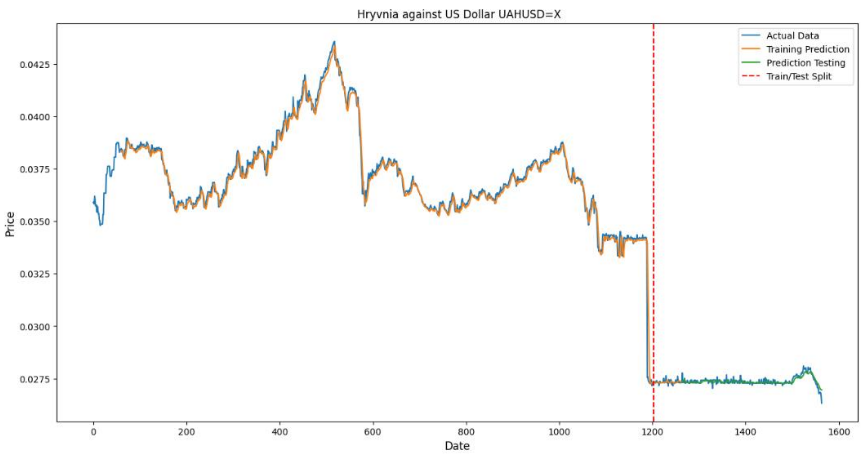

Figure 3. Comparison graph of train and test data of CNN-GRU model

The CNN-GRU hybrid model shows superior performance compared to other models, especially in terms of generalization on test data. With a moderate number of parameters (29,069), the model is able to strike a balance between complexity and generalization ability. The positive R^2 value on the test data (0.6899) indicates that the model is able to explain the variation in the data well.

The superiority of the CNN-GRU model can be explained through the synergy between CNN's ability in local feature extraction and GRU's ability in modeling sequential dependencies. The convolution layer serves as a feature extractor that identifies important patterns in the data, while the GRU layer processes the temporal information of the features.[8] . This combination is particularly suitable for currency exchange rate data that is influenced by various factors at different time scales.

This finding is consistent with the research of Livieris et al. (2020)[6] which shows that hybrid models often outperform single models in financial data prediction. In the context of the UAH/USD exchange rate, the hybrid approach proved to be able to capture the complexity of the data more effectively.

3.2.3. GRU Model

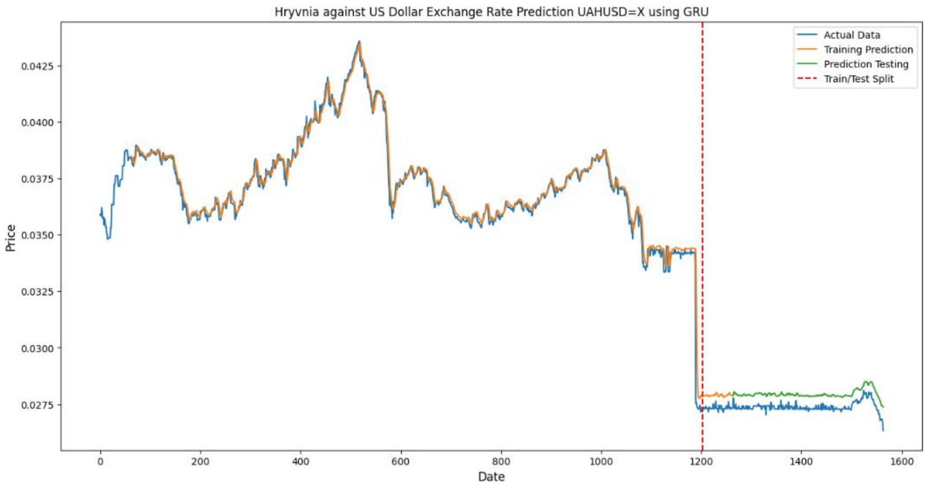

Figure 4. Comparison graph of train and test data of GRU model

The pure GRU model showed the worst performance on the test data despite having a competitive performance on the training data. The highly negative R^2 value (-6.3755) indicates that the model failed to generalize the learned patterns to new data. This contradicts the theoretical expectation that RNN architectures such as GRU should excel in modeling time series data .[7]

The failure of the GRU model in this study could be due to several factors. Firstly, although GRU is designed to handle long-term dependencies, it may struggle to capture complex non-linear patterns in the UAH/USD exchange rate without the aid of feature extraction as performed by CNN. Secondly, the high volatility and external factors affecting the Hryvnia exchange rate may create patterns that are difficult to identify by conventional GRU.

This finding is interesting to analyze further in the context of Fischer & Krauss (2018)[4] who found that the performance of RNN models can vary significantly depending on the characteristics of the financial data being modeled. For the case of the UAH/USD exchange rate, the pure GRU model structure appears to be less optimal than the hybrid approach.

3.3. Practical Implications

The results of this study have several practical implications for the development of currency exchange rate prediction systems, particularly for currencies from high volatility economies such as Ukraine:

1. Advantages of the hybrid approach: The CNN-GRU model shows significant potential in modeling and predicting currency exchange rates. The approach can be adapted for other currencies, especially from emerging economies that have similar volatility characteristics.
2. Computational efficiency: Although the CNN model has the highest number of parameters (384,869), the CNN-GRU model with significantly fewer parameters (29,069) is able to provide better results. This shows that an efficient and precise architecture can outperform more complex models.
3. Applications in trading and hedging: The high accuracy rate (99.69%) of the CNN-GRU model opens up application opportunities in algorithmic trading strategies and exchange rate risk management for businesses operating in Ukraine or transacting with Hryvnia.
4. Early warning system: High-accuracy predictive models can be integrated into an early warning system for exchange rate volatility, assisting monetary policymakers in anticipating market turmoil.

4. CONCLUSION

This research has compared the performance of three deep learning architectures, namely Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), and CNN-GRU hybrid model in predicting the exchange rate of Ukrainian Hryvnia (UAH) against the United States Dollar (USD) using data from Yahoo Finance for the period 2018 to 2023. Evaluation results with various metrics show that the CNN-GRU hybrid model provides the best performance compared to pure CNN and GRU models. The model achieved the highest accuracy on the test data of 99.69% and an R^2 value of 0.6899, which confirms the superiority of the hybridization approach in handling the complexity of financial time series data. Meanwhile, the CNN model performed adequately on the training data, but experienced a significant decline when applied to the test data, as reflected by the negative R^2 value. This shows the limited generalization ability of CNN to highly volatile exchange rate data. On the other hand, the pure GRU model performed the worst on the test data with a highly negative R^2 value, despite GRU being architecturally designed to manage sequential data. These findings highlight the challenges in modeling exchange rate dynamics that are influenced by many external factors. In addition, the CNN-GRU model proved to be more computationally efficient as it was able to achieve optimal performance with significantly fewer parameters compared to the CNN model, indicating the importance of proper architectural design over increasing model complexity alone. Despite the variation in R^2 values, the very high accuracy rates (more than 97%) of all three models indicate that deep learning approaches have great potential in currency exchange rate prediction, noting that generalizability remains a crucial aspect that needs to be considered in model development.

REFERENCES

[1] V. Shevchuk and R. Kopych, "Assessing fiscal sustainability in Ukraine: TVP and VAR/VEC approaches," *Entrep. Bus. Econ. Rev.*, vol. 6, no. 3, pp. 73-87, 2018.

[2] V. Shevchuk and R. Kopych, "Fiscal policy effects in Ukraine," *Argumenta Oeconomica Cracoviensia*, no. 18, pp. 33-50, 2018.

[3] A. Plastun, I. Makarenko, L. Khomutenko, Y. Belinska, and M. Domashenko, "Exploring the frequency of price overreactions in the Ukrainian stock market," *Invest. Manag. Financ. Innov.*, vol. 15, no. 3, pp. 157-168, 2018.

[4] 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.

[5] 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.

[6] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting," *Neural Comput. Appl.*, vol. 32, pp. 17351-17360, 2020.

[7] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Syst. Appl.*, vol. 103, pp. 25-37, 2018.

[8] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS One*, vol. 12, no. 7, p. e0180944, 2017.

[9] M. Gabrielzinho and G. Moraes, "Petrobras Stock Price Prediction Using Deep Learning Approach: Performance Comparison of CNN and CNN-GRU Methods," *Int. J. Artif. Intell. Informatics*, vol. 3, no. 1, pp. 30-36, 2025, doi: 10.33292/ijarlit.v3i1.44.

[10] F. Agustin and P. De Melin, "Comparison of GRU and CNN Methods for Predicting the Exchange Rate of Argentine Peso (ARS) against US Dollar (USD)," *Int. J. Artif. Intell. Informatics*, vol. 2, no. 1, pp. 9-16, 2024, doi: 10.33292/ijarlit.v2i1.31.

[11] F. U. Rehman and B. M. Al-Ghazali, "Evaluating the Influence of Social Advertising, Individual Factors, and Brand Image on the Buying Behavior Toward Fashion Clothing Brands," *Sage Open*, vol. 12, no. 1, 2022, doi: 10.1177/21582440221088858.

[12] M. Mandorino, A. Tessitore, C. Leduc, V. Persichetti, M. Morabito, and M. Lacome, "A New Approach to Quantify Soccer Players' Readiness Through Machine Learning Techniques," *Appl. Opt. Sci.*, 2023, doi: 10.3390/app13158808.

[13] H. Henderi, "Comparison of Min-Max Normalization and Z-Score Normalization in the K-Nearest Neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *Ijiis Int. J. Informatics Inf. Syst.*, 2021, doi: 10.47738/ijiis.v4i1.73.

[14] S. M. Kasongo and Y. Sun, "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00379-6.

[15] A. T. Setyoko, B. A. Saputra, E. Kristiningrum, R. Nurcahyo, M. Suryanegara, and I. Ali, "The Sustainability Performance of Maintenance 4.0 in the Indonesian Coal Mining Sector," 2024, doi: 10.21203/rs.3.rs-5422965/v1.

[16] V. R. Prasetyo, S. Axel, J. T. Soebroto, D. Sugiarto, S. A. Winatan, and S. D. Njudang, "Gold Price Prediction Based on Gold.org Data Using the Long Short Term Memory Method," *Sistemasi*, 2022, doi: 10.32520/stmsi.v11i3.1999.

[17] H. P. Wibowo, M. I. Anshori, and M. S. Haris, "The Discriminant Analysis Function Was Implemented to Predict the Presence of Diabetes," *Jesica*, 2024, doi: 10.47794/jesica.v1i2.10.

[18] M. Musfiroh, D. C. Rini Novitasari, P. K. Intan, and G. G. Wisnawa, "Application of Principal Component Analysis (PCA) and Long Short-Term Memory (LSTM) Methods in Predicting Daily Rainfall Predictions," *Build. Informatics Technol. Sci.*, 2023, doi: 10.47065/bits.v5i1.3114.

[19] M. F. Rizkilloh and S. Widiyanesti, "Cryptocurrency Price Prediction Using Long Short Term Memory (LSTM) Algorithm," *J. Resti (Engineering Systems and Technol. Information)*, 2022, doi: 10.29207/resti.v6i1.3630.

[20] L. Tian, "Forecast and Analysis for Stock Markets of the U.S., Canada, and Mexico Based on Time Series Forecasting Models," *Adv. Econ. Manag. Polit. Sci.*, vol. 13, no. 1, pp. 389-401, 2023, doi: 10.54254/2754-1169/13/20230759.

[21] G. Bekdaş, Y. Aydın, Ü. Işıkdağ, A. N. Sadeghifam, S. Kim, and Z. W. Geem, "Prediction of Cooling Load of Tropical Buildings With Machine Learning," *Sustainability*, vol. 15, no. 11, p. 9061, 2023, doi: 10.3390/su15119061.