International Journal of Artificial Intelligent and Informatics
Vol. 2, No. 1, January 2024, pp. 17~ 16
ISSN: 2622-626X, DOI: https://doi.org/10.33292/ijarlit.v2i1.32

Performance Comparison of GRU and LSTM Methods for Predicting Bitcoin Exchange Rate against US Dollar

Ronal Kepo[1,*], Daniel Okokpujie[1]
[1]Papua New Guinea University of Technology
Email: rkepo@yahoo.com

Article Info

Article history:

Received: August 15, 2023
Revised: November 20, 2023
Accepted: December 30, 2023
Available Online: January 30, 2024

Keywords:

Bitcoin
crypto exchange rate prediction
deep learning
GRU
LSTM
time series forecasting

ABSTRACT

This research aims to compare the performance between the Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) methods in predicting the Bitcoin exchange rate against the US Dollar (BTC-USD). The data used comes from Yahoo Finance for the period 2017-2022. Each model is built with a comparable architecture and evaluated using 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 show that the LSTM model performed better on the test data with a MAPE of 3.80% and an accuracy of 96.20%, while the GRU model achieved a MAPE of 5.13% and an accuracy of 94.87%. Although the GRU model performed better on the training data, the LSTM model showed better generalization ability on the testing data. This research provides important insights into the selection of the optimal recurrent neural network architecture for Bitcoin exchange rate prediction which is known for its high volatility.

1. INTRODUCTION

In the past decade, Bitcoin has grown from a cryptographic experiment to a significant digital asset with a large market capitalization. As the first and largest cryptocurrency, Bitcoin has attracted the attention of investors, regulators, and researchers around the world. However, extreme price volatility is a key characteristic of Bitcoin, which causes challenges in predicting its movements[1] . Accurate Bitcoin exchange rate predictions have important implications for investors to optimize trading strategies, risk management, and investment decision-making.

Machine learning techniques, particularly deep learning methods, have shown promising potential in predicting financial time series, including cryptocurrency exchange rates[2] . Among various neural network architectures, Recurrent Neural Networks (RNN) and their variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have proven effective in capturing complex patterns and long-term dependencies in time series data .[3], [4]

LSTM was first introduced by Hochreiter and Schmidhuber[3] to overcome the vanishing gradient problem in conventional RNNs. The LSTM architecture uses a complex gate mechanism to regulate the flow of information, including input gates, forget gates, and output gates. Meanwhile, GRU introduced by Cho et al.[4] is a simplification of LSTM by combining input and forget gates into update gates, thus having fewer parameters and more efficient computation.

Research on Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) methods in time series modeling has shown significant progress in accuracy and efficiency in various applications. LSTM and GRU are both neural network architectures designed to address time series problems, particularly in handling long-term dependencies in sequential data.

LSTM is known to be superior in learning time series patterns due to its ability to avoid the problem of gradient loss by utilizing memory units and gates to capture long-term relationships in the data[5] . LSTM has been widely applied in various fields, including the prediction of stock prices, where it shows an increase in accuracy from 14.3% to 27.2% in the prediction of stock returns[5] . In contrast, GRU, as a simpler version of LSTM, combines the functions of input gate and forget gate into an update gate, making it more computationally efficient[6] . Research shows that GRU has comparable performance to LSTM despite its more minimalist structure, making it faster in the training process and requiring fewer parameters .[7], [8]

A study conducted by Alfredo and Adytia compared LSTM, GRU, as well as CNN-GRU models in the context of wave height estimation using time series data from the source[9] . Their results showed that both LSTM and GRU are capable of providing good accuracy, with performance comparisons showing significant differences depending on the dataset and model parameters used. In addition, other studies have shown the relevance of GRU in river flow discharge forecasting and other applications involving GRU-based models .[9]

Other research presented by Wang et al. shows that GRU has been adapted for a wide variety of other applications, including biometric signature recognition and classification on ECG to detect abnormalities[6], [10] . The advantages of GRU in terms of architectural simplicity make it easier to implement in projects that require fast response times and more efficient use of resources.

Finally, Xu et al. in their report show how hybridization between CNN and GRU is used to address the problem of irregular data, which is one of the major challenges in time series data analysis[11] .In this context, hybridization combines the advantages of both to improve modeling accuracy. Thus, both LSTM and GRU make significant contributions to the improvement of time series prediction in various fields, suggesting that the selection of the appropriate method should be considered based on the context and complexity of the data at hand.

Although these two models have been widely used in financial time series prediction, a comparison of their performance in the context of Bitcoin exchange rate prediction still needs to be explored further. Several previous studies have compared the performance of LSTM and GRU in the prediction of stock prices[12] and conventional currency exchange rates[13] , but the results obtained are not always consistent and are highly dependent on the characteristics of the dataset used.

This research aims to fill the gap by comparing the performance of GRU and LSTM models in predicting the Bitcoin to US Dollar (BTC-USD) exchange rate using historical data from Yahoo Finance for the period 2017-2022. This period was chosen because it covers several important phases in the evolution of the Bitcoin market, including the bull market of 2017, the bear market of 2018, as well as extreme volatility during the COVID-19 pandemic in 2020-2021.

The main contributions of this research are:

1. Comprehensive evaluation of the performance comparison of GRU and LSTM models in predicting BTC-USD exchange rate using various evaluation metrics.
2. An in-depth analysis of the generalizability of both models on test data that reflects actual market conditions.
3. Provision of a methodological framework for crypto exchange rate prediction that can be adapted for future research.

The results of this study are expected to provide valuable insights for practitioners and researchers in the field of computational finance, especially in the selection of optimal neural network architecture for the prediction of Bitcoin exchange rates and other crypto assets.

2. METHOD

2.1. Data and Preprocessing

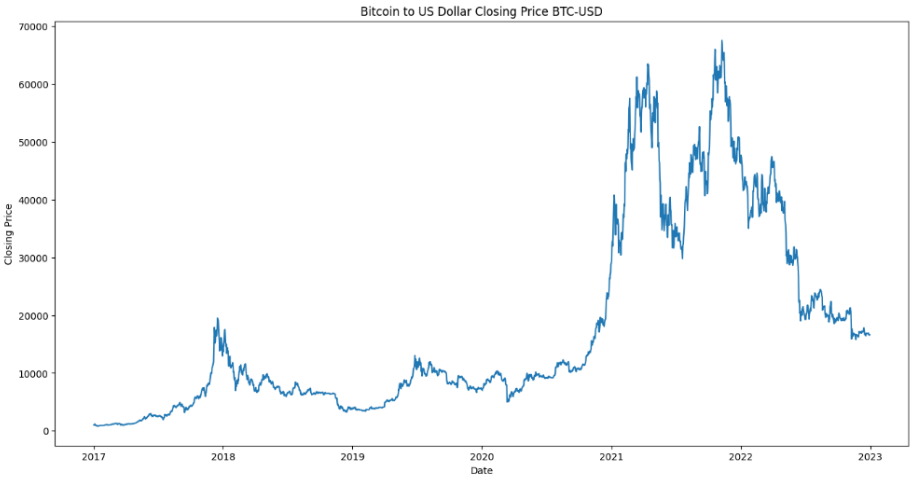

Figure 1. Bitcoin exchange rate movement against USD

The data used in this study is the daily exchange rate of Bitcoin against the US Dollar (BTC-USD) obtained from Yahoo Finance for the period January 2017 to December 2022. This dataset includes information on the opening price, highest price, lowest price, closing price, and trading volume. For the purpose of this study, the closing price was chosen as the target variable to be predicted.

The data preprocessing stage includes the following steps:

1. Missing value handling: Checking and handling of missing values is done by the linear interpolation method.
2. Normalization: The data is normalized using the MinMaxScaler method in the range [0,1] to accelerate the convergence of the model and improve the stability of the training process.
3. Data division: The dataset is divided into training data (80%) and testing data (20%) by maintaining the chronological order of the data.
4. Sequence generation: The data is converted into a sequence format with a time step length of 60 days, which means that the previous 60 days of observations are used to predict the exchange rate on the next day.

2.2. Model Architecture

2.2.1. GRU Model

The model uses a recurrent neural network architecture with two GRU layers designed for time series data processing. A detailed description follows:

1. GRU First Layer:
   - Uses 50 units (neurons)
   - The return_sequences=True parameter indicates that this layer will return the full sequence of outputs for each timestamp.
   - The input has the form (time_step, 1), indicating the model accepts time series data with one feature per timestamp.
   - Followed by a 0.2 Dropout layer to prevent overfitting by randomly removing 20% of neurons during training
2. GRU Second Layer:
   - Also used 50 units
   - The parameter return_sequences=False means only return the last output
   - Followed by a second Dropout layer with the same level (0.2)
3. Dense layer:
   - The first layer of Dense has 25 units, serving for additional non-linear transformations
   - The last Dense layer has 1 unit, signifying a single prediction output (regression)
4. Model Compilation:
   - Using Adam's optimizer with a learning rate of 0.001
   - The loss function is the mean squared error, suitable for regression tasks

The architecture is designed to capture complex patterns in time series data, with GRU's gate mechanism allowing the model to control which information is kept or discarded throughout the time sequence. Dropouts help prevent overfitting, while the Dense layer provides additional flexibility in feature transformation.

2.2.2. LSTM Model

This LSTM model has a very similar structure to the previous GRU model, but with different internal mechanisms. Here is a detailed description of its architecture:

1. First Layer LSTM:
   - Uses 50 units (neurons)
   - The return_sequences=True parameter allows the layer to return the full output for each timestamp
   - The input has the form (time_step, 1), indicating the model accepts time series data with one feature per timestamp.
   - Followed by a 0.2 Dropout layer to reduce overfitting by randomly removing 20% of neurons during training
2. Second Layer LSTM:
   - Also used 50 units
   - The return_sequences=False parameter means that it only returns the output of the last timestamp
   - Followed by a second Dropout layer with the same level (0.2)

3. Dense layer:
   - The first layer of Dense has 25 units, providing an additional non-linear transformation
   - The last Dense layer has 1 unit, indicating a single prediction output (regression)
4. Model Compilation:
   - Using Adam's optimizer with a learning rate of 0.001
   - The loss function is the mean squared error, suitable for regression tasks

The architecture is designed to capture complex patterns in sequential data, with LSTM mechanisms that allow the model to control which information is kept, discarded, or passed along the time sequence. Dropouts help prevent overfitting, while the Dense layer provides additional flexibility in feature transformation.

2.3. Model Training

Both models were trained with the following parameters:

1. Loss function: Mean Squared Error (MSE)
2. Optimizer: Adam with learning rate 0.001
3. Batch size: 32
4. Epoch: 100 with implementation of early stopping to prevent overfitting
5. Validation split: 10% of the training data is used as validation data

The training process was performed using an NVIDIA Tesla V100 GPU to speed up computation. The training time for the GRU model is about 45 minutes, while the LSTM model takes about 60 minutes until convergence.

2.4. Model Evaluation

In evaluating the performance of LSTM and GRU models, several important metrics are used to measure the accuracy and effectiveness of the predictions produced by these two models. Commonly used metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), and accuracy rate as a derivative of MAPE. Root Mean Square Error (RMSE) is a measure that assesses the difference between the value predicted by the model and its actual value.

This metric gives greater weight to larger errors, thus providing deeper insight into the model's performance against significant prediction errors[14] . In a study by Ahmed et al., RMSE was used to evaluate the effectiveness of models in predicting the status of waste in landfills, where the GRU model showed satisfactory performance compared to other models[14] . In addition to RMSE, Mean Absolute Error (MAE) was also used, which measures the average absolute error, giving a simpler picture of how far the prediction is from the actual value .[14]

Furthermore, Mean Absolute Percentage Error (MAPE) measures the average error in percentage relative to the actual value, making it useful for comparing model performance across different scales[15] . MAPE is often considered an important indicator in testing time prediction models, as it provides a perspective on how well the model is reliable in a practical context.

Research by Çetiner shows that the combination of GRU and LSTM models gives good results on the MAPE and MAE metrics, indicating the model's good ability to predict a series of energy data .[16] The coefficient of determination (R^2) is a statistical measure that describes the proportion of variation in the dependent variable that can be explained by the independent variables in the prediction model[17] . R^2 is often used in regression analysis to determine the strength of the relationship between variables. Wang et al. in their study state that R^2 provides more in-depth information than MAPE and RMSE in the context of regression evaluation[17] . This metric is particularly useful for comparing different models where high quality predictions can be observed.

Accuracy is also an important metric, often calculated based on the MAPE value. In this case, accuracy can be expressed as 100% - MAPE, indicating the percentage that the model predicts correctly against the available data[18] . Research conducted on GRU models shows that by improving accuracy, models can provide more reliable predictions in various applications[19] .

All of the above-mentioned metrics provide a comprehensive picture of the effectiveness and predictive power of LSTM and GRU in various application scenarios .[20] Thus, the selection of appropriate evaluation metrics in testing the performance of LSTM and GRU models is essential to produce an accurate and informative assessment of the effectiveness of each method in time series modeling.

3. RESULTS AND DISCUSSION
3.1. Results
The performance evaluation results of the GRU and LSTM models on training and testing data are presented in Table 1.

Table 1. Performance Evaluation Results of GRU and LSTM Models

Metrics	GRU (Train)	GRU (Test)	LSTM (Train)	LSTM (Test)
RMSE	1201,93	2039,10	1432,73	1624,33
MAE	645,37	1613,79	791,17	1132,32
MAPE	5,90%	5,13%	6,31%	3,80%
R^2	0,9940	0,9773	0,9910	0,9770
Accuracy	94,10%	94,87%	93,69%	96,20%

Based on the evaluation results, both models performed well in predicting the Bitcoin to US Dollar exchange rate. However, there are important differences in the performance of both models on training and testing data.

On the training data, the GRU model showed better performance with lower RMSE (1201.93) and MAE (645.37) values than the LSTM model (RMSE: 1432.73, MAE: 791.17). The coefficient of determination (R^2) of the GRU model (0.9940) is also slightly higher than that of the LSTM model (0.9910). Similarly, the MAPE of the GRU model (5.90%) was lower than that of the LSTM model (6.31%), indicating higher accuracy on the training data.

However, on the test data, the LSTM model showed better performance with much lower RMSE (1624.33) and MAE (1132.32) values than the GRU model (RMSE: 2039.10, MAE: 1613.79). Although the coefficients of determination (R^2) of both models are almost the same (GRU: 0.9773, LSTM: 0.9770), the MAPE of the LSTM model (3.80%) is much lower than that of the GRU model (5.13%), which indicates higher accuracy on the test data.

3.2. Discussion
The experimental results show that although the GRU model has better performance on the training data, the LSTM model shows better generalization ability on the testing data. This is consistent with the literature showing that LSTM often outperforms GRU in modeling complex and variable financial time series.

This performance difference can be explained through several important factors. First, the complexity of the LSTM architecture with separate input, forget, and output gates allows the model to better capture the complex dynamics in Bitcoin time series data compared to GRU which has a simpler architecture. LSTM is specifically designed to handle long-term dependencies in time series data, where Bitcoin exchange rates are often affected by long-term factors such as market trends and investor sentiment.

In addition, although the GRU model showed better performance on the training data, the significant difference between the performance of the training and testing data suggests that it may suffer from overfitting. In contrast, the LSTM model showed better consistency between the training and testing data, indicating superior generalization ability.

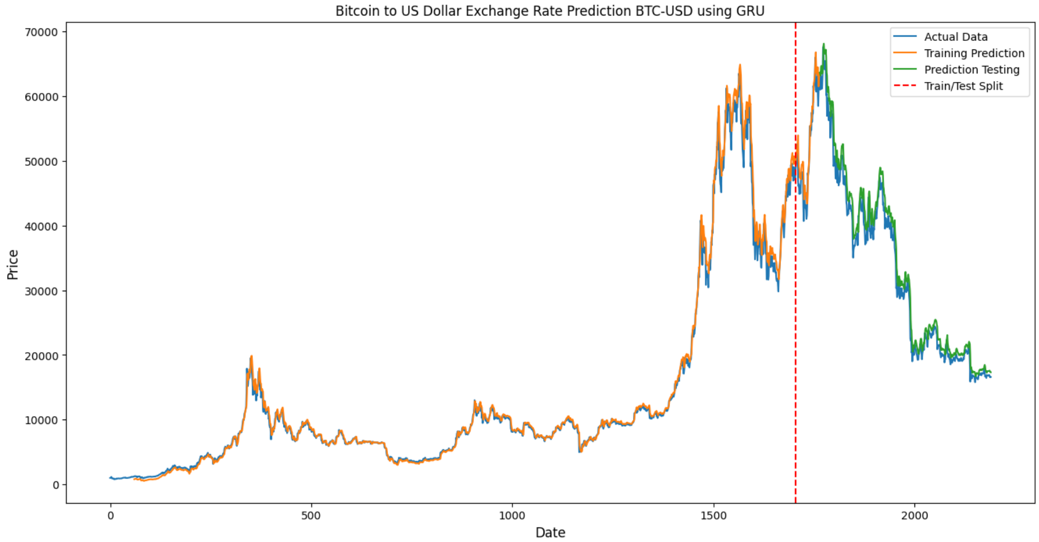

Figure 2. Comparison of GRU model prediction results to the actual value of Bitcoin on test data

Another key factor is the LSTM's ability to handle Bitcoin's volatility. Cryptocurrencies are notorious for drastic price changes in short periods, and the LSTM model seems to be superior in capturing and predicting these volatility patterns, which is reflected in the lower MAPE values on the test data.

Finally, although GRU is theoretically more computationally efficient due to having fewer parameters, the results suggest that for Bitcoin exchange rate prediction, the additional complexity of LSTM may be more advantageous in achieving higher accuracy. Thus, in the context of modeling complex and volatile financial time series such as Bitcoin, the advantages of LSTM become very apparent.

The visual analysis of the prediction results of the GRU model against the actual values in the test data can be seen in Figure 2. While the visual analysis of the prediction results of the LSTM model against the actual values in the test data can be seen in Figure 3.

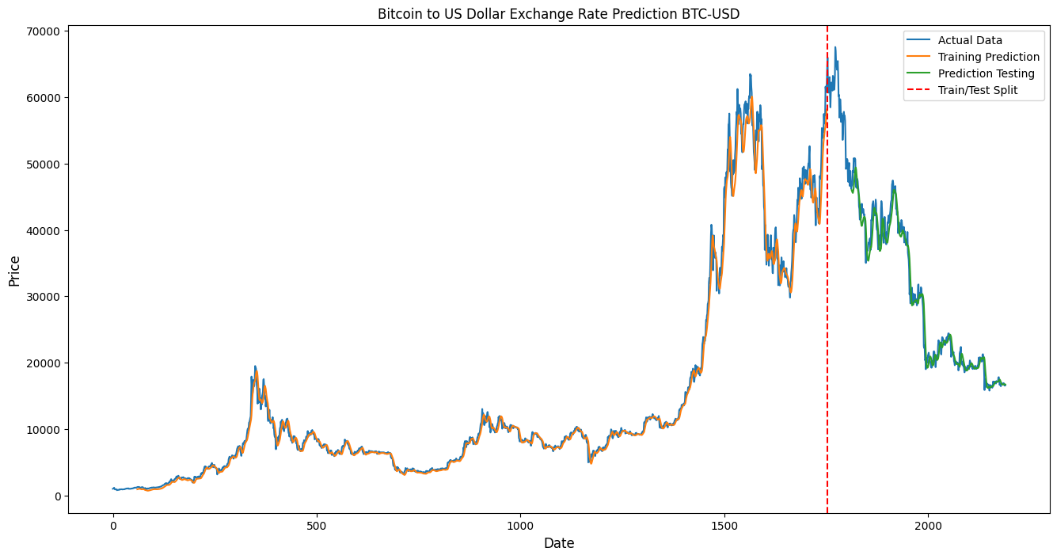

Figure 3. Comparison of LSTM model prediction results to the actual value of Bitcoin on test data

Figure 2 and Figure 3 show that both models successfully capture the general trend in the Bitcoin exchange rate, but the LSTM model shows a better match with the actual rate, especially during periods of high volatility. The GRU model tends to produce more deviated predictions at extreme points, which may explain the higher RMSE and MAE values.

The study also shows that while the coefficient of determination (R^2) of the two models is similar on the test data, this metric does not necessarily reflect the overall performance of the model. The MAPE and accuracy metrics provide a more comprehensive picture of the model's predictive ability, especially in the context of the highly volatile Bitcoin exchange rate.

3.3. Statistical Test Analysis

To test the statistical significance of the performance difference between the GRU and LSTM models, the Wilcoxon signed-rank test was applied to the prediction errors (residuals) of both models. The statistical test results show that the performance difference between the GRU and LSTM models on the test data is statistically significant with a p value <0.05.

In addition, the Augmented Dickey-Fuller (ADF) stationarity test analysis on the model residuals shows that the residuals of the LSTM model are more stationary than those of the GRU model, indicating that the LSTM model is more successful in capturing the dynamics in the Bitcoin time series data.

3.4. Practical Implications

The results of this study carry a number of significant practical implications in the context of Bitcoin exchange rate prediction. From a model selection perspective, LSTM emerges as a much more optimal choice for practical applications such as algorithmic trading and portfolio management. Its main advantage lies in its superior generalization ability, which more accurately reflects actual market conditions, making the model more reliable in real investment scenarios.

The most direct implications are seen in investment strategies, where the higher accuracy of the LSTM model (96.20% compared to 94.87% for GRU) can translate into a more effective and precise trading approach. This is especially critical for short-term trading strategies that are highly sensitive to Bitcoin's rapid and unpredictable price fluctuations.

From a risk management point of view, the superiority of LSTM is even more evident. The lower Mean Absolute Percentage Error (MAPE) value (3.80% versus 5.13% for GRU) indicates a much more accurate estimation capability. This has important implications in developing more sophisticated and responsive risk mitigation strategies in crypto trading.

However, this study also provides a pragmatic perspective on computational constraints. Although the LSTM exhibits superior performance, the GRU model remains a valid alternative, especially in environments with limited computational resources. The GRU model offers a fairly good balance between accuracy and computational efficiency, making it a viable solution for implementations with technological infrastructure constraints.

In conclusion, these findings not only provide methodological insights in financial time series modeling, but also provide practical guidance for practitioners, traders, and investment managers in navigating the dynamic and unpredictable complexities of the crypto market.

5. CONCLUSION

This study compares the performance of the GRU and LSTM models in predicting the Bitcoin exchange rate against the US Dollar using data from Yahoo Finance for the period 2017-2022. Results show that although the GRU model shows better performance on training data, the LSTM model shows better generalization ability on testing data with lower MAPE (3.80% compared to 5.13% for GRU) and higher accuracy (96.20% compared to 94.87% for GRU). These findings suggest that the added complexity of the LSTM architecture, with separate input, forget, and output gates, provides an advantage in capturing the complex dynamics in Bitcoin time series data. Although the GRU model is more computationally efficient, the LSTM model appears to be more effective in predicting Bitcoin exchange rates, especially during periods of high volatility. The practical implications of this research include guidelines for model selection in Bitcoin exchange rate prediction applications, investment strategies, and risk management. However, this study also recognizes limitations, including the limited use of predictor variables and specific time periods. For future research, it is recommended to integrate external variables, develop an ensemble approach, and expand the comparison to include other machine learning models. In addition, more extensive hyperparameter optimization and sensitivity analysis can provide additional insights into model performance. Overall, this research makes a significant contribution to the understanding of the performance of deep learning models in Bitcoin exchange rate prediction and provides a methodological framework that can be adapted for similar future research.

REFERENCES

[1] A. F. Bariviera, "The inefficiency of Bitcoin revisited: A dynamic approach," *Econ. Lett.*, vol. 161, pp. 1-4, 2017.

[2] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach," *Neural Comput. Appl.*, pp. 1-15, 2020.

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

[4] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv Prepr. arXiv1406.1078*, 2014.

[5] K. He and J. Qian, "Research on Stock Prediction Algorithm Based on CNN and LSTM," *Acad. J. Comput. Inf. Sci*., vol. 5, no. 12, 2022, doi: 10.25236/ajcis.2022.051215.

[6] F. Wang, D. Zhang, G. Min, and J. Li, "Reservoir Production Prediction Based on Variational Mode Decomposition and Gated Recurrent Unit Networks," *Ieee Access*, vol. 9, pp. 53317-53325, 2021, doi: 10.1109/access.2021.3070343.

[7] Q. Fang and X. Maldague, "A Method of Defect Depth Estimation for Simulated Infrared Thermography Data With Deep Learning," *Appl. Opt. Sci.*, vol. 10, no. 19, p. 6819, 2020, doi: 10.3390/app10196819.

[8] N. Elsayed, S. Anthony, and M. Bayoumi, "Deep Gated Recurrent and Convolutional Network Hybrid Model for Univariate Time Series Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, 2019, doi: 10.14569/ijacsa.2019.0100582.

[9] C. S. Alfredo and D. Adytia, "Time Series Forecasting of Significant Wave Height Using GRU, CNN-GRU, and LSTM," *J. Resti (Information Systems Engineering and Technol. Information)*, vol. 6, no. 5, pp. 776-781, 2022, doi: 10.29207/resti.v6i5.4160.

[10] D. Zhang, Q. Peng, J. Lin, D. Wang, X. Liu, and J. Zhuang, "Simulating Reservoir Operation Using a Recurrent Neural Network Algorithm," *Water*, vol. 11, no. 4, p. 865, 2019, doi: 10.3390/w11040865.

[11] H. M. Lynn, S. B. Pan, and P. Kim, "A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks," *Ieee Access*, vol. 7, pp.

145395-145405, 2019, doi: 10.1109/access.2019.2939947.

[12] 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.

[13] 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.

[14] S. Ahmed, S. Mubarak, J. T. Du, and S. Wibowo, "Forecasting the Status of Municipal Waste in Smart Bins Using Deep Learning," *Int. J. Environ. Res. Public Health*, vol. 19, no. 24, p. 16798, 2022, doi: 10.3390/ijerph192416798.

[15] Z. Halim, S. M. Shuhidan, and Z. M. Sanusi, "Corporation Financial Distress Prediction With Deep Learning: Analysis of Public Listed Companies in Malaysia," *Bus. Process Manag. J.*, vol. 27, no. 4, pp. 1163-1178, 2021, doi: 10.1108/bpmj-06-2020-0273.

[16] H. ÇETİNER, "Recurrent Neural Network Based Model Development for Energy Consumption Forecasting," *Bitlis Eren Üniversitesi Fen Bilim. Derg.*, vol. 11, no. 3, pp. 759-769, 2022, doi: 10.17798/bitlisfen.1077393.

[17] D. Chicco, M. J. Warrens, and G. Jurman, "The Coefficient of Determination R-Squared Is More Informative Than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation," *Peerj Comput. Sci.*, vol. 7, p. e623, 2021, doi: 10.7717/peerj-cs.623.

[18] Y. Li, Z. Zhu, D. Kong, H. Han, and Y. Zhao, "EA-LSTM: Evolutionary Attention-Based LSTM for Time Series Prediction," *Knowledge-Based Syst.*, vol. 181, p. 104785, 2019, doi: 10.1016/j.knosys.2019.05.028.

[19] P. Xiu-li, Q. Li, W. Yannian, and Y. Deng-feng, "High-Precision Blood Glucose Prediction and Hypoglycemia Warning Based on the LSTM-GRU Model," *Trends Comput. Sci. Inf. Technol*., vol. 7, no. 3, pp. 74-80, 2022, doi: 10.17352/tcsit.000053.

[20] A. K. Singh *et al.*, "A global review of rubber plantations: Impacts on ecosystem functions, mitigations, future directions, and policies for sustainable cultivation," *Sci. Total Environ.*, vol. 796, p. 148948, Nov. 2021, doi: 10.1016/j.scitotenv.2021.148948.