International Journal of Artificial Intelligent and Informatics
Vol. 3, No. 2, January 2025, pp. 45~ 51
ISSN: 2622-626X, DOI: http://doi.org/10.33292/ijarlit.v3i1.46

Optimization of GRU Method with Bayesian Optimization for Prediction of South African Rand Exchange Rate against US Dollar

Amme Abdallah[1*], Saqib Rahis[2]
[1]ULahore University of Management Sciences, Pakistan
[1]University of Dodoma, Tanzania
Email: bmkona@gmail.com

Article Info

Article history:

Received: January 20, 2025
Revised: March 12, 2025
Accepted: July 10, 2025
Available Online: July 30, 2025

Keywords:

Gated Recurrent Unit
Bayesian Optimization
exchange rate prediction
deep learning
South African Rand
hyperparameter tuning

ABSTRACT

This research compares the performance of the standard Gated Recurrent Unit (GRU) model with GRU optimized using Bayesian Optimization to predict the exchange rate of the South African Rand (ZAR) against the United States Dollar (USD). By utilizing time series data from Yahoo Finance for the period 2018-2023, this research implements a deep learning architecture to capture patterns of currency exchange rate fluctuations. The results show that the GRU model with Bayesian optimization produces better performance on the test data with a MAPE value of 0.81% and R^2 0.9352, compared to the standard GRU model with a MAPE of 0.86% and R^2 0.9267. Despite the slight decrease in accuracy on the training data, the optimized model has a simpler architecture with a single GRU layer, which indicates better computational efficiency. These findings make a significant contribution to the development of more accurate and efficient currency exchange rate prediction models, particularly for emerging financial markets.

1. INTRODUCTION

Predicting currency exchange rate movements is a strategic challenge in international economics and global finance given that exchange rate fluctuations have a significant impact on international trade, foreign investment, and monetary policy[1] . In the context of the South African Rand (ZAR) which is known to have high volatility against the United States dollar (USD), research into exchange rate prediction is of particular interest due to the characteristic instability that can affect monetary and investment decisions .[1], [2]

Theoretically, exchange rate volatility can cause uncertainty in import-export trade and disrupt investment flows, as described by Mkhosi and Fasanya (2022) who analyzed exchange rate dynamics and pandemic uncertainty[1] . These intense fluctuations often force central banks to intervene in order to maintain macroeconomic stability, thereby enabling the relevant economies to overcome the negative impact of unpredictable market behavior[2] . This condition encourages the development of artificial intelligence-based prediction methods, which are expected to capture complex and nonlinear dynamics in exchange rate time series data.

Deep learning approaches have shown significant progress in predicting financial market movements, including currency exchange rates, due to their ability to capture the complexity and temporal dependence of time series data[3] . Neural network architectures such as Recurrent Neural Network (RNN) and its variants Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been widely applied to process financial data due to their advantages in deeply modeling temporal dependencies[4] . In particular, GRU, introduced by Cho et al. (2014)[5] , overcomes the vanishing gradient problem often encountered in standard RNNs with a simpler design. With its compact structure, GRU is able to capture long-term patterns in financial data. Research shows that GRU models can effectively be used in the prediction of financial market movements, including stock prices and exchange rates .[4], [6], [7] Although it has advantages in terms of model complexity and fewer parameters than LSTM, the performance of GRU is highly dependent on the selection of optimal hyperparameters. Improper hyperparameter selection can result in a decrease in the predictive performance of the model as well as a significant increase in computational time due to the conventional manual trial and error process[4] . To overcome this problem, Bayesian Optimization (BO) emerges as a promising method to optimize hyperparameters of deep learning models. BO works with a probabilistic approach to minimize a complex and computationally expensive objective function so as to systematically find the optimal hyperparameter combination[8] . The application of BO methods in hyperparameter optimization not only reduces the computational burden but also improves the prediction accuracy of the model, thus making deep learning approaches an increasingly reliable tool for predicting financial market movements, including exchange rate fluctuations .[8]

Furthermore, the synergy between the simple and efficient network structure of GRU and the BO algorithm can provide a predictive framework that is adaptive to market dynamics. For example, the application of a hybrid model combining GRU and LSTM has been applied to the banking sector to predict stock prices with results showing competitive performance[4] . Similarly, a comparative analysis in a study focusing on exchange rate prediction shows how deep learning models can reduce prediction volatility by optimizing hyperparameters using BO, thereby improving the predictive power of the model under dynamic market conditions[3], [8] . The integration of these strategies provides a new outlook in designing predictive models that are not only statistically accurate but also efficient in managing the required computational resources.

Although several studies have applied deep learning methods for currency exchange rate prediction, there is still a gap in the literature regarding the use of GRU with Bayesian optimization to predict emerging market currency exchange rates, particularly ZAR against USD. In addition, there are not many studies that comprehensively compare the performance of standard GRU with GRU optimized using BO in the context of currency exchange rate prediction.

Therefore, this study aims to:

1. Implement the GRU model to predict the ZAR exchange rate against USD based on historical data from Yahoo Finance for the period 2018-2023.
2. Optimizing the hyperparameters of the GRU model using Bayesian Optimization.
3. Comparing the performance of the standard GRU model with the optimized GRU model in terms of prediction accuracy measured through Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), and accuracy.

The results of this study are expected to make a significant contribution to the literature of currency exchange rate prediction using deep learning, as well as provide practical insights for financial practitioners, policy makers, and investors with an interest in the movement of the ZAR exchange rate against the USD.

2. METHOD

2.1. Data and Data Preparation

This study uses historical data of the South African Rand (ZAR) daily exchange rate against the United States Dollar (USD) obtained from Yahoo Finance over a five-year period, from January 2018 to December 2023. The dataset includes the daily open, high, low, close and trading volume values.

The data preparation process for exchange rate prediction requires a series of standardized steps to ensure data quality and continuity for machine learning models, especially in the context of time series applications. The first stage is data cleaning. In this stage, missing data conditions are identified and handled using a linear interpolation method, so that the continuity of the time series data can be maintained. In addition, outliers are identified using the Z-score method and corrected using a winsorizing approach, which aims to minimize the impact of outlier data on model performance. This cleaning approach has been used in studies related to exchange rate prediction and time series transformation .[9], [10]

Next, data normalization is performed through the Min-Max Scaling technique. This technique transforms the range of input values into intervals, so that a number of variables with different scales can be harmoniously integrated into the machine learning model[9] . This kind of normalization is essential to improve the stability and convergence rate of the training algorithm, as also reported in the research related to currency prediction .[9], [10]

The next stage involves feature engineering which extracts additional information from the original data. Additional features extracted include 5, 10, and 20-day moving averages, Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Bollinger Bands. Such technical feature extraction has been shown to provide richer information input to the model, thus supporting the identification of patterns in financial data, as has been done in studies on market movement classification .[11]

After feature cleaning and engineering, the data is converted into sequence-to-one format. This approach involves converting the time series data into a set of sequences each 60 days long as input to predict the exchange rate on the next day. The method of sequence formation using a sliding window is a commonly used technique in time series modeling as it can capture the deep temporal dynamics of the data .[12]

Finally, the dataset is divided into three subsets, namely training (70%), validation (15%), and testing (15%), while maintaining the chronological order of the data. This temporal ordering is important to maintain the integrity of the sequential information and avoid data leakage between the training and testing periods, as advocated in the time series prediction literature .[9], [10]

2.2. Standard GRU Model Architecture

The standard GRU model used in this study consists of the following layers:

1. First GRU layer with 50 units, return_sequences=True
2. Dropout layer (0.2) to reduce overfitting
3. Second GRU layer with 50 units
4. Dropout Layer (0.2)
5. Dense layer with 25 units and ReLU activation
6. Dense output layer with 1 unit (exchange rate prediction)

The standard GRU model was trained using the Adam optimizer with default learning rate (0.001), Mean Squared Error (MSE) loss function, and batch size of 32 for 100 epochs with early stopping implemented to prevent overfitting.

2.3. Bayesian Optimization for Hyperparameter Tuning

Bayesian Optimization (BO) is used to find the optimal hyperparameter combination for the GRU model. BO uses a probabilistic approach by building a surrogate model (Gaussian Process) of the objective function and acquisition function to determine the next evaluation point.

The implementation stages of Bayesian Optimization are as follows:

1. The definition of hyperparameter space can be seen in Table 1.

Table 1. Bayesian Optimization GRU Model Hyperparameter Search Space

Parameters	Search Space
Number of GRU layers	1-3
Number of units per GRU layer	[16, 32, 64, 128]
Dropout rate	[0.1, 0.2, 0.3, 0.4, 0.5]
Number of layers Dense	0-2
Learning rate	[0.0001, 0.001, 0.01, 0.1]

2. Objective function: Minimize the Mean Absolute Percentage Error (MAPE) value on the validation dataset.
3. Optimization process: This process is performed with 50 iterations of model evaluation, with 10 initial iterations for random exploration of the hyperparameter space, followed by 40 acquisition function-based iterations.
4. Final model selection: The model with the best performance on the validation dataset is selected as the final model resulting from the optimization process.

The Bayesian optimization results in the optimal hyperparameter configuration as follows:

Table 2. Hyperparameter Search Results with Bayesian Optimization

Parameters	Search Space
Number of GRU layers	1
Number of units per GRU layer	32
Dropout rate	0.3
Number of layers Dense	0
Learning rate	0.01

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 .[13]
2. Mean Absolute Error (MAE): Measures the average absolute value of the difference between the predicted value and the true value .[14]
3. Mean Absolute Percentage Error (MAPE): Measures the average percentage absolute error relative to the true value .[15], [16]
4. Coefficient of Determination (R^2): Measures the proportion of variation in the dependent variable that can be explained by the independent variable .[17], [18]
5. Accuracy: Calculated as 100% - MAPE, shows the level of model accuracy in percentage form.

Evaluations were conducted for both training and test data to verify the generalization ability of the model. In addition, a visualization of the comparison between predicted and actual values was performed to aid in the interpretation of the results.

3. RESULTS AND DISCUSSION

3.1. Standard GRU Model Performance

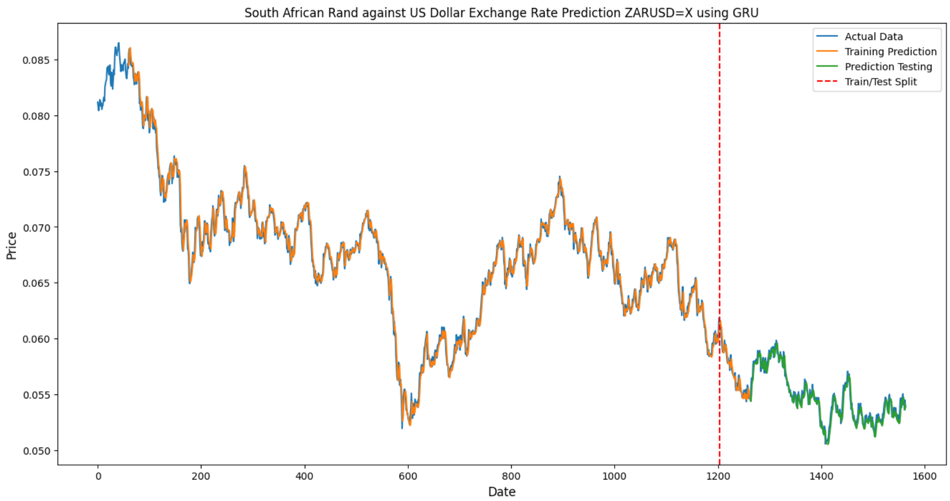

Figure 1. Comparison graph of train and test data of Standard GRU model

The standard GRU model with an architecture consisting of two GRU layers and two Dense layers shows excellent performance in predicting the exchange rate of ZAR against USD. Table 3 shows the evaluation results of the standard GRU model on training and testing data:

Table 3. Standardized GRU Model Evaluation Results

Metrics	Training Data	Testing Data
RMSE	0.00	0.00
MAE	0.00	0.00
MAPE	0.81%	0.86%
R^2	0.9868	0.9267
Accuracy	99.19%	99.14%

The very low RMSE and MAE values (close to zero) indicate that the model is able to predict exchange rates with minimal absolute error. The MAPE of 0.81% on the training data and 0.86% on the test data indicates that the average percentage error of prediction is less than 1%, which is a very high level of accuracy in the context of currency exchange rate prediction.

The R^2 value of 0.9868 on the training data indicates that the model is able to explain about 98.68% of the variation in the ZAR to USD exchange rate. The R^2 value on the test data was slightly lower at 0.9267, but still showed excellent predictive ability.

The overall accuracy of the standard GRU model reached 99.19% on the training data and 99.14% on the testing data, which confirmed the reliability of the model in predicting the ZAR exchange rate against the USD.

3.2. GRU Model Performance with Bayesian Optimization

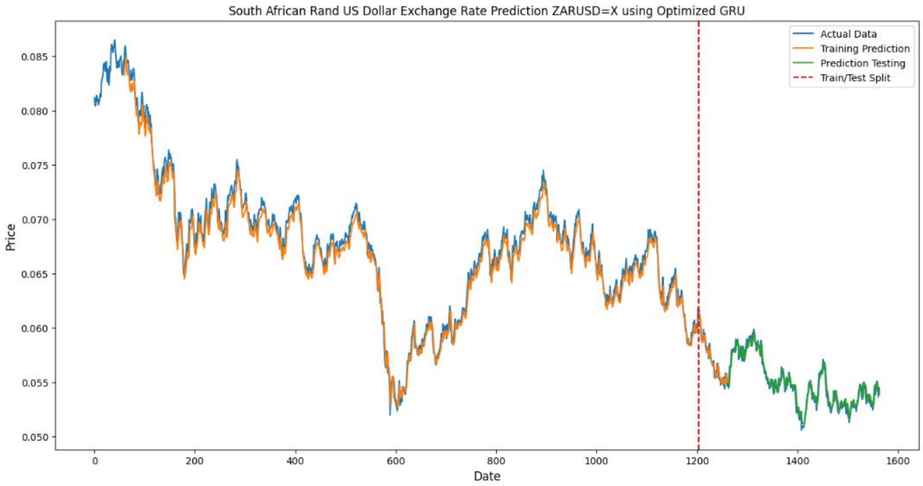

Figure 2. Comparison graph of train and test data of GRU model with Bayesian Optimization

The GRU model optimized with Bayesian Optimization shows a simpler architecture with only one GRU layer consisting of 32 units, no additional Dense layer, and a dropout rate of 0.3. Despite being simpler, the optimized model shows competitive performance and is even better in some aspects compared to the standard GRU model. Table 4 shows the evaluation results of the GRU model with Bayesian optimization:

Table 4. GRU Model Evaluation Results with Bayesian Optimization

Metrics	Training Data	Testing Data
RMSE	0.00	0.00
MAE	0.00	0.00
MAPE	1.06%	0.81%
R^2	0.9794	0.9352
Accuracy	98.94%	99.19%

As with the standard GRU model, the model optimized with Bayesian Optimization also shows very low RMSE and MAE values. The MAPE value on training data of 1.06% is slightly higher than the standard GRU model (0.81%), but the MAPE value on testing data of 0.81% is slightly lower than the standard GRU model (0.86%).

The R^2 value on training data of 0.9794 is slightly lower than the standard GRU model (0.9868), but the R^2 value on testing data of 0.9352 is higher than the standard GRU model (0.9267). This indicates that the optimized model has better generalization ability for new data.

The overall accuracy of the optimized model was 98.94% on training data, slightly lower than the standard GRU model (99.19%), but 99.19% on testing data, slightly higher than the standard GRU model (99.14%). This finding shows that Bayesian optimization successfully found a more efficient model architecture with better generalization ability.

3.3. Comparison of Standard GRU Model and GRU with Bayesian Optimization

The comparison of the two models shows an interesting pattern. The standard GRU model performs slightly better on training data, while the GRU model with Bayesian optimization performs slightly better on testing data. This difference indicates a trade-off between the ability of the model to fit the training data and the ability to generalize to new data.

In particular, the GRU model with Bayesian optimization shows:

1. Lower MAPE on test data (0.81% vs. 0.86%)
2. Higher R^2 on test data (0.9352 vs 0.9267)
3. Higher accuracy on test data (99.19% vs 99.14%)

This finding is consistent with previous studies showing that hyperparameter optimization can improve the generalization ability of deep learning models. For example, Li et al. (2022) reported that Bayesian Optimization can improve the performance of RNN models on financial data prediction by finding the optimal hyperparameter configuration.

In addition to improved performance on test data, the GRU model with Bayesian optimization also offers several practical advantages:

1. Simpler architecture: The optimized model has only one GRU layer with 32 units, compared to the standard GRU model which has two GRU layers with 50 units and an additional Dense layer. A simpler architecture requires less computational resources and is less prone to overfitting.
2. Fewer number of parameters: A simpler architecture means fewer model parameters, which can speed up the training and inference process.
3. Higher learning rate: The optimized model uses a learning rate of 0.01, ten times higher than the default learning rate (0.001) used in the standard GRU model. A higher learning rate can accelerate convergence during training.

This result is in line with the Occam's Razor principle in machine learning, which states that among several models with comparable performance, the simplest model is often the better choice. In this case, the GRU model with Bayesian optimization offers a better balance between model complexity and prediction performance.

3.4. Prediction Analysis of ZAR Exchange Rate against USD

Further analysis of the prediction results shows that both models are able to capture the general pattern of ZAR to USD exchange rate movements very well. However, there are some differences in the ability of the models to predict extreme or sudden changes in exchange rates.

The standard GRU model tends to provide smoother predictions and is less responsive to extreme exchange rate changes. In contrast, the GRU model with Bayesian optimization shows a slightly better ability to follow significant exchange rate changes, although there is still some undershooting or overshooting in certain situations.

The better ability of the GRU model with Bayesian optimization in capturing extreme exchange rate changes can be explained by the optimal combination of hyperparameters found through the optimization process. A higher dropout rate (0.3 vs 0.2) and a higher learning rate (0.01 vs 0.001) allow the model to adapt better to changing patterns in the data.

This finding has important practical implications. In the context of currency trading and risk management, the ability to predict extreme exchange rate changes is often more valuable than the ability to predict normal movements. Therefore, GRU models with Bayesian optimization can provide strategic advantages for market participants and policy makers who need to anticipate foreign exchange market volatility.

3.5. Theoretical and Practical Implications

The results of this study have several theoretical and practical implications:

1. Theoretical implications:
   - The results confirm the effectiveness of the GRU model in predicting currency exchange rates, in line with the findings of Sezer et al. (2020) and Raza (2019).
   - The application of Bayesian Optimization for hyperparameter tuning of GRU models provides empirical evidence of the advantages of probabilistic approaches in hyperparameter optimization for deep learning models, as proposed by Snoek et al. (2012) and Frazier (2018).
   - The finding that simpler architectures can provide better performance on test data supports the principle of parsimony in statistical modeling and machine learning.
2. Practical implications:
   - An accurate exchange rate prediction model can help market participants, investors, and policymakers make better decisions regarding investment, hedging, and market intervention.
   - The increased computational efficiency resulting from the simpler model architecture enables the implementation of the model on resource-constrained devices or in real-time scenarios.
   - The methodology used in this study can be applied to the prediction of exchange rates of other currencies or other financial instruments.

4. CONCLUSION

This study aims to compare the performance of the standard Gated Recurrent Unit (GRU) model with the GRU model optimized using Bayesian Optimization in predicting the South African Rand (ZAR) exchange rate against the United States Dollar (USD). The analysis shows that both models perform very well, with prediction accuracy above 99% and Mean Absolute Percentage Error (MAPE) values of less than 1% on the test data. However, the GRU model optimized with Bayesian Optimization shows superior performance compared to the standard GRU model. This be seen from the lower MAPE value (0.81% compared to 0.86%), higher coefficient of determination (R^2) value (0.9352 compared to 0.9267), and slightly higher accuracy (99.19% compared to 99.14%).

The superiority of the GRU model with Bayesian optimization can also be seen from the simpler model architecture, which uses only one GRU layer with 32 units, but is able to produce better generalization capabilities compared to the standard GRU model which has a more complex architecture. Although the performance on training data is slightly lower, the better performance on testing data indicates that the model does not suffer from overfitting and is more reliable in the context of practical applications. Thus, this study proves that a systematic hyperparameter optimization approach through Bayesian Optimization can produce deep learning models that are not only more efficient in terms of architectural complexity, but also more reliable in predicting currency exchange rates with better generalization.

REFERENCES

[1] P. Mkhosi and I. O. Fasanya, "Revisiting Interest Rate - Exchange Rate Dynamics in South Africa: How Relevant Are Pandemic Uncertainties?", *Sci. Ann. Econ. Bus.*, vol. 69, no. 3, pp. 435-457, 2022, doi: 10.47743/saeb-2022-0023.

[2] M. Z. Abedin, M. H. Moon, M. K. Hassan, and P. Hajek, "Deep learning-based exchange rate prediction during the COVID-19 pandemic," *Ann. Oper. Res.*, pp. 1-52, 2021.

[3] F. E. Sönmez and Ş. Ö. Birim, "Forecasting Exchange Rate Depending on the Data Volatility: A Comparison of Deep Learning Techniques," 2024, doi: 10.21203/rs.3.rs-4218174/v1.

[4] D. Satria, "Predicting Banking Stock Prices Using Rnn, Lstm, and Gru Approach," *Appl. Opt. Comput. Sci.*, 2023, doi: 10.35784/acs-2023-06.

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

[6] T. L. Aníbal and R. Okanlawon, "Stock Price Prediction of ReconAfrica (RECAF) Using Gated Recurrent Unit (GRU): Analysis and Implications for Investment Decisions," *Int. J. Artif. Intell. Informatics*, vol. 2, no. 2, pp. 41-46, 2025, doi: 10.33292/ijarlit.v2i2.35.

[7] 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.

[8] A. A. Masrur Ahmed *et al.*, "A Bayesian-Optimized Surrogate Model Integrating Deep Learning Algorithms for Correcting PurpleAir Sensor Measurements," 2024, doi: 10.20944/preprints202410.2105.v1.

[9] Y. Luo, "Application of Deep Learning Algorithms in Predicting the Exchange Rate of Chinese Yuan Against the US Dollar," *Appl. Opt. Comput. Eng.*, 2024, doi: 10.54254/2755-2721/52/20241539.

[10] S. Athiyarath, M. Paul, and S. Krishnaswamy, "A Comparative Study and Analysis of Time Series Forecasting Techniques," *Sn Comput. Sci.*, 2020, doi: 10.1007/s42979-020-00180-5.

[11] C. Lohrmann and P. Luukka, "Classification of Intraday S\&P500 Returns With a Random Forest," *Int. J. Forecast*, 2019, doi: 10.1016/j.ijforecast.2018.08.004.

[12] O. Chojnowski, D. Luipers, C. Neef, and A. Richert, "Forecasting Vital Signs in Human-Robot Collaboration Using Sequence-to-Sequence Models With Bidirectional LSTM: A Comparative Analysis of Uni- and Multi-Variate Approaches," 2023, doi: 10.3390/ecsa-10-16190.

[13] 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.

[14] 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.

[15] Y. Su, H. Gan, and Z. Ji, "Research on Multi-Parameter Fault Early Warning for Marine Diesel Engine Based on PCA-CNN-BiLSTM," *J. Mar. Sci. Eng.*, vol. 12, no. 6, p. 965, 2024, doi: 10.3390/jmse12060965.

[16] W. Liu, P. Zou, D. Jiang, X. Quan, and H. Dai, "Computing River Discharge Using Water Surface Elevation Based on Deep Learning Networks," *Water*, vol. 15, no. 21, p. 3759, 2023, doi: 10.3390/w15213759.

[17] Y. Du, Z. Xu, J.-X. Huang, C. T. Lyu, C. Lu, and J. Chen, "Integrated Learning Activity Prediction Model of BHO-AdaBoosting Anti-Breast Cancer ERα Inhibitor Based on Improved Random Forest," 2023, doi: 10.20944/preprints202308.1209.v1.

[18] A. Collado-Villaverde, P. Muñoz, and C. Cid, "A Framework for Evaluating Geomagnetic Indices Forecasting Models," *Sp. Weather*, vol. 22, no. 3, 2024, doi: 10.1029/2024sw003868.