Available online at https://ejournal.corespub.com/index.php/ijmsc/index

International Journal of Mathematics, Statistics, and Computing

e-ISSN 3025-0803

Vol. 3, No. 3, pp. 113-119, 2025

Comparison of Activation Functions in Recurrent Neural Network for Litecoin Cryptocurrency Price Prediction

Moch Panji Agung Saputra[1*], Astrid Sulistya Azahra[2], Dede Irman Pirdaus[3]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia*
[2]*Doctoral of Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia*
[3]*Informatics Study Program, Faculty of Technology and Informatics, Universitas Informatika dan Bisnis Indonesia, Bandung, Indonesia*

**Corresponding author email: moch16006@mail.unpad.ac.id*

Abstract

The rapid advancement of information technology and digitalization has significantly transformed the financial sector, particularly with the emergence of cryptocurrencies characterized by high price volatility and complex movement patterns. Accurate price prediction of these crypto assets is essential to support investment decision-making and risk management. This study aims to compare the performance of six activation functions ReLU, Tanh, Sigmoid, Softplus, Swish, and Mish in a Simple Recurrent Neural Network (RNN) model for predicting the price of Litecoin, a widely traded cryptocurrency. Using historical daily closing price data from May 2020 to April 2025, the data were preprocessed through Min-Max normalization and sliding window sequence formation to fit the RNN input requirements. Each activation function was applied in the RNN model under consistent training conditions, and model performance was evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2). Results indicate that the Swish activation function outperforms others by achieving the lowest RMSE of 4.58 and the highest R^2 score of 0.9578, demonstrating superior prediction accuracy and stable convergence. Tanh also showed competitive results, while Sigmoid and Softplus performed less effectively. In conclusion, Swish is recommended as the most suitable activation function for RNN-based cryptocurrency price forecasting due to its balance of accuracy and computational efficiency.

Keywords: Cryptocurrency price prediction, recurrent neural network, activation functions, Litecoin, time series forecasting.

1. Introduction

The development of information technology and digitalization has changed many aspects of life, including the financial sector (Okinrinola et al., 2024). One of the most important innovations is the emergence of crypto assets based on blockchain technology. Crypto assets have characteristics of high price volatility and complex movement patterns, thus requiring accurate prediction methods to assist in investment decision making and risk management (Dudek et al., 2024; Micu & Dumitrecus, 2022).

Indonesia has experienced significant growth in cryptocurrency adoption reflecting the digital transformation of the financial sector in developing countries. Data shows that Indonesia ranks third globally in cryptocurrency adoption rates with a penetration of 10.3% of the total population (28.5 million out of 277 million people), and ranks fifth in the world in the speculative cryptocurrency investment category with a contribution of 3.96% to total global speculative transactions. This phenomenon can be explained through three determinant factors, demographic structure with more than 50% of the population aged 18-39 years who have a high risk tolerance for alternative investment instruments.

Various methods of crypto asset price prediction have been developed, one of which is by using Recurrent Neural Network (RNN). RNN has special capabilities in handling time series data, because it can remember historical information and temporal patterns in the data (Sherstinsky, 2020; Wen & Ling, 2023). RNN is widely applied in financial time series modeling, including stock price prediction, as an effective tool to anticipate future price changes.

Research on activation functions in artificial neural networks has been widely conducted, considering the crucial function in determining how well the model is able to map complex relationships in the data. One relevant study was conducted by Firmansyah & Hayadi (2023), who compared the ReLU and Tanh activation functions in the Multilayer Perceptron (MLP) model in the Titanic dataset classification. The results of the study showed that ReLU consistently performed better than Tanh, in terms of average accuracy and precision.

Another study by Pipin et al., (2023) emphasized the potential use of Recurrent Neural Network models, especially the Long Short-Term Memory (LSTM) variant, in stock price prediction. In the study, RNN-LSTM was optimized using the Adaptive Moment Estimation (Adam) algorithm to improve prediction accuracy. By utilizing historical stock price data and other technical factors, this model is able to recognize temporal dependencies and complex nonlinear relationships in financial time series data. The test results showed that the model produced a low mean square error (MSE) value of 0.0109012 and a Mean Percentage Error (MPE) of only 1.74%, indicating that the model has good prediction accuracy. This study confirms the relevance and effectiveness of the RNN architecture, especially in the context of predicting dynamic and volatile financial assets.

Although various studies have explored the capabilities of Recurrent Neural Network (RNN), few studies have specifically compared various activation functions in the context of pure RNN for crypto asset price prediction. Activation functions play a central role in influencing the performance of neural network models, but in-depth comparisons of functions such as ReLU, Tanh, Sigmoid, Softplus, Swish, and Mish in RNN architecture are still limited, especially for crypto asset data such as Litecoin. Based on this gap, this study aims to conduct a comparative analysis of the six activation functions in the Recurrent Neural Network (RNN) model in predicting the price of the crypto asset Litecoin.

2. Literature Review

The use of artificial neural networks, particularly Recurrent Neural Networks (RNNs), in time series forecasting has attracted considerable attention due to their ability to capture temporal dependencies and nonlinear patterns in sequential data (Sherstinsky, 2020). Several studies have demonstrated the effectiveness of RNNs in various financial prediction contexts, such as stock prices, forex, and cryptocurrency markets, where traditional statistical models often fail to capture dynamic market behavior.

In the domain of cryptocurrency, previous studies have utilized deep learning methods such as RNN and its variants like LSTM and GRU for price forecasting. For instance, Choudhary et al. (2021) employed LSTM networks to predict Bitcoin prices and reported high prediction accuracy compared to traditional ARIMA models. Similarly, Patel and Thakkar (2022) integrated technical indicators and LSTM-based deep learning architectures for Ethereum price forecasting, showing the benefits of capturing long-term dependencies in volatile financial data.

In terms of activation functions, their selection critically influences neural network performance by affecting gradient propagation, convergence behavior, and the model's ability to learn complex patterns. Although commonly used functions like ReLU and Tanh have been extensively studied in general classification and regression tasks (Agarap, 2018), recent research emphasizes the importance of exploring newer activation functions like Swish and Mish, which exhibit smoother gradients and better non-linearity properties (Ramachandran et al., 2017; Misra, 2019).

Despite the growing use of RNN models in financial forecasting, few studies have provided direct, comprehensive comparisons of multiple activation functions within a pure RNN architecture particularly for cryptocurrency price prediction. Prior work by Verma & Gupta (2023) evaluated ReLU, Tanh, and Sigmoid in a GRU network for Bitcoin forecasting, showing varying performance depending on the evaluation metric used. However, their study did not include newer activation functions like Swish and Mish, and focused solely on GRU-based architecture.

Moreover, most comparative studies on activation functions have focused on classification datasets or static image data (e.g., MNIST, CIFAR), which differ significantly from financial time series due to their sequential nature and noise level. Therefore, evaluating activation functions such as Softplus, Swish, and Mish in an RNN setting for predicting a volatile asset like Litecoin remains a relatively unexplored area.

This study aims to address this gap by systematically comparing six activation functions ReLU, Tanh, Sigmoid, Softplus, Swish, and Mish under consistent RNN settings and training conditions. By focusing on Litecoin price data, the study not only contributes to deep learning-based cryptocurrency forecasting but also offers insights into the activation functions that yield the most robust and accurate performance for time series modeling in highly dynamic financial environments.

3. Methodology

3.1. Data Collection

The data used in this study is historical data on the daily closing price of the Litecoin (LTC) crypto asset against the United States dollar (USD). Data was taken from the Yahoo Finance site with the help of the yfinance library in the Python programming language. The data time range starts from May 1, 2020 to April 30, 2025, so it covers the last five years of Litecoin price movements. The data taken is focused on the Close column, which represents the daily closing price and is considered the most relevant in the context of price predictions in the financial market.

3.2. Data Pre-processing

After the data collection process is carried out, the next stage is data pre-processing which aims to improve the quality of input into the Recurrent Neural Network (RNN) prediction model. The first stage of pre-processing is the removal of missing values in the closing price data. These missing values can appear due to trading holidays or technical constraints in retrieval of historical data (John et al., 2019; Fernández-Navarro and Carbonero-Ruz, 2025). The removal of missing values is done so that the model does not receive invalid input that can interfere with the training process. The next stage is data normalization using the Min-Max Scaling method so that all values are in the range of 0 to 1. This normalization is important in machine learning because it can speed up the convergence process and increase the stability of weight calculations during the training process (Jo, 2019). The normalization formula used is as follows:

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where x is the original value of the closing price, $x_{\min}$ is the minimum value in the dataset, and $x_{\max}$ is the maximum value in the dataset. After the data is normalized, the process of forming windowed sequences is carried out using the sliding window approach. Each model input will consist of 60 previous closing price data to predict the closing price on the 61st day (Rouf et al., 2021). This process automatically transforms one-dimensional data into a three-dimensional form required by the RNN layer.

3.3. Model Development and Training

At this stage, a Recurrent Neural Network (RNN) model is built using the SimpleRNN architecture to predict the price of the Litecoin crypto asset (Tumpa & Maduranga, 2024). The model is built using the TensorFlow and Keras frameworks, with a two-layer architecture: one RNN layer with 64 units and one Dense layer as a single output. To test the effect of activation functions on prediction performance, the model is tested using six different activation functions, namely ReLU, Tanh, Sigmoid, Softplus, Swish, and Mish. To avoid overfitting, the Early Stopping technique is applied by monitoring the val loss value, and training will be stopped if there is no improvement after 10 consecutive epochs. The model is trained for a maximum of 100 epochs with a batch size of 32 and using 10% of the training data as validation data. All models are trained and evaluated using pre-processed historical data to ensure consistency of comparison between activations. The following is a summary of the model configuration used:

Table 1: Model configuration

Component	Specification
Architecture	SimpleRNN
Number of RNN Units	64
Activation Functions	ReLU, Tanh, Sigmoid, Softplus, Swish, Mish
Optimizer	Adam
Loss Function	Mean Squared Error (MSE)
Maximum Epochs	100
Batch Size	32
Regularization Technique	Early Stopping
Validation Split	10%
Time Steps	60
Input Shape	(60, 1)

3.4. Evaluation

To assess the performance of the Recurrent Neural Network (RNN) models using various activation functions, several regression evaluation metrics were employed, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2 score). These metrics provide a comprehensive understanding of both the magnitude and proportion of prediction errors.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|f_t - y_t| \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \times 100\% \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{5}$$

Each activation function ReLU, Tanh, Sigmoid, Softplus, Swish, and Mish was evaluated using these metrics on the test dataset after training. This multi-metric evaluation approach allows for a robust comparison in terms of accuracy, scale sensitivity, and explanatory power of the model. The activation function that consistently achieves lower RMSE, MAE, and MAPE, and a higher R^2 score, is considered to provide better predictive performance for the time series forecasting task.

4. Results and Discussion

4.1. Activation Function Performance Evaluation

Evaluation was conducted on six activation functions used in the RNN model, namely ReLU, Tanh, Sigmoid, Softplus, Swish, and Mish. This evaluation refers to the prediction performance metrics, namely Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2). In addition, training time and prediction time were also recorded to measure the efficiency of each model. Quantitative evaluation data are shown in Table 2.

Table 2: Evaluation results of all activation functions

Activation Function	RMSE	MAE	MAPE	R^2 score	Prediction time (s)
ReLU	4.65	3.05	3.30%	0.9564	0.90
Tanh	4.59	2.99	3.23%	0.9575	0.77
Sigmoid	4.80	3.15	3.40%	0.9536	0.71
Softplus	5.00	3.35	3.65%	0.9496	0.43
Swish	4.58	3.01	3.24%	0.9578	0.48
Mish	4.67	3.10	3.30%	0.9560	1.36

Based on the evaluation results of the six activation functions tested within the Recurrent Neural Network (RNN) model for predicting Litecoin price, the Swish activation function demonstrated the most optimal performance across nearly all evaluation metrics. It achieved the lowest RMSE value of 4.58, which indicates the smallest deviation between predicted and actual values. It also yielded a high R^2 score of 0.9578, suggesting a strong correlation and high explanatory power of the model when using this function. Although the MAE (3.01) and MAPE (3.24%) were slightly higher than those of Tanh, the differences were minimal and within an acceptable margin of variation. Additionally, Swish maintained a fast prediction time of 0.48 seconds, making it not only accurate but also computationally efficient.

The Tanh function followed closely, with an RMSE of 4.59, the lowest MAE of 2.99, and a competitive R^2 of 0.9575, demonstrating its strength in minimizing absolute errors. However, its performance gain over Swish was marginal and accompanied by a slightly longer prediction time (0.77 seconds). On the other hand, activation functions such as ReLU, Mish, Sigmoid, and especially Softplus showed inferior performance. Softplus recorded the highest RMSE (5.00) and the lowest R^2 score (0.9496), indicating weaker predictive accuracy and generalization ability. In conclusion, Swish is the best-performing activation function overall in this study, balancing high prediction accuracy with efficient computation.

4.2. Model Training Process

During the training process, observations were made on the loss function on the training and validation data to evaluate model convergence. The loss graphs shown in Figure 4.1 (a–f) provide a detailed visualization of the differences in training performance of each activation function.

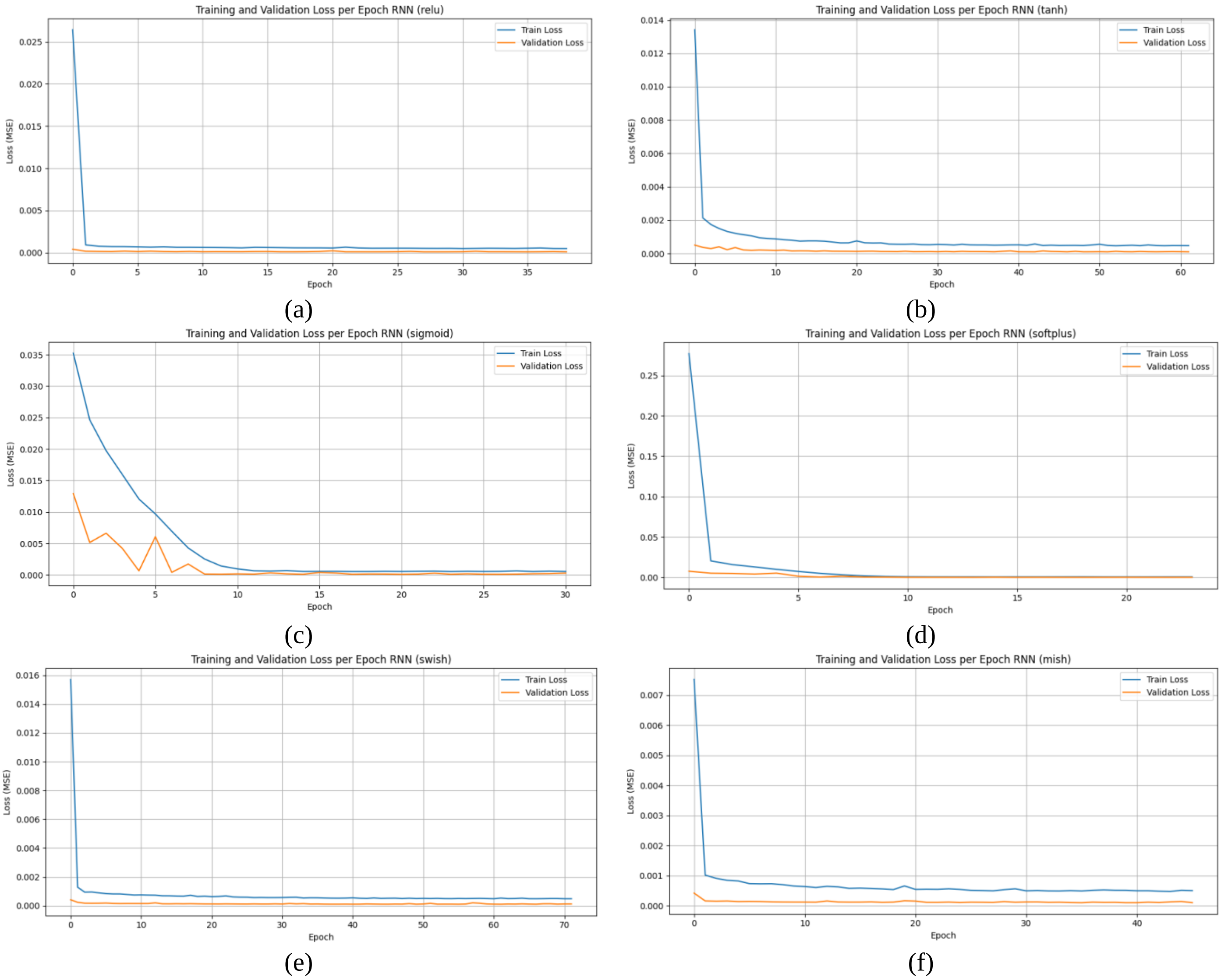

Figure 1: Training and validation loss per epoch for RNN using different activation functions

Based on the training and validation loss graphs for each activation function applied in the RNN model, Swish (e) demonstrates the most stable and consistent performance, with low loss values and smooth convergence, indicating its strong ability to capture complex temporal patterns in time-series data. Tanh (b) also performs well, showing gradual and stable loss reduction with minimal overfitting, making it suitable for modeling sequential dependencies. ReLU (a) achieves rapid convergence within a few epochs, but its sharp drop in loss could indicate overfitting or issues like the dying ReLU problem. Mish (f) follows closely behind Swish, delivering low and steady losses, which highlights its effectiveness for time-dependent prediction tasks. In contrast, Sigmoid (c) exhibits unstable learning, especially in early epochs, likely due to vanishing gradient problems, and Softplus (d), although quickly reducing its initially high training loss, stabilizes at higher loss values, suggesting weaker generalization. Overall, Swish (e) and Tanh (b) emerge as the most effective activation functions for RNN-based cryptocurrency prediction models based on their convergence behavior and loss minimization.

4.3. Comparison of Predictions and Actual Values

The comparison between the prediction results and the actual value of the crypto price (LTC-USD) is shown in Figure 4.2 (a–g).

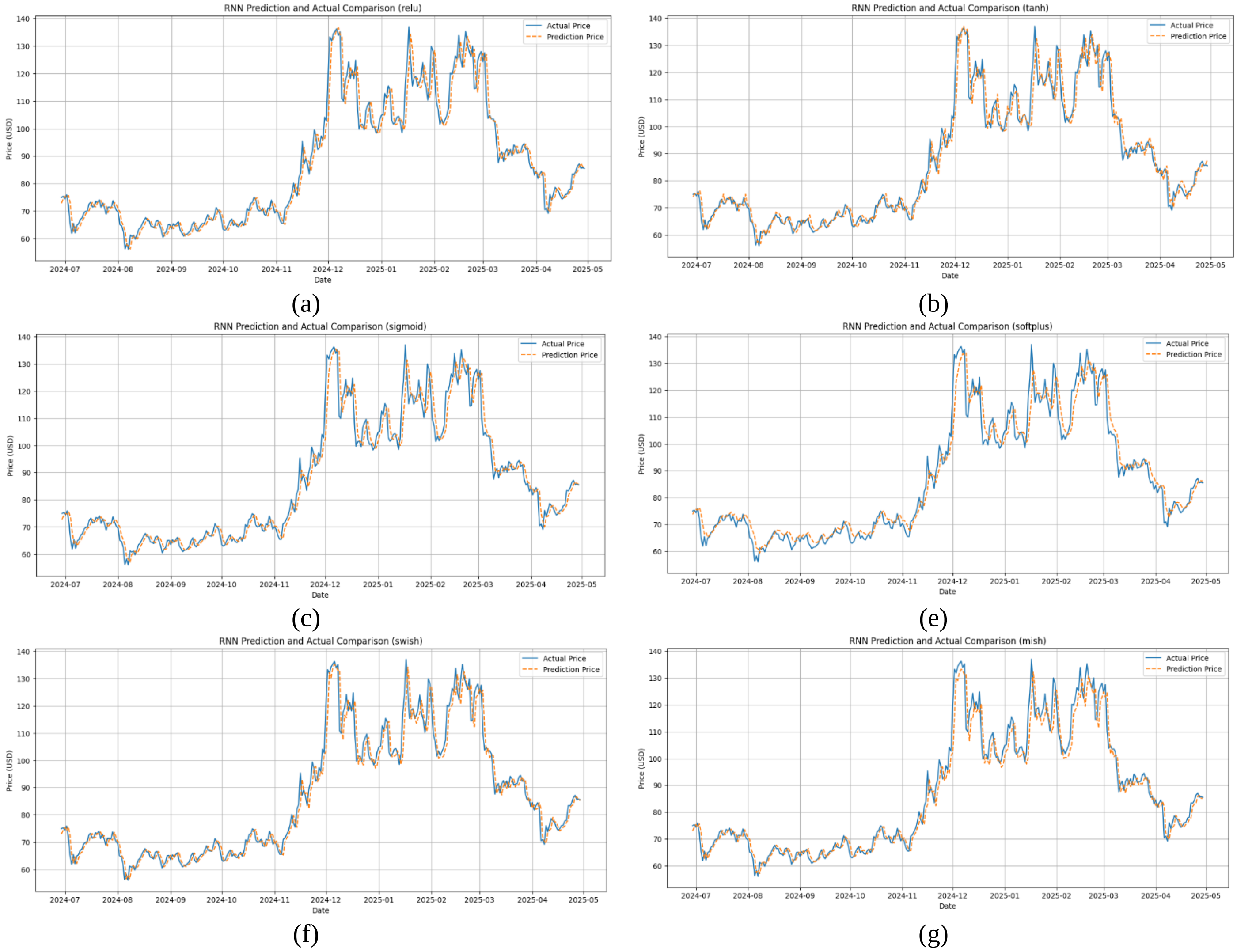

Figure 2: RNN prediction vs actual comparison for different activation functions

The visual comparisons between actual and predicted prices using different activation functions in the RNN model show distinct performance patterns. The Swish (f) activation function provides the closest match to actual values, with predicted price lines (orange) closely tracking actual price movements (blue) across all time periods. This indicates a high predictive capability with minimal deviation. Similarly, Tanh (b) and ReLU (a) also demonstrate strong performance, maintaining tight alignment with the actual price curve, particularly during periods of high volatility, such as December 2024 to February 2025.

Mish (g) also performs well, though in some periods, it shows slight lag or overshooting. Sigmoid (c) and Softplus (e) display slightly more noticeable deviations, especially during price spikes and drops, suggesting that these functions may struggle with capturing sharper transitions or sudden market changes. The Softplus (e) function, in particular, appears to underpredict peak values, which may lead to less accurate modeling in volatile markets. Swish (f) stands out as the most robust and precise activation function for cryptocurrency price prediction in this RNN configuration, followed by Tanh (b) and ReLU (a), while Softplus (e) and Sigmoid (c) appear less effective at capturing intricate market patterns.

5. Conclussion

Based on the experimental results and comprehensive evaluation of six activation functions ReLU, Tanh, Sigmoid, Softplus, Swish, and Mishwithin a Recurrent Neural Network (RNN) model for Litecoin cryptocurrency price prediction, several important conclusions can be drawn:

a. From the perspective of prediction accuracy (RMSE = 4.58, R^2 = 0.9578), Swish consistently outperformed the other activation functions. It showed a stable training process with smooth convergence, minimal loss, and high alignment between predicted and actual values. Swish also maintained computational efficiency with a fast prediction time (0.48 s), making it the most reliable activation function for time-series forecasting in this study.

Tanh ranked second with slightly better MAE but marginally lower overall performance, while Softplus and Sigmoid were the least effective.

b. Training loss graphs revealed that Swish and Mish exhibited the smoothest and most stable learning curves, indicating effective learning without overfitting. Tanh also maintained good convergence. In contrast, ReLU converged quickly but showed signs of instability, and Sigmoid experienced learning difficulty during early epochs due to vanishing gradient issues. Softplus failed to generalize well, stabilizing at higher loss values.

c. Visualization of predictions confirmed that the Swish function most accurately followed the real price trends of Litecoin across various market conditions, including volatile periods. Tanh and ReLU also performed well but with slightly more deviation. Mish showed promising results with minor overshoots, while Sigmoid and Softplus struggled to capture rapid market shifts. This reinforces the conclusion that Swish delivers the most accurate and stable predictions in RNN-based crypto price forecasting.

References

Agarap, A. F. (2018). Deep Learning using Rectified Linear Units (ReLU). *arXiv preprint arXiv:1803.08375*.

Akinrinola, O., Addy, W. A., Ajayi-Nifise, A. O., Odeyemi, O., & Falaiye, T. (2024). Predicting stock market movements using neural networks: A review and application study. *GSC Advanced Research and Reviews*, *18*(2), 297-311.

Choudhary, S., Garg, A., & Rajput, D. S. (2021). Forecasting Bitcoin Prices Using Deep Learning. *International Journal of Intelligent Systems and Applications*, 13(6), 30–40.

Dudek, G., Fiszeder, P., Kobus, P., & Orzeszko, W. (2024). Forecasting cryptocurrencies volatility using statistical and machine learning methods: A comparative study. *Applied Soft Computing*, *151*, 111132.

Fernández-Navarro, F. D. A., & Carbonero-Ruz, M. (2025). Enhancing financial time series forecasting through topological data analysis.

Firmansyah, I., & Hayadi, B. H. (2022). Comparison of ReLU and Tanh Activation Functions in Multilayer Perceptron. *JIKO (Jurnal Informatika dan Komputer)*, *6*(2), 200-206.

Jo, J. M. (2019). Effectiveness of normalization pre-processing of big data to the machine learning performance. *The Journal of the Korea institute of electronic communication sciences*, *14*(3), 547-552.

John, C., Ekpenyong, E. J., & Nworu, C. C. (2019). Imputation of missing values in economic and financial time series data using five principal component analysis approaches. *CBN Journal of Applied Statistics (JAS)*, *10*(1), 3.

Micu, R., & Dumitrescu, D. (2022). Study regarding the volatility of main cryptocurrencies. In *Proceedings of the International Conference on Business Excellence* (Vol. 16, No. 1, pp. 179-187).

Misra, D. (2019). Mish: A Self Regularized Non-Monotonic Neural Activation Function. *arXiv preprint arXiv:1908.08681*.

Patel, J., & Thakkar, P. (2022). Cryptocurrency Price Prediction using Deep Learning. *Procedia Computer Science*, 199, 754–761.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*.

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Electronics*, *10*(21), 2717.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.

Tumpa, S. N., & Maduranga, K. D. G. (2024). Utilizing RNN for real-time cryptocurrency price prediction and trading strategy optimization. *arXiv preprint arXiv:2411.05829*.

Verma, S., & Gupta, R. (2023). Performance Analysis of Activation Functions in GRU-based Bitcoin Price Forecasting. *Journal of Computational Finance and Economics*, 11(1), 42–53.

Wen, N. S., & Ling, L. S. (2023). Evaluation of cryptocurrency price prediction using lstm and cnns models. *JOIV: International Journal on Informatics Visualization*, *7*(3-2), 2016-2024.