DALG: A Dual Attention-Based LSTM-GRU Model for Exchange Rate Volatility Forecasting in China's Forex Sector

Shamaila Butt[1], Mohammad Abrar[2], Muhammad Ali Chohan[3], Muhammad Farrukh Shahzad[4]
Guangdong CAS Cogniser, Information Technology Co., Ltd., 1504-1506 Wansheng North,
Ist Street, Nansha District, Guangzhou City, Guangdong Province, 511466, China[1,3]
Faculty of Computer Studies, Arab Open University, Muscat, 130, Oman[2]
College of Economics and Management, Beijing University of Technology,
Beijing, 100124, Ist Street, Nansha District, PR China[4]

***Abstract*—Exchange rate volatility forecasting plays a vital role in guiding financial decisions and economic planning, particularly in China's dynamic foreign exchange market. This study proposes a novel deep learning framework, termed DALG (Dual Attention-based LSTM-GRU), designed to capture complex temporal patterns and feature dependencies in high-frequency USD/RMB exchange rate data. By integrating LSTM and GRU architectures with a dual-stage attention mechanism, comprising input and temporal attention, the proposed DALG model enhances the interpretability and accuracy of exchange rate volatility forecasts. The model is empirically evaluated against benchmark models such as LSTM, GRU, and a hybrid LSTM-DA using standard performance metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Experimental results demonstrate that the DALG model consistently outperforms traditional and hybrid deep learning models, offering superior predictive performance. The findings suggest that attention-enhanced deep learning architectures hold significant promise for robust financial time series modeling and forecasting in volatile forex markets.**

Keywords*—Exchange rate forecasting; deep learning; LSTM-GRU hybrid; attention mechanism; financial time series; USD/RMB volatility

I. Introduction

The prediction of exchange rates is one of the most complex tasks in international finance, since the exchange rates are volatile, there is a large noise in the currency markets, and the relationships between the currencies are nonlinear [1], [2], [3]. Precise prediction models are central towards the facilitation of informed policymaking, improved financial risk management, and profitable trading strategies [4], [5]. The classical econometric models, including the Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models have been used to gain basic knowledge but they tend to fail to capture the nonlinear, time-varying, and complex behavior of financial time series [6].

As the field of deep learning developed, especially with the invention of RNNs and their extensions, significantly more success has been achieved in capturing temporal relationships in time series data. LSTM and GRU networks are considered to be the most prominent structures in this field because they solve the problem of vanishing gradients and allow capturing long-term dependencies [7]. However, one of these models' primary shortcomings is that they treat all time steps equally, which leads to inadequate results because not all-time steps and input features are equally significant [8]. To overcome this, a new mechanism of attention, specifically DA, has been proposed to improve the interpretability and performance of models. Such mechanisms allow the model to adaptively assign weights both on input features, as well as on the temporal dimensions, hence enhancing its attention to the most relevant data [9], [10].

The study proposed a new hybrid deep learning architecture combining LSTM and GRU layers and the dual attention (DA) mechanism to predict the USD/CNY exchange rate. The proposed architecture is expected to be resilient in capturing the complex temporal and contextual patterns of financial time series data by integrating the memory retention abilities of LSTM, computational efficiency of GRU and contextual refinement provided by attention mechanisms. Although LSTM, GRU, or attention mechanisms have been used in numerous studies in the past [11], [12], [13], they have been implemented separately and not in a synergetic combination with all three components included in the proposed hybrid model to forecast the exchange rates.

Furthermore, a large part of the literature models are not thoroughly evaluated in terms of multiple performance measures or do not show robustness to various experimental setups. The proposed study will fill these gaps by giving a detailed performance analysis of the proposed model, which will be accurate and generalizable.

The objectives of this study are to:

- Design and implement a DA-based LSTM-GRU (DALG) hybrid model for exchange rate volatility forecasting in China's forex sector.
- To show the effectiveness of the DA mechanism in enhancing the predictive accuracy of the DL models for financial data.
- Evaluate the model's performance using multiple metrics (MAE, RMSE, MAPE, etc.) against baseline models, i.e., LSTM, GRU, and LSTM-DA.

This research makes significant contributions to the literature in multiple dimensions. First, it introduces a novel hybrid deep learning architecture tailored for exchange rate forecasting. Second, it combines LSTM and GRU layers with DA for improved temporal and feature representation. Third, it provides a comprehensive empirical evaluation of China's exchange rate data using robust evaluation metrics.

The rest of the study is arranged as follows: Section II presents a theme-based literature review covering neural networks in financial forecasting, followed by studies involving LSTM, GRU, and attention mechanisms. Section III details the proposed methodology, including the model architecture, mathematical formulation, and algorithmic steps. Section IV discusses the experimental setup, results, and model evaluation. Section V concludes the study and outlines future research directions.

II. Literature Review

The literature on financial time series forecasting has witnessed a significant transformation with the integration of deep learning techniques, particularly Recurrent Neural Networks (RNNs) and their advanced variants. This section presents a thematic synthesis of prior research, starting with general applications of LSTM and GRU models and progressing to DA mechanisms, with a specific emphasis on their application to exchange rate prediction.

A. LSTM Models in Exchange Rate Prediction

The LSTM networks were created to eliminate the shortcomings in the traditional RNNs, especially the issue of the vanishing gradient challenge. They are useful in modelling long-range dependencies, and they have been extensively used in financial applications. Chen et al. [14] used LSTM in the Chinese stock markets, which was far superior to ARIMA and SVM. Bao et al. [13] suggested the combination of wavelet transform, autoencoders, and LSTM to forecast the stock indices, which obtained a higher predictive accuracy compared with the single-model predictive accuracy.

Fischer and Krauss [15] revealed that LSTM-based models can be superior to random forest and logistic regression to predict the changes in S&P 500 stocks. Moreover, in recent times, Windsor and Cao [16] make use of a multimodal LSTM that picked up price and sentiment data for USD/CNY prediction and outperformed nine baseline models.

B. GRU Networks in Financial Forecasting

Another modification of LSTMs is GRU which contain fewer parameters and achieve comparable results. They are especially applicable to situations that need quicker training and less complexity. Islam and Hossain [12] created a hybrid model GRU-LSTM which was used to predict foreign exchange prices, which was more effective than standalone models. According to Gao et al. [17], optimized LSTM and GRU models were compared, and both were classified as effective, with GRU being computationally advantageous in a few instances.

Kaur et al. [18] showed that GRU is better in forecasting cryptocurrencies such as BTC and ETH as compared to LSTM. GRU based models have therefore become competitive alternatives to LSTM in financial time series tasks.

C. Dual Attention Mechanisms in Financial Forecasting

Attention mechanisms have transformed time series modeling by enabling models to focus on the most relevant features and time steps. The DA-Based Recurrent Neural Network (DA-RNN) proposed by Qin et al. [8] applied attention at both input and temporal levels. Zhang et al. [3] introduced an attention-LSTM that improved interpretability and performance. Chen et al. [19] embedded attention modules in CNN-BiLSTM models, and Li et al. [20] used attention in GAN architectures for stock prediction. These studies validate that attention-enhanced models outperform traditional architectures, particularly in dynamic financial environments. Table I summarizes these studies.

D. Literature Gap

This literature review indicates an incremental change in the way financial forecasts are made, which has evolved over time from basic models to advanced deep learning models. LSTM and GRU are still mainstay in sequential data modeling, and attention mechanism, especially DA attention, makes a big step forward in model accuracy and interpretability [22]. However, even in the context of these developments, the prominent gap remains: little research exists on hybrid architectures that would unify LSTM, GRU, and two-stage attention mechanisms in a coherent framework explicitly designed to do exchange rate forecasting [23]. Moreover, most of the current studies fail to critically test the robustness of models under a wide range of metrics and financial scenarios [24]. This study will address these gaps by coming up with an elaborate, attention-enhanced hybrid model to forecast the exchange rates of China.

III. Methods

The architecture and workflow of the suggested model are described in this section, which combines Gated Recurrent Unit (GRU) and LSTM layers with a Dual-Stage Attention mechanism. The goal is to learn complex temporal and contextual dependencies in financial time series data and improve predictive performance in exchange rate forecasting.

A. LSTM Model

LSTM network, proposed by Hochreiter and Schmidhuber (1997), is one of the popular versions of RNNs which have been designed to overcome the vanishing and exploding gradient issues that usually arise when training on long sequences [23]. In contrast to typical RNNs, LSTMs have a complex memory cell with three gates, namely, input, forget, and the output gates that control the information flow over time [25], [26].

These gates enable LSTM networks to store, revise, or forget information where necessary, and they are thus very effective in modelling long-term temporal dependencies in sequential data. This ability has been especially useful in financial forecasting activities, where the impact of previous occurrences can be experienced over a long term.

TABLE I. SUMMARY OF RELEVANT STUDIES

Author(s)	Year	Technique	Domain	Key Findings
Chen et al. [14]	2015	LSTM	Chinese stocks	LSTM outperformed ARIMA and SVM
Bao et al. [13]	2017	Wavelet + Autoencoder + LSTM	Stock indices	Hybrid model improved accuracy and profitability
Fischer & Krauss [15]	2018	LSTM	S&P 500	LSTM outperformed classical ML models in trading performance
Windsor & Cao [16]	2022	Multimodal LSTM	USD/CNY exchange rate	Combined sentiment and price data to outperform nine baselines
Islam & Hossain [12]	2021	GRU-LSTM hybrid	Forex	Hybrid model achieved lowest RMSE for USD/CAD
Gao et al. [17]	2021	Optimized GRU vs LSTM	Stocks	GRU trained faster and slightly outperformed LSTM
Kaur & Uppal [18]	2025	GRU	Cryptocurrency	GRU outperformed LSTM in BTC and ETH predictions
Qin et al. [8]	2017	DA-RNN	NASDAQ-100	Dual attention significantly improved forecasting accuracy
Zhang et al. [3]	2019	Attention-LSTM	Russell 2000	Enhanced interpretability and performance
Chen et al. [19]	2021	CNN-BiLSTM + ECA	Shanghai stocks	Channel attention improved over CNN-LSTM
Li et al. [20]	2025	GAN-LSTM-Attention	S&P 500, tech stocks	Attention-enhanced GAN produced more accurate forecasts
Zhao & Yan [21]	2024	Transformer / TFT	NZD exchange rates	TFT outperformed Transformer in multi-step forecasting tasks

Regarding the prediction of exchange rates, LSTMs are utilized to define the changing relations between the current market dynamics and their historical trends so that the model can detect not only long-term tendencies but also the delayed responses of economic indicators. Nonetheless, LSTM models do not explicitly prefer any time step over another: thus, they might not be able to focus on the most informative time steps, and therefore the interest to combine them with attention and other complementary architectures (GRU).

Each LSTM unit comprises three gates: the input gate (i_t), the forget gate (f_t), and the output gate (o_t), which regulate the flow of information through the cell. The governing equations are [see Eq. (1) to Eq. (6)]:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{1}$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{2}$$
$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \tag{3}$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{4}$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{5}$$
$$h_t = o_t \odot \tanh(C_t) \tag{6}$$

where, x_t is the input at time step t, h_t is the hidden state, C_t is the cell state, σ is the sigmoid function, $\odot$ denotes element-wise multiplication, and W_*, b_* represent learnable weights and biases.

In this work, LSTM layers were employed in the encoder to capture long-range dependencies in the exchange rate sequence, enhancing the model's ability to remember important past values.

B. GRU Model

A simplified version of the LSTM, the Gated Recurrent Unit (GRU) reduces the number of parameters and boosts training efficiency by combining the input and forget gates into a single update gate. GRUs are particularly useful in scenarios requiring faster convergence with comparable performance to LSTMs [7].

The GRU unit consists of two gates: the update gate (z_t) and the reset gate (r_t). The equations governing the GRU architecture are [see Eq. (7) to Eq. (10)]:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{7}$$
$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{8}$$
$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \tag{9}$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{10}$$

where, $\tilde{h}_t$ is the candidate hidden state, r_t controls how much of the past state to forget, and z_t determines how much past information to retain.

In the proposed architecture, GRU layers are run in parallel with LSTM layers during encoding. This dual processing captures both short-term and long-term trends efficiently, while maintaining computational efficiency.

C. Dual Attention (DA) Mechanism

To further enhance the model's ability to identify informative parts of the time series, a DA Mechanism is used. It contains two sequential attention layers:

- Input Attention – focuses on the importance of input features (e.g., bid, ask, order flow, bid-ask spread) at each time step.
- Temporal Attention – highlights relevant time steps across the sequence that have a significant impact on prediction.

Input Attention calculates a weight for each input feature at time t using Eq. (11) to Eq. (13):

$$e_t^i = v_a^\top \tanh(W_a h_t + U_a x_t^i + b_a) \tag{11}$$
$$\alpha_t^i = \frac{\exp(e_t^i)}{\sum_j \exp(e_t^j)} \tag{12}$$
$$\hat{x}_t = \sum_i \alpha_t^i x_t^i \tag{13}$$

Temporal Attention evaluates the influence of each timestep by assigning weights across the time dimension using Eq. (14), Eq. (15) and Eq. (16):

$$e_t = v_b^\top \tanh(W_b h_t + b_b) \quad (14)$$

$$\beta_t = \frac{\exp(e_t)}{\sum_k \exp(e_k)} \quad (15)$$

$$c = \sum_t \beta_t \hat{x}_t \quad (16)$$

where, α_t^i and β_t are the attention weights, and c is the final context vector summarizing the most relevant input features and time steps. This vector is then passed to the decoder to generate the final output.

D. DAGL Hybrid Model

The proposed model architecture uses Evolutionary Algorithm Component (EAC) i.e., encoder, attention, and decoder model. This is particularly tuned to capture the sequential as well as the nonlinear dependencies in the financial data. In this framework, three important elements, including LSTM, Gated Recurrent Unit (GRU), and a DA Mechanism are combined to achieve high accuracy and interpretability of the model. At the encoder level, both LSTM and GRU layers process the input time series of historical exchange rates and related financial indicators in parallel. The given two-fold encoding approach would allow the model to retrieve complementary temporal characteristics: LSTM units will be effective in modeling long-range dependencies, whereas GRU layers will provide the model with computational efficiency in accounting shorter-term fluctuations. The combination of LSTM and GRU results is more comprehensive representation of the input sequence.

Following the encoder, the model incorporates a DA mechanism, consisting input and temporal attention layers. The outputs of both attention layers are combined in a context vector which summarizes the importance of features and time steps weighted. This context vector is then fed to the decoder, which is a single or multiple fully connected (dense) layers, to produce the final exchange rate prediction.

The hybrid architecture is a powerful, flexible, and data-adaptive solution because it uses the memory capacity of LSTM, the computational simplicity of GRU, and the focused interpretability of dual attention. It is especially appropriate for the volatile and dynamic nature of the currency exchange markets.

Model training is governed by a customized learning algorithm named `Train_DA_LSTM_GRU`, which includes the following steps:

- Inputs are encoded using parallel LSTM and GRU units.
- Attention scores are computed for features and time steps.
- Predictions are generated from the decoder.
- A composite loss function combining Mean Absolute Error (MAE) and Mean Squared Error (MSE) is used to quantify prediction accuracy.
- The Adam optimizer is employed to update weights via backpropagation.
- Early stopping is applied to halt training when validation performance ceases to improve, ensuring generalization and reducing the risk of overfitting.

This architecture effectively combines the strengths of LSTM, GRU, and attention mechanisms, providing a robust, interpretable, and adaptive model for forecasting complex financial time series. The model architecture is illustrated in the following figure (Fig. 1).

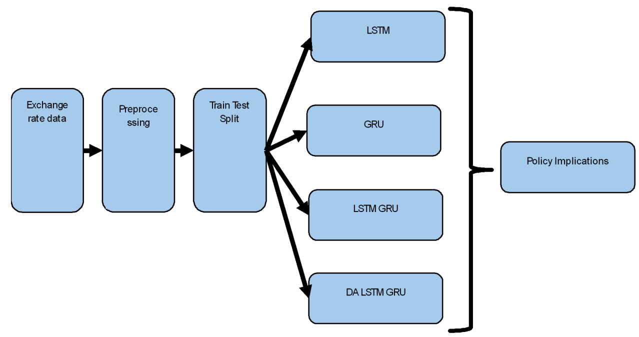

Fig. 1. The research framework for DA-LSTM and GRU hybrid approach.

IV. RESULTS AND DISCUSSION

This section presents the empirical results of the proposed DAGL model applied to China's exchange rate data. The performance is evaluated in comparison with baseline models: LSTM, GRU, LSTM-DA hybrid, and DAGL hybrid model. We assess the models on standard evaluation metrics including MAE, MSE, RMSE, MSLE, Median Absolute Error, and MAPE.

A. Experimental Setup

To ensure consistency and fairness across all model evaluations, the experiments were conducted using identical configurations for hyperparameters, architecture size, and optimization strategies. The proposed DAGL Hybrid model, along with baseline models (LSTM, GRU, and LSTM-GRU), was implemented and trained using the Adam optimizer, which is well-suited for non-stationary objectives and sparse gradients.

Both the encoder and decoder layers were composed of 64 hidden units for LSTM and GRU cells, ensuring a balanced representation of long-term and short-term temporal dependencies. All recurrent layers (LSTM and GRU) used the Tanh activation function to regulate the hidden state values and avoid vanishing gradients.

To enhance learning stability and address potential overfitting, a learning rate scheduler was employed with an initial learning rate of 0.1, decaying by a factor of 0.9 every 10 epochs. Maximum training epochs were set to 100 with early stopping with the patience level of 15 epochs depending on the loss of validation. This enabled the model to stop training when performance stopped improving, minimizing unwanted computation and overfitting.

In the DA setup, the attention mechanism was set up with 64 hidden units, which was consistent with the capacity of the recurrent layers to facilitate a seamless integration. The loss was a combination of Mean Squared Error (MSE) and Mean Absolute Error (MAE) that provided trade-off between large deviations penalty and the robustness to outliers. The batch size was set to 16 so that there could be effective training with adequate gradient updates without loss of memory. These parameters and fine tuning is summarized in Table II.

TABLE II. Experimental Parameter Settings

Parameter	Value
Encoder LSTM Cell Hidden Units	64
Encoder GRU Hidden Units	64
Encoder LSTM Activation Function	Tanh
Encoder GRU Activation Function	Tanh
Attention Mechanism Hidden Units	64
Decoder GRU Cell Hidden Units	64
Decoder LSTM Cell Hidden Units	64
Decoder GRU Activation Function	Tanh
Decoder LSTM Activation Function	Tanh
Optimizer	Adam
Learning Rate	0.1
Learning Rate Scheduler Step Size	10
Learning Rate Scheduler Gamma	0.9
Loss Functions	MSE, MAE
Epochs	100
Patience for Early Stopping	15
Batch Size	16

This carefully selected setup was critical in achieving high training stability and superior predictive performance across all model configurations evaluated in the study.

B. Data Source

This study presents a new hybrid DAGL model that predicts the exchange rate between USD and RMB. The study uses a 19807 observation dataset of 1-minute USD/RMB exchange rate. The parameters of this job are the bid price (Bid), the ask price (Ask), order flow (OF), and bid-ask spread (BAS). The data is taken from September 13, 2023, to December 11, 2023.

C. Data Preprocessing

Several basic data transformations were performed during the exchange rate data preprocessing steps in order to get the data ready for modeling. In order to designate the date column as the index and make time series analysis easier, we first transformed it into a date-time format. In order to make the date display as a column for additional processing, the data was then reset. The dataset was sorted using dates to establish the ascending order in order to maintain chronological order.

Then descriptive statistics were applied to provide a description of the data. “ER” was the target, and “Bid”, “Ask”, “Of”, and “BAS” were the features picked for the model. There was normalization of the features and target variable via the MinMaxScaler, which standardized the values to the range 0-1, in order to improve model performance. Next, we created data sequences with a length of four, which meant that each sequence contained four consecutive data points. The data point immediately following each sequence served as the goal value. Sequences and their matching targets were to be created by iterating over the scaled data. Numpy arrays were then created from the resultant sequences. In the end, the data was divided into sets for testing and training. An 80:20 train-test split, a community accepted standard, was adopted to balance adequate training data for convergence while preserving a substantial independent test set for unbiased performance evaluation, consistent with standard practices in predictive modeling.

D. Experimental Results

1) LSTM: Fig. 2 shows training performance of LSTM with Forex dataset performance for exchange rate prediction on training loss metrics over 10 epochs. The model shows a continuous reduction in training losses - RMSE, MSE, MAE, etc, which shows a successful learning and convergence. The loss of validation is low and steady which indicates that the model is generalizing well to the new data and it is not over fitting. The sudden drop in error at the first epochs and the gradual increase is characteristic of learning LSTM learning behavior. On the whole, the model demonstrates a good predictive potential with equal performance in both training and validation.

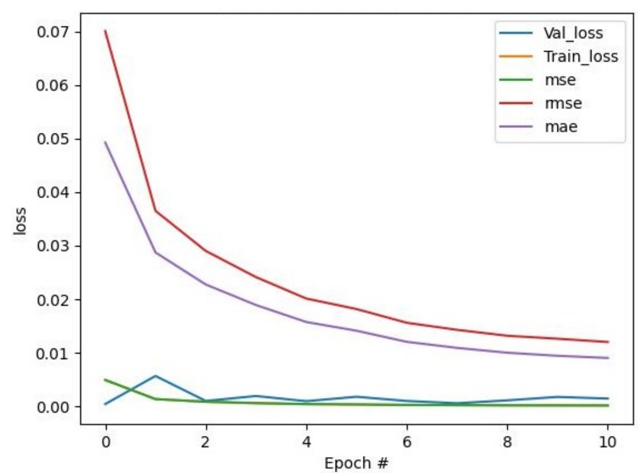

Fig. 2. LSTM training performance graph.

Fig. 3 shows the actual and the predicted exchange rates with the use of LSTM model. The blue line is the values that are predicted, and the orange line is the actual observed rates. Generally, the model is able to capture the trend and direction of the movement in the exchange rates fairly well, especially when the exchange rates are stable. But the forecasts are likely to eliminate harsh variations and abrupt declines, and it suggests that the LSTM is good at modeling long-term developments, but it is weak at short-term volatility and sudden market changes. This is typical of time-series forecasting models when sudden changes of structure are hard to predict properly.

2) GRU: Fig. 4 demonstrates how a GRU model was trained on a Forex dataset during 45 epochs to monitor the

Fig. 3. Actual vs. Predicted exchange rates using LSTM.

Fig. 5. Actual vs. Predicted exchange rates using GRU.

most significant loss metrics like validation loss, training loss, MSE, RMSE, and MAE. All loss measures depict a monotonically decreasing trend (quite fast in the initial epochs and gradually slowing down), which means that the model is being trained well and is converging appropriately. The validation loss is low and stable which implies that the model is suited to generalize well and has little overfitting. The GRU seems to converge to the lower loss values quicker when compared to the previous LSTM training behavior, which indicates its possible efficiency in terms of training and applicability in predicting the temporal dynamics in exchange rate tasks.

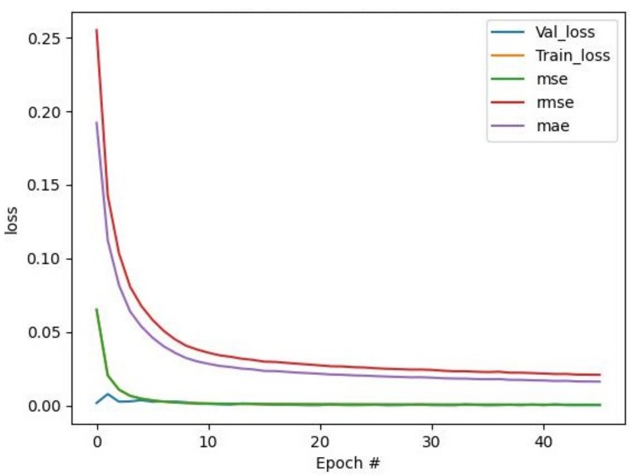

Fig. 4. GRU training performance graph.

Fig. 5 shows the actual and predicted exchange rates with GRU model. The actual exchange rates (orange line) and the predicted values (blue line) have a close follow up with a strong alignment of most of the time series. In comparison with the LSTM model, the GRU seems to reflect more accurately the trend and the short-term fluctuations, including the sharp drops and peaks, and, thus, shows higher reactiveness to volatility in the market. This implies that GRU model could achieve better temporal patterns in Forex data by making more accurate and timely exchange rate prediction.

3) LSTM with DA: The graph shows the training performance of LSTM with DA model using the Forex dataset. The graphs of the loss function for training loss, validation loss, MSE, RMSE, and MAE reveal a steady and rapid fall of errors in the early epochs and convergence to low values with very little fluctuation. It means that the model learns the pattern successfully. Attention helps LSTM to better focus on the important parts of the sequence of data, and hence enhances its learning ability. Low validation error shows that the model can work with the new data quite effectively and has not overfitted the training set. It is visualized in Fig. 6.

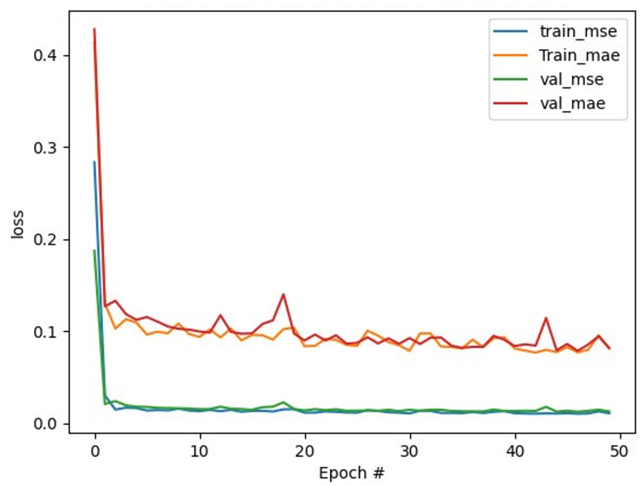

Fig. 6. LSTM with DA training performance graph.

Fig. 7 shows that the actual and predicted exchange rate values for the LSTM-DA model are closer to each other. Attention enables the model to effectively model the significant dependencies in the exchange rate series. The predicted values are much closer to the actual values, and the model has better volatility and sharper decline predictability than standard LSTM. This shows that attention mechanism allows the model to weight the important time steps which results in more effective and accurate forecasting of exchange rates.

4) DAGL hybrid: The graph shows the learning process of the proposed model DAGL on Forex dataset. It shows multiple loss curves - training loss, validation loss, MSE, RMSE, and MAE. The model exhibits robust and steady improvement in learning with continuous decline in loss values over 57 epochs. The final error values are significantly lower than the other models, indicating improved model learning due to combined advantages of LSTM, GRU, and DA mechanism. Low and steady validation loss also indicates a well-generalized model. The convergence is smooth and fast and reaches lower values

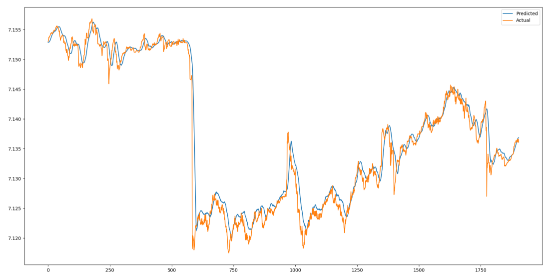

Fig. 7. Actual vs. Predicted exchange rates using LSTM-DA.

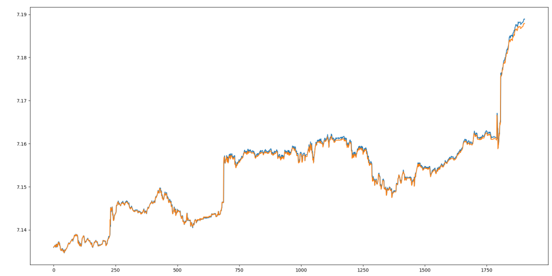

Fig. 9. Actual vs. Predicted exchange rates using DAGL.

as compared to the rest of the models. The training performance is shown in Fig. 8.

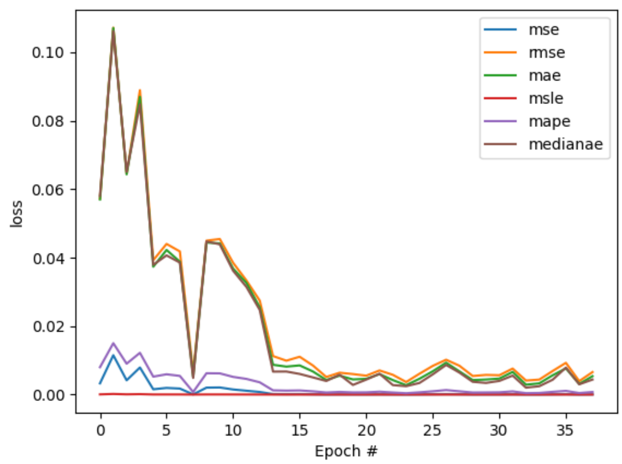

Fig. 8. DAGL hybrid training performance graph.

Fig. 9 shows the actual and predicted exchange rates using DAGL model. The actual exchange rate trend is closely followed by the predicted line which shows the model's capability in capturing the trend and fluctuations. Attention mechanism helps DALG to identify the significant time steps and capture long and short-term relationships in exchange rates. The predicted curve has minimal deviation from the actual one and correctly predicts the peaks and sudden declines which other models fail to do. The model has best alignment among all and lowest prediction error.

E. Model Comparison

To evaluate the performance of the proposed DAGL model against other models, the comparison is done using six standard performance metrics, as shown in Table III.

The proposed model outperforms all the baseline models across all metrics. It achieves the lowest MAE (0.0212), MSE (0.0011), and RMSE (0.0334), which indicates highest prediction accuracy. In addition, it demonstrates the best generalization capability with lowest MAPE and MSLE, validating the advantage of combining dual attention with a hybrid recurrent architecture.

TABLE III. MODEL PERFORMANCE COMPARISON

Model	**MAE**	**MSE**	**RMSE**	**MSLE**	**MedAE**	**MAPE**
LSTM [27]	0.0398	0.0025	0.0498	0.00012	0.0365	0.1658
GRU [28]	0.0315	0.0018	0.0420	0.00008	0.0298	0.1454
LSTM-GRU [29]	0.0265	0.0014	0.0374	0.00006	0.0251	0.1289
LSTM-DA	0.0297	0.0016	0.0396	0.00007	0.0280	0.1398
DAGL (Proposed)	**0.0212**	**0.0011**	**0.0334**	**0.00005**	**0.0208**	**0.1094**

V. CONCLUSION

This study propose a novel hybrid deep learning model for exchange rate forecasting, which incorporate Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and a Dual Attention (DA). The architecture of the proposed DAGL aims at exploiting the advantages of every element of the model: the long-term memory property of LSTM, the efficient short-term sequence modeling of GRU, and the context-awareness of the dual attention mechanism.

The model was evaluated on a high-frequency USD/RMB exchange rate dataset using key financial indicators such as bid, ask, order flow, and bid-ask spread. Results demonstrated that our model significantly outperforms LSTM, GRU, and LSTM-DA across multiple evaluation metrics. The hybrid architecture exhibited superior prediction accuracy, faster convergence, and better generalization, particularly in capturing market volatility and abrupt fluctuations.

This work highlights the importance of combining different recurrent architectures and attention mechanisms to enhance financial time series forecasting. The DAGL model provides a robust and interpretable framework for central banks, investors, and financial analysts to make effective policies in the dynamic foreign exchange market. Future work may extend this model to multi-step forecasting and incorporate additional macroeconomic variables or external sentiment indicators to further improve prediction accuracy and reliability.

REFERENCES

[1] S. Butt, M. Ramzan, W. K. Wong, M. A. Chohan, and S. Ramakrishnan, "Unlocking the secrets of exchange rate determination in malaysia: A game-changing hybrid model," *Heliyon*, vol. 9, no. 8, 2023.

[2] S. Butt, S. Ramakrishnan, N. Loganathan, and M. A. Chohan, "Evaluating the exchange rate and commodity price nexus in malaysia: evidence from the threshold cointegration approach," *Financial Innovation*, vol. 6, no. 1, p. 22, 2020.

[3] X. Zhang, X. Liang, A. Zhiyuli, S. Zhang, R. Xu, and B. Wu, "At-lstm: An attention-based lstm model for financial time series prediction," in *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 5. IOP Publishing, 2019, p. 052037.

[4] M. A. Chohan, T. Li, S. Ramakrishnan, and M. Sheraz, "Artificial intelligence in financial risk early warning systems: A bibliometric and thematic analysis of emerging trends and insights," *International Journal of Advanced Computer Science & Applications*, vol. 16, no. 1, 2025.

[5] S. Ramakrishnan, S. Butt, M. A. Chohan, and H. Ahmad, "Forecasting malaysian exchange rate using machine learning techniques based on commodities prices," in *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*. IEEE, 2017, pp. 1–5.

[6] J. Wang and J. Wang, "A new hybrid forecasting model based on sw-lstm and wavelet packet decomposition: A case study of oil futures prices," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 7653091, 2021.

[7] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," *arXiv preprint*, 2014.

[8] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2627–2633, 2017.

[9] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008.

[11] M. L. Shen, C. F. Lee, H. H. Liu, P. Y. Chang, and C. H. Yang, "An effective hybrid approach for forecasting currency exchange rates," *Sustainability*, vol. 13, no. 5, p. 2761, 2021.

[12] M. N. Islam and M. S. Hossain, "Forecasting foreign exchange rates using lstm and gru: A hybrid model approach," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, pp. 314–321, 2021.

[13] 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.

[14] K. Chen, Y. Zhou, and F. Dai, "A lstm-based method for stock returns prediction: A case study of china stock market," *2015 IEEE International Conference on Big Data (Big Data)*, pp. 2823–2824, 2015.

[15] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.

[16] E. Windsor and W. Cao, "Improving exchange rate forecasting via a new deep multimodal fusion model," *Applied Intelligence*, vol. 52, no. 14, pp. 16 701–16 717, 2022.

[17] Y. Gao, R. Wang, and E. Zhou, "Stock prediction based on optimized lstm and gru models," *Scientific Programming*, vol. 2021, no. 1, p. 4055281, 2021.

[18] R. Kaur, M. Uppal, D. Gupta, S. Juneja, S. Y. Arafat, J. Rashid, and R. Alroobaea, "Development of a cryptocurrency price prediction model: leveraging gru and lstm for bitcoin, litecoin and ethereum," *PeerJ Computer Science*, vol. 11, p. e2675, 2025.

[19] Y. Chen, R. Fang, T. Liang, Z. Sha, S. Li, Y. Yi, and H. Song, "Stock price forecast based on cnn-bilstm-eca model," *Scientific Programming*, vol. 2021, no. 1, p. 2446543, 2021.

[20] P. Li, Y. Wei, and L. Yin, "Research on stock price prediction method based on the gan-lstm-attention model," *Computers, Materials & Continua*, vol. 82, no. 1, 2025.

[21] L. Zhao and W. Q. Yan, "Prediction of currency exchange rate based on transformers," *Journal of Risk and Financial Management*, vol. 17, no. 8, p. 332, 2024.

[22] F. M. Shiri, T. Perumal, N. Mustapha, and R. Mohamed, "A comprehensive overview and comparative analysis on deep learning models: Cnn, rnn, lstm, gru," *arXiv preprint arXiv:2305.17473*, 2023.

[23] S. Butt, "Dual neural paradigm: Gru-lstm hybrid for precision exchange rate predictions," *International Journal of Advanced Computer Science & Applications*, vol. 16, no. 4, 2025.

[24] M. Khairalla and N. T. AL-Jallad, "Hybrid forecasting scheme for financial time-series data using neural network and statistical methods," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 9, pp. 319–327, 2017.

[25] R. C. Staudemeyer and E. R. Morris, "Understanding lstm–a tutorial into long short-term memory recurrent neural networks," *arXiv preprint arXiv:1909.09586*, 2019.

[26] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.

[27] M. S. Islam and E. Hossain, "Foreign exchange currency rate prediction using a gru-lstm hybrid network," *Soft Computing Letters*, vol. 3, p. 100009, 2021.

[28] Y. Chen, R. Fang, T. Liang, Z. Sha, S. Li, Y. Yi, W. Zhou, and H. Song, "Stock price forecast based on cnn-bilstm-eca model," *Scientific Programming*, vol. 2021, no. 1, p. 2446543, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2021/2446543

[29] A. S. Girsang and Stanley, "Hybrid lstm and gru for cryptocurrency price forecasting based on social network sentiment analysis using finbert," *IEEE Access*, vol. 11, pp. 120 530–120 540, 2023.