Machine Learning with Applications 20 (2025) 100648

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Prediction of foreign currency exchange rates using an attention-based long short-term memory network

Shahram Ghahremani *, Uyen Trang Nguyen

Department of Electrical Engineering and Computer Science, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3, Canada

ARTICLE INFO

Keywords:
Attention mechanism
LSTM
Forex
Currency exchange prediction

ABSTRACT

We propose an attention-based LSTM model for predicting forex rates (ALFA). The prediction process consists of three stages. First, an LSTM model captures temporal dependencies within the forex time series. Next, an attention mechanism assigns different weights (importance scores) to the features of the LSTM model's output. Finally, a fully connected layer generates predictions of forex rates. We conducted comprehensive experiments to evaluate and compare the performance of ALFA against several models used in previous work and against state-of-the-art deep learning models such as temporal convolutional networks (TCN) and Transformer. Experimental results show that ALFA outperforms the baseline models in most cases, across different currency pairs and feature sets, thanks to its attention mechanism that filters out irrelevant or redundant data to focus on important features. ALFA consistently ranks among the top three of the seven models evaluated and ranks first in most cases. We validated the effectiveness of ALFA by applying it to actual trading scenarios using several currency pairs. In these evaluations, ALFA achieves estimated annual return rates comparable to those of professional traders.

1. Introduction

The foreign exchange (forex or FX) market is the largest financial market in the world, with a daily trading volume of approximately $7.5 trillion (Forex, 2024). In this market, currencies are traded, and their exchange rates fluctuate continuously, making accurate predictions a challenging task. Furthermore, the forex market is highly complex, influenced by a wide range of factors, including national and global economies, political events, and societal changes (Hu et al., 2021). Forex market prediction has thus gained significant attention in financial technology (Baruník et al., 2017).

To predict forex rates, traders (both human and algorithmic) typically rely on one or a combination of the following types of data:

- **Forex market data**: This includes raw price data such as the low, high, open, and close prices of a currency pair. These values serve as the foundation for technical analysis. To assist traders in analyzing historical trends and forecasting future price movements, various technical indicators are derived from raw data. For example, the commonly used indicator, Moving Averages (MA), is computed using close prices over a period of time. Other common technical indicators include the Relative Strength Index (RSI), Rate of Change (ROC), and Bollinger Bands.
- **Macroeconomic indicators**: These factors typically encompass interest rates set by central banks, inflation rates, GDP growth, employment figures, and trade balances. Such indicators influence currency valuation by shaping economic expectations and monetary policies.
- **Market sentiment**: This reflects the collective attitudes and expectations of market participants, including buyers and sellers. Sentiment analysis is often derived from news reports, expert analyses, financial forecasts, and social media discussions. Traders leverage sentiment indicators to gauge market confidence and potential price trends.

In this article, we focus on forex rate prediction using the first type of data (forex market data) with the ultimate goal of building an ensemble model that captures all three data types. By first comprehensively developing and evaluating a machine learning model that predicts forex rates using only forex market data, we can isolate and understand factors that influence the behavior and performance of each individual base model (forex data, macroeconomic indicators, and market sentiment) in the future ensemble model.

We propose an LSTM model with an integrated attention mechanism to predict the closing price of the day (or the hour) of a currency in relation to another using forex market data. We name

* Corresponding author.
E-mail addresses: shg@yorku.ca (S. Ghahremani), unguyen@yorku.ca (U.T. Nguyen).

https://doi.org/10.1016/j.mlwa.2025.100648
Received 5 November 2024; Received in revised form 11 February 2025; Accepted 27 March 2025
Available online 11 April 2025
2666-8270/

the proposed model ALFA (attention-based LSTM for predicting forex rates). Attention mechanisms have significantly improved the prediction performance of many applications in computer vision and natural language processing due to their ability to rate the importance of input features and focus on the most relevant ones. Similarly, the attention mechanism in ALFA enables it to filter out irrelevant data and capture important temporal patterns in forex time series, improving the accuracy of exchange rate predictions. The detailed contributions of our work are as follows:

- First, we propose an attention-based LSTM model named ALFA, designed specifically for predicting forex rates. Although existing RNN models such as LSTM and GRU have been used for forex rate prediction (Ayitey Junior et al., 2022; Dash et al., 2023; Olanrewaju et al., 2023; Yıldırım et al., 2021), they do not incorporate attention mechanisms, which have been shown to enhance LSTM performance in various tasks (Chen et al., 2023; Cheng, 2023; Li, 2023; Lou et al., 2023). While LSTM models with attention mechanisms exist, they were primarily designed for predicting financial risks (Cheng, 2023), stock prices (Chen et al., 2023; Li, 2023), gold, and Bitcoin (Lou et al., 2023). These models are not directly applicable to forex rate prediction due to differences in underlying data, price movement patterns, and market dynamics. *To the best of our knowledge, ALFA is the first attention-based LSTM model specifically designed for predicting forex rates.* The proposed model is described in detail in Section 3.
- Second, we conducted comprehensive experiments to evaluate and compare ALFA's performance against baseline models used in previous work (Galeshchuk & Mukherjee, 2017), including GRU (Olanrewaju et al., 2023), LSTM (Ayitey Junior et al., 2022; Yıldırım et al., 2021), bi-LSTM (Dash et al., 2023; Lou et al., 2023), and stacked LSTM (Cheng, 2023), as well as state-of-the-art deep learning models such as TCN (Chen et al., 2023) and Transformer (Kantoutsis et al., 2024). To ensure the robustness of our experimental results, we tested multiple currency pairs with varying volatility levels: EUR/USD (low volatility), USD/JPY (medium volatility), and GBP/JPY (high volatility). Additionally, we analyzed the effects of different feature sets on model performance, including individual features, feature pairs, and technical indicators. Experimental results show that ALFA outperforms the baseline models in most cases across different currency pairs and feature sets, owing to its attention mechanism, which filters out irrelevant or redundant data. ALFA consistently ranks among the top three models and is the best-performing model in the majority of cases. These results and analyses are presented in Section 5.1.
- Third, we provide experimental results demonstrating that ALFA offers the best balance between predictive performance, model complexity (number of parameters), and inference time among the evaluated models. This is reported in Section 5.2.
- Fourth, we validate ALFA's effectiveness by applying it to real trading scenarios using multiple currency pairs (EUR/USD, USD/JPY, GBP/JPY) from June 10, 2024, to September 6, 2024, and evaluating various performance metrics (e.g., success rate, total return, average gain per trade). The results in Section 5.3 show that ALFA achieved an estimated annual return of 8.79% to 19% after fees, depending on the currency pair. This is comparable to the typical annual return for forex trading by professional traders, which ranges between 5% and 30% (Forex Training Group, 2025).

The remainder of this article is structured as follows: Section 2 reviews related work. Section 3 describes the proposed model. Section 4 details the experimental settings. Section 5 presents performance evaluations, experimental results, and analysis. Section 6 concludes the article and suggests future research directions.

2. Related work

Forex data is inherently noisy, non-stationary, and chaotic, making future trend prediction challenging (Baasher & Fakhr, 2011). Despite these difficulties, research has shown that Forex fluctuations can be predicted to some extent (Sirignano & Cont, 2019).

Traditional methods for predicting Forex rates often rely on stationary stochastic processes, assuming that time series data have constant statistical properties, such as mean and variance. For example, Ngan (2013) used the autoregressive integrated moving average (ARIMA) model to forecast exchange rates, while (Sidehabi & Tandungan, 2016) applied the adaptive spline threshold autoregression (ASTAR) model. However, these traditional statistical models struggle to capture the complex, nonlinear relationships present in multivariate Forex datasets, limiting their effectiveness for long time series.

In contrast, deep learning models have shown great potential in capturing the intricate dynamics of financial data. Galeshchuk and Mukherjee (2017) applied convolutional neural networks (CNNs) to predict Forex rate directions, finding that CNNs outperformed baseline models such as support vector machines (SVMs). However, despite their success, CNNs face challenges in effectively capturing temporal patterns in long Forex time series.

Recurrent neural networks, particularly long short-term memory (LSTM) and gated recurrent unit (GRU) models, have demonstrated superior performance in capturing long-term dependencies in time series data (Fazeli & Houghten, 2019; Fischer & Krauss, 2018; Ghahremani et al., 2021; Yıldırım et al., 2021). Yıldırım et al. (2021) proposed a hybrid LSTM model that combined macroeconomic and technical indicators to predict the EUR/USD pair, showing improved performance compared to other methods. Similarly, Olanrewaju et al. (2023) evaluated LSTM and GRU models for USD/NGN Forex prediction, with GRU showing better accuracy and lower error. Ayitey Junior et al. (2022) further improved performance by using stacked LSTMs, which outperformed single LSTM models for the AUD/USD pair. Additionally, Dash et al. (2023) explored bidirectional LSTMs combined with the DeepSense network for predicting the closing prices of the EUR/USD pair, illustrating the potential of advanced RNN architectures for Forex forecasting.

Despite their strengths, LSTM and GRU models face limitations when applied to multivariate and extended Forex datasets. These models can sometimes focus on irrelevant information, reducing their prediction accuracy. To address this issue, attention mechanisms have been introduced to enhance the capability of RNNs by allowing the model to focus on the most relevant features in the data (Chen et al., 2023; Cheng, 2023; Li, 2023; Lou et al., 2023).

Lou et al. (2023) applied a Bi-LSTM model with an attention mechanism to predict gold and Bitcoin prices. The attention mechanism helped the model identify critical time steps, improving accuracy. Similarly, Chen et al. (2023) developed a hybrid model combining Temporal Convolutional Networks (TCN) with LSTM, using causal self-attention for stock price prediction. Li (2023) employed an attention-based LSTM for stock selection, showing better performance than traditional LSTM models. Cheng (2023) applied an attention-embedded dual LSTM for financial risk early warning, achieving high accuracy in predicting financial risks for companies listed on China's New Third Board.

While the above models leverage attention mechanisms to enhance predictive performance, they were designed for predicting financial risks (Cheng, 2023) or prices of stocks (Chen et al., 2023; Li, 2023), gold, and Bitcoin (Lou et al., 2023), and are not readily applicable to Forex rate prediction. Forex markets are influenced by global economic events and policy changes, introducing unique dynamic patterns that differ from other financial assets such as stocks or Bitcoin. These distinctions require models that can effectively filter out irrelevant information and focus on the most critical features for accurate predictions.

Table 1
Summary of related work.

Authors	Deep learning methods	Attention-based	Prediction application	Data used
Galeshchuk and Mukherjee (2017)	CNN	No	Forex rates	USD/GBP, EUR/USD, JPY/USD
Yıldırım et al. (2021)	Hybrid LSTM	No	Forex rates	EUR/USD
Olanrewaju et al. (2023)	LSTM, GRU	No	Forex rates	USD/NGN
Ayitey Junior et al. (2022)	Stacked LSTM	No	Forex rates	AUD/USD
Dash et al. (2023)	Bi-LSTM	No	Forex rates	EUR/USD
Lou et al. (2023)	Bi-LSTM	Yes	Gold and Bitcoin prices	Gold and Bitcoin price data
Chen et al. (2023)	TCN and LSTM	Yes	Stock prices	Stock price data
Li (2023)	LSTM	Yes	Historical stock prices	Historical stock prices
Cheng (2023)	Dual LSTM	Yes	Financial risk early warning	Financial data (e.g., total assets, net asset income, net profit from sales)
Kantoutsis et al. (2024)	Transformer	Yes	Forex rates	URU/USD, BGP/USD

In this context, Transformer-based models have shown promise due to their self-attention mechanisms, which help focus on relevant temporal patterns in noisy Forex data. Kantoutsis et al. (2024) applied a Transformer model to high-frequency EUR/USD and GBP/USD trading, using an encoder with Exponential Moving Average (EMA) smoothing to improve predictive accuracy. However, Transformers can be computationally intensive, especially with large datasets, which may limit their efficiency in real-time Forex applications.

In this article, we propose the ALFA model, specifically tailored for predicting the closing rates in Forex data. The attention mechanism allows the model to focus on the most relevant features within the multivariate Forex dataset, improving predictive accuracy by reducing the impact of irrelevant data.

Table 1 provides a summary of the related works discussed above.

3. The proposed model ALFA

The proposed model consists of two main components: an LSTM layer and an attention mechanism, as illustrated in Fig. 1. We first preprocess the raw Forex data, i.e., clean, format, and normalize it. This processed data is then passed into the LSTM network, which extracts temporal features from the time series data.

Next, these temporal features are fed into the attention layer. The attention mechanism, positioned on top of the LSTM layer, identifies and assigns higher weights to the most relevant temporal features across all time steps, allowing the model to focus on the critical parts of the input data. The output from the attention layer is then passed to a multilayer perceptron (MLP), which performs the final regression step to predict Forex rates.

This architecture enables the model to dynamically prioritize important features and capture long-term dependencies within multivariate time series data, improving predictive accuracy. The following subsections provide a more detailed explanation of the LSTM and attention mechanism layers.

3.1. Long Short-Term Memory (LSTM) network

The Long Short-Term Memory (LSTM) network is a type of recurrent neural network (RNN) designed to recognize patterns in long time series data. Its architecture includes memory cells that allow the network to retain important information over time, making it well-suited for handling the sequential nature of Forex data. In our approach, the LSTM layer processes the time series and extracts high-level temporal features that capture the underlying dynamics of the Forex market.

Given an input sequence $X = (x^1, x^2, \dots, x^n)^T \in \mathbb{R}^{n\times T}$, where n is the number of input features and T is the number of time steps, the LSTM updates its states using the following computations for each time step t:

1. **Forget Gate**: Determines which information to discard from the cell state.
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t^i] + b_f) \tag{1}$$
2. **Input Gate**: Updates the cell state with new information.
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t^i] + b_i) \tag{2}$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t^i] + b_C) \tag{3}$$
3. **Cell State**: Updates the old cell state to produce the new cell state.
$$c_t = f_t * c_{t-1} + i_t * \tilde{C}_t \tag{4}$$
4. **Output Gate**: Determines what information from the cell state will be used in the output.
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t^i] + b_o) \tag{5}$$
$$h_t = o_t * \tanh(c_t) \tag{6}$$

In the above equations,

- h_{t-1} is the previous hidden state.
- x_t^i is the input at the current time step.
- f_t, i_t, o_t are the activations of the forget, input, and output gates, respectively.
- $\tilde{C}_t$ is the candidate cell state.
- c_t is the cell state at time t.
- h_t is the hidden state at time t.
- W and b are the weight matrices and bias vectors, respectively.
- σ is the sigmoid activation function.
- $*$ denotes element-wise multiplication.

Given a sequence of time steps in the Forex time series, the LSTM produces a sequence of hidden states $h_1, h_2, \dots, h_T$, where T is the number of time steps.

These equations describe the fundamental operations of an LSTM cell. The use of gates allows the LSTM to control the flow of information, making it capable of identifying and maintaining patterns over long sequences. This property is especially useful for processing Forex time series data, which often contains complex temporal patterns.

3.2. Attention mechanism

Building on the features extracted by the LSTM, the attention mechanism assigns an "attention weight" to each feature, determining its importance. This allows the model to focus on the most crucial features while reducing the influence of less relevant details. By emphasizing important trends and patterns in the data, the attention mechanism enhances overall prediction accuracy.

The attention mechanism works by assigning a score to each hidden state h_t, reflecting its relevance in the current context (see Fig. 2). These scores are then used to compute a context vector that summarizes the most important information from the sequence.

The attention scores for each time step are calculated as follows:
$$P_t = a(h_t), \tag{7}$$

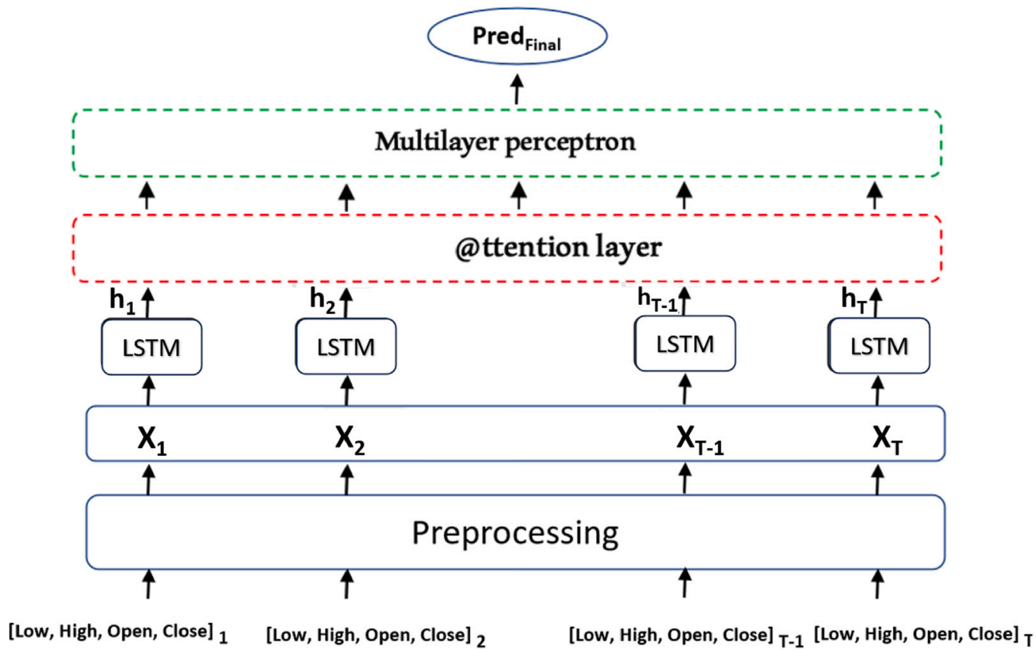


Fig. 1. The proposed architecture. X_i represents the input features with a window size of T.

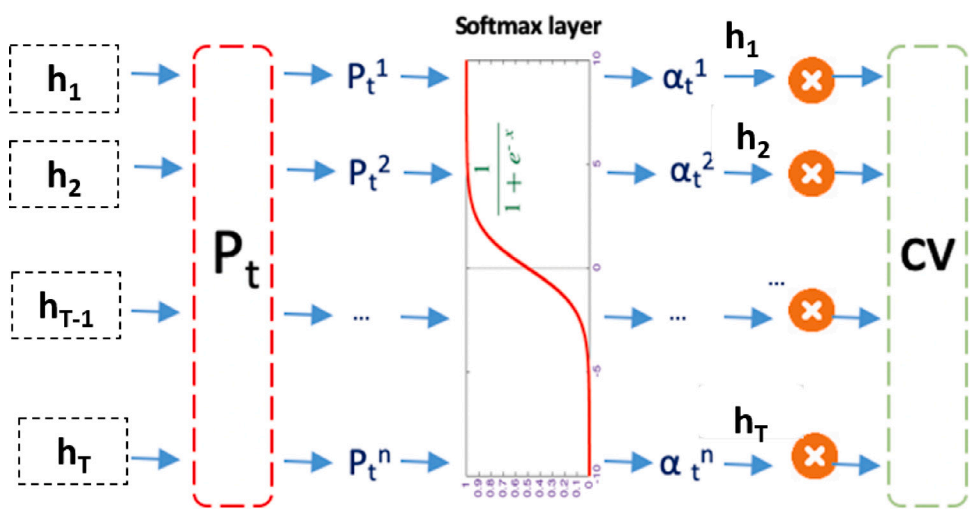


Fig. 2. Attention mechanism.

where a is a feed-forward neural network that produces a scalar value for each hidden state.

These scores are then normalized using the softmax function to produce the attention weights:

$$\alpha_t = \frac{e^{P_t}}{\sum_{j=1}^{T} e^{P_j}}. \tag{8}$$

The context vector CV is computed as a weighted sum of the hidden states:

$$CV = \sum_{t=1}^{T} \alpha_t h_t. \tag{9}$$

This context vector CV encapsulates the most relevant information from the sequence, as determined by the attention mechanism. It is typically passed to a subsequent layer, such as a multilayer perceptron (MLP), for the final prediction.

4. Experiment settings

In this section, we describe the experimental setup used to evaluate the performance of the proposed model. We outline the baseline models, evaluation scenarios, dataset, and evaluation metrics used to compare ALFA against the baseline models.

4.1. Baseline models

We compare ALFA with the following baseline models: LSTM, GRU, Bi-LSTM, Stacked LSTM, TCN, and Transformer. We selected LSTM, GRU, Bi-LSTM, and Stacked LSTM because they have been used in previous work on Forex rate prediction (Ayitey Junior et al., 2022; Dash et al., 2023; Olanrewaju et al., 2023; Yıldırım et al., 2021). TCN has been used for stock price prediction (Chen et al., 2023), while Transformer is a state-of-the-art deep learning model.

- **Long Short-Term Memory (LSTM)** (Li et al., 2019): LSTM is a specialized type of recurrent neural network (RNN) designed to effectively capture both short- and long-term dependencies in sequential data. It achieves this through a memory cell and three gating mechanisms: the input gate, forget gate, and output gate. These gates regulate the flow of information, allowing the model to retain relevant information over extended sequences while discarding less useful data. This capability makes LSTM particularly suitable for modeling complex temporal relationships in dynamic environments such as Forex markets, where patterns evolve over time.

- **Gated Recurrent Unit (GRU)** (Cho et al., 2014): The GRU is an advanced variant of the traditional RNN that simplifies the architecture of LSTMs while maintaining strong performance. It reduces computational complexity by combining the cell and hidden states and using only two gates: the update gate and the reset gate. The update gate controls the extent to which the previous state influences the current state, while the reset gate determines how much past information to forget. This streamlined design facilitates faster training and efficient learning of temporal dependencies, making GRUs effective for time-sensitive applications where computational efficiency is critical.
- **Bidirectional LSTM (Bi-LSTM)** (Sutskever et al., 2014): The Bidirectional LSTM is an extension of the LSTM model that processes input sequences in both forward and backward directions, allowing the model to capture context from both past and future time steps. In this study, we use the same hyperparameters as the single-layer LSTM to ensure consistent comparisons across models.
- **Stacked LSTM (St-LSTM)**: The Stacked LSTM model employs multiple LSTM layers to capture more complex patterns in sequential data. For this study, we use two LSTM layers to enhance the model's capacity for hierarchical feature learning, consistent with the parameters used in the single-layer LSTM setup.
- **Temporal Convolutional Network (TCN)** (Bai et al., 2018): TCN is designed to model sequential data through causal convolutions, ensuring that the prediction at any time step depends only on current and past inputs. It incorporates dilated convolutions to efficiently capture long-range dependencies, with the output aggregated via global pooling to produce a compact feature representation. We implemented a TCN with two layers of dilated convolutions, designed to match the scale of the ALFA model (in terms of the number of parameters). This ensures a fair comparison by keeping model complexity consistent.
- **Transformer** (Vaswani et al., 2017): The Transformer model employs self-attention mechanisms to effectively capture relationships within the entire sequence, regardless of the distance between elements. It consists of multiple stacked Transformer blocks, each containing self-attention and position-wise feed-forward layers. We implemented a Transformer with four stacked blocks, each combining multi-head self-attention and feed-forward layers. This architecture is scaled to match ALFA (in terms of the number of parameters) to ensure balanced and fair performance comparisons.

Further details of the model architectures and hyperparameters are provided in Table 2. (Note that the hyperparameters used for the baseline models were optimized for the datasets in this study (see Section 4.2) and may differ from those used in the papers cited in Table 2, as those studies used different datasets.)

4.2. Datasets and parameters

The dataset used for training and testing the ALFA model consists of hourly low, high, open, and close prices for three currency pairs: EUR/USD, USD/JPY, and GBP/JPY. These currency pairs represent different market conditions, including varying levels of volatility, trading volumes, liquidity, and economic influences, providing a comprehensive framework for evaluating the performance of the models across diverse scenarios. The dataset covers the period from March 2014 to April 2024, spanning ten years of historical data.

Below, we provide additional context on each currency pair:

- **EUR/USD:** The most traded currency pair globally, characterized by high liquidity and relatively low volatility compared to other pairs. It reflects the economic health of the European Union and the United States, making it a key indicator of global economic trends.
- **USD/JPY:** A major pair with high liquidity and moderate volatility (compared to EUR/USD). It serves as a key indicator of economic relations between the United States and Japan and is influenced by interest rate differentials, economic reports, and geopolitical stability.
- **GBP/JPY:** Known for its high volatility (compared to EUR/USD), this pair experiences large price swings driven by the differing economic landscapes of the UK and Japan. It is influenced by geopolitical events, central bank policies, and economic data releases.

For each currency pair, the dataset was divided into three subsets: training, validation, and testing. The first 60% of the time series was allocated for training, the next 20% for validation, and the final 20% for testing on unseen data. This sequential split preserves the temporal order and dependencies within the data, ensuring that the forecasting model remains unbiased and free from disruptions caused by random sampling.

Additionally, we experimented with different time step values, denoted by T, to determine the optimal historical window size for forecasting. A systematic search was conducted over the set of values $T = \{2, 4, 8, 16, 32\}$ to identify the best configuration for the model, allowing for an analysis of how varying lengths of historical data influence forecasting accuracy.

4.3. Evaluation metrics

The performance of the proposed model is evaluated using the following three metrics, which compare the predicted values with the actual values:

1. **Mean Absolute Error (MAE)**, computed as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i| \tag{10}$$

2. **Root Mean Square Error (RMSE)**, given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \tag{11}$$

3. **Mean Absolute Percentage Error (MAPE)**, defined as:

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{y_i} \tag{12}$$

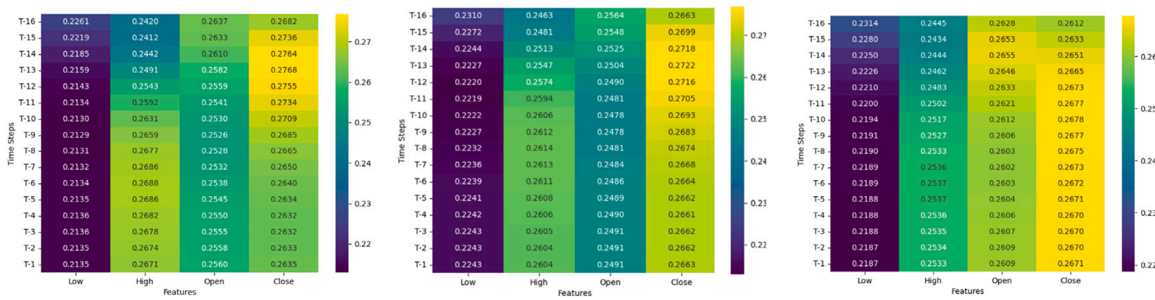
In the above formulas, y_i represents the true value, $\hat{y}_i$ denotes the predicted value at the ith test instance, and N is the total number of test instances.

The RMSE metric squares the errors before aggregation, assigning greater weight to larger errors. Given the importance of minimizing significant errors in our context, we use RMSE for model optimization.

In Section 5.1, we present and compare the performance of ALFA and the baseline models using the MAE metric, selected for its straightforward interpretation. The RMSE and MAPE results are provided in Appendices B and C, respectively, due to space limitations.

Table 2
Model architectures and hyperparameters.

Model	Used in reference	Architecture and hyperparameters
LSTM	Yıldırım et al. (2021)	LSTM (76 Units) → Dense (1 Unit)
GRU	Olanrewaju et al. (2023)	GRU (76 Units) → Dense (1 Unit)
Bi-LSTM	Dash et al. (2023)	Bi-LSTM (76 Units) → Dense (1 Unit)
St-LSTM	Ayitey Junior et al. (2022)	LSTM (76 Units) → LSTM (100 Units) → Dense (1 Unit)
TCN	Chen et al. (2023)	Dilated Conv1D (256, 3, 1) → Conv1D (76, 3, 4) → GAP → Dense (1)
Transformer	Kantoutsis et al. (2024)	4 Transformer Blocks (256 Dim, 0.1 Dropout) → Dense (76, 1 Unit)
ALFA	–	LSTM (76 Units) → Attention (100 Units) → Dense (1 Unit)



(a) EUR/USD (b) USD/JPY (c) GBP/JPY

Fig. 3. Attention weights of the ALFA model for EUR/USD, GBP/JPY, and USD/JPY trained with low, high, open, and close features.

4.4. Performance evaluation scenarios

We conducted comprehensive experiments to evaluate the performance of ALFA against six baseline models, dividing the experiments into three sets as follows.

Set 1: Predictive performance of ALFA vs. the baseline models

This is the most comprehensive set of experiments. We compared the MAE of ALFA against the six baseline models listed in Section 4.1.

The experiments were conducted for the three currency pairs listed in Section 4.2: EUR/USD, USD/JPY, and GBP/JPY.

For each currency pair, we trained and evaluated each model using the following 12 feature sets: {low}, {high}, {open}, {close}, {low, high}, {low, open}, {low, close}, {high, open}, {high, close}, {open, close}, {low, high, open, close}, and a set of technical indicators: {SMA, MACD, ROC, RSI, BB, CCI}.

Detailed mathematical definitions of these technical indicators are provided in Appendix A. The results of this evaluation are discussed in Section 5.1.

Set 2: Model complexity and running time

We compared the seven models in terms of model complexity (measured by the number of parameters) and average inference time. In general, a model with a high number of parameters may improve predictive performance at the cost of longer inference time. We examine this trade-off in Section 5.2.

Set 3: Effectiveness of ALFA in trading scenarios

We applied ALFA to execute trades using recent data for USD/JPY, EUR/USD, and GBP/JPY from June 10, 2024, to September 6, 2024. We evaluate the trading effectiveness of ALFA using metrics such as success rate, total return, and average gain per trade, which are defined and discussed in Section 5.3.

5. Results and discussion

In this section, we present and analyze the results from the three sets of experiments outlined above in Section 4.4.

5.1. Evaluation of predictive performance: ALFA vs. Baseline models

In this section, we analyze the predictive performance of the ALFA model against baseline models: TCN, Transformer, GRU, Bi-LSTM, St-LSTM, and LSTM, using various currency pairs and feature sets described in Section 4.4. Tables 3, 4, and 5 present the experimental results for the EUR/USD, USD/JPY, and GBP/JPY currency pairs, respectively. These tables report the mean absolute error (MAE) values with confidence intervals of 95%. For each set of feature (each row), the top three performers are highlighted with the color intensity indicating the ranks (darker shades indicating higher ranks).

Following are observations and findings drawn from the results:

- The results of ALFA are among the top three in all cases except two (the second row of Tables 3 and 4). Moreover, ALFA achieves the first rank in more cases than any other baselines. This suggests that ALFA can capture temporal dependencies and extract meaningful representations from the input data better than the other models, thanks to its attention mechanism.
- In general, models trained with pairs of features (rows 5 to 10 in the three tables) perform better than models trained with a single feature (rows 1 to 4). This shows that using single features may not yield good performance due to insufficient data provided to a model.
- However, models trained on all four features {low, high, open, close} (row 11 in the tables) do not always perform better than those trained with only two features. For example, in Table 4, ALFA performed better when trained with the {high, open} set than with the 4-feature set (MAE of 0.176 vs. 0.364).
- In general, sets of two or four features can provide good performance in most cases. However, the optimal feature set for one currency pair may not yield the best performance for another. For example, using the EUR/USD pair (Table 3), ALFA achieves the best MAE (0.887) with the feature set {low, close}. However, using the GBP/JPY pair (Table 5), ALFA achieves the best MAE (0.177) with the 4-feature set.
- Using all raw data (all four features) generally provides good performance. However, if one of the goals is to reduce the amount of data needed for input into a model by reducing the number of features, experiments and evaluations are required to determine

Table 3
MAE values of the seven models trained on the **EUR/USD** currency pair using different feature combinations. All numbers are presented in the format of $\times 10^{-3}$. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	2.017 ± 0.014	12.65 ± 0.050	1.199 ± 0.009	1.226 ± 0.009	1.177 ± 0.011	1.151 ± 0.009	1.011 ± 0.010
High	2.423 ± 0.018	8.321 ± 0.050	1.337 ± 0.009	1.231 ± 0.009	1.424 ± 0.011	1.057 ± 0.009	1.642 ± 0.009
Open	2.228 ± 0.016	14.89 ± 0.050	4.130 ± 0.010	1.350 ± 0.011	1.457 ± 0.011	1.337 ± 0.010	1.293 ± 0.010
Close	1.975 ± 0.014	7.983 ± 0.050	1.599 ± 0.007	1.121 ± 0.008	1.175 ± 0.009	1.011 ± 0.007	0.970 ± 0.008
Low, High	2.561 ± 0.020	3.524 ± 0.039	2.817 ± 0.008	1.155 ± 0.008	1.225 ± 0.010	1.014 ± 0.008	1.053 ± 0.009
Low, Open	2.514 ± 0.019	11.43 ± 0.058	3.562 ± 0.009	1.234 ± 0.009	1.342 ± 0.012	1.136 ± 0.009	1.034 ± 0.010
Low, Close	2.538 ± 0.019	18.66 ± 0.058	2.656 ± 0.007	1.030 ± 0.007	1.203 ± 0.010	0.931 ± 0.007	0.887 ± 0.008
High, Open	2.624 ± 0.020	9.781 ± 0.060	1.979 ± 0.009	1.311 ± 0.010	1.486 ± 0.012	1.073 ± 0.009	1.070 ± 0.009
High, Close	2.681 ± 0.020	15.43 ± 0.059	3.228 ± 0.008	1.162 ± 0.008	1.321 ± 0.011	0.933 ± 0.009	0.888 ± 0.007
Open, Close	2.779 ± 0.021	6.912 ± 0.060	1.542 ± 0.008	1.196 ± 0.008	1.249 ± 0.011	1.046 ± 0.007	0.918 ± 0.008
Low, High, Open, Close	2.734 ± 0.021	13.98 ± 0.060	1.049 ± 0.008	1.121 ± 0.008	1.383 ± 0.011	1.075 ± 0.007	0.905 ± 0.008
Technical Indicators	3.829 ± 0.032	17.35 ± 0.060	2.153 ± 0.009	1.541 ± 0.010	1.606 ± 0.011	1.353 ± 0.010	1.010 ± 0.008

Table 4
MAE values of the seven models trained on the **USD/JPY** currency pair using different feature combinations. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	0.562 ± 0.002	11.57 ± 0.094	0.770 ± 0.002	0.214 ± 0.001	0.931 ± 0.002	0.197 ± 0.001	0.437 ± 0.002
High	0.491 ± 0.002	11.57 ± 0.094	0.358 ± 0.002	0.267 ± 0.002	0.776 ± 0.002	0.205 ± 0.001	0.382 ± 0.002
Open	0.603 ± 0.002	11.57 ± 0.094	0.689 ± 0.002	0.349 ± 0.002	0.929 ± 0.002	0.376 ± 0.002	0.405 ± 0.002
Close	0.435 ± 0.002	11.57 ± 0.094	0.382 ± 0.002	0.191 ± 0.001	0.843 ± 0.002	0.274 ± 0.002	0.308 ± 0.002
Low, High	0.585 ± 0.002	29.99 ± 0.113	0.306 ± 0.001	0.346 ± 0.001	0.725 ± 0.002	0.393 ± 0.001	0.242 ± 0.002
Low, Open	0.712 ± 0.002	30.34 ± 0.114	0.435 ± 0.002	0.369 ± 0.001	0.795 ± 0.002	0.442 ± 0.001	0.243 ± 0.002
Low, Close	0.607 ± 0.002	29.43 ± 0.111	0.319 ± 0.001	0.386 ± 0.001	0.775 ± 0.002	0.382 ± 0.001	0.230 ± 0.002
High, Open	0.616 ± 0.002	31.20 ± 0.115	0.389 ± 0.002	0.411 ± 0.002	0.709 ± 0.002	0.478 ± 0.002	0.176 ± 0.001
High, Close	0.560 ± 0.002	30.99 ± 0.115	0.254 ± 0.001	0.417 ± 0.001	0.643 ± 0.002	0.316 ± 0.001	0.377 ± 0.002
Open, Close	0.520 ± 0.002	31.53 ± 0.116	0.309 ± 0.001	0.471 ± 0.001	0.711 ± 0.002	0.400 ± 0.001	0.294 ± 0.002
Low, High, Open, Close	0.561 ± 0.002	8.33 ± 0.028	0.364 ± 0.001	0.583 ± 0.002	0.641 ± 0.002	0.508 ± 0.002	0.364 ± 0.003
Technical Indicators	1.382 ± 0.003	26.37 ± 0.099	1.156 ± 0.003	0.767 ± 0.002	1.943 ± 0.005	0.846 ± 0.002	0.412 ± 0.002

Table 5
MAE values of the seven models trained on the **GBP/JPY** currency pair using different feature combinations. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	0.408 ± 0.002	11.925 ± 0.097	0.330 ± 0.002	0.280 ± 0.002	0.363 ± 0.002	0.334 ± 0.005	0.199 ± 0.004
High	0.337 ± 0.003	11.925 ± 0.097	0.265 ± 0.002	0.209 ± 0.002	0.400 ± 0.003	0.237 ± 0.004	0.192 ± 0.004
Open	0.459 ± 0.003	11.925 ± 0.097	0.392 ± 0.003	0.261 ± 0.002	0.470 ± 0.003	0.219 ± 0.005	0.223 ± 0.003
Close	0.322 ± 0.002	11.925 ± 0.097	0.261 ± 0.002	0.194 ± 0.002	0.250 ± 0.002	0.278 ± 0.005	0.217 ± 0.004
Low, High	0.382 ± 0.002	18.117 ± 0.079	0.233 ± 0.002	0.253 ± 0.002	0.291 ± 0.003	0.184 ± 0.002	0.214 ± 0.004
Low, Open	0.441 ± 0.003	12.934 ± 0.071	0.315 ± 0.002	0.287 ± 0.002	0.364 ± 0.003	0.205 ± 0.002	0.240 ± 0.004
Low, Close	0.347 ± 0.002	18.116 ± 0.079	0.239 ± 0.002	0.233 ± 0.002	0.284 ± 0.002	0.221 ± 0.002	0.196 ± 0.004
High, Open	0.360 ± 0.003	18.321 ± 0.079	0.260 ± 0.002	0.300 ± 0.003	0.429 ± 0.003	0.335 ± 0.003	0.214 ± 0.004
High, Close	0.320 ± 0.002	17.285 ± 0.078	0.230 ± 0.002	0.233 ± 0.002	0.297 ± 0.003	0.266 ± 0.002	0.193 ± 0.004
Open, Close	0.353 ± 0.003	12.119 ± 0.092	0.247 ± 0.002	0.285 ± 0.002	0.316 ± 0.003	0.269 ± 0.002	0.186 ± 0.004
Low, High, Open, Close	0.364 ± 0.003	7.217 ± 0.034	0.276 ± 0.002	0.324 ± 0.003	0.351 ± 0.003	0.333 ± 0.003	0.177 ± 0.004
Technical Indicators	0.385 ± 0.002	18.211 ± 0.079	0.418 ± 0.003	0.355 ± 0.003	0.494 ± 0.003	0.352 ± 0.003	0.296 ± 0.004

the optimal set of features for a specific currency pair. This recommendation can be illustrated by the above example, combined with the observation that ALFA achieves the best MAE (0.176) with the feature set {high, open} for the USD/JPY currency pair (Table 4).

- The set of technical indicators (TIs) gives lower performance than the other sets in most cases, most likely due to data redundancy. Since TIs are derived from raw forex data (e.g., close prices), they introduce duplicate information and thus noises in a model, leading to lower predictive performance. Interestingly, among the models trained on TIs, ALFA outperforms all baseline models, demonstrating its ability to filter out redundant information and focus on the most relevant features.
- Among the baseline models, GRU, Bi-LSTM and LSTM are among the best performers, thanks to the ability of recurrent neural networks (RNNs) to capture sequential dependencies in financial time series. ALFA also benefits from this ability of RNNs, and the attention mechanism enables it to outperform GRU, Bi-LSTM and LSTM in most cases.

To illustrate the effectiveness of the attention mechanism in ALFA, Fig. 3 shows the attention weights assigned to each feature when ALFA is trained with the feature set {low, high, open, close} using the EUR/USD, USD/JPY and GBP/JPY currency pairs. The“heat maps” show how the model assigns varying levels of importance to different features.

Across all three currency pairs, the closing price feature consistently receives the highest attention weights. Intuitively, human traders also use closing prices of the day (hour) more often than the other three features to predict prices for the next day (hour), in the stock market or forex market. Furthermore, most technical indicators (such as those defined in Appendix A) are also derived from closing prices.

The high price feature also holds moderate importance, particularly in the GBP/JPY and USD/JPY currency pairs. The high price feature may signal resistance levels and overbought conditions, which are relevant factors to consider in forex trading. The open price feature carries slightly more weight in the EUR/USD pair than in the other two, possibly due to its sensitivity to overnight news and session transitions, which influence early market sentiment. Meanwhile, the low price feature consistently receives the least attention, suggesting that downside movements are often corrected within the same session, making them less relevant for the prediction task.

Table 6
Model complexity and average inference time with confidence intervals of 95%.

Method	Number of parameters	Avg inference time (ms)
GRU	18,317	0.2934 ± 0.0254
LSTM	24,093	0.5223 ± 0.0355
ALFA	**45,093**	**0.5473 ± 0.0336**
Bi-LSTM	48,185	0.7867 ± 0.0844
Stacked LSTM	94,917	0.850 ± 0.0355
Transformer	51,889	2.269 ± 0.0455
TCN	59,545	4.272 ± 0.125

5.2. Model complexity and running time

Table 6 presents a comparative analysis of model complexity and average inference time for ALFA and the baseline models for the EUR/USD currency pair (corresponding to the experimental results in Table 3).

To measure inference time, each model was run to make predictions on the entire test set. The total inference time for the test set was then divided by the total number of test samples to obtain the average inference time per sample.

All experiments were conducted on a Windows workstation with the following configuration: NVIDIA GeForce RTX 3070 graphics card, AMD Ryzen 9 5900HX processor (3.3 GHz, 8 cores, 16 logical threads), 16 GB DDR4 RAM, and a 1 TB solid-state drive.

Among the models, GRU has the lowest number of parameters (18,317) and one of the fastest inference times (0.2934 ms). LSTM follows with 24,093 parameters and a slightly higher average inference time of 0.5223 ms, while ALFA, with 45,093 parameters, maintains a competitive inference time of 0.5473 ms. These results, along with those in Section 5.1, suggest that ALFA offers a balance between model complexity, inference time, and predictive performance.

Although ALFA and stacked LSTM have similar depths, ALFA requires fewer parameters (45,093 vs. 94,917) and achieves a lower average inference time (0.5473 ms vs. 0.850 ms).

The Transformer model, with 51,889 parameters, incurs a significantly higher average inference time of 2.269 ms, likely due to its computationally expensive self-attention mechanism. Meanwhile, TCN, with 59,545 parameters, records the highest average inference time (4.272 ms).

For hourly Forex trading, a form of intra-day trading, the latency requirement typically ranges from milliseconds to a few seconds (Electronic Trading Hub, 2024). With an average inference time of 0.5473 ms, ALFA meets this latency requirement for real-time trading. Deploying ALFA on industry-grade or optimized hardware, e.g., LSTM-accelerated GPUs, can further reduce the inference time. This enhancement could make ALFA a potential candidate for real-time trading, where sub-millisecond execution speeds are required (Xelera Technologies, 2023).

5.3. Effectiveness of ALFA in trading scenarios

In this section, we validate the effectiveness of ALFA using recent data for USD/JPY, EUR/USD, and GBP/JPY from June 10, 2024, to September 6, 2024. The model uses the full set of low, high, open, and close prices to generate predictions.

The feature set {low, high, open, close} may not always yield the best performance for a specific currency pair, as shown in Tables 3 and 5. However, we used this set to avoid the bias of cherry-picking the best feature set for demonstration and to show that even with a sub-optimal feature set, ALFA is still capable of generating profits in trading scenarios.

Algorithm 1 details the step-by-step process of executing buy and sell trades using the ALFA model. A trade is executed only when the difference between the predicted price and the current price falls within predefined lower and upper bound thresholds. The lower bound eliminates trades where the price difference is too small, making the trade unprofitable due to negligible gains. The upper bound prevents trades involving extreme price fluctuations, reducing exposure to unpredictable volatility and risks of loss.

Algorithm 1 Trading Algorithm

Initialize:
$initial\_investment \leftarrow 1 lot\_size(100000)$
$total\_profit \leftarrow 0$
$total\_trades \leftarrow 0$
$successful\_trades \leftarrow 0$
$w \leftarrow$ window size
for $t = w$ to T **do** ▷ Start from w to ensure full window
  $prediction[t+1] \leftarrow model(price\_history[t-w+1 : t])$
  $price\_diff \leftarrow prediction[t+1] - price[t]$
  if $lower\_bound \leq |price\_diff| \leq upper\_bound$ **then**
    $total\_trades \leftarrow total\_trades + 1$
    if $price\_diff > 0$ **then** ▷ Long trade (buy low, sell high)
      $gain \leftarrow \frac{price[t+1]-price[t]}{pip\_size}$
    else ▷ Short trade (sell high, buy low)
      $gain \leftarrow \frac{price[t]-price[t+1]}{pip\_size}$
    end if
    $gain \leftarrow gain - spread$ ▷ Deduct spread cost
    if $gain > 0$ **then**
      $successful\_trades \leftarrow successful\_trades + 1$
    end if
    $profit\_base\_currency \leftarrow gain \times pip\_size \times lot\_size$
    $total\_profit \leftarrow total\_profit + profit\_base\_currency$
  end if
end for
$Success\_Rate \leftarrow \frac{successful\_trades}{total\_trades} \times 100$
$Avg\_Gain\_Per\_Trade \leftarrow \frac{total\_profit}{total\_trades}$
Output: $Success\_Rate$, $Avg\_Gain\_Per\_Trade$, $total\_profit$

When the difference between the predicted price and the current price at time t falls within the defined threshold range, the algorithm determines whether to buy or sell based on the predicted price movement. A buy order is placed when the model predicts a price increase, while a sell order is placed when the model predicts a price decrease.

To account for transaction costs, we use spread fees, which are charged by trading platforms as their commission. A spread, in this case, refers to the difference between the market price and the execution price of a buy/sell order, with the trading platform retaining this difference as a transaction fee. Spread fees are a standard cost across many trading platforms.

Spread fees are measured in pips, as shown in Table 7 for each currency pair. A pip (percentage in point) is the smallest unit of price movement in Forex trading, typically equal to 0.0001 for EUR/USD and 0.01 for JPY-based pairs. After calculating the total pips gained from a trade, the spread fee is deducted to reflect the actual gain.

To evaluate the effectiveness of the ALFA model, we use the following key trading performance indicators:

- **Success rate**: The percentage of trades where the model correctly predicts the direction of price movement. A higher success rate suggests high predictive accuracy in identifying profitable trades.
- **Total gain**: The cumulative gain from all executed trades. If this number is negative, it indicates a loss.
- **Average gain per trade**: The average profit per trade after accounting for trading costs. If this number is negative, it indicates a loss.

In the trading scenarios analyzed in this section, the initial investment to place a buy or sell trade is one standard lot, equivalent to 100,000 units of the base currency. That is, the initial investment is:

Table 7
Performance of ALFA in real world trading scenarios with various currency pairs over a 3-month period.

Ticker	Success rate	Total trades	Successful trades	Total gain	Avg gain per trade	Spread (pips)
EUR/USD	62.82%	164	102	$2,264 US	$14.51 US	1
USD/JPY	60.59%	236	143	¥721,600	¥3,057.63	3.5
GBP/JPY	59.30%	199	118	¥538,700	¥2,707.04	1.5

Table 8
Estimated annual returns.

Currency pair	Initial amount	3-mo gain	3-mo return	12-mo return
EUR/USD	$103,056	$2,264	2.20%	8.79%
USD/JPY	¥15,194,336	¥721,600	4.75%	19.00%
GBP/JPY	¥18,793,000	¥538,700	2.87%	11.47%

- 100,000 euros (approximately 103,056 USD) for the EUR/USD pair.
- 100,000 USD (approximately 15,194,336 yen) for the USD/JPY pair.
- 100,000 British pounds (approximately 18,793,000 yen) for the GBP/JPY pair.

Table 7 presents the success rate and total gain for each currency pair over the 3-month period from June 10, 2024, to September 6, 2024. The model demonstrates consistent success rates across all three pairs, with EUR/USD achieving the highest at 62.82%, followed by USD/JPY at 60.59% and GBP/JPY at 59.30%. This indicates that ALFA effectively predicts market direction with reasonable accuracy.

The higher success rate in EUR/USD can be attributed to several factors. As one of the most liquid currency pairs, EUR/USD typically exhibits more stable price movements and lower volatility compared to GBP/JPY and USD/JPY. This stability reduces erratic price swings and market noise, allowing the ALFA model to make more accurate predictions. Additionally, EUR/USD tends to have tighter spreads, which minimizes execution costs and enhances the likelihood of profitable trades.

In contrast, GBP/JPY and USD/JPY are generally more volatile, which can introduce greater unpredictability and make precise forecasting more challenging. Sudden price spikes and increased market noise in these pairs may contribute to a slightly lower success rate, even though they offer larger price movements that could result in higher potential profits per trade.

We extrapolated the total gains over the three months of trading given in Table 7 to estimate the annual returns (by multiplying the 3-month gains by four). The estimated annual returns are shown in Table 8. The returns range from 8.79% to 19%, comparable to the typical annual return for forex trading, which falls in the range of 5% to 30% for profitable traders (Forex Training Group, 2025).

6. Conclusion

In this study, we propose ALFA, an attention-based LSTM model for forex rate prediction. We conducted comprehensive experiments to evaluate and compare the performance of ALFA against several models used in previous work (GRU, LSTM, Bi-LSTM and stacked LSTM) and state-of-the-art deep learning models (TCN and Transformer). Following are the results and findings of our study:

- ALFA consistently ranks among the top three of the seven models evaluated, and ranks first in the majority of cases, across different currency pairs and feature sets. This advantage can be attributed to its attention mechanism, which filters out irrelevant or redundant data to focus on important features.
- Among the baseline models, GRU, Bi-LSTM and LSTM are among the best performers, thanks to the ability of recurrent neural networks (RNNs) to capture sequential dependencies in financial time series.
- In general, sets of two or four features can provide good performance in most cases. Experiments and evaluations are required to determine the optimal set of features for a specific currency pair.
- Technical indicators (TIs) generally give lower performance than raw data features, likely due to data redundancy. ALFA can mitigate this disadvantage of TIs (e.g., if only TI data are available for training a model) by using its attention mechanism to filter out redundant information and focus on the most relevant features.
- ALFA offers the best balance between model complexity, inference time and predictive performance among the seven models evaluated.
- The average inference time of ALFA (0.5473 ms), as measured in our experiments on a personal computer, meets the latency requirement for real-time trading (typically ranging from milliseconds to a few seconds).
- In our simulations of trading scenarios, ALFA yielded estimated annual returns of 8.79% to 19%, comparable to the typical annual return for forex trading, which falls in the range of 5% to 30% for profitable traders.

In summary, the above experimental results highlight the potential of ALFA as a forex trading tool in a comprehensive and risk-managed investment framework.

Future research directions include the following:

- Incorporating macroeconomic indicators (e.g., interest rates, inflation rates, GDP) into the model could provide a broader market context and thus potentially more accurate forex rate prediction.
- Combining ALFA with an NLP model for analysis of market sentiment (captured in news reports, expert analyses and social media discussions) could enhance forecasting accuracy.

7. Ethical implications

This article presents an academic exercise demonstrating the potential and capability of AI in financial applications. The forex market is highly volatile and risky, with no guaranteed returns. Therefore, traders should not rely solely on the proposed model for decision-making. Financial markets are influenced by multiple factors, including macroeconomic conditions and market sentiment, which are not incorporated into the current model and will be explored in future research. These factors, particularly unforeseen "black swan" events, may lead to significant investment losses. To mitigate risks, AI and automated trading should be integrated into a comprehensive framework that includes expert knowledge, market analysis, and robust risk management strategies, such as stop-loss orders, to minimize potential financial losses.

While ALFA is well-suited for various trading applications, it may be less optimized for high-frequency trading (HFT), where price movements are highly volatile and driven by real-time news and sentiment shifts. HFT requires ultra-low latency execution and advanced sentiment analysis techniques, which go beyond ALFA's reliance on historical patterns and technical indicators. As such, ALFA is best applied in medium- to long-term trading strategies rather than ultra-short-term market reactions.

Additionally, ALFA's performance may vary across different market conditions, particularly in exotic currency pairs that experience higher volatility due to lower liquidity, geopolitical events, and economic fluctuations. For instance, pairs like USD/TRY (US Dollar/Turkish Lira) are often influenced by external shocks, necessitating additional risk assessment and caution when using predictive models in such markets.

To ensure transparency and responsible use, the ALFA model code is publicly available on GitHub under an Apache License, along with comprehensive documentation detailing its methodology, performance, and limitations. This openness allows users to fully understand the model's capabilities and make informed financial decisions, reinforcing ethical and responsible AI deployment in Forex trading applications.

CRediT authorship contribution statement

Shahram Ghahremani: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Writing – review & editing. **Uyen Trang Nguyen:** Conceptualization, Supervision, Validation, Formal analysis, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN-2018-05911 and the York University PhD Fellowship. The authors thank NSERC, Canada and York University, Canada for their support of the research in this article.

Appendix A. Technical indicators

Technical indicators are mathematical calculations based on historical price, volume, or open interest data of financial instruments. They are designed to identify patterns, trends, and potential market reversals, providing insights into price movement behavior over time. These indicators help traders and analysts make informed decisions by highlighting key aspects such as momentum, volatility, trend strength, and market cycles. Technical indicators are widely applied across various financial markets, including stocks, commodities, futures, and Forex.

Several studies have employed technical indicators as features for training machine learning models (Alade & Okafor, 2024; Yıldırım et al., 2021). To evaluate the effectiveness of technical indicators in predicting future Forex prices, we trained models using seven widely used technical indicators: Moving Average (MA), Moving Average Convergence Divergence (MACD), Rate of Change (ROC), Momentum, Relative Strength Index (RSI), Bollinger Bands (BB), and Commodity Channel Index (CCI). Each of these indicators captures different characteristics of price dynamics. They are derived from mathematical formulas applied to price data, typically involving open, close, high, low, and volume values.

The mathematical formulas for these technical indicators are provided below:

- **Moving Average (MA):** The Moving Average is the average of the **closing** prices over a specified period:

  $$MA = \frac{\sum_{i=1}^{n} \mathbf{Close}_i}{n}$$

  where n is the period (10 in this case). The purpose of MA is to smooth out price data to identify trends and filter out short-term fluctuations.
- **Moving Average Convergence Divergence (MACD):** MACD measures the difference between two Exponential Moving Averages (EMAs), commonly used for trend and momentum analysis:

  $$MACD = EMA_{\text{short-term}} - EMA_{\text{long-term}}$$

  where the short-term EMA is computed over 12 periods and the long-term EMA over 26 periods. A signal line (a 9-period EMA of MACD) is used to generate buy/sell signals.
- **Rate of Change (ROC):** ROC calculates the percentage change in price over a specific period to measure momentum:

  $$ROC = \frac{\mathbf{Close}_t - \mathbf{Close}_{t-n}}{\mathbf{Close}_{t-n}} \times 100$$

  where t is the current time and n is the period (2 in this case).
- **Momentum:** Momentum measures the price difference over a defined period, indicating the strength of a trend:

  $$\text{Momentum} = \mathbf{Close}_t - \mathbf{Close}_{t-n}$$

  where n is the period (4 in this case).
- **Relative Strength Index (RSI):** RSI evaluates the magnitude of recent price changes to identify overbought or oversold conditions:

  $$RSI = 100 - \frac{100}{1 + RS}$$

  where

  $$RS = \frac{\text{Average Gain (10 periods)}}{\text{Average Loss (10 periods)}}$$

  The average gain is the mean of all positive **close** price movements over the last 10 periods. RSI ranges from 0 to 100, where values above 70 indicate overbought conditions and below 30 indicate oversold conditions.
- **Bollinger Bands (BB):** Bollinger Bands consist of a moving average (middle band) and two standard deviation bands (upper and lower) above and below the MA:

  $$\text{Middle Band} = MA_{20}$$

  $$\text{Upper Band} = MA_{20} + (2 \cdot \sigma_{20})$$

  $$\text{Lower Band} = MA_{20} - (2 \cdot \sigma_{20})$$

  where σ_{20} is the standard deviation of the **closing** price over 20 periods.
- **Commodity Channel Index (CCI):** CCI measures the deviation of the price from its average over a specific period to identify overbought or oversold levels:

  $$CCI = \frac{\text{Typical Price} - SMA_{20}}{0.015 \cdot \text{Mean Deviation}}$$

  where

  $$\text{Typical Price} = \frac{\mathbf{High} + \mathbf{Low} + \mathbf{Close}}{3}$$

  SMA is the Simple Moving Average of the Typical Price, and Mean Deviation is the average absolute deviation from the SMA.

Appendix B. Evaluation of predictive performance of ALFA and the baseline models using the RMSE metric

While MAE provides a direct measure of absolute differences between predicted and actual values, RMSE is often preferred in financial forecasting due to its ability to penalize larger errors more heavily. This characteristic is particularly important in Forex prediction, where outlier movements and volatility spikes can significantly impact trading strategies. Since RMSE squares the deviations before averaging, it places greater emphasis on larger prediction errors, making it a more suitable metric for applications where risk management and precision are critical.

Table B.9
RMSE values of the seven models trained on the **EUR/USD** currency pair using different feature combinations. All numbers are presented in the format of $\times 10^{-3}$. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	2.747 ± 0.123	10.203 ± 0.102	1.632 ± 0.081	1.671 ± 0.080	1.731 ± 0.100	1.612 ± 0.082	1.544 ± 0.096
High	3.281 ± 0.154	20.672 ± 0.095	1.887 ± 0.080	1.797 ± 0.080	2.067 ± 0.088	1.608 ± 0.078	1.642 ± 0.077
Open	3.044 ± 0.146	7.956 ± 0.120	16.248 ± 0.085	1.956 ± 0.097	2.096 ± 0.102	1.901 ± 0.090	1.876 ± 0.094
Close	2.615 ± 0.098	13.417 ± 0.110	3.189 ± 0.051	1.632 ± 0.052	1.684 ± 0.062	1.413 ± 0.050	1.399 ± 0.051
Low, High	3.571 ± 0.187	24.531 ± 0.108	25.739 ± 0.090	1.583 ± 0.058	1.855 ± 0.078	1.500 ± 0.054	1.524 ± 0.058
Low, Open	3.483 ± 0.180	18.362 ± 0.098	19.846 ± 0.104	1.748 ± 0.074	1.931 ± 0.099	1.598 ± 0.074	1.547 ± 0.083
Low, Close	3.521 ± 0.188	22.157 ± 0.107	26.342 ± 0.109	1.435 ± 0.052	1.732 ± 0.077	1.368 ± 0.049	1.318 ± 0.052
High, Open	3.614 ± 0.199	6.784 ± 0.093	4.016 ± 0.076	1.925 ± 0.078	2.145 ± 0.096	1.617 ± 0.077	1.603 ± 0.071
High, Close	3.664 ± 0.186	15.673 ± 0.105	11.214 ± 0.100	1.706 ± 0.052	1.907 ± 0.075	1.411 ± 0.051	1.322 ± 0.049
Open, Close	3.806 ± 0.204	9.378 ± 0.110	2.427 ± 0.057	1.677 ± 0.053	1.798 ± 0.075	1.469 ± 0.050	1.348 ± 0.051
Low, High, Open, Close	3.798 ± 0.215	21.893 ± 0.114	1.444 ± 0.049	1.531 ± 0.052	1.991 ± 0.087	1.472 ± 0.049	1.339 ± 0.051
Technical Indicators	19.847 ± 0.109	14.287 ± 0.100	3.332 ± 0.060	2.098 ± 0.060	2.262 ± 0.079	1.892 ± 0.061	1.458 ± 0.054

Table B.10
RMSE values of the seven models trained on the **USD/JPY** currency pair using different feature combinations. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	0.612 ± 0.002	16.630 ± 3.751	0.829 ± 0.003	0.279 ± 0.002	0.965 ± 0.004	0.265 ± 0.002	0.516 ± 0.003
High	0.545 ± 0.003	16.630 ± 3.751	0.412 ± 0.003	0.331 ± 0.002	0.816 ± 0.004	0.275 ± 0.002	0.468 ± 0.003
Open	0.664 ± 0.003	16.630 ± 3.751	0.748 ± 0.004	0.418 ± 0.003	0.967 ± 0.005	0.444 ± 0.003	0.492 ± 0.004
Close	0.479 ± 0.002	16.630 ± 3.751	0.427 ± 0.002	0.245 ± 0.002	0.875 ± 0.004	0.339 ± 0.002	0.398 ± 0.003
Low, High	0.640 ± 0.003	33.235 ± 5.453	0.352 ± 0.002	0.384 ± 0.002	0.767 ± 0.004	0.431 ± 0.002	0.324 ± 0.002
Low, Open	0.775 ± 0.003	33.590 ± 5.547	0.480 ± 0.002	0.410 ± 0.002	0.834 ± 0.004	0.481 ± 0.002	0.356 ± 0.003
Low, Close	0.658 ± 0.003	32.647 ± 5.296	0.360 ± 0.001	0.417 ± 0.002	0.811 ± 0.004	0.415 ± 0.002	0.304 ± 0.002
High, Open	0.676 ± 0.003	34.461 ± 5.777	0.438 ± 0.002	0.458 ± 0.003	0.762 ± 0.005	0.521 ± 0.003	0.256 ± 0.003
High, Close	0.613 ± 0.003	34.247 ± 5.721	0.302 ± 0.002	0.452 ± 0.002	0.696 ± 0.004	0.357 ± 0.002	0.424 ± 0.002
Open, Close	0.572 ± 0.003	34.790 ± 5.864	0.350 ± 0.002	0.507 ± 0.002	0.758 ± 0.004	0.437 ± 0.002	0.353 ± 0.002
Low, High, Open, Close	0.616 ± 0.003	9.074 ± 0.503	0.399 ± 0.002	0.631 ± 0.003	0.707 ± 0.004	0.556 ± 0.003	0.505 ± 0.005
Technical Indicators	1.426 ± 0.007	29.243 ± 4.244	1.234 ± 0.007	0.818 ± 0.003	2.050 ± 0.018	0.890 ± 0.003	0.508 ± 0.003

Table B.11
RMSE values of the seven models trained on the **GB/JPY** currency pair using different feature combinations. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	0.496 ± 0.007	16.210 ± 4.035	0.414 ± 0.005	0.384 ± 0.006	0.452 ± 0.006	0.421 ± 0.009	0.307 ± 0.008
High	0.447 ± 0.017	16.210 ± 4.035	0.375 ± 0.015	0.334 ± 0.015	0.502 ± 0.017	0.330 ± 0.009	0.307 ± 0.010
Open	0.563 ± 0.016	16.210 ± 4.035	0.494 ± 0.015	0.381 ± 0.014	0.569 ± 0.016	0.336 ± 0.009	0.342 ± 0.010
Close	0.420 ± 0.010	16.210 ± 4.035	0.340 ± 0.008	0.284 ± 0.008	0.345 ± 0.009	0.360 ± 0.008	0.298 ± 0.008
Low, High	0.474 ± 0.010	20.190 ± 2.856	0.321 ± 0.008	0.355 ± 0.009	0.408 ± 0.010	0.282 ± 0.008	0.214± 0.006
Low, Open	0.529 ± 0.010	15.210 ± 2.006	0.403 ± 0.008	0.384 ± 0.008	0.479 ± 0.010	0.311 ± 0.008	0.335 ± 0.009
Low, Close	0.439 ± 0.008	20.190 ± 2.856	0.316 ± 0.006	0.318 ± 0.007	0.396 ± 0.008	0.312 ± 0.007	0.295 ± 0.008
High, Open	0.467 ± 0.016	20.380 ± 2.890	0.368 ± 0.014	0.414 ± 0.015	0.541 ± 0.016	0.443 ± 0.016	0.321 ± 0.010
High, Close	0.418 ± 0.012	19.390 ± 2.710	0.318 ± 0.009	0.342 ± 0.011	0.427 ± 0.012	0.368 ± 0.011	0.298 ± 0.009
Open, Close	0.461 ± 0.011	15.980 ± 3.634	0.331 ± 0.008	0.388 ± 0.010	0.437 ± 0.011	0.377 ± 0.010	0.282 ± 0.008
Low, High, Open, Close	0.467 ± 0.011	8.180 ± 0.646	0.381 ± 0.009	0.469 ± 0.010	0.530 ± 0.012	0.452 ± 0.010	0.275 ± 0.008
Technical Indicators	0.476 ± 0.013	20.280 ± 2.874	0.513 ± 0.009	0.483 ± 0.016	0.625 ± 0.022	0.472 ± 0.017	0.383 ± 0.009

Tables B.9, B.10, and B.11 present the RMSE evaluation results for EUR/USD, GBP/USD, and USD/JPY, comparing the ALFA model with baseline models. The results reveal trends similar to those observed in MAE-based evaluations.

The ALFA model consistently outperforms all baseline models across most feature sets, demonstrating its superior ability to capture complex temporal dependencies and minimize forecast errors. Notably, LSTM and Bi-LSTM remain the strongest competitors, outperforming TCN, Transformer, and GRU, further reinforcing the effectiveness of recurrent architectures in Forex forecasting.

Appendix C. Evaluation of predictive performance of ALFA and the baseline models using the MAPE metric

While MAE and RMSE provide absolute error values, MAPE is scale-independent, making it more suitable for comparing prediction errors across different currency pairs with varying price ranges. This property is particularly useful in Forex trading, where exchange rates fluctuate across different magnitudes. However, MAPE can be sensitive to very small actual values, potentially leading to large percentage errors. Therefore, it is important to interpret results in the context of the specific dataset.

Tables C.12, C.13, and C.14 present the MAPE evaluation results for EUR/USD, GBP/USD, and USD/JPY, comparing the ALFA model with baseline architectures. The findings are largely consistent with those observed using MAE and RMSE. The ALFA model continues to demonstrate superior performance across most feature sets, achieving lower MAPE values compared to baseline models, confirming its strong ability to capture relative changes in price movements.

Data availability

Data will be made available on request.

Table C.12
MAPE(%) values of the seven models trained on the **EUR/USD** currency pair using different feature combinations. All numbers are presented in the format of $\times 10^{-3}$. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	1.651 ± 0.012	12.402 ± 0.099	1.129 ± 0.008	1.156 ± 0.008	1.113 ± 0.010	1.084 ± 0.008	0.952 ± 0.009
High	2.018 ± 0.014	22.361 ± 0.105	1.264 ± 0.008	1.163 ± 0.008	1.357 ± 0.009	0.996 ± 0.008	1.020 ± 0.008
Open	1.874 ± 0.013	9.872 ± 0.093	4.102 ± 0.009	1.280 ± 0.010	1.384 ± 0.010	1.257 ± 0.009	1.223 ± 0.009
Close	1.696 ± 0.011	19.204 ± 0.102	1.560 ± 0.007	1.073 ± 0.007	1.120 ± 0.008	0.948 ± 0.007	0.919 ± 0.007
Low, High	2.121 ± 0.015	26.507 ± 0.115	2.796 ± 0.007	1.089 ± 0.007	1.225 ± 0.009	0.954 ± 0.007	1.274 ± 0.007
Low, Open	2.205 ± 0.016	8.981 ± 0.091	3.535 ± 0.008	1.173 ± 0.009	1.277 ± 0.010	1.068 ± 0.008	0.975 ± 0.008
Low, Close	2.109 ± 0.015	17.345 ± 0.108	2.635 ± 0.007	0.972 ± 0.007	1.146 ± 0.009	0.875 ± 0.006	0.837 ± 0.009
High, Open	2.093 ± 0.015	21.876 ± 0.112	1.936 ± 0.008	1.252 ± 0.008	1.418 ± 0.010	1.010 ± 0.008	1.010 ± 0.008
High, Close	2.289 ± 0.016	10.674 ± 0.095	3.212 ± 0.007	1.113 ± 0.007	1.263 ± 0.009	0.880 ± 0.007	0.838 ± 0.007
Open, Close	3.169 ± 0.024	14.789 ± 0.108	1.476 ± 0.007	1.140 ± 0.007	1.190 ± 0.009	0.980 ± 0.007	0.868 ± 0.007
Low, High, Open, Close	2.277 ± 0.016	27.593 ± 0.120	0.988 ± 0.007	1.051 ± 0.007	1.323 ± 0.010	1.011 ± 0.006	0.855 ± 0.007
Technical Indicators	2.205 ± 0.024	23.190 ± 0.110	2.092 ± 0.008	1.479 ± 0.008	1.540 ± 0.009	1.293 ± 0.008	0.959 ± 0.007

Table C.13
MAPE(%) values of the seven models trained on the **USD/JPY** currency pair using different feature combinations. All numbers are presented in the format of $\times 10^{-3}$. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	4.923 ± 0.018	87.770 ± 0.611	6.770 ± 0.024	1.848 ± 0.012	8.000 ± 0.019	1.675 ± 0.012	3.616 ± 0.017
High	4.283 ± 0.017	87.770 ± 0.611	3.139 ± 0.015	2.337 ± 0.014	6.568 ± 0.016	1.697 ± 0.012	3.165 ± 0.017
Open	5.288 ± 0.021	87.775 ± 0.611	6.046 ± 0.022	3.081 ± 0.017	7.938 ± 0.019	3.311 ± 0.017	3.315 ± 0.016
Close	3.801 ± 0.015	87.770 ± 0.611	3.362 ± 0.014	1.672 ± 0.011	7.149 ± 0.016	2.256 ± 0.012	2.535 ± 0.016
Low, High	5.112 ± 0.019	267.700 ± 1.096	2.681 ± 0.013	2.967 ± 0.012	6.076 ± 0.015	3.374 ± 0.012	1.934 ± 0.012
Low, Open	6.233 ± 0.022	270.700 ± 1.104	3.810 ± 0.015	3.196 ± 0.013	6.704 ± 0.016	3.808 ± 0.013	1.894 ± 0.015
Low, Close	5.304 ± 0.019	262.752 ± 1.081	2.800 ± 0.012	3.315 ± 0.011	6.526 ± 0.015	3.286 ± 0.012	1.895 ± 0.012
High, Open	5.371 ± 0.020	278.200 ± 1.121	3.401 ± 0.014	3.550 ± 0.014	5.922 ± 0.017	4.052 ± 0.013	1.479 ± 0.012
High, Close	4.878 ± 0.018	276.400 ± 1.117	2.220 ± 0.012	3.558 ± 0.012	5.341 ± 0.015	2.714 ± 0.011	3.131 ± 0.012
Open, Close	4.545 ± 0.017	281.100 ± 1.128	2.714 ± 0.012	4.049 ± 0.013	5.942 ± 0.015	3.439 ± 0.012	2.512 ± 0.013
Low, High, Open, Close	4.886 ± 0.018	70.700 ± 0.229	3.150 ± 0.012	4.850 ± 0.014	5.272 ± 0.017	4.223 ± 0.013	2.890 ± 0.020
Technical Indicators	11.878 ± 0.028	235.400 ± 0.966	10.166 ± 0.034	6.711 ± 0.022	17.019 ± 0.053	7.360 ± 0.022	3.603 ± 0.022

Table C.14
MAPE(%) values of the seven models trained on the **GBP/JPY** currency pair using different feature combinations. All numbers are presented in the format of $\times 10^{-3}$. Highlighted numbers indicate the top three performers for each feature, with the color intensity indicating the ranks.

Feature	TCN	Transformer	GRU	Bi-LSTM	St-LSTM	LSTM	ALFA
Low	2.704 ± 0.016	75.146 ± 0.531	2.256 ± 0.016	1.830 ± 0.015	2.449 ± 0.016	2.131 ± 0.239	1.256 ± 0.029
High	2.281 ± 0.018	75.146 ± 0.531	1.800 ± 0.017	1.407 ± 0.016	2.661 ± 0.018	1.500 ± 0.016	1.210 ± 0.029
Open	3.091 ± 0.020	75.146 ± 0.531	2.659 ± 0.019	1.766 ± 0.017	3.138 ± 0.019	1.365 ± 0.018	1.404 ± 0.032
Close	2.117 ± 0.015	75.146 ± 0.531	1.734 ± 0.013	1.283 ± 0.012	1.646 ± 0.014	1.775 ± 0.285	1.375 ± 0.025
Low, High	2.540 ± 0.017	123.740 ± 0.595	1.572 ± 0.014	1.665 ± 0.014	1.898 ± 0.016	1.470 ± 0.264	1.352 ±0.292
Low, Open	2.935 ± 0.017	88.069 ± 0.511	2.123 ± 0.015	1.903 ± 0.015	2.386 ± 0.017	1.780 ± 0.015	1.517 ± 0.029
Low, Close	2.274 ± 0.015	123.740 ± 0.595	1.630 ± 0.013	1.542 ± 0.013	1.838 ± 0.015	1.280 ± 0.287	1.238 ± 0.027
High, Open	2.414 ± 0.018	125.217 ± 0.599	1.760 ± 0.016	1.976 ± 0.017	2.825 ± 0.019	2.217 ± 0.017	1.287 ± 0.031
High, Close	2.115 ± 0.016	117.664 ± 0.576	1.539 ± 0.013	1.514 ± 0.014	1.911 ± 0.017	1.744 ± 0.015	1.221 ± 0.028
Open, Close	2.309 ± 0.017	77.477 ± 0.521	1.630 ± 0.013	1.856 ± 0.015	2.049 ± 0.016	1.755 ± 0.015	1.175 ± 0.026
Low, High, Open, Close	2.395 ± 0.017	48.195 ± 0.213	1.798 ± 0.015	2.058 ± 0.017	2.215 ± 0.020	2.149 ± 0.016	1.115 ± 0.026
Technical Indicators	2.565 ± 0.017	124.424 ± 0.597	2.779 ± 0.017	2.456 ± 0.021	3.291 ± 0.023	2.398 ± 0.020	1.878 ± 0.030

References

Alade, T., & Okafor, O. (2024). A novel FIG-LSTM ensemble machine learning technique for currency exchange rate forecasting. In *Proceedings of the IEEE Canadian conference on electrical and computer engineering*. http://dx.doi.org/10.1109/ccece59415.2024.10667208.

Ayitey Junior, M., Appiahene, P., & Appiah, O. (2022). Forex market forecasting with two-layer stacked Long Short-Term Memory neural network (LSTM) and correlation analysis. *Journal of Electrical Systems and Information Technology*, *9*(1), 14. http://dx.doi.org/10.1186/s43067-022-00054-1.

Baasher, A. A., & Fakhr, M. W. (2011). Forex trend classification using machine learning techniques. *1*, In *Proceedings of the 11th WSEAS international conference on applied computer science* (pp. 41–47). World Scientific and Engineering Academy and Society (WSEAS) Stevens Point ..., Issue: 1.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

Baruník, J., Kočenda, E., & Vácha, L. (2017). Asymmetric volatility connectedness on the forex market. *Journal of International Money and Finance*, *77*, 39–56, Publisher: Elsevier.

Chen, L., Li, X., & Xie, Z. (2023). Two-stage attentional temporal convolution and LSTM model for financial data forecasting. *12700*, In *International conference on electronic information engineering and data processing* (pp. 122–130). SPIE.

Cheng, X. (2023). An attention embedded DUAL-LSTM method for financial risk early warning of the three new board-listed companies. *PeerJ Computer Science*, *9*, Article e1271, Publisher: PeerJ Inc..

Cho, K., Van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1724–1734).

Dash, S., Sahu, P. K., & Mishra, D. (2023). Forex market directional trends forecasting with Bidirectional-LSTM and enhanced DeepSense network using all member-based optimizer. *Intelligent Decision Technologies*, (Preprint), 1–32, Publisher: IOS Press.

Electronic Trading Hub (2024). Understanding trading latencies. URL: https://electronictradinghub.com/. (Accessed 19 January 2024).

Fazeli, A., & Houghten, S. (2019). Deep learning for the prediction of stock market trends. In *2019 IEEE international conference on big data (big data)* (pp. 5513–5521). IEEE.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669, Publisher: Elsevier.

Forex (2024). *Foreign exchange committee releases FX volume survey results*: *Technical Report*, Federal Reserve Bank of New York, URL: https://www.newyorkfed.org/medialibrary/Microsites/fxc/files/2024/Volume-Survey-Results_07222024. (Accessed 10 February 2025).

Forex Training Group (2025). What kind of returns can be expected from FX trading? Forex Training Group. URL: https://forextraininggroup.com/what-kind-of-returns-can-be-expected-from-fx-trading/. (Accessed 11 February 2025).

Galeshchuk, S., & Mukherjee, S. (2017). Deep networks for predicting direction of change in foreign exchange rates. *Intelligent Systems in Accounting, Finance and Management*, *24*(4), 100–110, Publisher: Wiley Online Library.

Ghahremani, S., Yau, D. K., Albrethsen, J., Sidhu, R., & Cheung, N.-M. (2021). Time delay attack detection using recurrent variational autoencoder and K-means clustering. In *2021 IEEE PES innovative smart grid technologies-Asia* (pp. 1–5). IEEE.

Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, *4*(1), 9, Publisher: MDPI.

Kantoutsis, K. T., Mavrogianni, A. N., & Theodorakatos, N. P. (2024). Transformers in high-frequency trading. *Journal of Physics*, *2701*(1), Article 012134. http://dx.doi.org/10.1088/1742-6596/2701/1/012134.

Li, Z. (2023). Research on LSTM multi-factor quantitative stock selection strategy based on attention mechanism. *12566*, In *Fifth international conference on computer information science and artificial intelligence* (pp. 844–852). SPIE.

Li, Y., Zhu, Z., Kong, D., Han, H., & Zhao, Y. (2019). EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowledge-Based Systems*.

Lou, J., Cui, L., & Li, Y. (2023). Bi-LSTM price prediction based on attention mechanism. arXiv. arXiv:2212.03443 [cs, q-fin].

Ngan, T. M. U. (2013). Forecasting foreign exchange rate by using ARIMA model: A case of VND/USD exchange rate. *Methodology*, *2014*, 2015.

Olanrewaju, M. O., Luka, S., & Echobu, F. O. (2023). Prediction of forex prices on USD/NGN using deep learning (LSTM and GRU) techniques. *Journal of Information Systems and Informatics*, *5*(4), 1609–1622.

Sidehabi, S. W., & Tandungan, S. (2016). Statistical and machine learning approach in forex prediction based on empirical data. In *2016 international conference on computational intelligence and cybernetics* (pp. 63–68). IEEE.

Sirignano, J., & Cont, R. (2019). Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance*, *19*(9), 1449–1459, Publisher: Taylor & Francis.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, *27*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*: *vol. 30*.

Xelera Technologies (2023). Low-latency machine learning inference for high-frequency trading. URL: https://www.xelera.io. (Accessed 19 January 2025).

Yıldırım, D. C., Toroslu, I. H., & Fiore, U. (2021). Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. *Financial Innovation*, *7*, 1–36, Publisher: Springer.