Applying Two-Layer Stacked Long Short Term Memory Neural Network on AUD/USD Currency Pair

Aaron Pagaygay[1], Lara Gabrielle Lim [1,*] and Angelyn Lao[1]
[1] *Department of Mathematics and Statistics, De La Salle University*
**Corresponding Author: lara_gabrielle_lim@dlsu.edu.ph*

Abstract: Given that foreign exchange (Forex) is one of the most significant markets allowing for international trading and transactions, various methods have been explored to predict the Forex trends more accurately. In this paper, we build on Forex market forecasting with Two-Layer Stacked Long Short-Term Memory neural network (TLS-LSTM) and correlation analysis. The paper applies the TLS-LSTM to the AUD/USD currency pair which was chosen by computing if its time series data is persistent through the Hurst exponent. Further, correlation analysis is conducted between AUD/USD and other currencies to show positive and negative correlations which could also be a basis for buying and selling the currencies.
Key Words: Forex, Time series, Stacked LSTM neural network, Hurst exponent, Correlation analysis

1. INTRODUCTION

The foreign exchange (Forex) market is a form of exchange for the global trading of international currencies. As has been true for decades, the markets remain decentralized with high liquidity and continuous trading (Vyklyuk, Vuković, & Jovanović, 2013). As Forex markets affect output and employment (through real exchange rates), inflation (through the cost of imports and commodity prices), and international capital flows (through the risks and returns of different assets), this shows how significant its importance is to the world economy (Rime, 2003). Large investors such as central banks and investment firms are in the market to manage their portfolios and prevent exchange risks, while individuals (or traders) mostly benefit from short-term currency rate changes (Ayitey, Appiahene, & Appiah, 2022). Currently, over 95% of all daily Forex transactions involve only eight currencies, belonging to economic areas with stable governments, respected central banks, and relatively low inflation. (Petropoulos, Chatzis, Siakoulis, & Vlachogiannakis, 2017).

With effective time series forecasting in the Forex market, investors and traders can participate in trend trading, which is a method of attempting to capture profits by analyzing the trend of a currency pair in a specific direction. When the trend is heading upward, traders can take a *buy* position and if it is heading downward, can take a *sell* position. A *buy* position means that a trader buys the currency expected to rise so that if its value rises, profits are gained from the market when the trader closes this position by selling at a higher value. Conversely, a *sell* position means that a trader sells the currency expected to fall so that if its value falls, profits are gained from the market when the trader closes this position by buying at a lower value. With trend trading, it takes time for a trader to realize a profit. As a result, it is important and beneficial for traders to anticipate the direction of a currency pair's movement as well as its correlation with other pairs in the market (Ayitey, Appiahene, & Appiah, 2022). Once traders are informed on the predicted direction of a currency pair, traders will also be able to take buy or sell positions on other strongly correlated currency pairs without having to go through these processes. This allows traders better avenue and options (such as take action, hedge, diversity or double position advantage) to maximize their profits in the Forex market (Ramadhani, Jondri, & Rismala, 2016).

Since correctly predicting a currency's behavior is key to the gains of investors and traders, it is crucial to anticipate the trend. For this purpose, the two-layer stacked Long Term Short Memory (TLS-LSTM) neural network is selected and examined for its performance to forecast currency trends. The Long Short-Term Memory (LSTM) was invented by Hochreiter and Schmidhuber in 1997. This particular type of recurrent neural network has been developed to effectively handle sequential data and address the issue of vanishing gradients commonly seen in traditional neural network structures. LSTM has become state-of-the-art in many fields and is still being developed further to improve certain aspects (Dautel, Härdle, Lessmann, & Seow, 2020).

Before subjecting any Forex market data to the TLS-LSTM, the Hurst exponent is used to determine whether the currency data is trending or not. In time series forecasting, it is key that a time series is predictable since all methods are expected to fail if it is random (Qian & Rasheed, 2004). The statistical method in which Harold Edwin Hurst invented, is called the Hurst exponent. It provides concrete information on correlation and persistence which makes it an excellent index for studying complex processes such as the financial time series (Raimundo & Okamoto Jr., 2018). For this reason, the AUD/USD currency pair was selected because its Hurst exponent suggests that it is a trending time series.

In this study, we follow the methodology of the paper by (Ayitey, Appiahene, & Appiah, 2022) and the goal is to anticipate the trend rather than estimate the exchange rate price; that is, the aim is to be able to forecast the direction of AUD/USD based on daily measured rates in price and volume. With this, the TLS-LSTM neural network is thoroughly examined to see how it works, how it can be improved for currency trading forecasting, and how to use the model to anticipate the trend to make long-term gains. Comparisons on efficiency are also made against other artificial neural networks like Multilayer Perceptron (MLP) and single-layer LSTM (Vanilla LSTM or SL-LSTM). Given our aim to reproduce and improve on the results made in (Ayitey, Appiahene, & Appiah, 2022), we adopted most of the parameters indicated in their study. These parameters are used in the TLS-LSTM, as well as the baseline models MLP and SL-LSTM.

This paper is organized as follows: Section 2 focuses on the discussion of the methods of the study. In Section 3, we present a detailed discussion of the results of the time series forecasting using TLS-LSTM. Lastly, in Section 4, we have the conclusion of the study.

2. METHODOLOGY

2.1 Framework

In this section, we discuss the conceptual framework of the study. We follow the methods of (Ayitey, Appianhene, & Appiah, 2022). First, we ensure that the dataset chosen is trending using the Hurst exponent. We then establish the efficiency of a two-layer stacked Long Short-Term Memory Neural Network to forecast the trend of the chosen data, where the obtained predicted trend can assist a trader whether to take buy or sell positions in the FOREX market. Also, correlation analysis between other currency pairs is done so that traders can consider buying or selling other currencies based on the obtained trend prediction. The proposed conceptual framework is as shown in Figure 1.

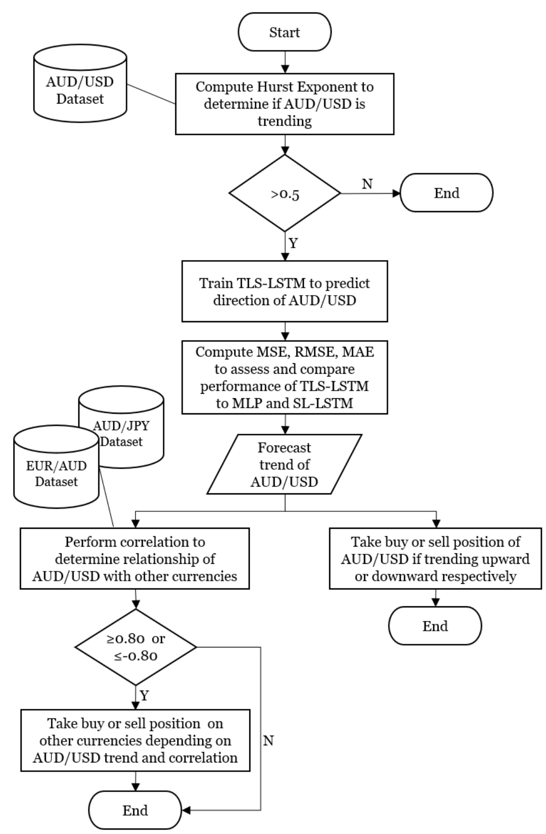

Figure 1. Proposed Conceptual Framework on Applying TLS-LSTM to AUD/USD Dataset

This adapts and intends to improve the traditional machine learning problem-solving approach to the Forex market trend prediction. The framework aims to satisfy implementation needs, and to analyze the relevance, topic matter, and context of the problems recognized. During their extensive literature review, one major gap identified is that many researchers arbitrarily chose a dataset for prediction without clear reason. Another gap is that the vanilla LSTM has been extensively used in Forex forecasting but there was a need to increase its performance, expected to be achieved by adding another layer to the single layer. Lastly, it is realized that movement of currency pairs affect one another, hence the need to see how changes in trend of one currency affects others. Addressing these gaps builds toward the goal of effectively predicting the currency trends so that traders earn profit from the market.

To be able to test and run machine learning models, modern machine learning platforms with strong processing capabilities are utilized by the study in (Ayitey, Appiahene, & Appiah, 2022). In consideration of the research goal and objectives, an online platform that connects to supercomputing platforms is chosen for the study since these platforms have strong processing capabilities. They have recognized that extensive machine learning, data science libraries and significant online computing power is preferred since machine learning implementations are usually computationally intensive. For our own analysis, we used a free offline platform due to limited access on online platforms. However, it is noted that running machine learning models are slower.

2.2 Hurst Exponent

The Hurst exponent provides a measure for long-term memory and fractality of a time series. It ranges between 0 and 1 and can be calculated by rescaled range (R/S analysis). The Hurst exponent H classifies a time series into three categories: a random series if $H = 0.5$, an anti-persistent series if $0 < H < 0.5$, and a persistent series $0.5 < H < 1$. An anti- persistent series has a characteristic of *mean-reverting*, which means an up value is more likely followed by a down value, and vice versa. A persistent series is *trend reinforcing*, which means the direction (up or down compared to the last value) of the next value is more likely the same as the current value. The strength of mean-reverting increases as H approaches 0 and the strength of trend reinforcing increases as H approaches 1 (Qian & Rasheed, 2004). The Hurst exponent was computed using the hurst and pandas package in python.

2.3 Long Short-Term Memory Neural Network

An *artificial neural network (ANN)*, or simply *neural network*, is a data processing model based on the way biological nervous systems process data. It consists of a pool of simple processing units which communicate by sending signals to each other over a large number of weighted connections (Jurafsky & Martin, 2023). The basis of a neural network is a single computational unit. It takes a set of real numbers as input, performs some computation on the set, and produces an output. At its core, a neural unit takes a weighted sum of its inputs, with an additional term in the sum called a bias. Given a set of inputs $x_1, \dots, x_n$, a unit has a set of corresponding weights $w_1, \dots, w_n$ and a bias b, so the weighted sum z is given by

$$z = b + \sum_i w_i x_i.$$

If a *feedforward neural network (FFNN)*, a multilayer neural network in which the units are connected with no feedback connections, is extended to include feedback connections, it is called a *recurrent neural network (RNN)*. However, one of the disadvantages of early RNN architectures was their limited memory capacity, caused by the vanishing or exploding gradient problem that has become evident when the information contained in past inputs must be retrieved after a long time interval. To address these concerns, different solutions have been proposed.

One of these solutions is by utilizing the *Long Short Term Memory (LSTM) neural network*, which is a recurrent neural network with the capability for both long and short term memory. The LSTM was specifically designed to deal with vanishing gradients and allow the network to learn much longer-range dependencies. It has three mechanisms that control the memory: input gate, forget gate and output gate, which are responsible for updating the cell state c_k at every time step k where k is an integer such that $0 \leq k \leq t - 1$. The structure of an LSTM cell unit is shown in Figure 2.

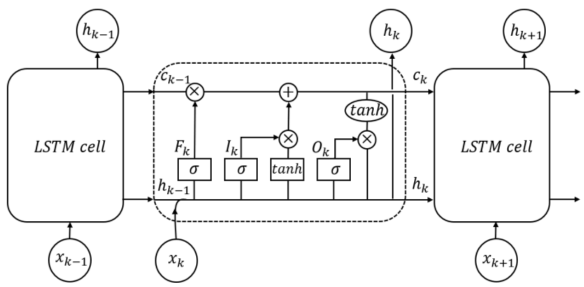

Figure 2. The structure of an LSTM cell
Note. Adapted from Lee et al. (2019).

In Figure 2, x_k and h_k are denoted by the *input* and the *output* of the memory cell, respectively, at time k. The *forget gate* F_k determines how much of the memory in the previous cell c_{k-1} is put into the memory of the cell c_k, meanwhile, the *input gate* I_k determines how much new information will enter the memory. On the other hand, the *output gate* O_k that determines how much updated information is to output for the next LSTM unit in the network. Thus, the LSTM unit cell function is defined as follows:

$$h_k = O_k \bullet tanh\left(c_k\right)$$

such that

$$c_k = F_k \bullet c_{k-1} + I_k \bullet tanh\left(W_k^{\ c} x_k + W_h^{\ c} h_{k-1} + b_c\right)$$
$$F_k = \sigma\left(W_x^{\ F} x_k + W_h^{\ F} h_{k-1} + b_F\right)$$
$$I_k = \sigma\left(W_x^{\ I} x_k + W_h^{\ I} h_{k-1} + b_I\right)$$
$$O_k = \sigma\left(W_x^{\ O} x_k + W_h^{\ O} h_{k-1} + b_O\right)$$

where $W_k^{\ c} x_k$, $W_h^{\ c} h_{k-1}$, $W_x^{\ F} x_k$, $W_h^{\ F} h_{k-1}$, $W_x^{\ I} x_k$, $W_h^{\ I} h_{k-1}$, $W_x^{\ O} x_k$, $W_h^{\ O} h_{k-1}$ represents the weight matrices, b_c, b_F, b_I, b_O represents the bias vector, and the sigmoid function σ is given by $\sigma(z) = \left(1 + e^{-z}\right)^{-1}$. And, the prediction x_k is determined by h_{k-1} via

$$x_k = W_S h_{k-1} + b_s$$

where W_S is the weight and b_s is the bias. Here, h_{k-1} is dependent on $x_1, \dots, x_{k-1}$ in the input layer and $h_1, \dots, h_{k-2}$ output layer for each LSTM cell.

The more common LSTM only has a single layer but the focus of this study is to see the performance of a Stacked LSTM neural network which is a variation of a single hidden Layer LSTM model that contains numerous buried LSTM layers, each having many memory cells. In particular, we will be using a Two-Layer Stacked LSTM (TLS-LSTM). The architecture of the proposed TLS-LSTM neural network is given in Figure 3. The TLS-LSTM model was implemented using the TensorFlow package in python.

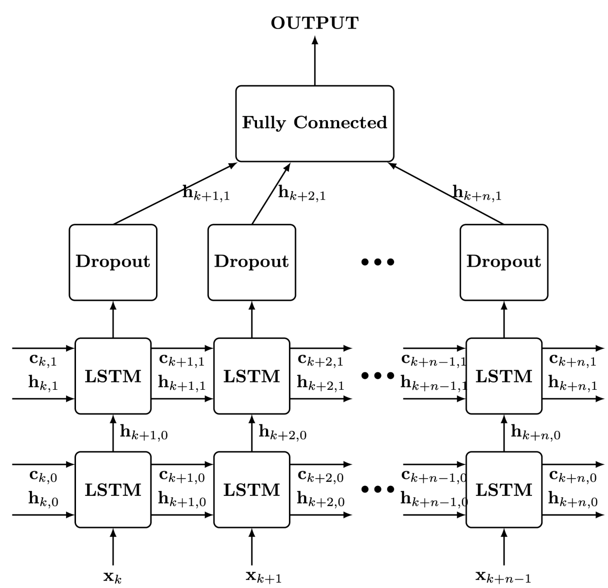

Figure 3. The Architecture of a Two-Layer Stacked LSTM Neural Network with Dropout

2.4 Dataset Features

We have obtained the data on AUD/USD currency by request from the authors of the paper. The dataset has 1772 records on the daily exchange rate with a period spanning from May 1, 2013 to March 2, 2020. This was collected from FXpro (broker) through the Metatrader 4 trading platform. The data features used for the TLS-LSTM include 'Date,' 'Open,' 'Close,' 'High,' 'Low,' and 'Volume.' *Date* is the time the price happened in the market, that is, a 24-hour period for this study. *Open* is the price of the first trade for the period, or the starting contract price for the time. *Close* is the price of the last trade during the period, or the price of the most recent trade made within the period. *High* is the highest price that the asset traded during the period. *Low* is the asset's lowest price during the period. *Volume* is the total quantity of assets traded within the specified period. The relationship between pricing and volume is critical, for example. Price increases are accompanied by increases in volume.

2.5 Dataset Preprocessing

To analyze if the dataset is viable for forecasting, the Hurst exponent estimate is applied on the target feature which is the 'Closing Price'. This was implemented using the Hurst module in python. In Ayitey et al. (2022), the dataset is split as follows: 75% for training and 25% for testing. Furthermore, to have the same degree of values in the dataset, the Standard Scaler is used to scale all values to have zero mean and unit variance. Both of these preprocessing procedures were inherited in this study.

2.6 Dataset Processing

The study focuses on the comparison of the performance of the models MLP, SL-LSTM and TLS-LSTM with their respective parameters.

The *hidden units per layer* used on any of the models is 128. To introduce nonlinearity, we used the *activation function ReLU* for each layer, that is, for the weighted sum z of an ANN, $ReLU(z) = max(z, 0)$. The *time step* or number of data in a sequence to be fed on an LSTM cell is set to 10. The *optimizer*, which is a strategy to reduce losses by modifying the parameters of a neural network, is *Adam*. We adopted 0.001 as the *learning rate* of Adam, as this is a good default setting (Kingma & Ba 2015). *Dropout* is included for regularization to prevent overfitting after the two LSTM layers and is set to 0.2. The *batch size*, which is the total number of training examples in a single batch, is set to 20. The *epoch*, representing the frequency of introducing the training dataset to the neural network, is set to 32. The v*alidation split*, which is the percentage of data in the training set used to evaluate the model during training, is 0.1. The *loss function* which measures the discrepancy between the actual and predicted values of the time series is the Mean Squared Error (MSE).

The *evaluation metrics* used to measure performance of the models MLP, SL-LSTM and TLS-LSTM are Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

Since we intend to essentially replicate and find improvement on the results of (Ayitey, Appiahene, & Appiah, 2022), most of the parameters were adopted from the paper. The baseline models MLP and SL-LSTM are also given the parameters of the TLS-LSTM.

2.7 Pearson's Correlation Coefficient

To know if it is advised to trade on other currency pairs based on the trend of AUD/USD, we perform correlation analysis. We obtain the closing price data of AUD/JPY, EUR/AUD, AUD/CAD and AUD/NZD to match the 1772 records of AUD/USD by date from www.finance.yahoo.com and www. exchangerates.org.uk.

We use the *Pearson correlation coefficient* ρ which is a measure of the strength and nature of any existing linear relationship between two variables X and Y, defined by

$$\rho = \frac{\sigma_{XY}}{\sigma_X^2 \sigma_Y^2},$$

where σ_{XY} denotes the covariance of X and Y, and σ_X^2 and σ_Y^2 the variance of X and Y, respectively. The coefficient value ranges from -1 to $+1$. If $0 < \rho \leq 1$, the correlation is positive and the closer it is to 1, the stronger the positive correlation. If $-1 \leq \rho < 0$, the correlation is negative and the closer it is to -1, the stronger the negative correlation. There is a perfect positive correlation if $+1$, negative correlation if -1 and no correlation if ρ is 0. For this study, we computed the correlation using scikit-learn and pandas package in python. Moreover, as indicated in (Ayitey, Appiahene, & Appiah, 2022) to possibly make good profit on trading with currency pairs, the target value of correlation value is at least 0.80 or less than -0.80 which indicates strong positive and negative correlations, respectively.

3. RESULTS AND DISCUSSION

3.1 Hurst Exponent

We applied the Hurst exponent estimate on the AUD/USD dataset with 1772 records from May 1, 2013 to March 2, 2020.

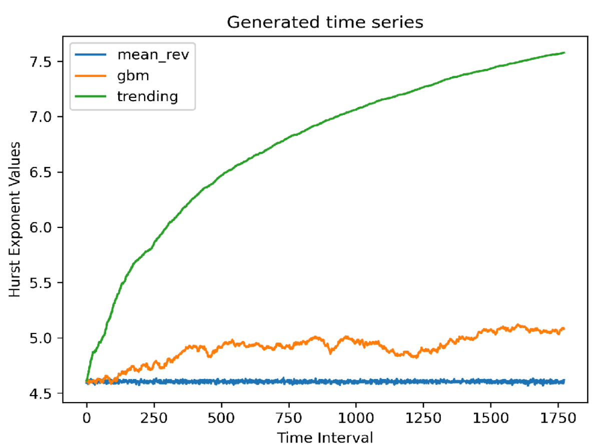

Figure 4. A graph of generated time series for Hurst exponent with 1772 records of AUD/USD

In Figure 4, the slope indicates the Hurst exponent value. The target feature used was the Closing price and the Hurst exponent obtained for the dataset was 0.5827. This falls in $0.5 < H < 1$ which means the time series is persistent; that is, it is trending and good for forecasting. In comparison, the Hurst exponent estimate in (Ayitey, Appiahene, & Appiah, 2022) is 0.6026 which falls under the same classification.

Knowing the predictability of the dataset, Figure 5 shows the connection between the AUD/USD currency and its trend line determined by the Closing price, that is, the AUD/USD currency and its trend line have a high positive correlation. Thus, an AUD/USD trader may decide based on the behavior of the trend.

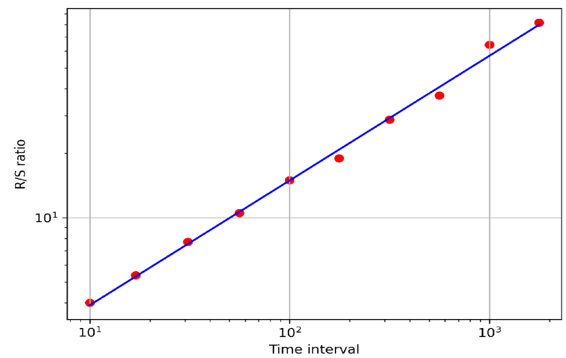

Figure 5. Scatter plot of AUD/USD currency connection with 1-day intervals and its trend line

3.2 Two-Layer Stacked LSTM

Using the parameters given in Section 2.4, we have found the behavior of the actual trend against the predicted trend. The proposed two-layer stacked LSTM is compared with the baseline models, single-layer MLP and single-layer LSTM, based on the evaluation metrics. However, some details in (Ayitey, Appiahene, & Appiah, 2022) are not indicated. These are the learning rate and the time steps used for the TLS-LSTM, and the parameters of the baseline models used in the evaluation metrics. Thus, certain adjustments have been made. That is, we set the learning rate and time steps for the TLS-LSTM into 0.001 and 10, respectively, and performed the baseline models using the same parameters with TLS-LSTM. Furthermore, instead of the MinMax Scaler, we used the Standard Scaler.

From the values of the evaluation metrics (MSE, RMSE, and MAE) in Table 1, it shows that the TLS-LSTM model outperforms both MLP and SL-LSTM. In comparison, the obtained evaluation metric of TLS-LSTM is lower than of TLS-LSTM in (Ayitey, Appiahene, & Appiah, 2022). However, we do note that in their model comparisons, the date range used is April

1, 2013 to December 30, 2020. This is a difference of around 10 months of data.

Table 1. Proposed TLS-LSTM comparison with baseline models based on evaluation metrics

Model	MSE	RMSE	MAE
MLP	0.000320819	0.017198858	0.017911411
SL-LSTM	0.000014135	0.002901306	0.003759686
TLS-LSTM	**0.000013843**	**0.002886714**	**0.003720649**

In Figure 6, we obtain the plot for the actual values versus the predicted values of AUD/USD Closing price on the test data. It is seen here that the model follows the trend of the closing price quite accurately, reflecting the very low MSE of the model.

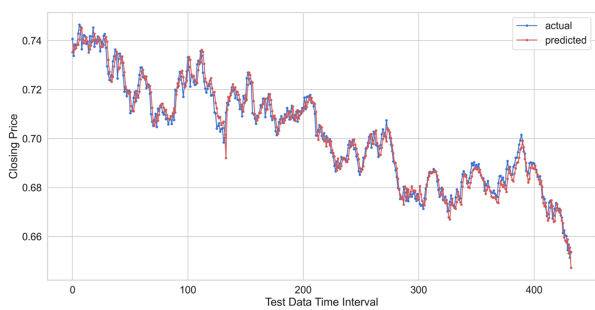

Figure 6. AUD/USD Closing Price – Graph of Actual Values vs Predicted Values

If we obtain the plot of future expected rates and trends of AUD/USD from the model, it is important to note the peaks (local maximums) and valleys (local minimums) of the plotted line, which are the ideal times for traders to take sell and buy positions on AUD/USD, respectively. On peaks where the trend is expected to pull downward in the next days, a trader can sell USD so that they can utilize its highest value with the prospect that once the USD price becomes low, the trader can buy it back for a lower price. Conversely, a trader can take a buy position when AUD/USD is at a valley, then sell the USD at a higher price on later dates. However, it is at the trader's discretion and preference when they will buy and sell, depending on the amount of gain the trader wants to receive, considering the forecasted trend and perceived confidence level on the model.

3.3 Correlation analysis

Table 2 shows the results of Pearson's correlation analysis of AUD/USD with the other currencies using the Closing price as target feature. We note that it is only recommended to trade if the correlation value is greater than or equal to 0.80 and is less than or equal to -0.80. With this, it is shown that it is only recommended to trade AUD/JPY based on its correlation to AUD/USD.

Table 2. Pearson's correlation coefficient on AUD/USD to other Currencies Closing Prices

	AUD/USD
AUD/JPY	0.800018973
EUR/AUD	-0.638947426
AUD/CAD	0.535342051
AUD/NZD	0.579551291

4. CONCLUSIONS

One of the most substantial markets for international trading is foreign exchange (FOREX). For a trader to successfully predict the Forex trends, a variety of techniques have been investigated. A notable method is through neural networks. The AUD/USD currency pair, which has 1772 records on the daily exchange rate for the period from May 1, 2013, to March 2, 2020, was the main subject of this study. As it turns out, the Hurst exponent of the dataset AUD/USD falls under the classification of a persistent time series which indicates that dataset was trending and good for forecasting. This dataset was forecasted using the two-layer stacked LSTM model, and it was shown that this model outperforms Multilayer Perceptron (MLP) and single-layer LSTM (SL-LSTM) through various evaluation metrics. This shows that the TLS-LSTM has achieved increased performance to forecast the currency pair trends, so that there is more confidence in using the forecasted result to take buy or sell positions in the market. Further, a correlation analysis was done between the dataset of AUD/USD with AUD/JPY, EUR/AUD, AUD/CAD and AUD/NZD, and we see that it is reasonable to take buy or sell positions in AUD/JPY in consideration of AUD/USD trends. In the future, we wish to explore more possible improvements on the forecasting model and how it behaves with different parameters and data, such as a longer date range and other data features.

5. ACKNOWLEDGMENTS

We acknowledge the authors M. Ayitey, P. Appiahene and O. Appiah of the paper *Forex market forecasting with two-layer stacked Long Short-Term Memory neural network (LSTM) and correlation analysis* for providing the dataset.

A.P. would like to thank the Department of Science and Technology - Science Education Institute (DOST-SEI) Accelerated Science and Technology Human Resource Development Program-National Science Consortium (ASTHRDP-NSC) for their support. The funding agency did not have any additional role in the study design, data collection, analysis, decision to publish, or preparation of manuscript.

6. REFERENCES

Ayitey, 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*.

Dautel, A. J., Härdle, W. K., Lessmann, S., & Seow, H.-V. (2020). Forex exchange rate forecasting using deep recurrent neural networks. *Digital Finance 2(1–2)*, 69-96.

Jurafsky, D. & Martin, J. H. (2023). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Third Edition Draft.

Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Kingma, D., Ba, J (2015). Adam: a method for stochastic optimization. In 3rd international conference on learning representations, ICLR 2015,San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings

Lee, C.-I. , Chang, C.-H. & Hwang, F.-N. (2019). Currency Exchange Rate Prediction with Long Short-Term Memory Networks Based on Attention and News Sentiment Analysis, 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), pp. 1-6.

Petropoulos, A., Chatzis, S. P., Siakoulis, V., & Vlachogiannakis, N. (2017). A stacked generalization system for automated FOREX portfolio trading. *Expert Systems with Applications Vol. 90*, 290-302.

Raimundo, M. S., & Okamoto Jr., J. (2018). Application of Hurst Exponent (H) and the R/S Analysis in the Classification of FOREX. *International Journal of Modeling and Optimization, Vol. 8, No. 2.*

Ramadhani, I., Jondri, & Rismala, R. (2016). Prediction of multi currency exchange rates using correlation analysis and backpropagation. *2016 International Conference on ICT For Smart Society (ICISS).*

Rime, D. (2003). New Electronic Trading Systems in the Foreign Exchange Markets. In *New Economy Handbook* (pp. 471-504). Academic Press.

Qian, Bo & Rasheed, Khaled. (2004). Hurst exponent and financial market predictability. Proceedings of the Second IASTED International Conference on Financial Engineering and Applications.

Vyklyuk, Y., Vuković, D., & Jovanović, A. (2013). Forex predicton with neural network: usd/eur currency pair. *Actual problem of economics, Issue 10 (148)*, 251-261.