Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 8 – No. 1, December 2025 – www.caeaccess.org

A Ridge-Regularized Deep Learning Stacking Ensemble for Multivariate Daily Exchange Rate Forecasting in Emerging Markets: Evidence from the Ghana Forex Market

Ernest Kwame Ampomah	Ezekiel Mensah Martey	Emmanuel Asante
University of Mines and Technology (UMaT)	University of Mines and Technology (UMaT)	University of Mines and Technology (UMaT)
Tarkwa - Ghana	Tarkwa - Ghana	Tarkwa - Ghana

ABSTRACT
Forecasting exchange rate remains a major challenge due to nonlinear dynamics, structural volatility, and complex cross-currency interactions. Although deep learning models have demonstrated strong predictive capability, however, the individual architectures often specialize in limited temporal patterns and may overfit volatile financial series. This study proposes a hybrid multivariate stacking ensemble that integrates Long Short-Term Memory (LSTM), Temporal Convolutional Network (TCN), and Multi-Task Learning (MTL) models with a ridge-regression meta-learner to enhance predictive accuracy and stability. Daily EUR/GHS, GBP/GHS, and USD/GHS exchange rates from January 2010 to September 2025 were modeled using a sliding-window, multi-input, multi-output configuration. Performance was benchmarked against standalone deep learning models and a Vector Auto-Regression (VAR) baseline. Results show that the proposed hybrid multivariate stacking ensemble (TCN+LSTM+MTL) model achieves the lowest mean RMSE (0.7488), MAE (0.5285), MAPE (3.75 percent), and SMAPE (3.64 percent), representing approximately 89.7 percent and 92.7 percent RMSE reduction compared to VAR and LSTM, respectively. The findings confirm that combining specialized deep architectures with regularized meta-learning significantly improves forecasting accuracy and robustness in volatile financial markets and offers insights for future research in cross-architecture fusion and meta-learning for financial econometrics.

Keywords
Exchange Rate Forecasting, Multivariate Modeling, Stacking Ensemble, Long Short-Term Memory, Temporal Convolutional Network, Multi-Task Learning, Ridge Regression, Meta-Learning, Deep Learning, Financial Time Series.

1. INTRODUCTION
Financial markets are very important in the development of the global economy. The forex (FX) market is an integral component of the financial market. As international trades continue to exist, exchange rates remain a crucial link between national economies [1]. They influence macroeconomic fundamentals, stability of capital flows and international transactions. Hence, forecasting exchange rate dynamics and trends is vital for investors and multinational firms to make informed financial decisions, optimize portfolio strategies, and manage exchange rate risks effectively [2]. Forecasting of exchange rate has attracted significant interest in academic research and economic analysis [3]. Many scholars have carried out research studies on exchange rate behavior [4-6]. Exchange rate time series studies is a major aspect of time-series research. This involves the assessment and forecast of fluctuation ranges and trend changes in the exchange rate dynamics.

Various techniques have been employed to model exchange rates. Fundamental analysis relies on economic theory to identify variables that influence exchange rates in the long term, such as trade imbalances, interest rate differentials, and overall economic conditions of countries [7]. Technical analysis focuses on recognizing patterns in historic data to forecast future movements, often disregarding underlying economic fundamentals [8]. Econometrics and time series techniques such as vector autoregressive (VAR), autoregressive integrated moving averages (ARIMAs) and generalized autoregressive conditional heteroscedastic (GARCH) models, analyze historical data to capture temporal dependencies and volatility structures in exchange rate dynamics [9]. However, these conventional economic approaches are limited in their ability to capture the nonlinearities, structural shifts, and long range dependencies characteristics associated with exchange rate time series [10]. These data are highly nonlinear, nonstationary, and volatile, making accurate forecast of exchange rate fluctuations extremely complicated. Exchange rate fluctuations are influenced by both domestic and international economic conditions, domestic and international economic conditions, global market sentiment, political developments, and other external influences. Additionally, factors such as market participants' psychological expectations, geopolitical tensions, and financial crises often have significant impact on currency valuations, introducing volatility and complexity into exchange rate data [11]. The limitations of traditional economic models to forecast exchange rates have prompted the need for alternative approaches in modeling exchange rate dynamics [12]. Recent advancements in machine learning (ML) and deep learning (DL) techniques are proving to be promising in improving forecasting accuracy [13]. ML and DL models are gaining widespread adoption for their ability to analyze large datasets and capture complex patterns to enhance exchange rate forecasting [14]. DL models use several neurons and hidden layers to analyze sequential data, enabling them to generate better forecasts than economic models [15]. They provide computational power and functional flexibility necessary to detect latent patterns in complex datasets [16]. Aydin &

Cavdar [17] showed in their study that Artificial Neural Networks (ANNs) provided predictive performance that is superior to VAR models in exchange rate forecasting. Dautel et al. [18] systematically compared LSTM networks and Gated Recurrent Units (GRUs) with traditional econometric models like VAR. The study concluded that the deep learning models outperformed traditional models in exchange rate forecasting. García et al. [19] compared the predictive performance of ARIMA and LSTM on foreign exchange forecasting tasks. The outcome of the study showed that LSTM offers superior predictive accuracy. Zhao et al. [20] compared the performance of ARIMA, LSTM, and Gated Recurrent Unit (GRU) models in exchange rates forecasting. The findings from the study indicated that the LSTM model delivers better forecasting accuracy than ARIMA, which underscores the potential of DL models in financial time series forecasting.

Although individual DL models have achieved considerable success in financial forecasting, the effectiveness of individual DL models remains limited as they are susceptible to issues such as overfitting and the inability to fully exploit complementary information across diverse architectures. When individual DL models are trained on highly volatile financial time series data like exchange rates, they can learn both meaningful patterns and random noise resulting in poor generalization, and hence poor performance on unseen data [21-22]. Standalone DL model restricts learning process to a narrow feature domain. For instance recurrent models like LSTM gives priority to long-term temporal dependencies, convolutional models such as Temporal Convolutional Network (TCN) put emphasize on short-term local patterns, and Multi-Task Learning (MTL) networks leverage useful information contained across multiple related tasks [23-24]. This specializations of DL architectures implies that individual model may not be able to capture diverse and interacting dependencies found in multivariate financial data holistically. Consequently, lack of cross-architectural integration limits ability of individual models to leverage complementary information that might improve robustness and adaptability across varying market conditions reducing the potential to achieve generalizable financial predictions [25].

In recent years, the use of hybrid models that combine ML and DL techniques has increased significantly in FX market prediction. Hybrid models have demonstrated improved performance. Lin et al. [26] proposed a hybrid model that integrated complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and multilayer long short-term memory (MLSTM) networks and the findings showed that the hybrid model effectively captured complex correlations in exchange rate data, and improved performance. Islam and Hossain [27] integrated GRU and LSTM neural network and found the hybrid model to be superior to simple moving average, LSTM, and GRU models. He et al. [28] proposed an ensemble model that integrated ARMA, CNN, and LSTM to capture both linear and nonlinear features in financial time series. The results showed that the ensemble model performed better than the individual models, including LSTM and CNN. Sina et al. [29] performed a systematic review that analyzed twenty one (21) studies. The study concluded that hybrid forecasting methods perform better individual models. Gu et al. [30] created a hybrid model that integrated LSTM and GRU within an AdaBoost framework. The results from the study indicated that the ensemble model produced an improved performance in forecasting exchange-rate compared to standalone models.

This study builds on prior research by developing and evaluating hybrid multivariate ensemble stacking framework designed to augment forecasting of daily exchange rate. The proposed framework integrates forecasts from deep learning models including Temporal Convolutional Network (TCN), Long Short-Term Memory (LSTM), and Multi-Task Learning using ridge regression as a meta-learner. While each of these DL architecture has shown a strong capabilities across diverse time series forecasting tasks, prior research has rarely explored the effectiveness of integrating their complementary inductive biases which are causal dilated convolutions, sequential memory, and shared multi-task representations within a unified multivariate stacking framework designed for exchange rate forecasting. Hence, there is a limited empirical evidence regarding which heterogeneous model combinations yield the most substantial forecasting improvements. Also, existing research provides limited insight into how a simple yet robust technique like ridge regression can serve as an effective meta-learner for aggregating outputs from complex DL models while ensuring generalization, computational efficiency, and interpretability. To address these gaps the current study systematically integrating TCN, LSTM, and MTL models within a ridge regression based stacking ensemble to achieve an improved predictive performance in exchange rate forecasting.

2. METHOD

2.1 Dataset Description

This study conducted an empirical analysis using daily closing prices of three (3) major currency pairs - EUR/GHC, GBP/GHC, and USD/GHC obtained from the Bank of Ghana database (https://www.bog.gov.gh). Each currency pair data range from 4th January 2010 to 17th September 2025, comprising 3894 entries. The dataset was partitioned sequentially into training, validation, and test sets in a 70:15:15 ratio.

2.2 Data Standardization

The z-score technique illustrated in (1) was applied to scale the data to ensure zero mean and unit variance before model training. The transformation was subsequently inverted to recover values on the original scale for model evaluation.

$$\grave{x} = \frac{x-\mu}{\sigma} \tag{1}$$

μ is mean of the training data

σ denote standard deviation of the training data.

2.3 Sliding Window

The sliding window approach with a lookback window L and forecast horizon H was used to convert the exchange rate time series data into a supervised MIMO (multi-input multi-output) learning tasked. For each time step t, an input window of length L was used to predict the next H steps as formulated in (2) and (3).

$$X_t = [x_{t-L}, x_{t-L+1}, \dots, x_{t-1}] \tag{2}$$

$$Y_t = [x_t, x_{t+1}, \dots, x_{t+H-1}] \tag{3}$$

$X_t \in \mathbb{R}^M$ and M = number of currency pairs = 3.

L =30, and H = 1 was used in this study. The resulting supervised dataset is defined in (4):

$$D = \{(X_t, Y_t)\}_{t=L}^{T-H} \tag{4}$$

2.4 Baseline Model

Vector auto-regression (VAR) model was implemented as the baseline statistical forecasting framework. It is a multivariate

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 8 – No. 1, December 2025 – www.caeaccess.org

time series model designed to capture dynamic linear interdependencies across various endogenous variables. It was introduced by Sims [31]. VAR model extends univariate autoregressive (AR) framework to systems of equations. It models each variable as a linear function of its own past values and the past values of all other variables in the system as shown in (5) and (6). VAR is very popular and widely used in macroeconomics, finance, and forecasting. It enables flexible, data-driven representation of dynamic interactions without any strong a priori structural assumptions imposition. Given M-dimensional time series.

$$X_t = [x_{1t}, x_{2t}, ... , x_{Mt}]' \qquad (5)$$

VAR model is expressed as:

$$X_t = c + \sum_{i=1}^{p} A_i X_{t-1} + \varepsilon_t \qquad (6)$$

Where c represents $M \times 1$ vector of intercept terms, $A_i (i = 1, \ldots, p)$ denotes $M \times M$ coefficient matrices characterizing linear lag-i interdependences among endogenous variables, p is optimal lag order determined based on Akaike Information Criterion (AIC), and ε_t denotes white-noise error term.

2.5 Deep Learning Models

Three DL architectures comprising Long Short-Term Memory (LSTM), Temporal Convolutional Network (TCN), and Multi-Task Learning (MTL) network were developed to forecast exchange rate. These architectures are designed to capture non-linear temporal dependencies and cross-series interactions in the exchange rate market. The models were trained on standardized input sequences generated by the sliding window transformation of historical exchange rate data. Model parameters were optimized using the Adam optimizer to minimize the mean squared error (MSE) loss function. The models were trained to minimize MSE with L2 regularization. The resulting loss function as shown by (7) penalizes both large prediction errors and excessively large parameter magnitudes to ensure good generalization.

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{t=1}^{N} \|y_t - \hat{y}_t\|_2^2 + \lambda \sum_{w \epsilon \Theta} \|w\|_2^2 \qquad (7)$$

2.5.1 Long Short-Term Memory (LSTM)

LSTM architecture is an extension of conventional recurrent neural networks (RNNs) that incorporate memory cells and gating mechanisms to alleviate vanishing and exploding gradient problems to enable long-range sequence modeling. Each LSTM cell maintains two internal states, namely a cell state and a hidden state. Together, the two internal states control long and short-term memory propagation across time [32]. For a time step t, given the input vector, x_t, the input gate, forget gate, and output gate are computed using (8)–(10). The candidate cell state is generated with (11). Then the cell state and hidden state are updated using (12) and (13), respectively:

$$i_t = \sigma(W_i x_t + U_i h_t + b_i) \qquad (8)$$
$$f_t = \sigma\left(W_f x_t + U_f h_t + b_f\right) \qquad (9)$$
$$o_t = \sigma(W_o x_t + U_o h_t + b_o) \qquad (10)$$
$$\tilde{c}_t = tanh(W_c x_t + U_c h_t + b_c) \qquad (11)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \qquad (12)$$
$$h_t = o_t \odot tanh(c_t) \qquad (13)$$

Where:

- i_t, f_t, and o_t are input, forget, and output gates respectively.
- $\tilde{c}_t$ is candidate memory,
- c_t is cell state update, and h_t represents hidden states update,
- σ(·) denotes sigmoid activation function and tanh(·) hyperbolic tangent function,
- W_* and U_* are the input and recurrent weight matrices, and b_* is the corresponding biases.
- $\odot$ represents element-wise multiplication

The implemented LSTM model architecture, as shown in Fig. 1, used a deep sequential architecture. It is composed of two stacked LSTM layers, followed by batch normalization, dropout, and fully connected dense mappings. The input tensor has a shape of (L, M), where L is the number of lookback time steps and M=3 is the number of exchange rate pairs. The architecture consists of two LSTM layers, each followed by a batch normalization block to stabilize hidden state activations and improve training efficiency by reducing internal covariate shift. The first LSTM layer consists of 128 hidden units with an output sequence $H^{(1)} \in R^{L \times 128}$. The second LSTM Layer has 64 hidden units, generating a final temporal embedding $h^{(1)} \in R^{64}$. A dropout layer with a rate of 0.25 was employed to reduce overfitting by randomly deactivating neurons during training. A dense layer with 128 neurons and ReLU activation serves as a fully connected layer, converting the temporal embedding into a nonlinear feature space. The final output layer is a linear dense mapping with M units with a forecast as indicated in (14). The overall network defines a nonlinear mapping function as shown in (15).

$$\hat{y}_{t+1} = W_0^{(f)} h^{(2)} + b_0^{(f)} \qquad (14)$$

$W_0^{(f)}$ and $b_0^{(f)}$ are learnable weight matrix and bias vector respectively.

$$\hat{y}_{t+1} = f_{LSTM}(X_t, \Theta) \qquad (15)$$

Where Θ represents all learnable weights and biases of the network

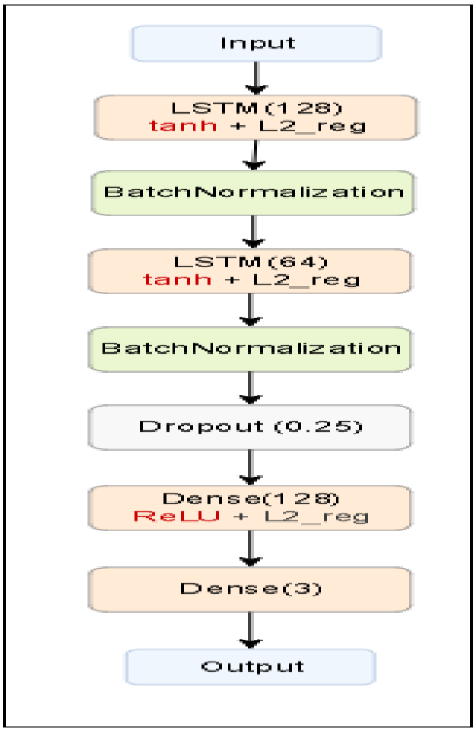

Fig 1: LSTM Architecture

2.5.2 Temporal Convolutional Network (TCN)

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 8 – No. 1, December 2025 – www.caeaccess.org

TCN is a convolutional architecture designed specifically for sequential and time-series modeling. They effectively capture long-range temporal dependencies while ensuring parallelized sequence processing by using dilated convolutional filters with controlled receptive fields [33]. Each TCN layer applies 1D causal convolution to ensure that predictions at time t only depend on inputs up to t. For a given layer l, the convolution output is shown in (16).

$$h_t^{(l)} = \phi\left(\sum_{k=0}^{K-1} W_k^{(l)}\, h_{t-k}^{(l-1)} + b^{(l)}\right) \qquad (16)$$

Where:

- K is the kernel size (set to 3),
- $W_k^{(l)}$ are the convolutional filters,
- $\phi(\cdot)$ is the ReLU activation,
- $h_t^{(0)} = x_t$

The TCN architecture implemented as shown in Fig. 2 consists of processed sequential inputs of shape (L, M) followed by a Conv1D layer with a kernel size of 3 and 64 filter with a ReLU activation and an L2 weight regularizer to ensure smooth training and avoid overfitting. Batch normalization is then applied to stabilize feature distributions and accelerate convergence. This is followed by a second Conv1D layer with 64 filters, a kernel size of 3, and ReLU activation, and a second Batch normalization to boost training stability. A dropout layer with a rate of 0.25 is then applied to mitigate overfitting. A Global Average Pooling (h_{GAP}) as indicated in (17) is employed to aggregate temporal information by computing the mean across time steps. The aggregated information is passed to a dense layer comprised of 128 neurons, ReLU activation, and L2 regularization. The architecture has a final output layer with a linear dense consisting of M units, which generates the multi-output forecasting as shown in (18).

$$h_{GAP} = \frac{1}{L}\sum_{t=1}^{L} h_t^{(L)} \qquad (17)$$

$$\hat{y}_{t+1} = W_0^{(f)} h_{GAP} + b_0^{(f)} \qquad (18)$$

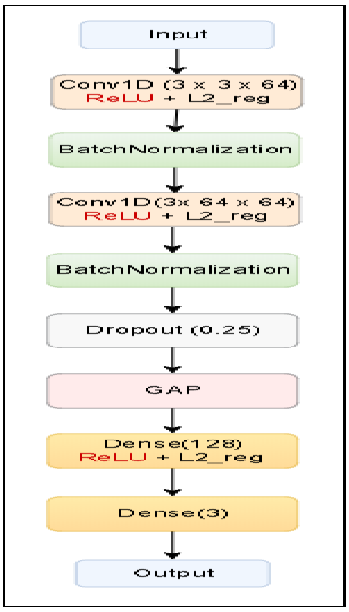

Fig. 2: TCN architecture

2.5.3 Multi-Task Learning (MTL) Network

MTL is a DL paradigm that is designed to enable training of a single model to perform multiple related tasks simultaneously. Instead of training independent models for each tasks, it exploits shared representations to achieve efficient learning and generalization. It learns a common latent representation across all series and incorporates distinct task-specific output heads which facilitates effective information sharing and preserving task-level specialization [34].

The model takes the exchange rate series as input and outputs the one-day-ahead exchange rates. The MTL Architecture is shown in Fig 3. The architecture consists of a Shared Dense layer (h_s) as shown in (19) and it comprised of 256 units with ReLU activation and L2 regularization. The shared dense layer is followed by batch normalization and dropout, followed by three separate dense layers with 128 units for task-specific branches: for each currency pair (m=1, 2, and 3), the model used a dedicated task-specific branch as shown in (20). The three scalar outputs are then combined to form the multi-output prediction vector ($\hat{y}_{t+1}$) as indicated in (21).

$$h_s = ReLU(W_s z_t + b_s) \qquad (19)$$

z_t denotes the input feature vector, W_s and b_s are the corresponding weight matrix and bias term, and h_s represents the shared latent representation.

$$h_m = ReLU(W_m h_s + b_m) \qquad (20)$$

$$\hat{y}_{t+1}^{(m)} = w_m^T + b'_m \qquad (21)$$

Where h_m is the task-specific hidden features, and $\hat{y}_{t+1}^{(m)}$ is the predicted value for task m.

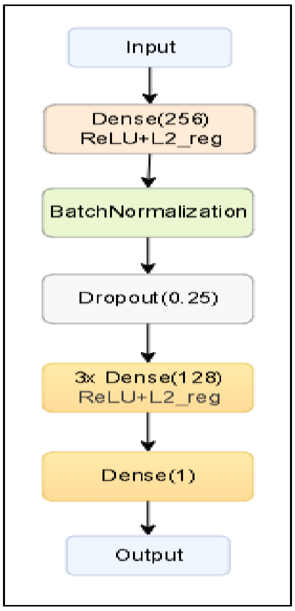

Fig. 3: MTL Architecture

2.6 Stacking Framework

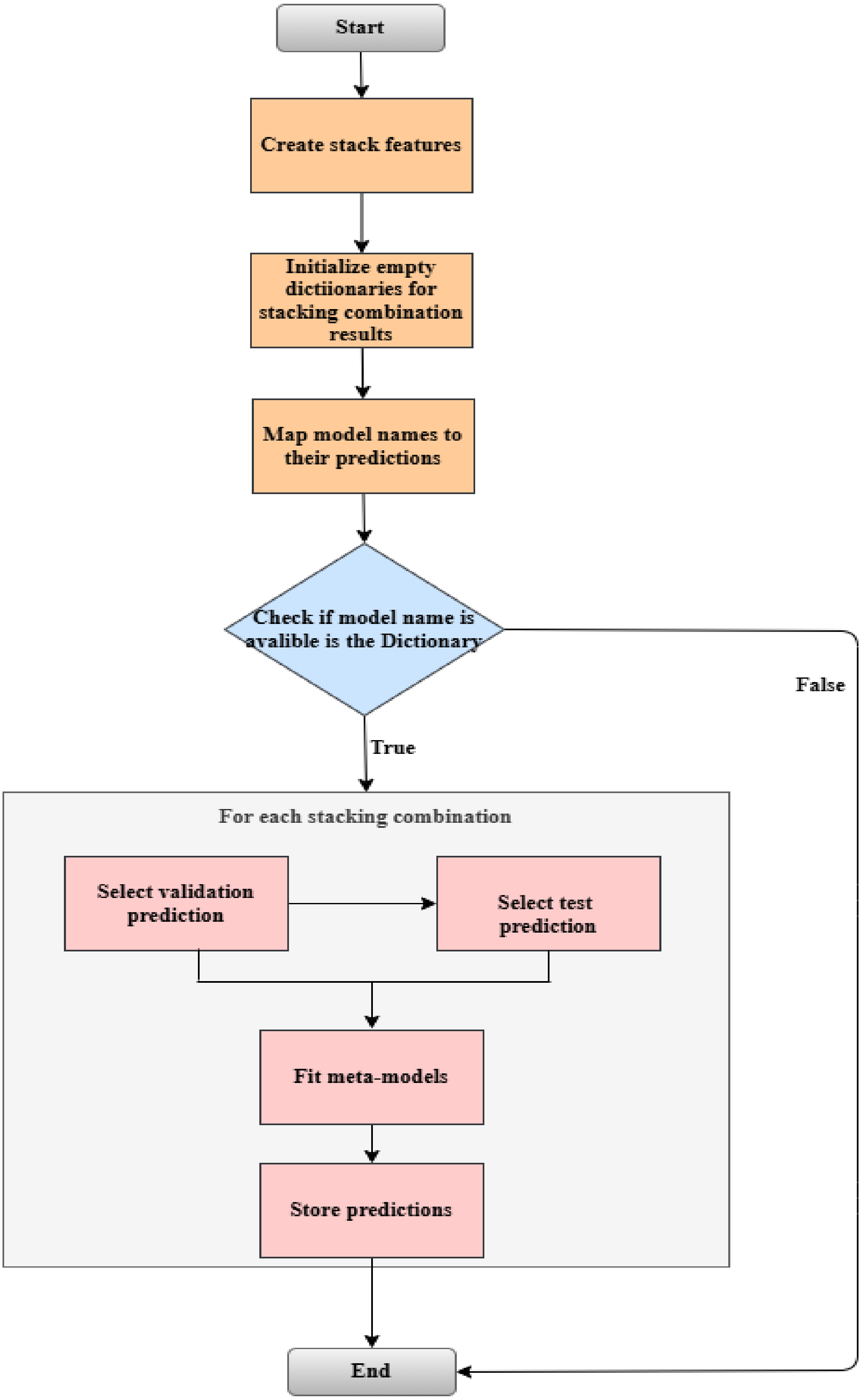

Fig. 4: Stacking approach

A stacking framework is an ensemble learning architecture that combines multiple predictive models to achieve better performance than any single model. In a stacking framework, various base learners are first trained independently on the same dataset. Each of the base learner generate its own predictions, which are treated as meta-features representing

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 8 – No. 1, December 2025 – www.caeaccess.org

different perspectives on the data. The meta-features are used as inputs to a meta-learner, which learns to optimally combine the outputs of the base learners. The meta-learner identifies systematic patterns in the predictive behavior of each base model across instances, and learns to assign appropriate weights or transformations to their predictions to generate a more accurate final prediction [35]. The stacking framework in this work is shown in Fig 4.

Letting $Y_t \in \mathbb{R}^M$ denote the target vector of M currency pairs at time t. Then for each model $k \in \{1, 2, 3\}$ which corresponds to the LSTM, TCN, and MTL networks, one-step ahead forecast on the validation set as shown in (22).

$$\hat{Y}_t^{(k)} = f_k(X_t; \theta_k) \quad (22)$$

$f_k(\cdot)$ denotes the nonlinear mapping learned by the kth deep network with parameters θ_k and X_t is the multivariate lagged input vector constructed via a sliding window mechanism. The base models' predictions on the validation dataset were concatenated column-wise to form the stacking feature matrix, which served as the input to the meta-learner as defined in (23).

$$Z_t = \left[\hat{Y}_t^{(1)}, \hat{Y}_t^{(2)}, \hat{Y}_t^{(3)}\right] \in \mathbb{R}^{M\times 3} \quad (23)$$

Ridge regression model was used as meta-learner to learn a linear combination of the predictions of the base model that minimized the MSE on the validation data. Ridge regression was selected as the meta-learner because of its favorable bias - variance trade off and numerical stability when combining predictions from correlated deep learning architectures.

For each currency pair $i \in \{1, \dots, M\}$ the meta-learner estimated coefficients $\beta_i = [\beta_{i1}, \beta_{i2}, \beta_{i3}]^T$ by solving (24)

$$\min_{\beta_i} \sum_{t=1}^{T_{val}} \left(y_{t,i} - Z_t\beta_i\right)^2 + \alpha\|\beta_i\|_2^2 \quad (24)$$

$y_{t,i}$ is the observed value of currency pair i at time t, T_{val} is the number of validation samples,

α = 1.0 is the ridge regularization coefficient that controls model complexity. The trained ridge regression models were subsequently applied to the test dataset to generate final ensemble predictions as expressed in (25):

$$\hat{y}_{t,1}^{(stack)} = Z_t\hat{\beta}_i,\ t \in \text{Test set} \quad (25)$$

Multiple stacking configurations were constructed and evaluated, including TCN+LSTM, TCN+MTL, LSTM+MTL, and TCN+LSTM+MTL.

2.7 Model Evaluation

The performance of the models were evaluated on the test set using multiple evaluation metrics:

Root Mean Squared Error (RMSE)

RMSE quantifies the square root of the average squared differences between predicted and actual values. It assigns bigger weights to larger errors. It is defined by (26).

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t - y_t)^2} \quad (26)$$

Mean Absolute Error (MAE)

MAE determines the average magnitude of prediction errors without consideration of their direction as expressed by (27). It provides a direct interpretation of absolute deviations:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|\hat{y}_t - y_t| \quad (27)$$

Mean Absolute Percentage Error (MAPE)

MAPE expresses prediction accuracy in percentage terms through normalization of absolute errors with respect to the actual values: It is expressed by (28).

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \quad (28)$$

Symmetric Mean Absolute Percentage Error (SMAPE)

SMAPE provides a scale-independent alternative to MAPE. It symmetrically normalize absolute errors with respect to both predicted and actual values as defined by Equation 29.

$$SMAPE = \frac{100}{N}\sum_{t=1}^{N}\frac{|\hat{y}_t - y_t|}{(|y_t| + |\hat{y}_t|)/2} \quad (29)$$

3. RESULTS AND DISCUSSION

This section presents a comprehensive evaluation of the proposed ridge-regularized deep learning stacking ensemble for multivariate daily exchange rate forecasting. Model performance is assessed across three major currency pairsUSD/GHS, EUR/GHS, and GBP/GHS using four complementary evaluation metrics: RMSE, MAE, MAPE, and SMAPE. Comparative analysis is conducted against a traditional econometric baseline (VAR), standalone deep learning models (LSTM, TCN, and MTL), and multiple stacking configurations to demonstrate the robustness and effectiveness of the proposed framework.

3.1 Overall Model Performance

Table 1 reports the detailed forecasting performance of all models across individual currency pairs, while Table 2 summarizes the mean performance across all three exchange rates. Figures 5–8 illustrate the mean RMSE, MAE, MAPE, and SMAPE values, respectively, enabling visual comparison of model ranking consistency across evaluation metrics.

Across all metrics and currency pairs, the proposed Stack_TCN+LSTM+MTL ensemble consistently achieves the lowest forecasting errors. Specifically, it records the smallest mean RMSE (0.7488), MAE (0.5285), MAPE (3.75%), and SMAPE (3.64%), demonstrating superior predictive accuracy and stability relative to both standalone models and alternative ensemble configurations. This consistent dominance across scale-dependent and scale-independent metrics confirms the robustness of the proposed stacking framework.

In contrast, the traditional VAR model exhibits substantially higher error levels, reflecting its limited capacity to model nonlinear dynamics and regime shifts that characterize foreign exchange markets in emerging economies. Similarly, standalone deep learning models, while outperforming VAR in most cases, show inferior performance compared to ensemble-based approaches, highlighting the limitations of relying on a single architectural bias.

3.2 Performance across Individual Currency Pairs

USD/GHS Exchange Rate

For the USD/GHS series, the Stack_TCN+LSTM+MTL ensemble achieves an RMSE of 0.8915, significantly outperforming VAR (6.59) and standalone LSTM (9.21). Percentage-based error metrics further reinforce this advantage, with SMAPE reduced to 4.78% compared to over 60% for VAR and more than 100% for LSTM. This demonstrates the ensemble's ability to track sharp movements and volatility clustering commonly observed in the USD/GHS market.

EUR/GHS Exchange Rate

Similar performance trends are observed for the EUR/GHS exchange rate. The proposed ensemble achieves an RMSE of 0.5687 and MAPE of 2.77%, representing substantial improvements over both statistical and standalone deep learning baselines. The consistently low error values indicate strong generalization and reduced sensitivity to abrupt market fluctuations.

GBP/GHS Exchange Rate

For GBP/GHS, which exhibits relatively higher volatility and irregular fluctuations, the ensemble maintains robust performance with an RMSE of 0.7863 and SMAPE of 3.41%. Standalone models, particularly LSTM and MTL, struggle to adapt to these dynamics, recording significantly higher errors. This highlights the advantage of combining heterogeneous models that capture both short-term and long-term dependencies.

3.3 Comparison of Standalone and Ensemble Models

Among the standalone models, TCN emerges as the strongest performer, outperforming LSTM and MTL across all metrics.
This suggests that temporal convolutional structures are particularly effective in capturing short-term dependencies and local temporal patterns in exchange rate data. However, despite its relative strength, TCN alone remains inferior to all stacking configurations, indicating that convolutional representations benefit substantially from integration with recurrent and multi-task learning architectures.

The LSTM model records the poorest performance among deep learning approaches, with the highest mean RMSE (10.19) and SMAPE (101.93%). This behavior suggests overfitting and limited adaptability to rapidly changing market conditions, particularly when trained in isolation on highly volatile financial time series.

The MTL model demonstrates mixed performance, performing competitively for USD/GHS but deteriorating for EUR/GHS and GBP/GHS. While shared representation learning offers some generalization benefits, it is insufficient on its own to handle complex cross-series interactions without complementary temporal modeling.

Table 1. Evaluation results of each model across all currency pairs

Model	Currency Pair	RMSE	MAE	MAPE	SMAPE
VAR	USD/GHS	6.5921	6.3101	47.3500	62.6878
VAR	EUR/GHS	6.8154	6.5389	45.0837	58.7443
VAR	GBP/GHS	8.3674	8.0005	46.8195	61.7433
TCN	USD/GHS	2.8697	2.7310	20.6311	23.1667
TCN	EUR/GHS	4.2518	4.0555	27.9057	32.6461
TCN	GBP/GHS	6.4401	6.1460	35.9446	44.1611
LSTM	USD/GHS	9.2105	8.9899	68.2903	104.1396
LSTM	EUR/GHS	9.8288	9.6184	67.0919	101.3317
LSTM	GBP/GHS	11.5298	11.2471	66.6234	100.3155
MTL	USD/GHS	2.3179	2.2299	17.0247	18.6476
MTL	EUR/GHS	6.3914	6.2717	43.8697	56.3003
MTL	GBP/GHS	9.2325	9.0099	53.3915	73.0258
Stack_TCN+LSTM	USD/GHS	1.4927	1.1028	9.1393	8.4037
Stack_TCN+LSTM	EUR/GHS	1.0307	0.7066	5.2653	5.0399
Stack_TCN+LSTM	GBP/GHS	1.2214	0.8556	5.2930	5.1383
Stack_TCN+MTL	USD/GHS	1.2106	0.9415	7.3416	6.8989
Stack_TCN+MTL	EUR/GHS	0.5632	0.3726	2.6804	2.6332
Stack_TCN+MTL	GBP/GHS	0.8043	0.5989	3.5301	3.5169
Stack_MTL+LSTM	USD/GHS	1.1321	0.9120	6.4916	6.7583
Stack_MTL+LSTM	EUR/GHS	1.0112	0.9078	6.1636	6.3752
Stack_MTL+LSTM	GBP/GHS	1.8064	1.5952	9.0508	9.5523
Stack_TCN+LSTM+MTL	USD/GHS	0.8915	0.6173	5.0518	4.7807
Stack_TCN+LSTM+MTL	EUR/GHS	0.5687	0.3884	2.7690	2.7293
Stack_TCN+LSTM+MTL	GBP/GHS	0.7863	0.5798	3.4283	3.4103

Table 2. Mean metric measure across all currency pairs for each model

Model	RMSE	MAE	MAPE	SMAPE
Stack_TCN+MTL+LSTM	0.7488	0.5285	3.7497	3.6401
Stack_TCN +MTL	0.8594	0.6377	4.5174	4.3496
Stack_TCN+LSTM	1.2482	0.8883	6.5659	6.1940
Stack_MTL+LSTM	1.3166	1.1384	7.2353	7.5619
TCN	4.5205	4.3108	28.1604	33.3247
MTL	5.9806	5.8372	38.0953	49.3246
VAR	7.2583	6.9499	46.4177	61.0585
LSTM	10.1897	9.9518	67.3352	101.9289

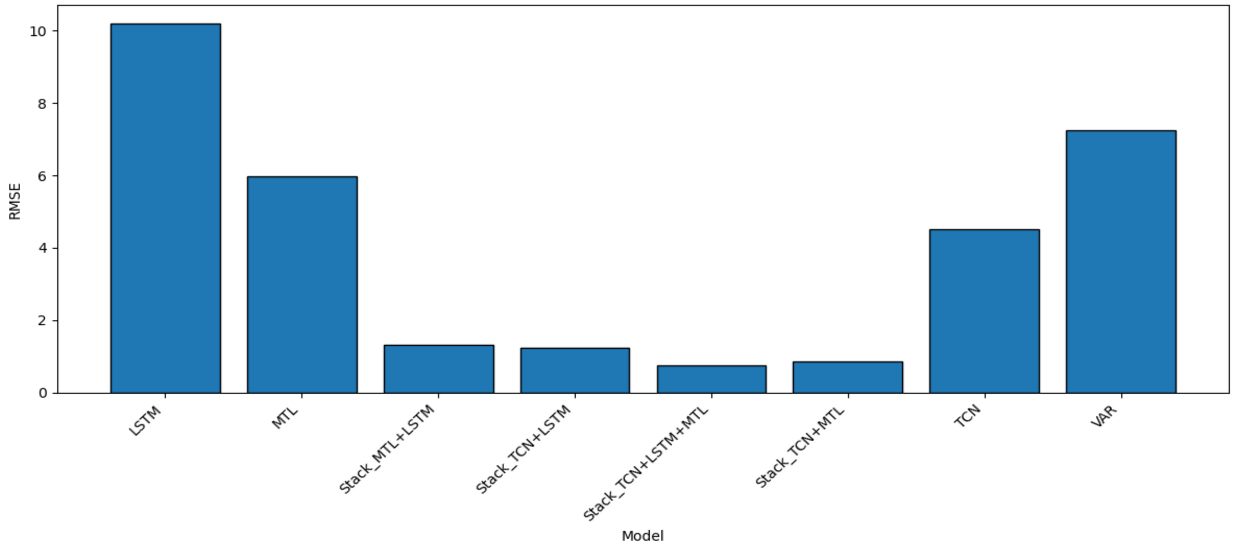

Fig 5: Mean RMSE across all currency pairs for each model

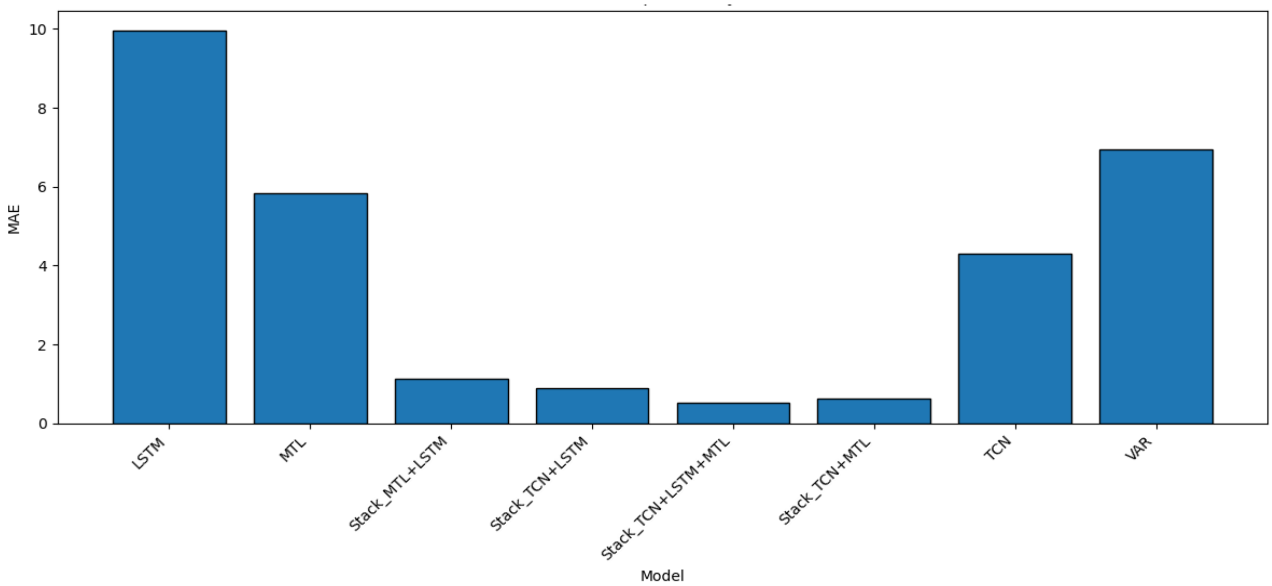

Fig 6: Mean MAE across all currency pairs for each model

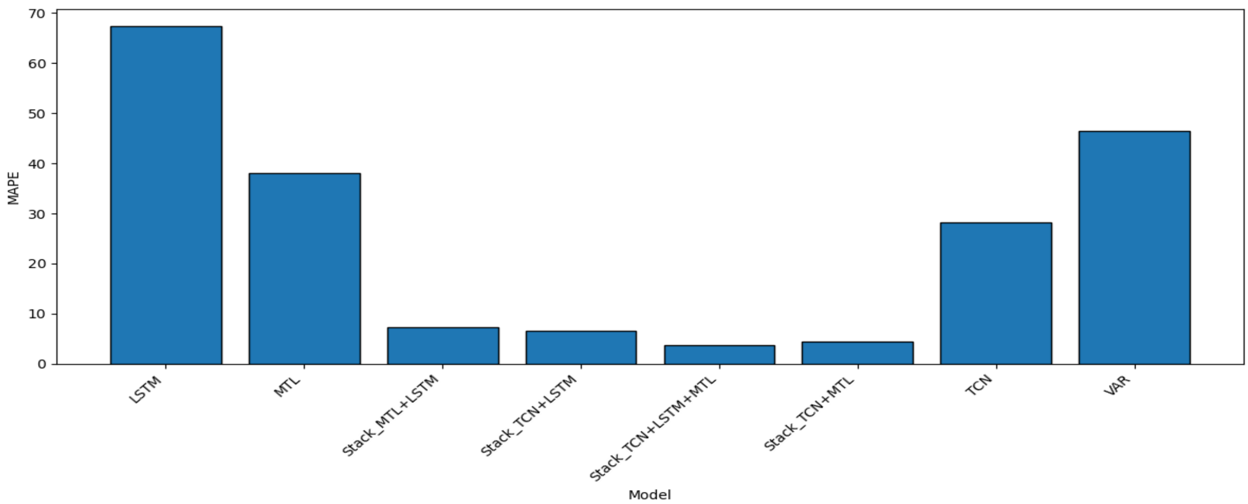

Fig 7: Mean MAPE across all currency pairs for each model

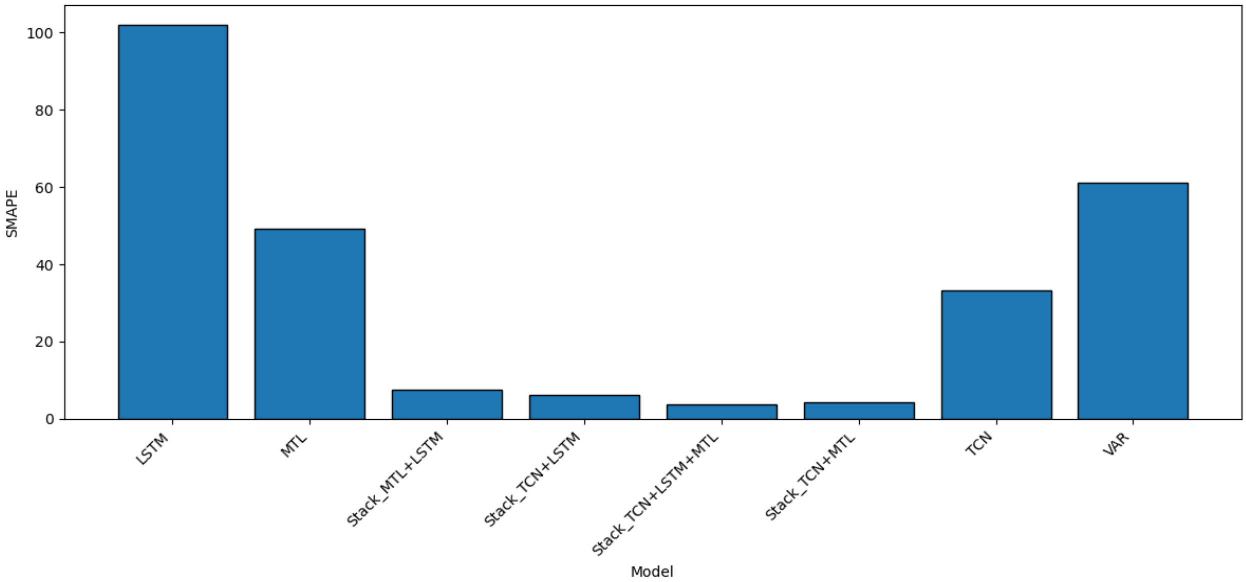

Fig 8: Mean SMAPE across all currency pairs for each model

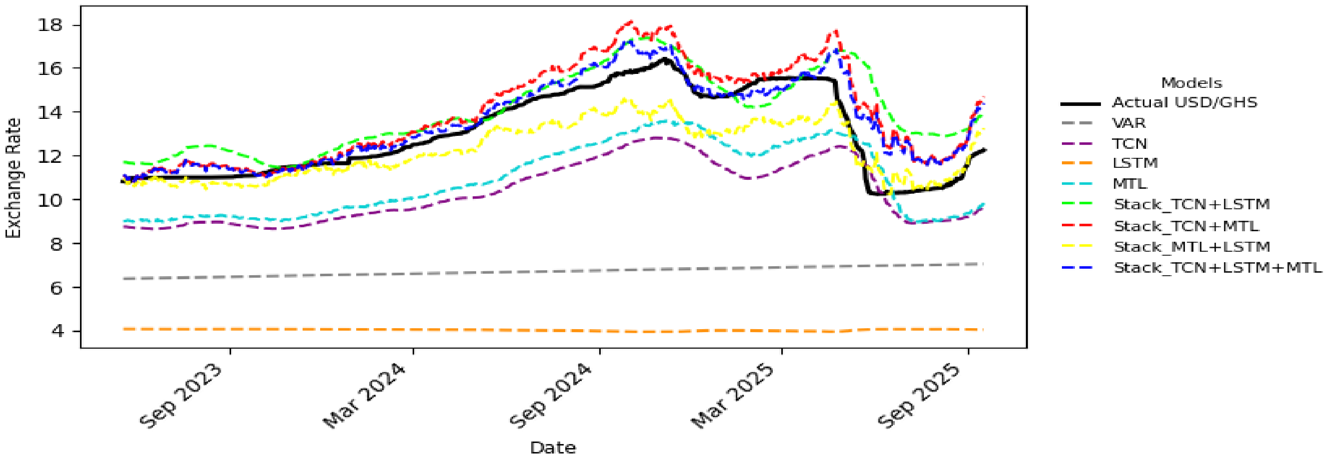

Fig 9: USD/GHS one-day ahead forecast by all models

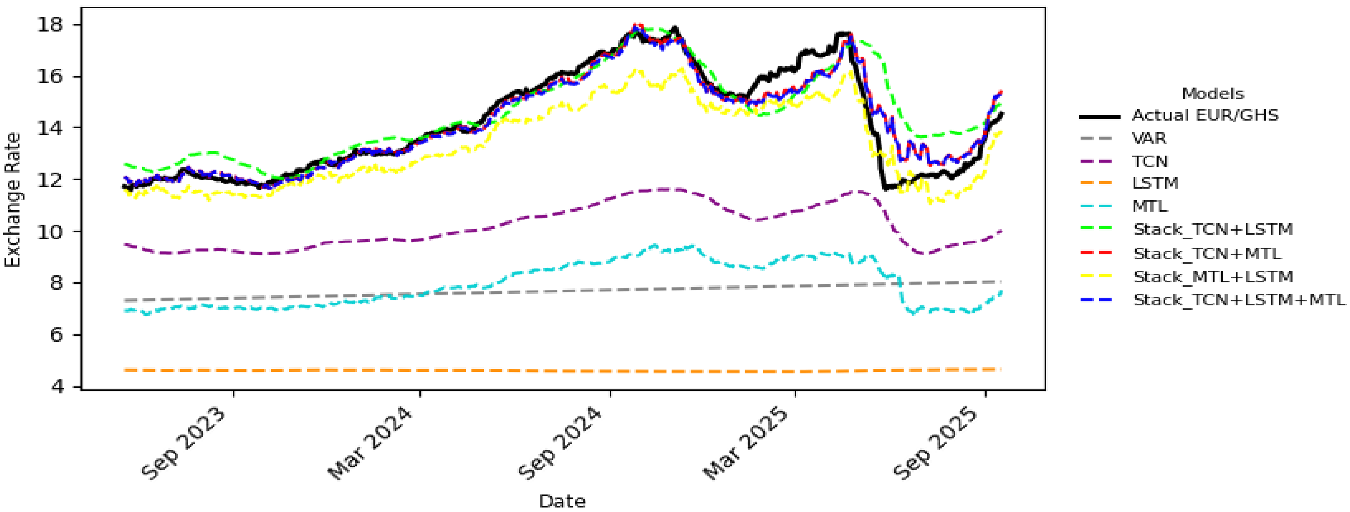

Fig 10: EUR/GHS one-day ahead forecast by all models

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 8 – No. 1, December 2025 – www.caeaccess.org

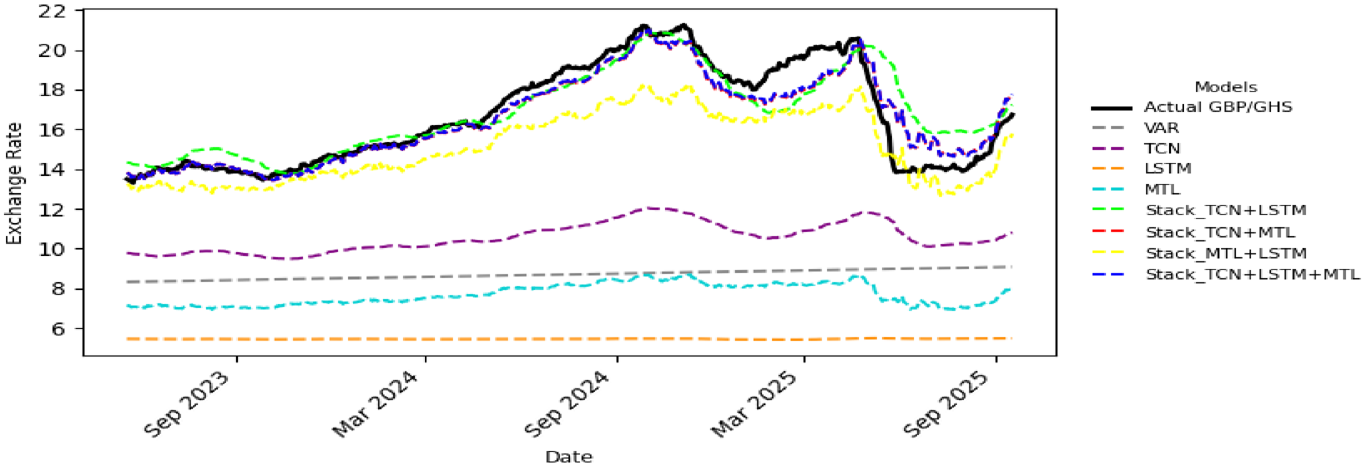

Fig 11: GBP/GHS one-day ahead forecast by all models

3.4 Effectiveness of the Stacking Framework

The stacking ensembles consistently outperform their constituent base learners, confirming the effectiveness of the ensemble learning strategy. Among the evaluated configurations, Stack_TCN+MTL and Stack_TCN+LSTM rank second and third, respectively, while Stack_MTL+LSTM shows comparatively weaker performance due to the absence of convolutional temporal feature extraction.

The superior performance of Stack_TCN+LSTM+MTL can be attributed to the complementary inductive biases of its base learners. Specifically, TCN captures short-term and local temporal dependencies through causal dilated convolutions, LSTM models long-term sequential dependencies via gated memory mechanisms, and MTL enhances generalization by leveraging shared information across related currency pairs. The ridge regression meta-learner further stabilizes predictions by optimally weighting base model outputs while mitigating overfitting through L2 regularization.

Compared to VAR and standalone LSTM, the proposed ensemble achieves approximately 89.7% and 92.7% reductions in RMSE, respectively, underscoring its substantial predictive advantage.

3.5 Forecast Trajectory Analysis

Figures 9–11 present the one-day-ahead forecast trajectories for USD/GHS, EUR/GHS, and GBP/GHS. The stacked ensembles exhibit close alignment with observed exchange rate movements, effectively tracking both trend changes and short-term fluctuations. In contrast, standalone models display noticeable lag and deviation during periods of sharp price movements, particularly under high volatility conditions.

These visual results corroborate the quantitative findings and further demonstrate the ensemble's ability to maintain stability and accuracy across diverse market regimes.

3.6 Discussion and Implications

The results provide strong empirical evidence that ridge-regularized stacking ensembles significantly enhance multivariate exchange rate forecasting performance in emerging market contexts. By integrating heterogeneous deep learning architectures within a unified framework, the proposed model overcomes the limitations of individual learners and achieves improved robustness, accuracy, and generalization.

From a practical perspective, the improved forecasting accuracy has important implications for central banks, financial institutions, and fintech firms operating in volatile currency environments. Enhanced exchange rate predictions can support more effective monetary policy analysis, risk management, hedging strategies, and automated trading systems.

4. CONCLUSION

This study proposed a ridge-regularized deep learning stacking ensemble for multivariate daily exchange rate forecasting, integrating Temporal Convolutional Networks (TCN), Long Short-Term Memory (LSTM), and Multi-Task Learning (MTL) architectures within a unified meta-learning framework. The proposed approach was designed to address the nonlinear dynamics, volatility, and cross-currency interdependencies that characterize foreign exchange markets in emerging economies, using the Ghana forex market as a representative case study.

Comprehensive empirical evaluation on USD/GHS, EUR/GHS, and GBP/GHS exchange rates demonstrates that the proposed Stack_TCN+LSTM+MTL ensemble consistently outperforms both traditional econometric and standalone deep learning models across all evaluation metrics. The ensemble achieved the lowest mean RMSE (0.7488), MAE (0.5285), MAPE (3.75%), and SMAPE (3.64%), representing substantial error reductions of approximately 89.7% relative to the VAR baseline and 92.7% relative to standalone LSTM. These results confirm that individual deep learning architectures, while effective in isolation, are limited in their ability to fully capture the diverse temporal and cross-series patterns inherent in multivariate exchange rate data.

The superior performance of the proposed framework is attributable to the complementary inductive biases of its base learners. TCN effectively captures short-term and local temporal dependencies, LSTM models long-term sequential behavior through gated memory mechanisms, and MTL enhances generalization by exploiting shared representations across related currency pairs. The ridge regression meta-learner plays a critical role in stabilizing ensemble predictions by optimally weighting base model outputs while mitigating overfitting through L2 regularization, thereby ensuring robustness under varying market conditions.

The superior performance of the proposed framework is attributable to the complementary inductive biases of its base learners. TCN effectively captures short-term and local temporal dependencies, LSTM models long-term sequential behavior through gated memory mechanisms, and MTL

enhances generalization by exploiting shared representations across related currency pairs. The ridge regression meta-learner plays a critical role in stabilizing ensemble predictions by optimally weighting base model outputs while mitigating overfitting through L2 regularization, thereby ensuring robustness under varying market conditions.

Beyond methodological contributions, the findings have important practical implications for policymakers, central banks, financial institutions, and fintech platforms operating in volatile currency environments. Improved multivariate exchange rate forecasts can support enhanced monetary policy analysis, more effective risk management and hedging strategies, improved foreign exchange pricing, and more reliable automated trading and decision-support systems. The results are particularly relevant for emerging markets, where exchange rate volatility poses significant challenges to economic planning and financial stability.

Despite its strong performance, this study has several limitations that suggest directions for future research. First, the framework relies solely on historical exchange rate data and does not incorporate exogenous macroeconomic indicators, geopolitical variables, or market sentiment information, which may further enhance predictive performance. Second, model evaluation was conducted in an offline forecasting setting; future studies should explore real-time and streaming implementations. Finally, future work will explore transformer and attention-based architectures, multi-horizon forecasting strategies, and the integration of macroeconomic and text-based features

5. REFERENCES

[1] Liu, S., Huang, Q., Li, M., &Wei, Y. (2024). A new LASSO-BiLSTM-based ensemble learning approach for exchange rate forecasting. *Engineering Applications of Artificial Intelligence*, 127, 107305.

[2] Alexandridis, A. K., Panopoulou, E., & Souropanis, I. (2024). Forecasting exchange rate volatility: An amalgamation approach. *Journal of International Financial Markets, Institutions and Money*, 97, 102067.

[3] Liu, J. (2021). Impact of uncertainty on foreign exchange market stability: Based on the LT-TVP-VAR model. *China Finance Review International*, *11*(1), 53-72.

[4] Mate, C., & Jimeńez, L. (2021). Forecasting exchange rates with the iMLP: New empirical insight on one multi-layer perceptron for interval time series (ITS). *Engineering Applications of Artificial Intelligence*, *104*, 104358.

[5] Cao, H., Lin, F., Li, Y., & Wu, Y. (2021). Information flow network of international exchange rates and influence of currencies. *Entropy*, *23*(12), 1696.

[6] Antwi, A., Kyei, K. A., & Gill, R. S. (2020). The use of mutual information to improve value-at-risk forecasts for exchange rates. *IEEE Access*, *8*, 179881-179900.

[7] Gamaliy, V., Shalimova, N., Zhovnovach, R., Zahreba, M., & Levchenko, A. (2018). Exchange rates: The influence of political and economic events. A fundamental analysis approach. *Banks and Bank Systems*, *13*(4), 131-142.

[8] Picasso, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, *135*, 60-70.

[9] Rubio, L., Palacio Pinedo, A., Mejía Castaño, A., & Ramos, F. (2023). Forecasting volatility by using wavelet transform, ARIMA and GARCH models. *Eurasian Economic Review*, *13*(3), 803-830.

[10] Beckmann, J., Czudaj, R., & Feess, E. (2020). Forecasting with VAR models under structural breaks and regime shifts. *Journal of Applied Econometrics,* 35(3), 333–356.

[11] Han, C., Wang, Y., & Xu, Y. (2020). Nonlinearity and efficiency dynamics of foreign exchange markets: evidence from multifractality and volatility of major exchange rates. *Economic research-Ekonomska istraživanja*, *33*(1), 731-751.

[12] Rossi, B. (2013). Exchange rate predictability. *Journal of economic literature*, *51*(4), 1063-1119.

[13] Hassani, H., & Silva, E. S. (2015). Forecasting with big data: A review. *Annals of Data Science*, *2*(1), 5-19.

[14] Rodrigues, P. C., Awe, O. O., Pimentel, J. S., & Mahmoudvand, R. (2020). Modelling the behaviour of currency exchange rates with singular spectrum analysis and artificial neural networks. *Stats*, *3*(2), 137-157.

[15] Liu, Yang. 2019. Novel volatility forecasting using deep learning–Long Short Term Memory Recurrent Neural Networks. *Expert Systems with Applications* 132: 99‑109.

[16] Kim, J., Kim, H., Kim, H., Lee, D., & Yoon, S. (2025). A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges. *Artificial Intelligence Review*, *58*(7), 1-95

[17] Aydin, A. D., & Cavdar, S. C. (2015). Comparison of prediction performances of artificial neural network (ANN) and vector autoregressive (VAR) Models by using the macroeconomic variables of gold prices, Borsa Istanbul (BIST) 100 index and US Dollar-Turkish Lira (USD/TRY) exchange rates. *Procedia Economics and Finance*, *30*, 3-14.

[18] Dautel, A. J. (2020). Forex exchange rate forecasting using deep recurrent neural networks. *Computational Economics, 56*(1), 1–22.

[19] García, F., Guijarro, F., Oliver, J., & Tamošiūnienė, R. (2023). Foreign exchange forecasting models: ARIMA and LSTM comparison. *Engineering Proceedings*, *39*(1), 81.

[20] Zhao, L., Zhang, X., & Li, Y. (2024). Prediction of currency exchange rate based on ARIMA, LSTM, and GRU models. *Journal of Risk and Financial Management, 17*(8), 332.

[21] Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, *54*(8), 6391-6438.

[22] Madhulatha, T. S., & Ghori, D. M. A. S. (2025). Deep neural network approach integrated with reinforcement learning for forecasting exchange rates using time series data and influential factors. *Scientific Reports*, *15*(1), 29009

[23] Lara-Benítez, P., Carranza-García, M., Luna-Romera, J. M., & Riquelme, J. C. (2020). Temporal convolutional networks applied to energy-related time series forecasting. *applied sciences*, *10*(7), 2322..

[24] Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(12), 5586–5609.

[25] Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information*, *15*(12), 755.

[26] Lin, Hualing, Qiubi Sun, and Sheng-Qun Chen. 2020. Reducing Exchange Rate Risks in International Trade: A Hybrid Forecasting Approach of CEEMDAN and Multilayer LSTM. *Sustainability* 12: 2451.

[27] Islam, M. S., & Hossain, E. (2021). Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. *Soft Computing Letters*, *3*, 100009.

[28] He, K., Yang, Q., Ji, L., Pan, J., & Zou, Y. (2023). Financial time series forecasting with the deep learning ensemble model. *Mathematics*, *11*(4), 1054.

[29] Sina, L. B., Secco, C. A., Blazevic, M., & Nazemi, K. (2023). Hybrid forecasting methods—a systematic review. *Electronics*, *12*(9), 2019.

[30] Gu, J., Zhang, S., Yu, Y., & Liu, F. (2025). AB-LSTM-GRU: A novel ensemble composite deep neural network model for exchange rate forecasting. *Computational Economics*, *66*(2), 1767-1791.

[31] Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica,* 48(1), 1–48.

[32] Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, *15*(9), 517.

[33] Fan, J., Zhang, K., Huang, Y., Zhu, Y., & Chen, B. (2023). Parallel spatio-temporal attention-based TCN for multivariate time series prediction. *Neural Computing and Applications*, *35*(18), 13109-13118.

[34] Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. *IEEE transactions on knowledge and data engineering*, *34*(12), 5586-5609.

[35] Ghasemieh, A., Lloyed, A., Bahrami, P., Vajar, P., & Kashef, R. (2023). A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients. *Decision analytics journal*, *7*, 100242.