IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

HyBiLSTM: Multivariate Bitcoin Price Forecasting using Hybrid Time Series Models with Bidirectional LSTM

Anny Mardjo[1] and Chidchanok Choksuchat[2]

[1]College of Digital Science, Prince of Songkla University, Thailand.
[2]Division of Computational Science, Faculty of Science, Prince of Songkla University, Thailand.

Corresponding author: Chidchanok Choksuchat (e-mail: chidchanok.ch@psu.ac.th).

"For researcher load: this work was supported by the Digital Science for Economy, Society, Human Resources Innovative Development and Environment project funded by Reinventing Universities & Research Institutes under grant no. 2046735, Ministry of Higher Education, Science, Research and Innovation, Thailand."

ABSTRACT Despite their popularity in recent studies, most hybrid models that exploit the advantages of both classical time series and deep learning models were conducted in univariate forecasting context. For econometric domain which exogenous factors play a crucial role, more studies in multivariate forecasting is essential and should be encouraged. Thus, contributing to hybrid multivariate forecasting literature, a hybrid model named HyBiLSTM was proposed. The algorithm began with ARIMAX GARCHX model forecasting, followed by second forecasting of model residual using Grey Wolf Optimizer based hyperparameters Bidirectional LSTM model. With residuals instead of original multivariate features, LSTM can avoid to processes each feature independently and therefore, reducing convergence complexity and execution time. The final forecasting results was compounded from both models. Three quantitative measurements, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), were used to evaluate the established models using the historical daily runoff social and economic based data (01/07/2019-31/12/2022). The findings showed 1) the addition of exogenous factors improved the performance of ARIMA and GARCH models; 2) BiLSTM outperformed other LSTM variants when was integrated with ARIMAX GARCHX model; 3) Using SHAP, Bitcoin price was influenced by stock price, Twitter volume, gold price, and Twitter sentiment index; and 4) structural break had significant effect on forecasting. Other than expanding the literatures regarding hybrid models in multivariate context, this study provides practical contribution for investors by analyzing the factors that the investors can use as an early warning for Bitcoin price fluctuation.

INDEX TERMS LSTM, ARIMAX, GARCHX, Bitcoin, SHAP, Bidirectional LSTM

I. INTRODUCTION

Unlike stocks which are backed by underlying companies' assets and earnings, cryptocurrencies do not have any intrinsic value. Their high volatility, as shown during 2021 when Bitcoin lost more than half of its values before it fully recovered a few months, has made them high investment risk. Despite this, their premises of decentralization, investment diversification, potential high returns and 24/7 trading access still attracted growing numbers of investor to invest in various cryptocurrencies. Since its debut in 2009 as the world's first decentralized currency, Bitcoin continues to function as the most valued cryptocurrency. Electric car maker Tesla's Bitcoin holdings and El Salvador's Bitcoin adoption as legal currency in 2021 strengthen Bitcoin potential as digital asset and currency. To minimize the risk and gain the highest return, a dependable and precise forecasting model that reflected Bitcoin characteristics, such as high volatility, non-stationarity, non-linearity and dependency on numerous factors, is vital.

Its dependency on varied factors have increased studies of multivariate Bitcoin price forecasting in recent years. Multivariate statistical time-series modelling includes the Autoregressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and their variants [1], [2], [3]. Despite their ease of interpretation, implementation and fast computing time,

they fall behind Deep Learning (DL) models in capturing non-stationary and non-linear behavior of Bitcoin price. Long Short-term Memory (LSTM) was particularly popular as it overcame the primary drawback of RNN, known as diminishing gradient problem, by incorporating a hidden memory layer to retain long-term dependencies that allows it to learn patterns more precisely in sequential datasets [1]. Other studied DL models covers Adaptive Neuro-Fuzzy Inference Systems [2]; Gated Recurring Unit)GRU([3], [4]; Recurrent Neural Network (RNN) [5], [6]; and Multilayer Perceptron [5], [6].

Despite its superiority, LSTM has its disadvantage in processing multiple features. It processes each feature independently, with its own set of weights and biases using the same computational graph at a given timestep. Then, it produces a hidden state for each feature which will be combined to use as an input for the next timestep. This increase convergence complexity and processing time. To harnesses the unique advantages of classical time series models and LSTM to predict multivariate time series data, a hybrid approach combining ARIMAX GARCHX and variant of LSTM (standard LSTM, Bidirectional LSTM, and attention-based layer LSTM) was proposed. Lack of hybrid model studies in multivariate forecasting also motivated us to propose this model.

Multivariate forecasting requires multiple input data and although many factors have been known to influence Bitcoin price, this study prioritized social and economic based factors. This decision was based on the wide accessibility and comprehensibility of social and economic based factors data among the general public, therefore, increases the relevance and practicality of this study. Past studies in social-based factors include Google search trends, tweet sentiments, tweet volume, Wikipedia search trends, online communities' comments [7], [8], [9], [10], [11], [12], [13], [14], [15] while economic-based factors covers Federal funds rate, stock market indexes, gold price, oil price, foreign exchange rate, consumer price index, US dollar index, general commodity index, US Dollar supply, US GDP, Bonds, producer price index, employment rate) [7], [8], [9], [10], [11], [16], [17], [18].

The key novel contributions from this paper include:

1) in-depth comparison and evaluation of hybrid multivariate time series models like ARIMAX and GARCHX with LSTM variants (i.e. standard LSTM, BiLSTM, and attention-based layer LSTM). To our knowledge, this is the first study to provide such in-depth comparison for all of these models.

2) the use of SHAP to comprehend the effects of exogeneous factors on Bitcoin price forecasting model leading to new detailed knowledge (or reinforce existing understanding) of complex Bitcoin price dynamics.

The remaining sections of this paper are structured as follows: Section II provides a comprehensive overview of the literature reviews related to our work. In Section III, we introduce the methodology employed in our research. Section IV presents the results and discussions. Lastly, in Section V, we offer our concluding remarks and explore potential directions for future research.

II. RELATED WORKS

A. HYBRID TRADITIONAL STATISTICAL MODEL OF TIME SERIES DATA

Hybrid model of time series data is the combination of two or more time series forecasting models to improve the predictions. By blending the strengths of different modelling approaches, hybrid models aim to mitigate the weaknesses inherent in individual models. Most extensive prior literature in hybrid models are univariate forecasting model and although univariate model offers simplicity and more reliability compared to multivariate models, using a single factor to analyze complicated econometric problems is insufficient. The incorporation of exogenous factors has been found to extend modelling capability and prediction accuracy [19]. Multivariate models using DL generally require more processing complexity and longer training times which deemed them to be less desirable solution. On the other hand, classical time series model such as ARIMAX and GARCHX requires less parameters and easier to implement but has lower forecasting accuracy than DL Having a hybrid model that utilized both of their advantages should be a desirable solution. Following [20], our hybrid model assumed that a time series y_t is the combination of linear component L_t and non-linear component N_t, as specified in (1).

$$y_t = L_t + N_t \tag{1}$$

Using two level forecasting procedures, the first level utilized ARIMAX GARCHX model to forecast the linear component of time series data $\hat{L}_t$ while the second level analysed the residuals of that model e_t (2) using variants of LSTM model to forecast the non-linear pattern $\hat{N}_t$. Then, both linear and nonlinear forecasts were combined to gain final prediction (3). Fig. 1 depicts our proposed model.

$$e_t = y_t - \hat{L}_t \tag{2}$$

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{3}$$

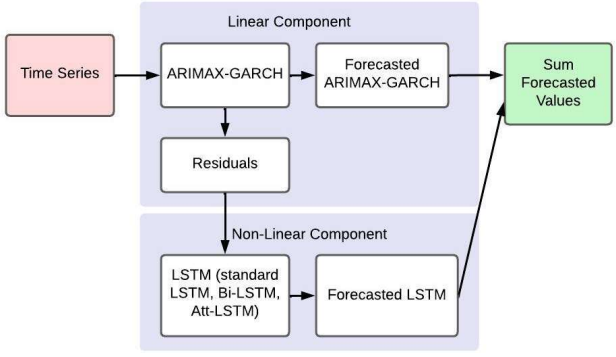

FIGURE 1. Proposed Hybrid Model

B. MULTIVARIATE HYBRID TIME SERIES FORECASTING MODEL

A "Multivariate hybrid time series forecasting model" integrates multiple forecasting techniques and uses multiple variables (or series) simultaneously to predict one or more outcomes. For past recorded influences of external factors on Bitcoin price, Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) model was used. The general form of ARIMAX (*p,d,q*) model [6] can be written as:

$$\phi(B)(1-B)^d Y_t = \theta_0 + \theta(B)\varepsilon_t + \sum_{i=1}^{r} a_i x_{it} \qquad (4)$$

where B is the backward shift operator; $(1-B)^d$ is the differencing operator of order d to produce stationarity of the dth differenced data; $\phi(B) = \left(1 - \phi B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$ is a moving average polynomial with order p; $\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right)$ is an autoregressive polynomial with order q; $\phi_1 \ldots \phi_p$ = the parameters of the autoregressive part of model; $\theta_0 \ldots \theta_q$ = the parameters of the moving average part; d is a number of times of order differencing; ε_t is error term; x_t is the value of the independent variable X at time t; a, ϕ, θ = coefficients of the factors, autoregressive terms, moving average terms; and r refers to number of factors.

As ARIMAX linearly models the data, GARCHX model [21] was added to capture heteroscedastic variances of highly volatile Bitcoin price. The addition of X on standard GARCH model in both the mean and variance of the model has been reported to improve model forecasting [22]. The mean and variance model can be written as (5) and (6).

$$\phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t + \sum_{s=0}^{l}\sum_{k=1}^{r} Y_{ks}\, x_{k,t-s} \qquad (5)$$

$$\sigma^2{}_t = a_0 + \sum_{i=1}^{p} a_i \in^2{}_{t-1} + \sum_{j=1}^{q} b_j \sigma^2{}_{t-j} + \sum_{s=1}^{l}\sum_{k=1}^{r} Y_{ks}\, x_{k,t-s} \qquad (6)$$

where r is the number of exogenous variables; l is the lag length of the exogenous variables; a, b = coefficients of the autoregressive terms, moving average terms; Y_k are the effects of the exogenous variables on the conditional variance of the residuals.

C. DEEP LEARNING MODEL: LSTM, BILSTM AND ATTENTION-BASED LAYER LSTM

Deep learning models outperform traditional statistical models in situations of complex non-linear patterns and long-term dependencies in the data. Despite its popularity in DL studies, RNNs cannot preserve long sequences of input and suffers the long-term dependency problem [23]. To overcome this, LSTM with three gates mechanism and memory cells C_t of hidden layer node was proposed by Hochreiter and Schmidhuber [24]. It starts by updating the candidate for the new cell states $\tilde{C}_t$ as a function of hidden state at previous timestep h_{t-1} and input vector at the current timestep x_t using (7)

$$\tilde{C}_t = tanh(W_c h_{t-1} + U_c x_t + b_c) \qquad (7)$$

where W_c, U_c represent the weights of networks and b_c are bias variable value. Then, it computes forget f_t and input i_t gates to determine how much contents from the previous cell C_{t-1} will be erased and how much of values of the new candidate cell states $\tilde{C}_t$ combined into the new cell state C_t, using (8), (9) and (10) respectively.

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_f + b_f\right) \qquad (8)$$

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \qquad (9)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \qquad (10)$$

where W_f, W_i, U_f, U_i are the weights of networks; b_f, b_i are bias variable values. Output hidden states h_t will then be filtered by output gate $o_t$using (11) and (12) to decide value of cell state C_t that will go to output.

$$o_t = \sigma(W_0 h_{t-1} + U_o x_t + b_o) \qquad (11)$$

$$h_t = o_t \otimes tanh(C_t) \qquad (12)$$

where W_0, are weights of the networks; b_o are bias variable values. Two activation functions were used here, namely the sigmoid (σ) and the tanh. The prediction of the next event $\hat{x}_{t+1}$ can be calculated using

$$\hat{x}_{t+1} = g(V \,.\, h_t) \qquad (13)$$

where V is output layer weight matrix and g can be any activation function which matches the type of the target in data. Two layers stacked LSTM (Fig. 2) was proposed as the first model as two layers LSTM structure was found to improve the accuracy of the model prediction [25].

The second proposed LSTM model was two layers stacked BiLSTM (Fig. 3). BiLSTM model employed two separate hidden layers with reverse direction for both previous data and future data [26]. The output values of each hidden layers are calculated independently and then, merged for the final output. Using this approach, both past and future information contained in the dataset can be preserved.

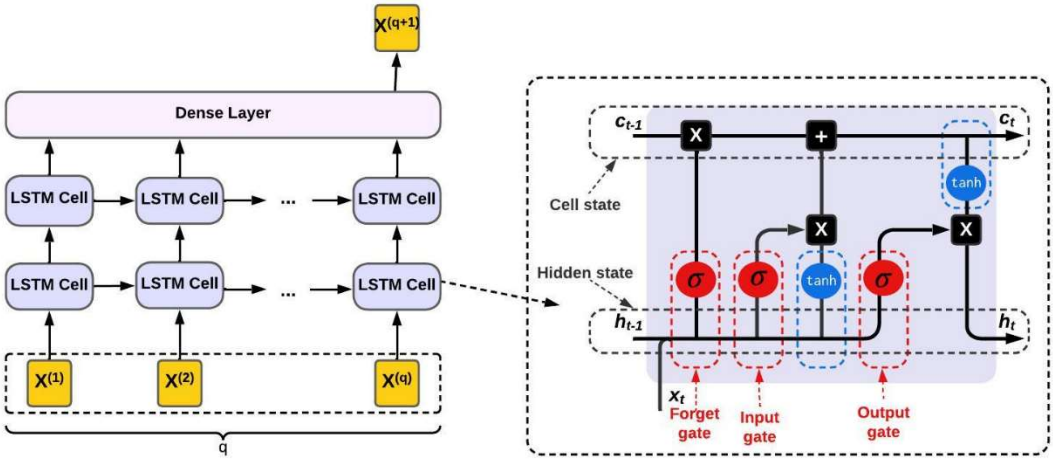

FIGURE 2. **Proposed 1st model: Two Layers Stacked Standard LSTM (LSTM).**

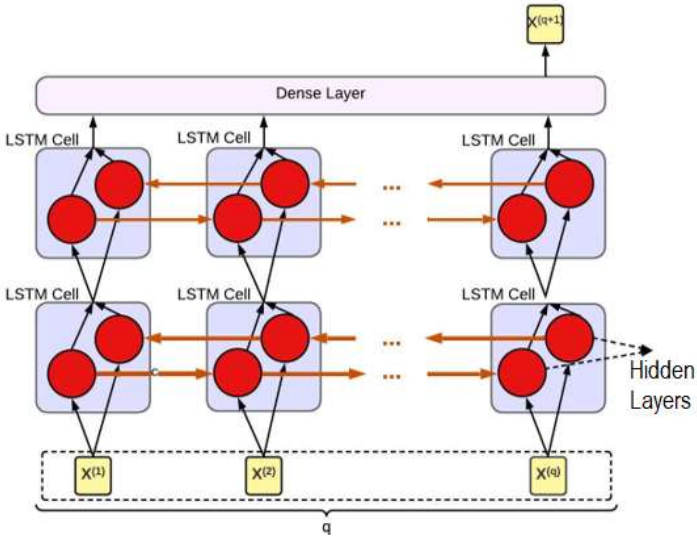

FIGURE 3. **Proposed 2nd model: Two Layers Stacked Bidirectional LSTM (BILSTM).**

LSTM accuracy tends to decrease the longer the length of the input and output sequences is as the hidden state of the model gets overwritten repeatedly, reducing the ability of initial inputs to influence the later outputs in the sequence. Attention mechanisms [27] seek to correct this by weighting the hidden states of all available time steps $h_1, \ldots, h_t$ instead of the last one hidden state h_t [28]. By assigning different weights to different hidden state, the model can focus on most relevant parts of the input sequence for making predictions. Then, the output o_t can be computed as a weighted sum of hidden states as in (14).

$$o_t = \sum_{i=1}^{t} \alpha_i^t \cdot h_i \tag{14}$$

where attention weight α_i^t is computed as the following

$$\alpha_i^t = \frac{exp(score(h_i,\ q_t))}{\sum_{j=1}^{t} exp(score(h_i,\ q_t))} \tag{15}$$

while the following (16) for the score function used Dot-product attention algorithm, selected because of its computing efficiency and popularity

$$score(h_i,\ q_t) = h_i^t W_a q_t \tag{16}$$

where W_a is the weight matrix and the previous time-step's output o_{t-1} is used for the query term q_t. For our proposed third (Fig. 4) and fourth models (Fig. 5), another LSTM and application layers were added above proposed first and second models. This combination of the attention-based layer o_t and the LSTM output h_{t-1}as the next layer's input for the current timestep results in

$$X_t = concat(o_t, h_{t-1}) \tag{17}$$

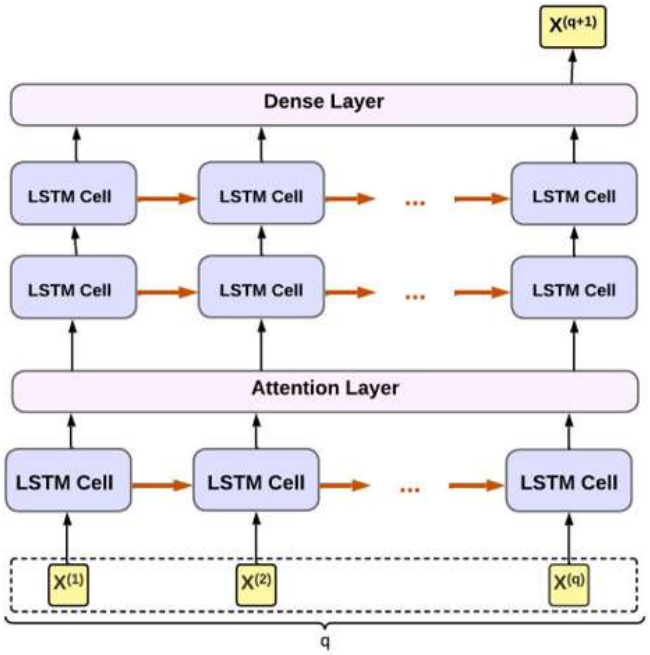

FIGURE 4. **Proposed 3rd model: Two Layers Stacked Attention layer LSTM (ATT-LSTM).**

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3386029

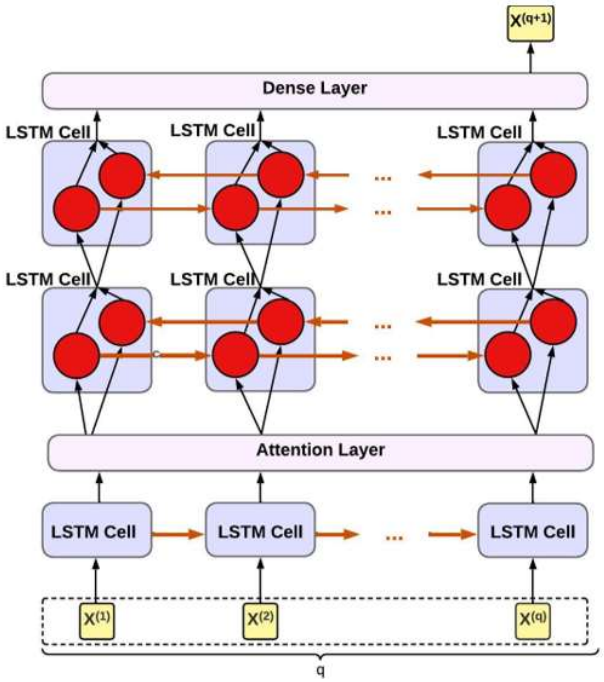

FIGURE 5. **Proposed 4th model: Two Layers Stacked Attention layer BiLSTM (ATT-BiLSTM).**

D. LSTM PARAMETERS OPTIMIZATION: GREY WOLF OPTIMIZER (GWO)

To optimize LSTM parameters, GWO which is a Bio-inspired computing (BIC) algorithm rooted in the natural behavior of animals, birds, insects, and other organisms, was used. While BIC algorithms may require significant time and resources due to parameter optimization and iterative processes, they possess the ability to discover unknown patterns and rely less on mathematical modelling or exhaustive training [29]. Based on the social structure and hunting nature of grey wolves, GWO was selected among BIC algorithms because of its past performance in varied domains and simplicity [30], [31], [32]. Holding the highest hierarchy, alpha (α) wolf makes decisions for the pack, followed by the beta (β) wolf acting as the alpha's advisor and enforcer of pack discipline. The lowest-ranking member, the omega (Ω) wolf, submits to other wolves, while the delta (δ) wolf commands the omega and obeys the alpha and beta. The GWO algorithm follows these steps: 1) each wolf measures its distance from α, β, and δ using (18) to (23), and 2) updates its position using (24) [33].

$$D_\alpha = |2r_2 . X_\alpha - X_i| \tag{18}$$

$$D_\beta = |2r_2 . X_\beta - X_i| \tag{19}$$

$$D_\delta = |2r_2 . X_\delta - X_i| \tag{20}$$

$$X_1 = X_\alpha - (2a \cdot r_1 - a) \cdot D_\alpha \tag{21}$$

$$X_2 = X_\beta - (2a \cdot r_1 - a) \cdot D_\beta \tag{22}$$

$$X_3 = X_\delta - (2a \cdot r_1 - a) \cdot D_\delta \tag{23}$$

$$X_1(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{24}$$

where X_α, X_β, and X_δ are the positions of the search agents (grey wolves); D_α, D_β, and D_δ are distances between the current position i and the search agents (α, β, δ); $r1$ and $r2$ are randomly generated numbers for boundary search space. Fig. 6 depicted GWO used in our study.

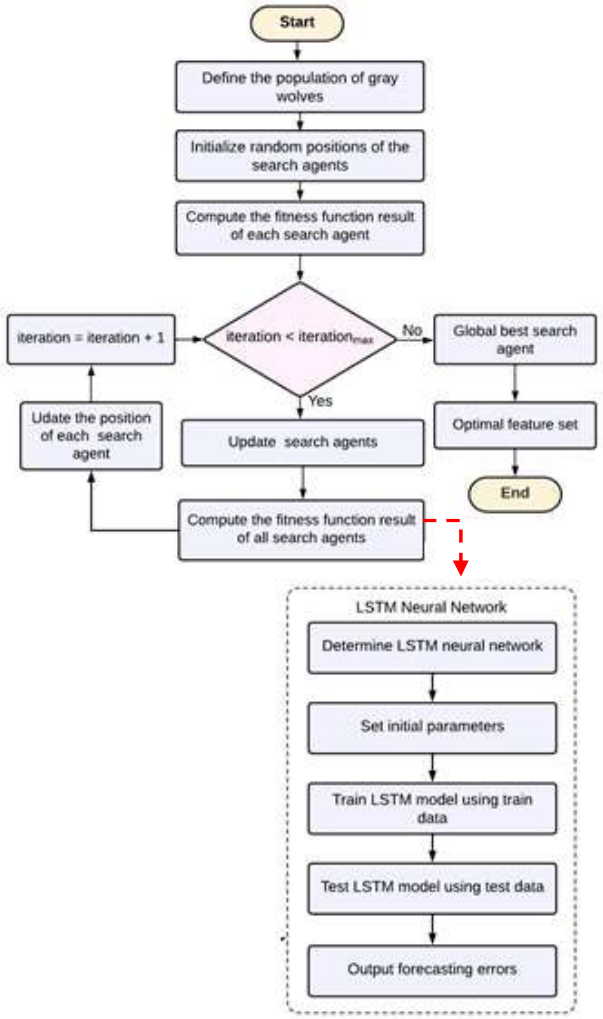

FIGURE 6. **illustrates the GWO algorithm used in conjunction with the proposed LSTM models.**

E. SHAPLEY ADDITIVE EXPLANATIONS (SHAP)

Another important contribution of this paper is the usage of SHAP for model explanation to find the influencing factors of Bitcoin price. SHAP was utilized because of its 3 advantages. First, SHAP is a local Interpretable Machine Learning (IML) technique that can be adjusted to become constant global explanations [34]. Therefore, explanations of

both single operating points and general trend identification can be obtained. Second, Shapley values, computed through the average marginal contribution of each feature to all feature coalitions with that feature, give insights into the impact of a feature on the model outcome and therefore, use to modify the model. Third, SHAP additive feature attribution method enable connection to other IML techniques, like LIME and DeepLIFT. The SHAP model can be represented as a linear combination of the binary variables in the following (25).

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i' \qquad (25)$$

where g is an explanatory model, $z^i \epsilon \{0,1\}^M$ is the coalition vector, M is the maximum number of features, the *i*th feature has a contribution ($z' = 1$) or not ($z' = 0$). ϕ is the SHAP value of the *i*th feature, representing the contribution of the *i*th feature and can be calculated according to the following (26).

$$\phi_i(f, x) = \sum_{S \subset N} \frac{|S|!\,(M - |S| - 1)!}{M!} [f_x(S \cup \{i\}) - f_x(S)] \qquad (26)$$

where N is the set of all features, $|S|$ represents the number of features in feature subset S excluding the *i*th feature; $f_x(S)$ represents the result of the machine learning model f training in feature subset S.

III. METHODOLOGY

Research methodology consists of data life cycle management until dealing with forecasting model.

A. DATA EXTRACTION

As in [2], historical Twitter volume was obtained from bitinfocharts.com using the keyword "Bitcoin" (case insensitive). Google search volume was gathered using the Gtrends library. The daily closing prices of Bitcoin, oil price, gold price, and U.S. stock market indexes (S&P 500, NASDAQ, and Dow Jones Industrial Average) were collected using either the *Quantmod* or *Quandl* libraries. Following prior research (Chen, 2023) to consider less than four-year date range limit and Bitcoin price bubbles, the data was collected for the period between 01/07/2019 and 31/12/2022.

Using Twitter API, we collected 2,000 tweets four times a day with keyword "Bitcoin". HyVADRF (Hybrid VADER-Random Forest) and GWO algorithms [4] (Fig. 7) was used to pre-process and classify sentiments of the tweets. The VADER algorithm eliminates the need for laborious, time intensity, error prone, and costly manual labelling while the TF-IDF implementation in this algorithm restricts the terms used for classification calculations to those relevant to the data context, resulting in a simpler and faster processing model. The GWO-Random Forest (RF) in the model is appropriate, considering the advantages of supervised over unsupervised sentiment analysis [5]. HyVADRF algorithm is divided into three distinct processes: 1) generating data labelling, 2) selecting the best machine learning model, and 3) performing model optimization using GWO (Fig. 7).

Algorithm: HyVADRF (Hybrid VADER-Random Forest)
Input: Twitter dataset *T*
Output: Optimized classifying machine learning model
Process:

To generate data labelling:
  for *t* ∈ *T*
  Clean *t* from URL links, *hashtag* symbols, and irrelevant tweets using RegEx.
  Calculate sentiment Compound Value (*cv*) using VADER library.
  if *cv* > 0.05
    class = "positive"
  elseif cv <= 0.05 and *cv* >= -0.05
    class = "neutral"
  else
    class = "negative"
  end if
  end for

To select the best machine learning model:
Create Vector Corpus (*V*) from the *labelled tweets*
Clean *v* by removing punctuations, numbers, stopping words, and stemming
Create a Document Term Matrix (*M*) from *V* using TF-IDF.
Clean *M* by removing sparse terms using a 98% term sparsity threshold.
Divide *M* into 70% for train set (*N*) and 30% for the test set (*D*).
for *j* ∈ (NB,DT,RF,SVM)
  Train *N* using *j* model.
  Test the trained model using *D*
end for
Select the best performance-trained model based on accuracy, precision, recall, and F-score.

To perform model optimization using Grey Wolf Optimizer GWO):
Initialize GWO parameters (i.e. *maxIteration*).
while *i=0 < maxIteration*
  Get *min.node.size, num.trees, mtry and sample.fraction* values from GWO.
  Execute *ML* train model with *min.node.size, num.trees, mtry and sample.fraction* values using *N*.
  Execute *ML* with *D* to get accuracy.
end while
Get optimum result of *min.node.size, num.trees, mtry and sample.fraction* values.
Execute *min.node.size, num.trees, mtry,* and *sample.fraction* values with *N* and *D*.

FIGURE 7. Algorithm for Hybrid VADER – Random Forest

These sentiments were then converted into continuous data for the daily Twitter sentiment using (27), following [36].

$$SENT_t = \frac{N_{positive}}{N_{positive} + N_{negative}} \qquad (27)$$

where $N_{positive}$ is the total positive tweets on day t and $N_{negative}$ is total negative tweets on day t. The daily twitter sentiments were then combined with other data (Fig. 8).

B. DATA PREPROCESSING

Preprocessing of the original dataset was essential to address varying data recording methods, missing values, and non-uniform scales among the features. As these daily data will be used for future time series prediction, which generally depends on the historical data; missing or incomplete data was handled using linear interpolation from the *imputeTS* library of R. Linear interpolation had been found to give better results compared to mean method in substituting time series missing values problem [37]. It followed by data decomposition to observe its seasonality. Outliers are commonly occurred in time series within cryptocurrency, commodity and finance domains [38], as such, descriptive statistics of the data were computed to identify extreme outliers. These outliers were retained in order to preserve the

dynamics of each series caused by exceptional events [38]. However, as dominant patterns from variances of significant outliers can hide trends, the logarithmic transformation was used to normalize the data.

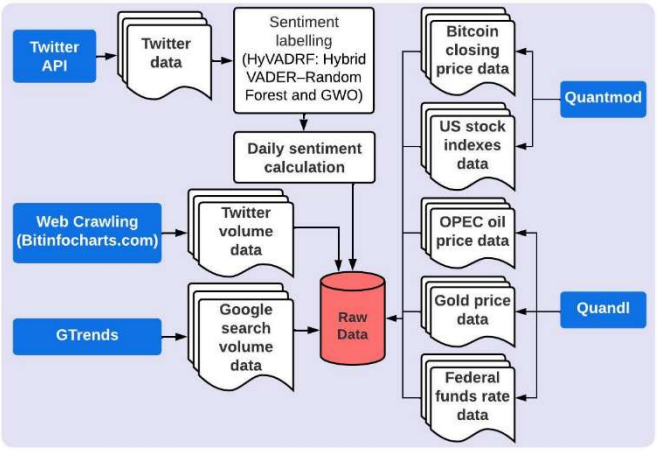

FIGURE 8. Process for data collection

The dataset was split into approximately 85%, accounting 20% of it was used as LSTM validation test, for train set and 15% for test set. Then, considering the datasets have different kind of measurements and no negative values, they were scaled into values (0,1). As stationarity has significance influence for the efficiency of a time-series model [39], the data stationary need to be confirmed using Augmented Dicker Fuller (ADF) Test for a unit root test. If non-stationary was found, differencing was conducted to de-trend the data. The *d* parameter for differencing was selected using ACF plot pattern. If the pattern featured a first-order moving average process with a parameter of -0.5, it implied over differencing. After the stationery condition was filled, data correlations were tested with Pearson correlation coefficient. Our selected factors had been found to influence Bitcoin price in past studies, therefore, Principal Component Analysis (PCA) was utilized to evaluate the possibility of merging highly correlated factors into one factor, rather than removing them (Fig. 9).

C. FORECASTING MODELS

The development and validation of the proposed model was separated into 2 levels: linear and non-linear components (Fig. 10). Level 1 started by finding the best candidates of ARIMAX (*p, d, q*) models. For its *p* and *q* parameters, the values were derived by traversing $p \in (0,5)$ and $q \in (0,5)$. The candidates ARIMAX model, selected based on AIC and BIC scores, were then trained with test data to select the best ARIMAX model using the lowest RMSE, MAE, and MAPE scores. Afterwards, ARCH Lagrange Multiplier (LM) Test of the best ARIMAX model residuals was tested. If ARCH effect was indicated, the addition of GARCHX model to ARIMAX model could overcome the non-stationary in residuals and improve model performance.

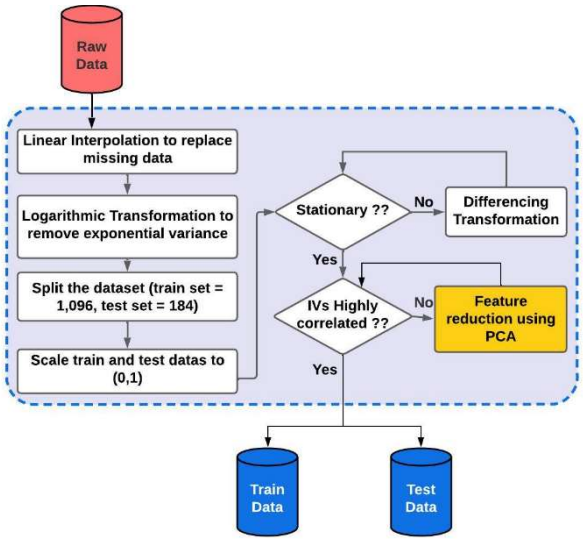

FIGURE 9. Data Preprocessing

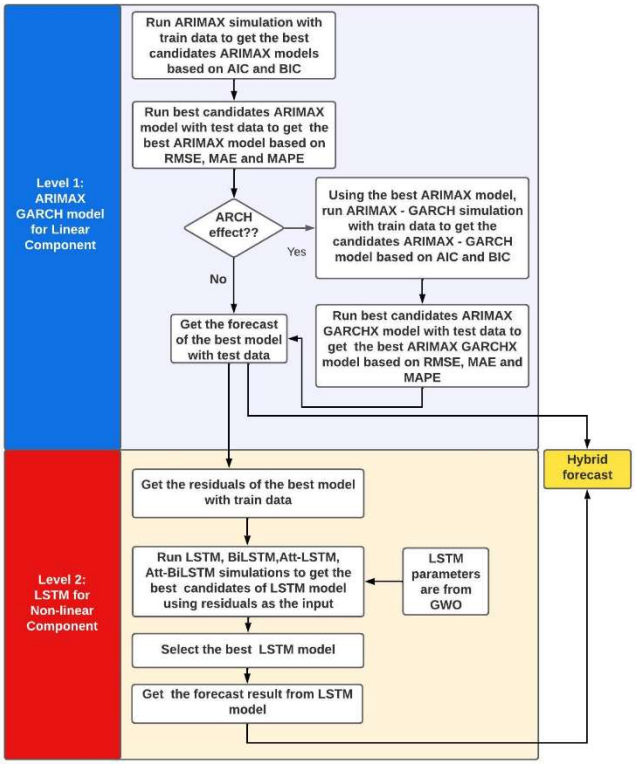

FIGURE 10. Methodology

The *p* and *q* parameters of the GARCHX (*p, q*) model were obtained by iterating $p \in (0,5)$ and $q \in (0,5)$. The residuals from the fitted ARIMAX-GARCHX model was used as the input for LSTM models in Level 2. Lags for LSTM, identified using partial autocorrelation function (PACF) plot [25], were incorporated as the model parameter for LSTM, BiLSTM, ATT_LSTM and ATT-BiLSTM in R Keras package. Different combinations of hyperparameters (epoch, hidden nodes, dropout rate, learning rate, and batch size,) were selected through Grey Wolf Optimization

algorithm. Forecasted residuals based on LSTM model and forecasted values based on ARIMAX GARCHX were combined to get the final forecast. In addition, SHAP algorithm was applied to obtain global and local explanation of ARIMAX GARCHX model.

D. PERFORMANCE MEASUREMENTS

For preliminary data analysis, the normal distribution of the data was measured using descriptive statistics (skewness and kurtosis). The non-stationary of the data was observed using ADF test. The p-value from ADF test should be less than 0.05 indicating a stationary series. The candidates ARIMAX and GARCHX models were selected based on the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values using (28) and (29).

$$AIC = \frac{-2}{N} * (log - likelihood) + 2 * k/N \quad (28)$$

$$BIC = -2 * (log - likelihood) + ln(N) * k \quad (29)$$

where k is the number of model parameters; N is the number of examples in the training dataset and *log-likelihood* is a measure of model fit. The optimal ARIMAX, GARCHX and LSTM were selected according to the lowest Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) indicating that the model's predicted values are closer to the true values [40]. MAE, RMSE, and MAPE were calculated using the following:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (30)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|^2} \quad (31)$$

$$MAPE = \sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \quad (32)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, y_{mean} is the mean of actual value, and n is a total number of data points.

For the best performance ARIMAX model, residual diagnostics was performed to evaluate the assumption of non-autocorrelated random stationary using the Ljung-Box Q-statistic test. The returned p-value should be greater than 0.05, indicating residuals independency and model fitness. ARCH LM Test to discover ARCH effect was also performed. The returned p-value less than 0.05 signified variance heteroscedasticity and GARCHX model could be added to improve ARIMAX performance. Residuals of the best ARIMAX GARCHX models was used as the input data for LSTM (LSTM, BiLSTM, ATT-LSTM, and ATT-BiLSTM) to create a hybrid model. The best hybrid models were chosen based on the lowest MAE, RMSE and MAPE. Finally, SHAP values were calculated to explain the model. The global explanation used the average absolute SHAP values of a factor and the bigger the SHAP value, the more important the factor for the model. The local explanation used individual observational data to analyze its effect on the model.

E. MACHINE SPECIFICATION

The specification is Lenovo IdeaPad S1145-14IIL, Processor Intel™ Core ™ i5-1035G1 CPU @ 1.00GHz. The system specifications also include a processor running at 1.19 GHz, with 20.0 GB of installed RAM (19.8 GB usable), and a 64-bit operating system with an x64-based processor. The software used is RStudio version 1.4.1717, and R's libraries are utilized in this setup: {Keras, TensorFlow, metaheuristicOpt, fastshap, dplyr, imputeTS, rugarch, forecast, Quantmod, Quandl, GTrends}.

IV. RESULTS

The collected data was ranging from 01/07/2019 to 31/12/2022. Using Quantmod library, we collected 1,280 records of Bitcoin closing price, 884 records each for Dow Jones, NASDAQ and S&P500 stock indexes. *Quandl* library was applied to obtain 906 records of OPEC oil price, 886 records of gold price and 1,280 records of Federal funds rate. 1,280 records of Bitcoin tweets volume were scraped from Bitinfocharts.com while 1,280 records of Google search trends were gathered using *Gtrends* library. 2,000 tweets were collected for each date using Twitter API and binary labelled using HyVADRF [30]. These binary labelled tweets for each date were then transformed into continuous value to get daily sentiment index. In total, 2,131,900 tweets were collected and pre-processed for 1,280 Bitcoin tweets' daily sentiment index. The final total dataset was 1,280 records.

A. DATA INTERPRETATION

Decomposition of timeseries was conducted to analyze the seasonality and the trends of Bitcoin price (Fig.10). The monthly average Bitcoin price plot displays the absence of monthly (a) or yearly (b) seasonality while Bitcoin closing price plot showed the characteristics of non-existent trend. As depicted in Fig 11a, the data curve for year 2019 starts from month of July as our collected data is from 01/07/2019 to 31/12/2022.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3386029

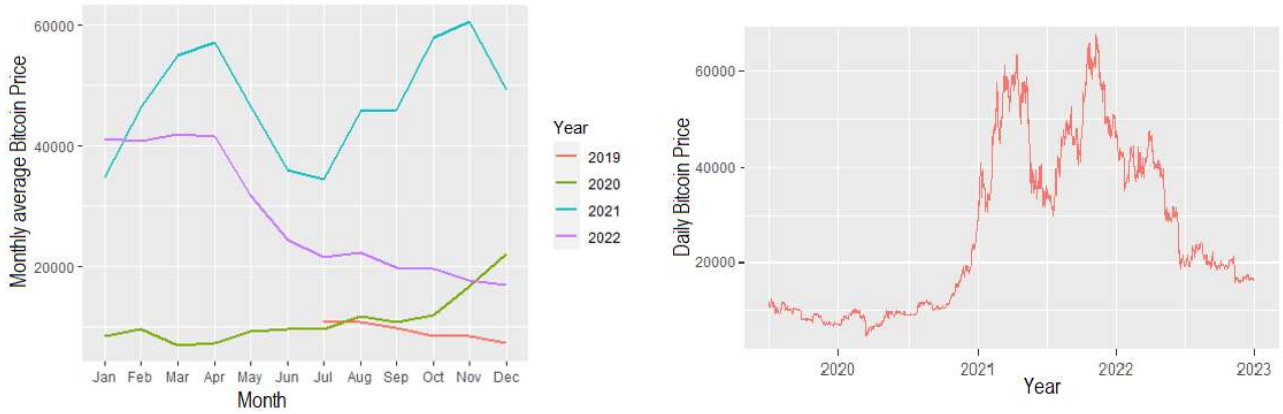

FIGURE 11.(a) monthly average Bitcoin price; (b) yearly Bitcoin price

TABLE I
DATASET DESCRIPTIVE STATISTIC

Statistics	Twitter Sentiment Index (TSI)	Twitter Volume (TV)	Google Trends (GT)	Oil Price (OIL)	Gold Price (GO)
Mean	0.3239	84,421	52.11	69.24	1,745
Median	0.3100	88,890	49.00	66.72	1,780
Maximum	0.5500	363,566	100.00	128.27	2,067
Minimum	0.2100	445	22.00	12.22	1,389
Skewness	0.7636	0.6347	0.6339	0.1472	-0.6466
Kurtosis	2.8637	3.1988	2.6588	2.5293	2.6091
Statistics	Dow Jones stock index (DJI)	NASDAQ stock index (NA)	S&P 500 stock index (SP)	Federal Funds Rate (FFR)	Bitcoin Price (BTC)
Mean	30,679	11,687	3,741	0.8873	26,069
Median	31,063	11,622	3,822	0.1000	19,979
Maximum	36,800	16,057	4,797	4.3300	67,567
Minimum	18,592	6,861	2,237	0.0400	4,971
Skewness	-0.4079	-0.0868	-0.1404	1.2275	0.5788
Kurtosis	2.3177	1.8535	1.9022	3.5381	1.9790

Other than gold price and stock indexes, the skewness values of data are positive which means that they have longer right tails compared to theirs left tails (Table I). Twitter volume and federal funds rate with kurtosis greater than three forms leptokurtic and fat tails, signifying a relatively high number of extreme values. In contrast, other variables have platykurtic distributions and thinner tails, indicating fewer extreme values. The log-transformation was performed for the whole dataset to get more normalized data before it was split into 1,096 records for train set and 184 records for test set (Fig. 12). As the order of the data points in time series is crucial, randomly shuffling the data could disrupt the temporal order, dependencies within the time series and unintentionally infer the trend of future data. Both datasets were then transformed into uniform values (0-1) using Min-Max scaling.

The original sequence of BTC price has a unit root in the three models while the three models of the first-order difference sequence all passed the ADF test (Table II). Similarly, all the exogenous factors succeeded the ADF test after the first-order difference series whilst only original sequence of Twitter sentiment index, Google Trends and Dow Jones stock index indicate otherwise (Table III).

The ACF plots first-order and second-order differences on the original data was analyzed to get *d* parameter (Fig. 13).

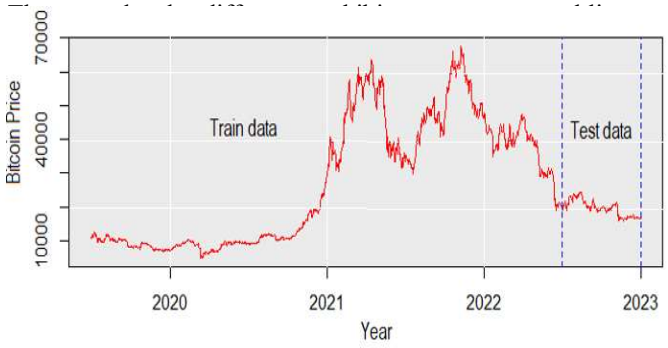

FIGURE 12. Dataset Proportion

TABLE II
ADF TEST RESULTS OF BTC PRICE

Model settings	original sequence		first-order difference sequence	
	ADF statistic	p-value	ADF statistic	p-value
No intercept term, no trend	-0.158	0.598	-10	0.01
Intercept term with no trend	-0.982	0.708	-10	0.01
Intercept term with a trend	-0.656	0.974	-10	0.01

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3386029

Table III
ADF TEST RESULTS OF EXOGENOUS FACTORS*

Exogenous factors	original sequence		first-order difference sequence	
	ADF statistic	p-value	ADF statistic	p-value
TSI	-3.791	0.194	-13.398	0.01
TV	-4.699	0.010	-13.357	0.01
GT	-5.783	0.010	-12.649	0.01
DJI	-1.857	0.639	-9.134	0.01
NA	-0.135	0.990	-9.537	0.01
SP	-1.319	0.867	-9.141	0.01
GO	-2.011	0.574	-11.060	0.01
OIL	-2.214	0.488	-8.500	0.01
FFR	-0.267	0.990	-9.049	0.01

*based on the intercept term with a trend

A correlation test was conducted to analyze the relationship of exogenous factors DJI, NA and SP stock indexes exhibit high correlation matrixes (Fig. 13). To eliminate information redundancy in highly correlated data, Principal Component Analysis (PCA) was employed to identify uncorrelated variables known as principal components (PCs), which are obtained through linear combinations of the original variables [42]. Significant items for these PCs can then be identified and the possibility of combining them into a composite variable can be considered. The first PC (PC1) is the most significant since it contains almost 32.76% of the total information of the data (Fig. 15a). The variables contributing the most to PC1 are DJI, NA, and SP with loadings above a threshold of 0.50 which used to identify significant items in the components (Fig. 15b) [43]. Consequently, a composite variable (STO) was created using PCA loading values for these indexes. PCA analysis was conducted and its result in Fig 16a shows that other than PC1 and PC2, all PCs have higher variances than previous PCA analysis (Fig15a). Assuming PCs with low variance as noise and the higher the variation in a PC, the more information that PC carries; this new dataset with composite variable improve the signal-to-noise ratio and could aid to identify the underlying pattern in the data. Multicollinearity testing was fulfilled by having VIF score below 4.0 [44] (Fig. 16b). As PCA was intended to fix multicollinearity problem rather than to reduce dimensionality, other PCs from PCA were not examined if multicollinearity testing has been satisfied. Furthermore, all correlation test values of exogenous factors are below 0.90 signifying no multicollinearity issues in the data (Fig. 17). All factors correlated with Bitcoin price with the strongest correlations existed between Bitcoin price and stock indexes (-0.33) and gold price (0.16).

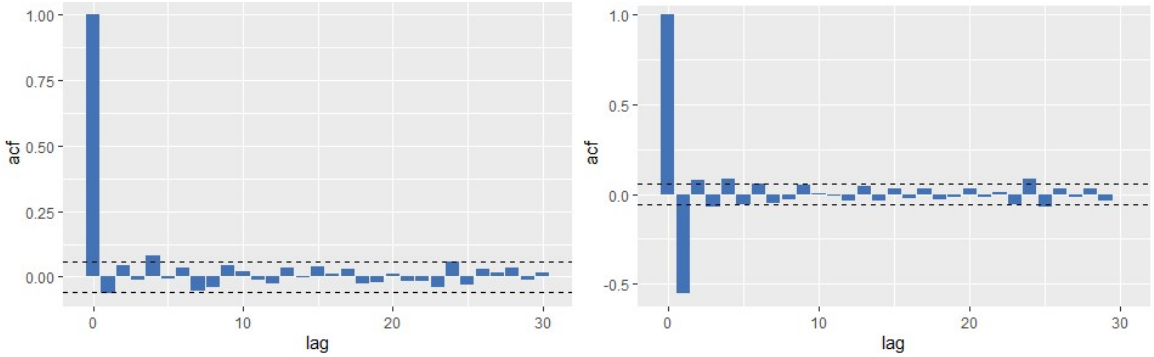

FIGURE 13. (a) first-order difference; (b) second-order difference of ACF plot

FIGURE 14.Correlation test before PCA

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3386029

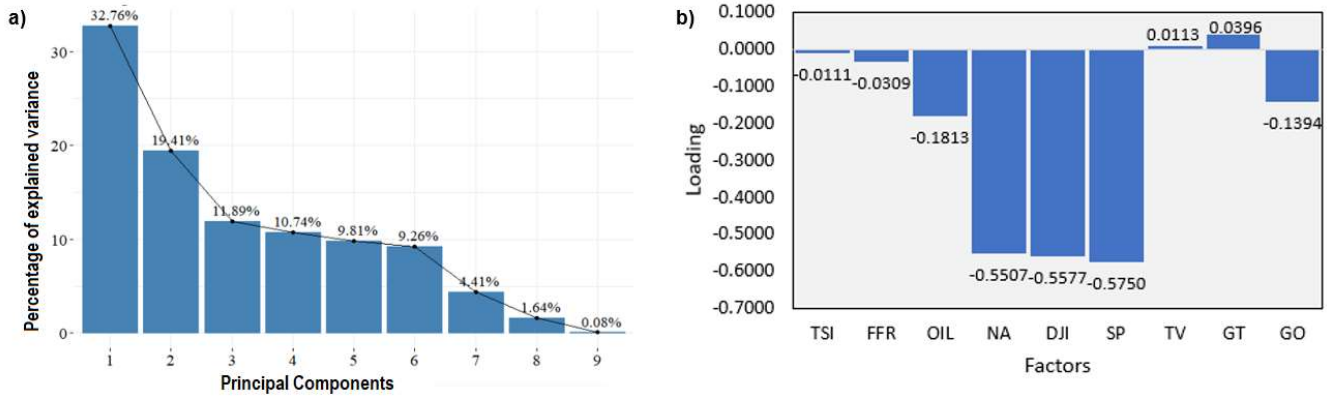

FIGURE 15. Principal component analysis (PCA) of exogeneous factors. (a) Scree plot of principal components and (b) Contribution of variables to 1st Principal Component observed.

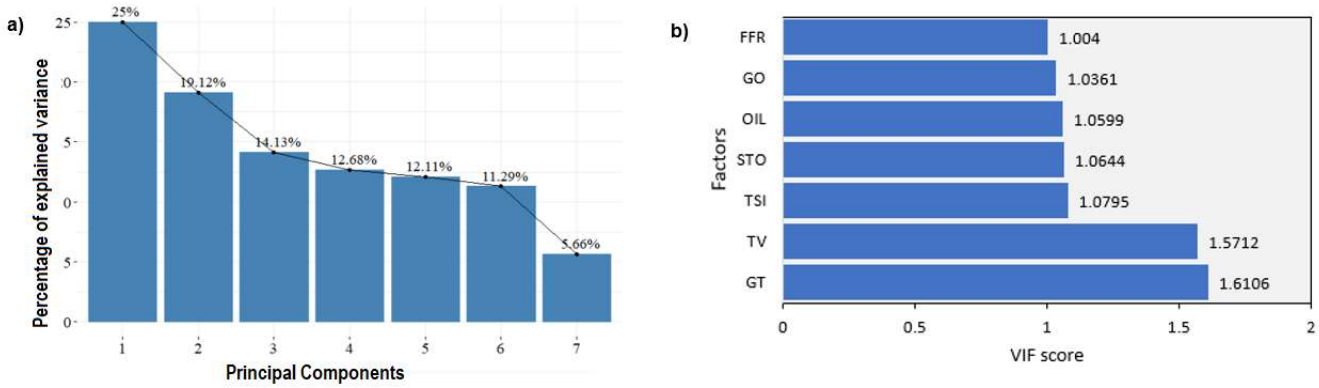

FIGURE 16. After combining factors using PCA factor loading, (a) Scree plot of principal components b) VIF scores of exogeneous factors

FIGURE 17.Correlation test after PCA

B. FITTING MODELS WITH ARIMAX GARCHX

Using a function that traverses $p \in (0,5)$ and $q \in (0,5)$, various combination of p and q orders for ARIMA and ARIMAX models were explored with ARIMAX(1,1,1) had the best performance among these models (Table IV). This result supported by ACF and PACF plots. From the ACF plot in Fig 18a, there was potential influences of the first and fourth order moving averages, while the PACF plot in Fig 18b indicated the likelihood of the first and fourth order autorepression. The Ljung-Box test on the residuals to assess the adequacy of the model was satisfied for lags 10, 15, and 20. The ARCH LM test to justify the inclusion of GARCHX in our model revealed the presence of an ARCH effect (Table V). Through various simulations to explore $p \in (0,5)$ and $q \in (0,5)$, different GARCH and GARCHX models were integrated into the ARIMAX(1,1,1) model. The result using train data indicated that ARIMAX(1,1,1) sGARCHX(1,3)

yielded the lowest AIC and BIC while the outcomes from the test data demonstrated that ARIMAX(1,1,1) sGARCHX(1,3) achieved the lowest RMSE, MAE, and MAPE (Table VI). Furthermore, confirming previous findings [42], the inclusion of exogenous factors in ARIMA and GARCH models enhances their prediction accuracy.

TABLE IV.
TENTATIVE MODELS

Models	AIC	BIC	Error Forecasting		
			RMSE	MAE	MAPE
ARIMA(1,1,1)	-5,224.27	-5174.86	2894.99	2346.00	11.44
ARIMAX(1,1,1)	-6,217.40	-6,162.15	2150.79	1924.29	9.92

Note: AIC & BIC are using training data (n= 1,096); MAE, RMSE and MAPE are using the test data (n= 184)

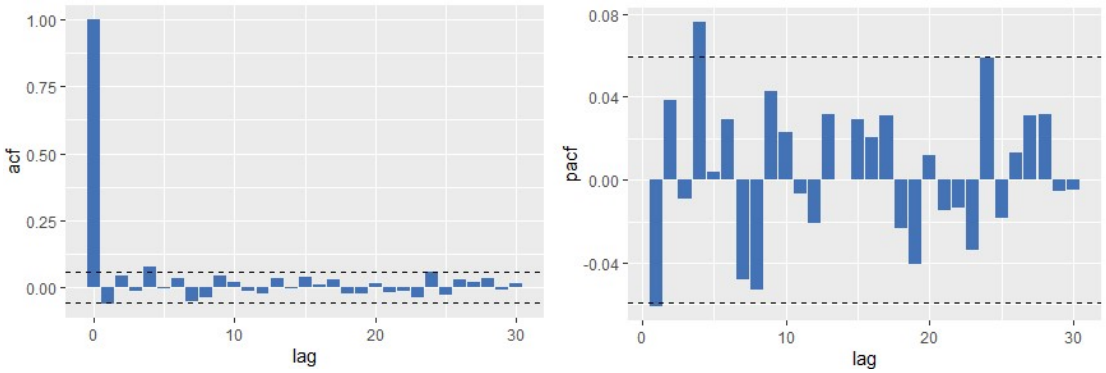

FIGURE 18.(a) Autocorrelation; (b) Partial autocorrelation plots of BTC price

TABLE V.
RESIDUAL TESTS FOR ARIMAX (1, 1, 1) MODEL

Lags	Ljung–Box Q Test		ARCH LM Test	
	Statistic	p-value	Statistic	p-value
Up to lag 10	9.80611	0.4574421	43.078	0.00000482
Up to lag 15	11.611583	0.7081592	45.904	0.00005508
Up to lag 20	16.150464	0.7072469	48.654	0.00034410

TABLE VI.
RESULT OF TENTATIVE MODELS

Models	AIC*	BIC*	Error Forecasting		
			RMSE**	MAE**	MAPE**
ARIMAX(1,1,1) sGARCH(1,1)	-5.7593	-5.7000	2202.93	2011.85	10.94
ARIMAX(1,1,1) sGARCH(1,3)	-5.7686	-5.7001	2204.96	2037.87	10.98
ARIMAX(1,1,1) sGARCHX(1,1)	-5.8208	-5.7296	2190.53	2016.68	10.61
ARIMAX(1,1,1) sGARCHX(1,3)	-5.9142	-5.8138	2129.75	1961.95	10.47

Note: AIC & BIC are using training data (n= 1,096); MAE, RMSE and MAPE are using the test data (n= 184)

C. FITTING MODELS WITH LSTM

Using the residuals from ARIMAX GARCHX model as the input layer and Bitcoin daily closing price (BTC) as the output layer, 4 proposed LSTM models was constructed. As in [45], the window size for LSTM model was selected from the lags which are above 95% confidence interval in PACF plot. Based on PACF plot of our model residuals (Fig. 19), lag value=4 was used as the window size. Following [46], our LSTM model used ReLU activation in hidden layers and Linear activation in final dense output layer while Adam was used as optimizer.

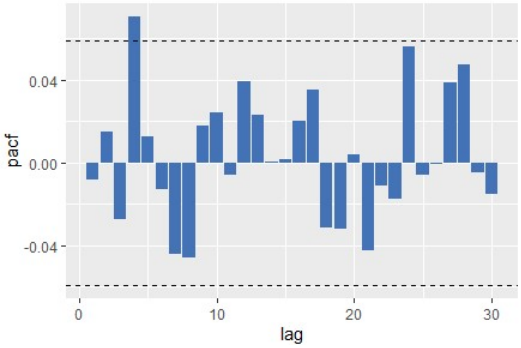

FIGURE 19. Partial autocorrelation plots of ARIMAX-GARCH residuals

To optimize LSTM hyperparameters, GWO with 80 iterations and a pack of 10 wolves was performed [47] on the training dataset. The complete details of the tuned LSTM parameters are displayed in Table VII. Compared to ARIMAX GARCHX (AG) model, the prediction performances of hybrid models, i.e. ARIMAX GARCHX LSTM (AG-LSTM); ARIMAX GARCHX BiLSTM (AG-BiLSTM); ARIMAX GARCHX ATT-LSTM (AG-AttLSTM); ARIMAX GARCHX ATT-BiLSTM (AG-AttBiLSTM), are better (Fig. 20). This signifies the ability of multivariate LSTM to extract features from nonlinear component of time series data, given the data is adequately pre-processed, thereby improving its predictive capability.

Of the models tested (Table IV, Table VI, and Fig. 20), the most accurate estimation was given by the AG-BiLSTM model with significantly lower RMSE and MAE values compared to other models.

As depicted in Fig. 21, the forecasting performance worse after the sudden drop close to day 128 of test data. Traditional econometric models, such as ARIMA, simulate and predict time series data fluctuations under the assumption that the fluctuations are linear and there is no structural break in them. This is contrary to the real data, as such the impacts of structural breaks are neglected and ARIMAX GARCHX performed worse after the impact of extreme events. It subsequently influenced LSTM performance, which used the residuals from ARIMAX GARCHX.

TABLE VII.
LSTM HYPERPARAMETERS TUNED USING GWO

Hyperparameter	Range	Optimized value			
		LSTM	BiLSTM	Att-LSTM	Att-BiLSTM
lr.rate	[0.00001, 0.1]	0.0002	0.0003	0.0003	0.0003
batch size	[1,2,4,8,12,16, 36]	36	1	1	1
nodes	[1, 128]	50	54	42	29
epochs	[35, 50]	47	30	40	48
dropout.rt	[0.1, 0.9]	0.39	0.37	0.39	0.47

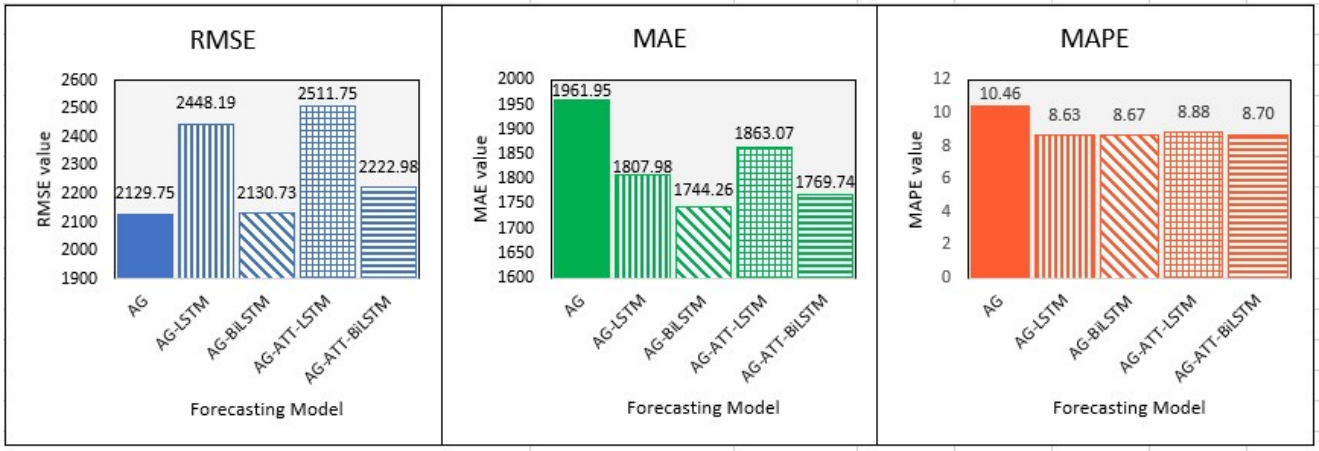

FIGURE 20. **Comparison chart results using test data**

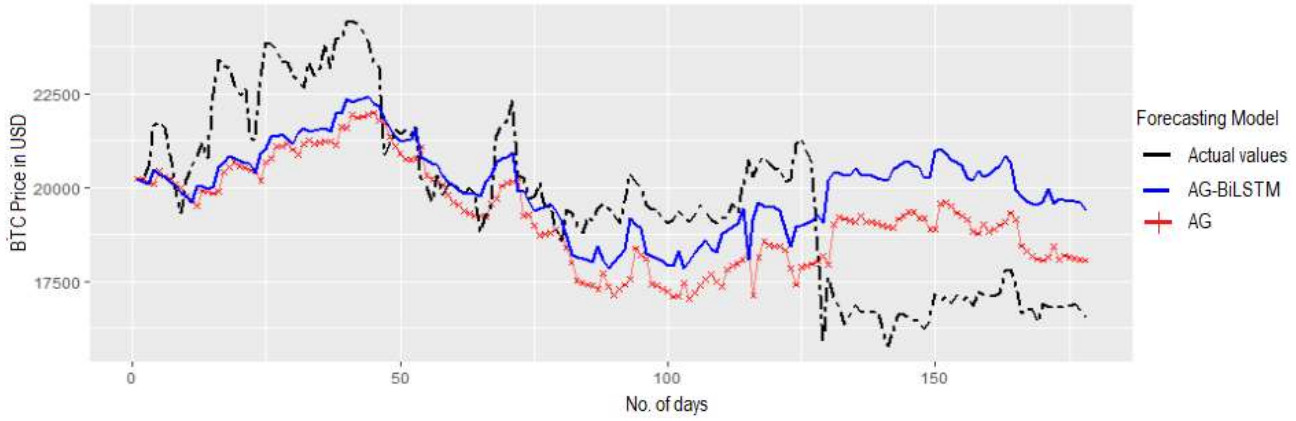

FIGURE 21. **Forecasting plot using test data**

D. MODEL EXPLANATION USING SHAP

SHAP values were calculated using ARIMAX GARCHX model as LSTM model used the residuals. Using the global explanation, the most important factor is STO, TV, GO, and TSI (Fig 22a). In contrast, and despite previous studies reporting of the effect of economic factors on Bitcoin price; GT, FFR, and OIL have minimal impact on the model output (positions 5 to 7 and mean SHAP values close to 0). These results directly informed a model modification using the removal of these factors that can be utilized in the future. A more in-depth explanation of how the specific value of each feature contributes to the model output is depicted in Fig 22b. For STO, GT and FFR, high values result in high negative SHAP values and vice versa, indicating negative relationships with Bitcoin price. High values of GO, and TSI lead to high positive SHAP values and vice versa, signifying their positive relationship with Bitcoin price. TV and OIL display no clear trend. Low values of STO and FFR resulting in high model output are understandable considering people are more likely to invest in Bitcoin if they deem other investment and savings unprofitable. However, high GO caused high SHAP values seems contradictory and needs further study.

Dependence plot of a specific factor reveals the impact a single factor has on the predictions made by the model. Because of limited space, only the four most important factors are explained. In Fig 23a, STO is taken as the primary factor and the corresponding SHAP values are given on the y-axis while GO, the secondary factor, is plotted on the color axis, capturing interaction effects between STO and GO. When STO value <0, the impact on the model estimation of Bitcoin price is positive. Inversely, when STO >0, the impact becomes negative. Therefore, stock price has a negative relationship with Bitcoin price. Furthermore, the presence of high positive gold price on low stock price values contributes to that relationship and vice versa. High Twitter volume has a positive relationship with Bitcoin price. However, it may be reversed when it has significant high positive gold price (Fig. 23b). Both gold price and twitter sentiment index' dependence plots show similar trends, indicating their positive relationships with Bitcoin price (Fig. 23c-d). The relationship of gold price and prediction outputs are influenced mostly by positive oil price values while the relationship of twitter sentiment index and prediction outputs are influenced mostly by negative twitter volume values.

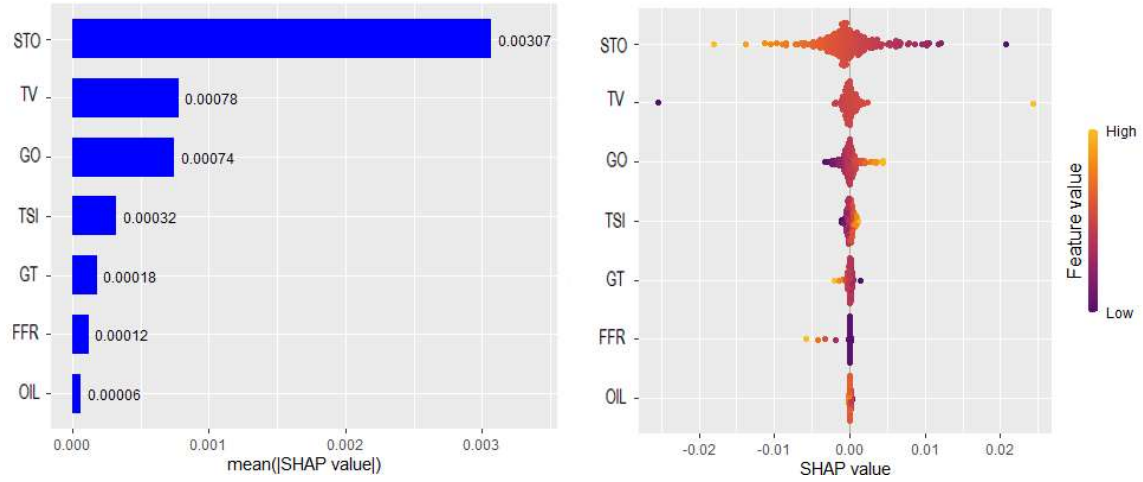

FIGURE 22. SHAP analysis of model data. (a) Overall importance of each feature on the model output, and (b) Influence of each feature value on the model output.

V. CONCLUSION

This paper demonstrates a successful attempt to predict the closing price of Bitcoin by incorporating exogenous factors. Hybrid classical time series models (ARIMAX, GARCHX) and deep learning (LSTM variants) models were employed, with the hybrid ARIMAX GARCHX BiLSTM model exhibiting the best performance. Future research to strengthen the validity of our best selected model could explore other best model selection methods, such as Model Confidence Set procedure. In-depth exploration of the performance of our model would also be very helpful. Further research might compare forecasts from tentative models using Diebold-Mariano test.

We acknowledge that this model is not comprehensive as it does not consider other potential exogenous factors that could influence Bitcoin price. Therefore, future studies could enhance the model's performance by incorporating additional factors identified in previous studies. Nonetheless, our systematic data pre-processing procedures for addressing non-stationarity, the simplicity of our prediction approach, and the successful tuning of LSTM hyperparameters are notable advantages that yield close predicted values in non-sudden unexpected event.

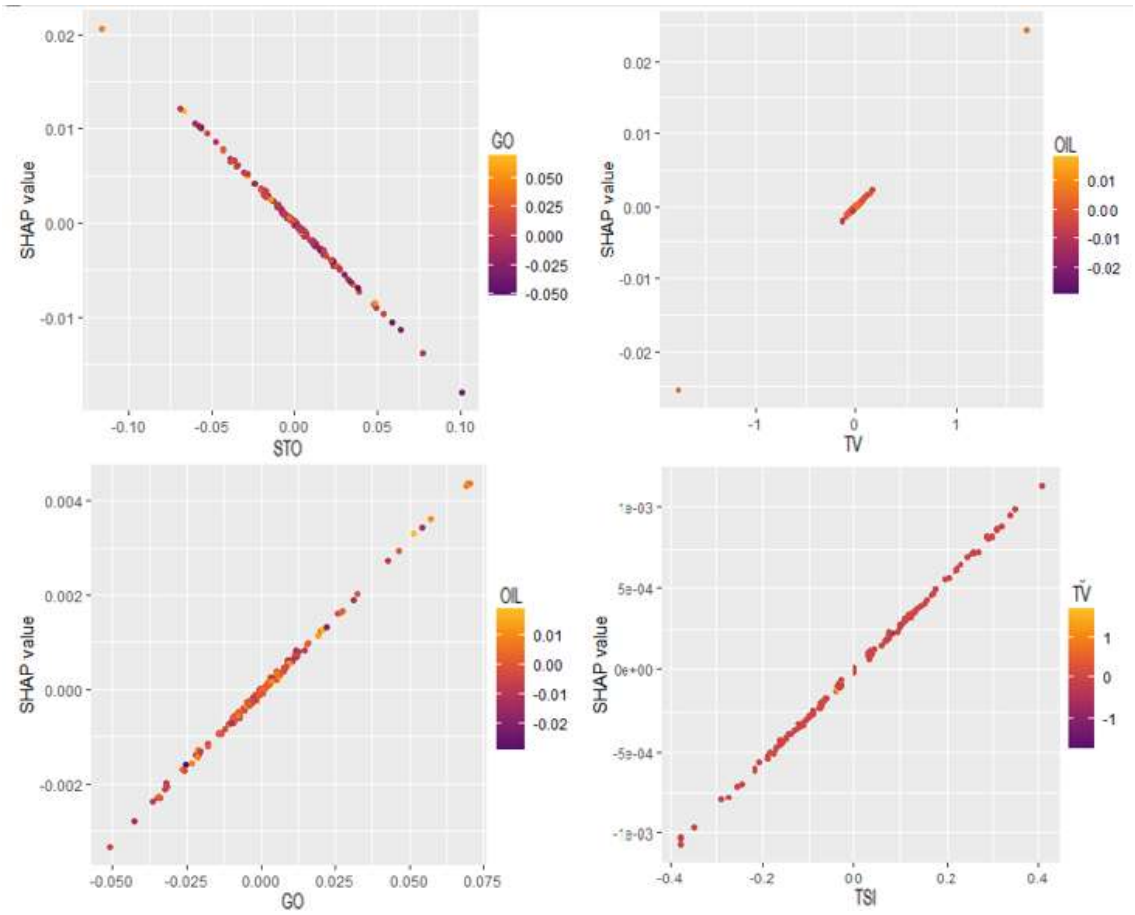

FIGURE 23. Dependence plot(a) Stock price, (b) Twitter volume, (c) Gold price, and (d) Twitter sentiment index.

The weakness of our model after structural break event gives the direction for future studies to identify structural break in time series data using techniques, such as the Iterative Cumulative Sum of Squares algorithm (ICSS) and the Chow tests, and incorporate structural break in forecasting model to improve forecasting accuracy. A recent study highlights the growing interest among investors in alternative cryptocurrencies such as Ethereum, Ripple, Litecoin, Stellar, and Dash (Ji et al., 2019). Hence, it is important to develop a robust model that is not specific to a single cryptocurrency and further research can explore the application of this systematic process and algorithm for modelling other cryptocurrencies. Additionally, since GWO is effective in tuning LSTM hyperparameters, it would be worthwhile to investigate its potential for tuning other LSTM hyperparameters (number of layers, window size, optimizer, activation function, and loss function). Lastly, to aids the understanding of how exogenous factors influence model predictions, exploring alternative factor importance techniques, such as using DALEX and LIME, can help shed light on the implications of these factors.

APPENDIX

TABLE VIII
ACRONYMS

Symbol	Description
ACF	AutoCorrelation Function
ADF	Augmented Dickey–Fuller
ARIMA	Auto-Regressive Integrated Moving Average
AIC	Akaike Information Criterion
ARMA	Autoregressive Moving Average
BIC	Bayesian Information Criterion
BiLSTM	Bidirectional LSTM
BTC	Bitcoin
DJI	Dow Jones Index
FFR	Federal Funds Rate
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GT	Google Trends
GWO	Grey Wolf Optimization
LSTM	Long Short-Term Memory
NA	NASDAQ Index
PACF	Partial AutoCorrelation Function
R	R is a programming language and open-source environment that is utilized for traditional statistical, machine learning, and deep learning computing and graphics.
SHAP	Shapley Additive exPlanations

REFERENCES

[1] A. A. Abdelhamid *et al.*, "Robust Speech Emotion Recognition Using CNN+LSTM Based on Stochastic Fractal Search Optimization Algorithm," *IEEE Access*, vol. 10, pp. 49265–49284, 2022, doi: 10.1109/ACCESS.2022.3172954.

[2] W. Chen, H. Xu, L. Jia, and Y. Gao, "Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants," *International Journal of Forecasting*, vol. 37, no. 1, pp. 28–43, 2021, doi: 10.1016/j.ijforecast.2020.02.008.

[3] A. Dutta, S. Kumar, and M. Basu, "A Gated Recurrent Unit Approach to Bitcoin Price Prediction," *JOURNAL OF RISK AND FINANCIAL MANAGEMENT*, vol. 13, no. 2. MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND, Feb. 2020. doi: 10.3390/jrfm13020023.

[4] S. Hansun, A. Wicaksana, and A. Q. M. Khaliq, "Multivariate cryptocurrency prediction: comparative analysis of three recurrent neural networks approaches," *Journal of Big Data*, vol. 9, no. 1, p. 50, Apr. 2022, doi: 10.1186/s40537-022-00601-7.

[5] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," *Entropy*, vol. 21, no. 6, 2019, doi: 10.3390/e21060589.

[6] H. Jang and J. Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information," *IEEE Access*, vol. 6, pp. 5427–5437, 2017, doi: 10.1109/ACCESS.2017.2779181.

[7] X. Li and C. A. Wang, "The technology and economic determinants of cryptocurrency exchange rates," *Decision Support Systems*, vol. 95, pp. 49–60, Mar. 2017, doi: 10.1016/j.dss.2016.12.001.

[8] O. Poyser, "Exploring the determinants of Bitcoin's price: an application of Bayesian Structural Time Series," 2017.

[9] S. Palamalai, B. Maity, and K. Kumar, "Macro-Financial Parameters Influencing Bitcoin Prices: Evidence from Symmetric and Asymmetric ARDL Models," *Review of Economic Analysis*, vol. 13, Oct. 2021, doi: 10.15353/rea.v13i3.3585.

[10] T. Panagiotidis, T. Stengos, and O. Vravosinos, "On the determinants of bitcoin returns: A LASSO approach," *Finance Research Letters*, vol. 27, no. C, pp. 235–240, 2018, doi: 10.1016/j.frl.2018.03.016.

[11] T. Panagiotidis, T. Stengos, and O. Vravosinos, "The effects of markets, uncertainty and search intensity on bitcoin returns," *International Review of Financial Analysis*, vol. 63, pp. 220–242, May 2019, doi: 10.1016/j.irfa.2018.11.002.

[12] S. Dastgir, E. Demir, G. Downing, G. Gozgor, and C. K. M. Lau, "The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger causality test," *Finance Research Letters*, vol. 28, pp. 160–164, Mar. 2019, doi: 10.1016/j.frl.2018.04.019.

[13] D. Shen, A. Urquhart, and P. Wang, "Does twitter predict Bitcoin?," *Economics Letters*, vol. 174, Nov. 2018, doi: 10.1016/j.econlet.2018.11.007.

[14] J. Abraham, D. W. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis," 2018.

[15] Y. B. Kim *et al.*, "Predicting fluctuations in cryptocurrency transactions based on user comments and replies," *PloS one*, vol. 11, no. 8, p. e0161197, 2016.

[16] E. Bouri, P. Molnár, G. Azzi, D. Roubaud, and L. I. Hagfors, "On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?," *Finance Research Letters*, vol. 20, no. C, pp. 192–198, 2017.

[17] Y. Zhu, D. Dickinson, and J. Li, "Analysis on the influence factors of Bitcoin's price based on VEC model," *Financial Innovation*, vol. 3, Dec. 2017, doi: 10.1186/s40854-017-0054-0.

[18] S. Pyo and J. Lee, "Do FOMC and macroeconomic announcements affect Bitcoin prices?," *Finance Research Letters*, vol. 37, p. 101386, Nov. 2020, doi: 10.1016/j.frl.2019.101386.

[19] J. Wang, F. Ma, E. Bouri, and Y. Guo, "Which factors drive Bitcoin volatility: Macroeconomic, technical, or both?," *Journal of Forecasting*, Nov. 2022, doi: 10.1002/for.2930.

[20] P. Zhang, "Zhang, G.P.: Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing 50, 159-175," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003, doi: 10.1016/S0925-2312(01)00702-0.

[21] R. J. Brenner, R. H. Harjes, and K. F. Kroner, "Another Look at Models of the Short-Term Interest Rate," *Journal of Financial and Quantitative Analysis*, vol. 31, no. 1, pp. 85–107, 1996.

[22] M. Yeasin, K. Singh, A. Lama, and R. Paul, "Modelling Volatility Influenced by Exogenous Factors using an Improved GARCH-X Model," vol. 74, pp. 209–216, Dec. 2020.

[23] M. M. Patel, S. Tanwar, R. Gupta, and N. Kumar, "A Deep Learning-based Cryptocurrency Price Prediction Scheme for Financial Institutions," *Journal of Information Security and Applications*, vol. 55, p. 102583, Dec. 2020, doi: 10.1016/j.jisa.2020.102583.

[24] L. Yu, S. Liang, R. Chen, and K. K. Lai, "Predicting monthly biofuel production using a hybrid ensemble forecasting methodology," *International Journal of Forecasting*, vol. 38, no. 1, pp. 3–20, Jan. 2022, doi: 10.1016/j.ijforecast.2019.08.014.

[25] A. N. Huynh, R. C. Deo, D.-A. An-Vo, M. Ali, N. Raj, and S. Abdulla, "Near Real-Time Global Solar Radiation Forecasting at Multiple Time-Step Horizons Using the Long Short-Term Memory Network," *Energies*, vol. 13, no. 14, 2020, doi: 10.3390/en13143517.

[26] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997, doi: 10.1109/78.650093.

[27] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate." 2016.

[28] J. S. Bridle, "Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition," in *Neurocomputing*, F. F. Soulié and J. Hérault, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 227–236.

[29] A. A. Sekh, D. P. Dogra, S. Kar, P. P. Roy, and D. K. Prasad, "ELM-HTM guided bio-inspired unsupervised learning for anomalous trajectory classification," *Cognitive Systems Research*, vol. 63, pp. 30–41, Oct. 2020, doi: 10.1016/j.cogsys.2020.04.003.

[30] A. Mardjo and C. Choksuchat, "HyVADRF: Hybrid VADER–Random Forest and GWO for Bitcoin Tweet Sentiment Analysis," *IEEE Access*, vol. 10, pp. 101889–101897, 2022, doi: 10.1109/ACCESS.2022.3209662.

[31] J. Hu, T. Zhou, S. Ma, D. Yang, M. Guo, and P. Huang, "Rock mass classification prediction model using heuristic algorithms and support vector machines: a case study of Chambishi copper mine," *Scientific Reports*, vol. 12, no. 1, p. 928, Jan. 2022, doi: 10.1038/s41598-022-05027-y.

[32] J. Batra, R. Jain, V. A. Tikkiwal, and A. Chakraborty, "A comprehensive study of spam detection in e-mails using bio-inspired optimization techniques," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100006, Apr. 2021, doi: 10.1016/j.jjimei.2020.100006.

[33] J.-S. Pan, P. Hu, and S.-C. Chu, "Novel Parallel Heterogeneous Meta-Heuristic and Its Communication Strategies for the Prediction of Wind Power," *Processes*, vol. 7, no. 11, 2019, doi: 10.3390/pr7110845.

[34] R. I. Hamilton and P. N. Papadopoulos, "Using SHAP Values and Machine Learning to Understand Trends in the Transient Stability Limit." 2023.

[35] J. Chen, "Analysis of Bitcoin Price Prediction Using Machine Learning," *Journal of Risk and Financial Management*, vol. 16, p. 51, Jan. 2023, doi: 10.3390/jrfm16010051.

[36] H. Mao, S. Counts, and J. Bollen, "Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data." 2011.

[37] N. Mohamed Noor, M. M. A. B. Abdullah, A. S. Yahaya, and N. Ramli, "Comparison of Linear Interpolation Method and Mean Method to Replace the Missing Values in Environmental Data Set," *Materials Science Forum*, vol. 803, pp. 278–281, Aug. 2014, doi: 10.4028/www.scientific.net/MSF.803.278.

[38] I. E. Livieris, N. Kiriakidou, S. Stavroyiannis, and P. Pintelas, "An advanced CNN-LSTM model for cryptocurrency forecasting,"

Electronics (Switzerland), vol. 10, no. 3, pp. 1–16, 2021, doi: 10.3390/electronics10030287.

[39] I. E. Livieris, S. Stavroyiannis, E. Pintelas, and P. Pintelas, "A novel validation framework to enhance deep learning models in time-series forecasting," *Neural Computing and Applications*, vol. 32, pp. 17149–17167, 2020.

[40] L. Luo and C. Qi, "An analysis of the crucial indicators impacting the risk of terrorist attacks: A predictive perspective," *Safety Science*, vol. 144, p. 105442, Dec. 2021, doi: 10.1016/j.ssci.2021.105442.

[41] M. C. Chang and D. A. Dickey, "RECOGNIZING OVERDIFFERENCED TIME SERIES," *Journal of Time Series Analysis*, vol. 15, no. 1, pp. 1–18, Jan. 1994, doi: 10.1111/j.1467-9892.1994.tb00173.x.

[42] M. Waqar, H. Dawood, P. Guo, M. B. Shahnawaz, and M. A. Ghazanfar, "Prediction of Stock Market by Principal Component Analysis," *2017 13th International Conference on Computational Intelligence and Security (CIS)*, pp. 599–602, 2017.

[43] E. Nzayisenga and Y. Zhu, "The Import Trade Forecasting Model Based on PCA: Evidence from Rwanda," *Open Journal of Statistics*, vol. 10, pp. 678–693, Jan. 2020, doi: 10.4236/ojs.2020.104042.

[44] J. F. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd Edition. Thousand Oaks, CA: Sage Publications Inc., 2017.

[45] T. Okedi and A. Fisher, "Time series analysis and Long Short-Term Memory (LSTM) network prediction of BPV current density," *Energy & Environmental Science*, vol. 14, Apr. 2021, doi: 10.1039/D0EE02970J.

[46] M. Roondiwala, H. Patel, and S. Varma, "Predicting Stock Prices Using LSTM," *International Journal of Science and Research (IJSR)*, vol. 6, Apr. 2017, doi: 10.21275/ART20172755.

[47] M. M. Eid *et al.*, "Meta-Heuristic Optimization of LSTM-Based Deep Network for Boosting the Prediction of Monkeypox Cases," *Mathematics*, vol. 10, no. 20, 2022, doi: 10.3390/math10203845.

ANNY MARDJO is a doctoral student in the College of Digital Science at the Prince of Songkla University (Hat Yai campus), Thailand. Presently, she is working as a Lecturer in the Faculty of Commerce and Management at Prince of Songkla University (Trang Campus), Thailand. Her research activities are in the field of internet-based applications and data analysis. She holds a Master's degree in Information Technology from Swinburne University, Melbourne, Australia, and a Bachelor degree in Business Administration from RMIT University, Melbourne, Australia.

CHIDCHANOK CHOKSUCHAT received her Ph.D. in computer and information science from the Silpakorn University. She is currently an assistant professor in the Division of Computational Science with the Faculty of Science, Prince of Songkla University (Hat Yai campus), Thailand. Her research interests include parallel processing, ontology engineering, web services, linked open data, data science toolkits, and the Internet of Things. She is also currently with the Office of Digital Innovation and Intelligent Systems at the Prince of Songkla University, Thailand.