Received: November 22, 2024. Revised: December 17, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Predictive Analytics System for Time Series Stock Data Using LSTM with Residual Unit and Attention Mechanism

Khin Nyein Myint[1]* **Myo Khaing[1]**

[1]*University of Computer Studies, Mandalay, Myanmar*
* Corresponding author's Email: khinnyeinmyint@ucsm.edu.mm

Abstract: Stock price prediction is a challenging research topic because of non-linearity, significant noise and volatility of time series data. Deep learning techniques enable to learn complex and non-linear patterns of sequential time series data. Long Short-Term Memory (LSTM) is a technique which is designed to handle time series data. While LSTM model is used to extract temporal dependencies of stock data, the performance can be limited by noisy data and the challenge of capturing intricate patterns. In this research, LSTM-based framework with residual unit and attention mechanism is proposed to enhance the temporal dependencies and important features of stock price movements. Residual unit with skip connection captures more complex patterns and representations in stock price data and reduces the over-fitting problem to noisy time series data. LSTM with attention focuses on the significant time stamps which enhances the model prediction performance. The proposed system is experimented on five datasets: Apple (AAPL), Bitcoin, Ethereum, Litecoin and GOLD_PRICE. To prove the effectiveness of the model, the proposed system is compared with LSTM and Bidirectional LSTM (Bi-LSTM) models. Experimental results show that the proposed system outperforms baseline models such as LSTM, Bi-LSTM, LSTM+Bi-LSTM and state-of-the-art methods in term of error rates such as mean square error, root mean square error and mean absolute error.

Keywords: Time series data, Stock price prediction, LSTM, Attention mechanism, Residual unit.

1. Introduction

Predictive analytics is a pillar of science for forecasting future trends in time series data. Time series data forecasting is a process of predicting the future values or trends by using collected historical data overtime. A large amount of data often contains unreliable and repetitive information that cannot be used directly for predictive analysis. Stock price data are generally both non-stationary and non-linear. Understanding both characteristics are essential for selecting the right models and techniques for analysis and prediction. In previous stock price forecasting works, the prediction methods are mainly focused on statistical indicators and traditional neural network. For forecasting stock price, the research [1] proposed univariate singular spectral analysis using Hadamard transform. This method facilitates to choose window length in time and improves the performance of standard singular spectrum analysis method. A technical indicator is proposed in the paper [2] based on non-linear support vector machine prediction model. This method is experimented on real-time National Stock Exchange (NSE) stock market data on daily and weekly basis. The presented framework achieved high accuracy as compared to the other frameworks. The technique requires historical stock data and social media comments. A stock price prediction system [3] is constructed to predict next day stock market using weighted moving average, min-max normalization, and box-cox transformation techniques, multilayer perceptron (MLP). The results demonstrate that MLP is better for time series forecasting analysis. Deep convolutional neural network has a strong ability to learn deep features and is suitable for complex non-linear time series stock data. A deep learning prediction model that uses LSTM with residual unit and attention mechanism is applied to achieve reliable prediction results. LSTM network is used to make predictions for sequence of

Received: November 22, 2024. Revised: December 17, 2024.

data with complex and non-linear structure. It is effective in stock price prediction because of the ability to capture both short-term and long-term dependencies. The research [4] explored deep-learning models for NSE stock market prediction system. The authors proved that convolutional neural network outperformed than existing linear model such as autoregressive integrated moving average (ARIMA). Time series forecasting system in [5] compares the performances of LSTM and Bi-LSTM models. This paper observed that Bi-LSTM model is better than ARIMA and LSTM models for time series data prediction. To extract the effective feature for stock price forecasting, a model is designed by applying Principal Component Analysis (PCA) and LSTM network structure [6]. In the model, PCA is used to extract principal components of technical indicators related to stock prices and LSTM is used to predict the prices. Attention-based deep convolutional neural network [7] is designed using efficient channel attention (ECA) module in squeeze and excitation network (SENet). The results shown that the ECA-Net improves the prediction result and decrease model complexity. Attention mechanism provides the most significant past data points that leads to more accurate predictions. LSTM network can improve the performance during model training but reduces the stability of network in testing. Therefore, the performance of the model varies one training by others. The advantages of LSTM network and attention mechanism are combined to improve the training and testing performance for stock price prediction system. The aim of the paper is to develop a time series stock price prediction system using LSTM model with residual unit and attention mechanism. The objectives are to extract crucial temporal features from the time series stock data and to reduce error rates for training and testing of prediction on both univariate and multivariate time series stock data.

The contributions of the paper are summarized in the followings:

- An enhanced LSTM model is developed for stock price prediction by combining residual unit, attention mechanism and LSTM architecture.
- Residual unit and LSTM are used as prior feature extractors to extract most important local and temporal features.
- The features are enhanced by using attention mechanism and Bi-LSTM network. Soft attention learns and assigns dynamically higher attention weights to key time stamp. Bi-LSTM architecture is used to capture temporal patterns from past and future stock prices at each time step.
- The model is tested on both univariate and multivariate time series stock data. The LSTM model with attention extracts the important features and reduce error rate on univariate and multivariate time series stock data.

The rest of the paper is composed of four sections. Section 2 discusses the previous related works of time series forecasting. Section 3 describes the techniques of the proposed time series prediction system. In section 4, experimental results are shown and discuss the detail of network testing. The final section describes the conclusion of this prediction system.

2. Literature review

Several techniques for stock price prediction have been developed by scholars. In this paper, a review of existing works is discussed for time series stock market data analysis.

The research paper [8] presented predictive analytics system to make decision buying or selling trend of stock using Moving Average (MA) technical indicator and LSTM network. The authors proposed an prediction system using ARIMA and LSTM network in [9]. These studies are the fundamental concepts for scholars to analysis of time series stock data movement.

The three deep learning approaches such as MLP, LSTM and gated recurrent unit (GRU) models are compared to analyse the performance in making stock price prediction. The evaluation results show that LSTM exhibited the best predictive ability due to capturing patterns for long-term dependencies. LSTM model can be selected for forecasting time series stock data. The research work tested on only one stock dataset AAPL with one variable.

The stock price predictive model is developed in the research [10] by using LSTM and recurrent neural network (RNN) to make appropriate decisions for investors. The important features are selected using mutual information (MI) feature selection approach. MI calculates the value of cross entropy loss to minimize the optimization. The proposed model handles better than traditional time series models and reduces the error computational cost for multivariable testing. The model used a number of stock movement features that tend to feature redundancy and gradient vanishing problem.

By combining the different deep learning networks, the model can predict the future stock prices with accurate results. The research of stock market analysis [11] presented experimental results of RNN, LSTM and three variants of GRU on the NIFTY 50 index (TA1 and TA2). The evaluation shown that GRU model provides the lowest error rate compared to the RNN and LSTM model with technical indicator TA1. The model is tested on only one variable such as closing prices.

The research work in [12] developed a multivariate time series data prediction model based on MLP, feed forward attention mechanism and LSTM network. LSTM is insufficient to extract various degrees of attention of stock data. Therefore, MLP is used to map the multivariate initial sequences into another latent dimensional space. It can adjust the feature space for multiple variables. The feed forward attention mechanism assigns the weights on the feature mapping. LSTM network is used to make final predictions. The combination of three approaches effects in prediction for multivariate time series data.

An enhanced LSTM model [13] is designed for stock price forecasting. A residual unit is integrated in the existing LSTM network. The residual unit is employed as a feature extractor to identify the dependencies between stock prices. The residual features are fed into LSTM model to predict future stock prices. It performs better than LSTM, CNN, GRU and Bi-LSTM models. This research work achieves the mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE) of 2762.82, 37.34, 52.56 on CSI 300 Index dataset and 1350.16, 28.08, and 36.74 on SSE Composite Index dataset respectively. The model is overfit that it is not stable in training progress on stock prices datasets. It is tested on only univariate feature. The model decreases performance of prediction on multiple variables evaluation.

The research paper [14] explored the important of Bi-directional feature extraction for predicting the future stock price. The prediction system [15] analysed four network architectures such as RNN, LSTM, GRU and Bi-LSTM on stock datasets. The study found that Bi-LSTM model is more efficient in stock price prediction than other three models. The paper is only tested on existing frameworks. The performance of the network structure is not stable on training and testing data.

The attention based LSTM network [16] is developed for forecasting multi-variable time series data. This research used mixture attention mechanism to enhance the network training and prediction. The network is tested on three stock datasets Apple (AAPL), Amazon (AMZN), and Microsoft (MSFT). The results shown that the network with attention improved prediction accuracy for multivariate stock data. The research paper [17] explored a stock price prediction model by combining LSTM network and neural prophet (NP) with regressors. This architecture is tested on MCX dataset. It improved the prediction results with less errors.

A stock price prediction model is designed based on Bi-GRU network and attention mechanism [18]. This paper tested LSTM network, GRU network, Bi-LSTM network and bidirectional GRU (Bi-GRU) network with attention on BYD company dataset. Bi-GRU model reduces the complexity of model and attention allocated the corresponding weights for different time steps. It mainly effected on the feature extraction of stock price prediction and improved the prediction speed and model stability.

The deep learning model for gold price prediction [19] is developed using convolutional network and bi-directional LSTM networks with gird search parameter tuning algorithm. This research tested four networks stacked LSTM, convolutional network (CNN), CNN-LSTM and CNN-Bi-LSTM. For training the prediction model, grid search method is applied to tune hyperparameters. The results demonstrated that CNN-Bi-LSTM is better model than other three models. This model obtains 27.77 of MAE and 37.94 of RMSE error value on GOLD-PRICE dataset. It is only tested on one variable and the search algorithm tend to increase time complexity. The paper [20] presented a deep learning model based on LSTM network with Bayesian network optimization for financial marketing gold price prediction system. The research work in the paper [21] also used deep LSTM network with artificial rabbits optimization algorithm. The optimization algorithms improved the model training and predicted more accurately time series market.

A time series prediction model is constructed based on multi-scale residual network with encoder-decoder. This study demonstrates the effectiveness of using multiscale convolutions. The presented framework reduced the loss of data. It is tested on different time series data. The network structure tends to information redundancy and too many parameters [22]. For forecasting cryptocurrency prices, the research [23] conducted and compared LSTM, GRU, and Bi-LSTM model. These models are tested on Bitcoin, Litecoin, and Ethereum datasets. Evaluation results illustrated that Bi-LSTM model performs better other models. It attains the error value 1029.36, 83.59, and 8.02 in term of RMSE. This work only experimented on pre-trained models and overfitted on testing data.

The research work in [24] explored the performance analysis of different deep learning models such as MLP, LSTM, and GRU network. This study compared the results on AAPL dataset and discussed the findings that LSTM model performed with the best prediction capability. The experimental

Received: November 22, 2024. Revised: December 17, 2024.

results show that LSTM network achieves error rate such as 11.70 of MSE, 37.96 of MAE and 6.16 of RMSE on AAPL dataset. Therefore, LSTM can be chosen for stock price prediction task. In the stock price forecasting system [25], a convolutional attention Bi-LSTM network is developed to extract the important features and key information from time series stock data. Efficient channel attention is embedded into Bi-LSTM network to improve the network ability. This network effectively predicts for time series stock price data. This framework is tested on two datasets. The error values are achieved 1956.03, 28.34, 44.22 on SSE Shanghi Composite Index and 0.028, 0.10 and 0.162 on China Unicom datasets in term of MSE, MAE and RMSE respectively.

3. Methodology

In the research, an attention-based LSTM model is developed to predict time series stock price. The proposed model consists of a residual unit, an LSTM network, an attention mechanism and a Bi-LSTM network. The overall architecture of the proposed network is illustrated in Figure. 1. The first two sub-networks are applied to extract crucial temporal features from scaled stock dataset. The output features from LSTM are enhanced by using attention unit and Bi-LSTM network to improve the predictive result of time series data. The model captures both past and future information learns important weights and improves the model training.

3.1 Preprocessing

Preprocessing is a vital step for prediction of time series stock data which include noise, missing values, and various features. Data normalization is the process of scaling individual samples to have unit norm and changes all features to same scale. It scales the data to fall within a small, specified range of data. It is a critical process to improve the predictive accuracy of deep learning model. Time series data have characteristics such as high dimensionality, excessive noise, data imbalance, etc. Feature selection and transformation reduce the irrelevant variables, computational cost, and error rates and improve the model performance.

Stock prices are both linear and non-linear patterns which can change overtime and tend to different relationships between features and target variables. In the study, mutual information (MI) is used for feature selection for ranking variables in stock dataset. The greater the MI scores are considered as the stronger correlation. The MI scores are calculated as in Eq. (1).

$$I(X;Y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} p(x,y) log\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (1)$$

where,

*I(X;Y)*is the mutual information of *X* and *Y*,

p(x,y) is the joint probability distribution of *X* and *Y*,

p(x) is the marginal probability distribution of *X*,

p(y) is the marginal probability distribution of *Y*,

The logarithm is usually taken in base 2(if measuring MI in bits).

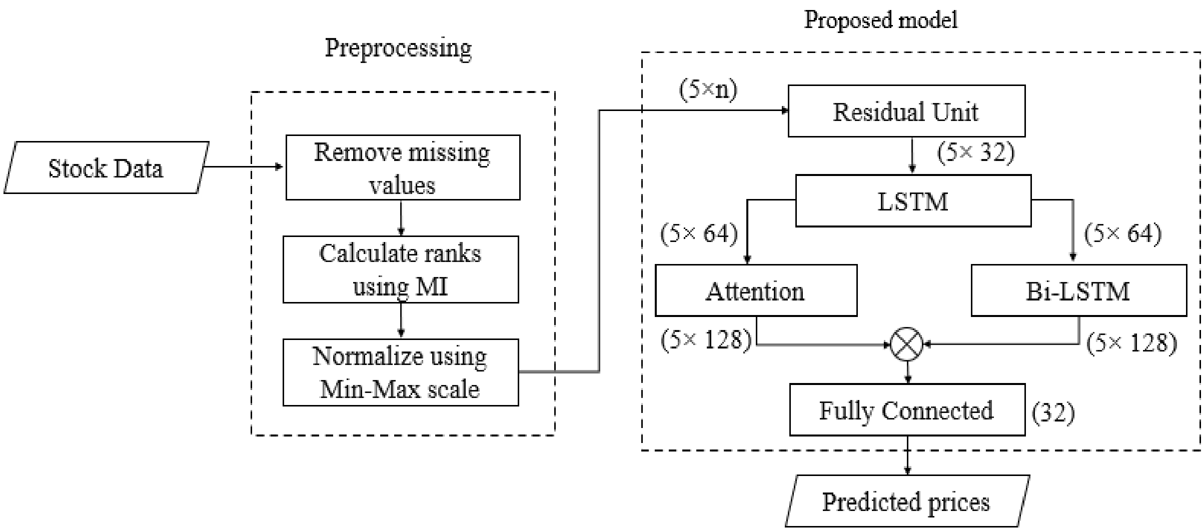

Figure. 1 Overall architecture of the Proposed System

Data normalization is an essential step in the pre-processing of data for machine learning models, and it is a feature scaling technique. It is the process of scaling individual samples to have unit norm and changes all features to same scale. Min-max normalization performs a linear transformation on the original data and gets all the scaled data in the range as in Eq. (2).

$$v' = \frac{v - min_F}{max_F - min_F} \tag{2}$$

where,

v is the original value of feature F

v' is the normalized value in the range [0, 1]

min_F is the minimum value of the feature

max_F is the maximum value of the feature

3.2 Network Layer

In developing deep neural network, layers are the building blocks that process data, transforming input with a series of computations to output predictions.

3.2.1 Convolutional layer

In the system, convolution operation is used to extract patterns from historical stock data. It slides a filter or kernel over the time steps of the stock data. Convolution layer in attention unit enhances the feature weights to detect the important pattern. For univariate testing, one dimensional convolutional layer is used to predict the stock prices as shown in Eq. (3). To test stock data with multiple variables, Eq. (4) is used to extract convolutional features for the prediction system.

$$Conv1D(X) = \sum_{i=1}^{n}(X_i \times W_i) + b_i \tag{3}$$

$$Conv2D(X) = \sum_{j=1}^{c}\sum_{i=1}^{n}\left(X_{ij} \times W_{ij}\right) + b_{ij} \tag{4}$$

where,

X is input feature map.

W is filter weights

n is number of time stamp

c is number of dimensions

b is bias

3.2.2 Batch normalization (BN) layer

BN layer normalizes each input channel using the mean and standard deviation of the current batch of inputs. This layer reduces the sensitivity and speed up on network training. The BN layer normalizes the features by subtraction the minibatch mean and dividing the minibatch standard deviation as shown in Eq. (5), (6), (7) and (8).

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{5}$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_{i} - \mu_B)^2 \tag{6}$$

$$\bar{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{7}$$

$$y_i = \gamma \bar{x}_i + \beta \tag{8}$$

where,

γ and β are scaling and shifting factors

x_i denotes the input pixels

m is the batch size

μ_B and σ_B^2 are mean and variance of the batch

3.2.3 Rectified Linear Unit (ReLU) activation layer

ReLU activation layer is an activation function that can apply an element-wise function which gives an output x if x is positive and zero. The network model used this function shown in Eq. (9) because it achieves high performance and easier to train (accelerate the learning).

$$A(x) = \max(0, x) \tag{9}$$

3.3 Residual Unit

Residual unit with skip connection reduces vanishing gradient problem and improves the training progress of stock price prediction. It learns the features by combining input and residual output mapping instead of direct transformations. It improves the model ability to capture complex hierarchical features in the data. The residual features can be extracted using Eq. (10).

$$Res(X) = ReLU(ReLU\big(BN\big(Conv(Conv(X))\big)\big) + Conv(X)) \tag{10}$$

where,

X is input feature map.

$Conv$ is Convolution operation.

BN is Batch Normalization

$ReLU$ is Rectified linear unit activation function

A residual unit is used as prior feature instructor to extract local patterns in time series stock data. It is more stable for deep network training. The residual

Received: November 22, 2024. Revised: December 17, 2024.

features can be extracted with the following Figure. 2.

```
def Residual_Unit (ip, f)
   Conv1D with ip and f/2 are used to get feature RF^1 ∈ R^(f/2×n)
   Conv1D with RF^1 and f are used to get feature RF^2 ∈ R^(f×n)
   BN with RF^2 to get adjust feature RF^3 ∈ R^(f×n)
   ReLU with RF^3 to substitute all negatives with zero to get feature RF^4 ∈ R^(f×n)
   Conv1D with ip to reduce parameter RF^5 ∈ R^(f×n)
   ElementWise_Add RF^6= RF^4 ⊕ RF^5 to get residual_out RF^6 ∈ R^(f×n)
   ReLU with RF^6 to substitute all negatives with zero to produce resit_out ∈ R^(f×n)
   return resit_out
```

Figure. 2 Layers of Residual Unit

In Figure. 2, ip is the input feature map with size of (*d*, *n*), n is the timestamp and d is single or multi features, and f is the number of filters. The first layer is convolution layer with (1×1) filter size and $\frac{f}{2}$ filters and get RF[1] output with size of ($\frac{f}{2}$ ×n). The next layer is convolution layer with filter of size (3×3) and f filters. This layer outputs RF[2] feature map. BN layer and ReLU layer produce RF[3] and RF[4] feature map respectively. The skip connection performs that the input ip is convolved with (1×1) filter to add the features. The final layer of residual unit is ReLU activation layer which substitute zero instead of negative values and produces resit_out feature map. This research uses one residual unit.

3.4 LSTM

LSTM network [26] is designed for processing sequential data as it propagated forward. It is composed of three gates: forget gate, input gate, and output gate. Forget gate decides what information should be thrown away or kept through the sigmoid function. It passes the hidden state and current input into the tanh function to squish values between -1 and 1. Forget gate filter out irrelevant historical stock data. It focuses on the most crucial trends that impacts future stock prices. Eq. (11) illustrates the process of forget gate.

$$f_t = \sigma(W_f \,.\, [h_{t-1}, x_t] + b_f) \tag{11}$$

where,

f_t is the forget gate output

W_f *and* b_f are the weights and bias for the forget gate

h_{t-1} is the previous hidden state, and x_t is the input at current time step

The function of the input gate is to determine how much current information is added to the information flow that is calculated using Eq. (12) and Eq. (13). It incorporates recent stock prices as relevant new information into the model memory.

$$i_t = \sigma(W_i \,.\, [h_{t-1}, x_t] + b_i) \tag{12}$$

$$\widetilde{C_t} = \tanh\left(W_c \,.\, [h_{t-1,} x_t] + b_c\right) \tag{13}$$

where,

i_t is the input gate output and $\widetilde{C_t}$ is the candidate cell state

W_i and b_i are the weights and bias for the input gate

The LSTM updates the cell state for the output of the current cell and transfer it to the next cell as shown in Eq. (14). The output gate combines the current input and cell state to determine the output of the current LSTM cell shown in Eq. (15) and Eq. (16). In the standard LSTM network, sigmoid is used as the gating function and the tanh is used as the output activation function.

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{14}$$

$$O_t = \sigma(W_o \,.\, [h_{t-1}, x_t] + b_o) \tag{15}$$

$$h_t = O_t \odot \tanh(C_t) \tag{16}$$

where,

f_t , i_t *and* O_t the forget, input and output gate vectors respectively.

3.5 Attention Mechanism

Attention mechanism focuses on most relevant information from input sequence. It learns and assigns dynamically higher attention weights to key time stamps as illustrated in Fig. 3. Attention block computes the attention scores using SoftMax function shown in Eq. (17).

$$Att(X) = SoftMax(ReLU(BN(Conv(X)))) \tag{17}$$

where,

X is input feature map.

Conv is Convolution operation.

BN is Batch Normalization

ReLU is Rectified linear unit activation function

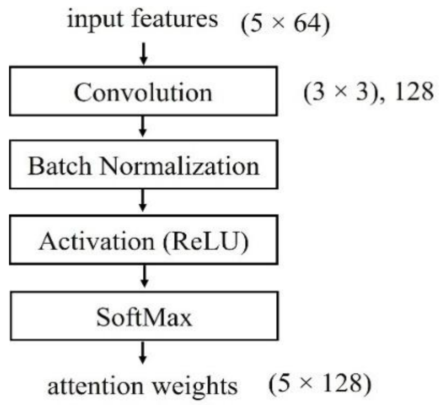

Figure. 3 Attention Unit

Def Attention_Unit(ip, f)
Conv1D and BN with ip and f are used to get feature $AF^1 \epsilon R^{f \times n}$
ReLU with AF^1 to substitute all negatives with zero $AF^2 \epsilon R^{f \times n}$
Extract SoftMax attention weight with AF^2 to get $AF^3 \epsilon R^{f \times n}$
return AF^3

Figure. 4 Layers of Attention Unit

By integrating attention unit into LSTM, the model captures patterns more effectively than traditional LSTM model. It captures different dependencies between distant time steps. Attention scores can identify which features in the sequence relate to each other. The algorithm of attention mechanism is illustrated in Figure. 4.

In Figure. 4, ip is the input feature map with size of (d, n), n is the timestamp and d are single or multi features, and f is the number of filters. The convolution layer performs on the input feature map with f filters. The BN is batch normalization layer and ReLU is activation process. After the activation, the probability value of each feature is calculated by using SoftMax to get attention weight. The final output is AF^3 feature map.

3.6 Bi-LSTM

The Bi-directional LSTM network is an extension of standard LSTM architecture that improves performance on sequential prediction tasks by processing the data in both forward and backward directions. The network capture information from both past and future contexts, providing a more comprehensive understanding of the sequence. The forward LSTM can be utilized to learn about the input sequence's prior data, whilst the reverse LSTM can be utilized to learn about the input sequence's subsequent data. The H_t hidden state for Bi-LSTM in the t time incorporates $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ as forward and backward as Eq. (18), Eq. (19) and Eq. (20)

$$\overrightarrow{h_t} = LSTM(h_{t-1}, x_t, c_{t-1}),\ t \in [1, T] \tag{18}$$

$$\overleftarrow{h_t} = LSTM(h_{t+1}, x_t, c_{t+1}),\ t \in [1, T] \tag{19}$$

$$H_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{20}$$

4. Experimental results and discussion

An enhanced residual attention LSTM model is developed for predicting time series stock data. This model is tested on five datasets. The datasets are tested on standard LSTM network, Bi-LSTM model, LSTM-Bi-LSTM framework. Then, the results are compared with the proposed LSTM model. In the proposed network, 32 filters are input to residual unit, the parameter of LSTM network is set to 64. The attention mechanism is performed with 128 filters. The feature enhancement using Bi-LSTM model is applied with parameter 64. The enhanced feature maps from Bi-LSTM are multiplied by attention weights. The fully connected layer with 32 units is used in the prediction system. The final layer is prediction layer with ReLU activation function.

4.1 Dataset collection

The daily time series of five different stocks index market information are extracted for system implementation. The datasets such as AAPL, BTC, ETH and LTC are downloaded from Yahoo Finance website [27]. GOLD-PRICE dataset is downloaded from the research paper [19].

Each daily record consists of seven variables that describe the daily trade stock data see in Fig. 5. The Date column contains the list of the daily stock trading dates. The Open column represents the value of opening stock price exchange for the first trade for that day. The Close column indicates the final price of the stock of the trading for that day. Adj Close column reflects the true price of that stock which is commonly used when analysis of stock returns. High and low columns are the rule of price strategies by stock companies.

Date	Open ($)	High($)	Low($)	Close($)	Adj Close($)	Volume
11/22/2010	124.1538	124.1538	121.3846	123.3077	92.55427	8541871
11/23/2010	121.7692	122.2308	120.2308	121.2308	90.99533	7273370
11/24/2010	121.6154	122.9231	121.3846	122.6154	92.03465	5512533
11/26/2010	121.3846	122.1538	121.1538	121.5385	91.2263	2267213
11/29/2010	121	123.4615	120.6154	122.8462	92.20786	7346014

Figure. 5 Sample stock data for 5 days

Table 1. Datasets and time

Dataset	Year (dd/mm/yyyy)
AAPL	(3/3/2014) – (1/3/2024)
Bitcoin (BTC)	(1/1/2018) – (2/1/2023)
Ethereum (ETH)	(1/1/2018) – (2/1/2023)
Litecoin (LTC)	(1/1/2018) – (31/12/2022)
GOLD-PRICE	(29/12/1978) – (4/6/2021)

Table 2. Mutual information score on AAPL dataset

Variable	Score
Close	3.62
High	3.52
Low	3.50
Adj Close	3.47
Open	3.38
Volume	0.61

Volume column is the number of shares or contracts traded in a security or an entire market during a given period of time. Sample of dataset is shown in Table1.

4.2 Evaluation metrices

The proposed stock price prediction system is measured with mean square error (MSE), mean absolute error (MAE) and root mean square error (RMSE). MAE is the average of the squared differences between the actual and predicted prices. It emphasizes the larger errors due to the squaring operation as Eq. (21). MAE is the average of the absolute differences between the actual and predicted stock prices in Eq. (22). RMSE is calculated as the square root of MSE. It provides an error metric on the same scale as the original stock price as Eq. (23).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \quad (21)$$

$$MAE = \sum_{i=1}^{n}\left|Y_i - \widehat{Y}_i\right| \quad (22)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2} \quad (23)$$

where,

n = the number of observations
Y_i = the actual value
$\widehat{Y}_i$ = the predicted value

4.3 Hardware and software specifications

The experimentations of the study are implemented on a Laptop with Window 10 operating system. Hardware devices are Intel Core^{TM} 6500U 2.5 GHz with Turbo Boost up to 3.1 GHz, graphic NVIDIA GeForce 940M with 2GB Dedicated VRAM. 12GB DDR3 L Memory and 512 SSD. For developing LSTM model based on residual and attention units, authors used the Keras library, Tensor Flow library in the backend with Python 3.12 version. The hyperparameters are specified time-stamps=5, Adam optimizer, batch-size=32, epochs=200, learning-rate = 0.0001.

4.4 Results on stock price datasets

For experiment on stock price data with one variable, the most significant variable is selected from multiple dimensions such as open, high, low, close, adjust close and volume.

Mutual information (MI) is used to determine the important features in a dataset. It is measure of the interdependence between two variables, which indicates how much information is shared between two variables. MI is a statistical measure that captures both linear and non-linear relationships between variables. The MI score of Close is highest score of 3.62 as shown in Table 2 and Fig. 6.

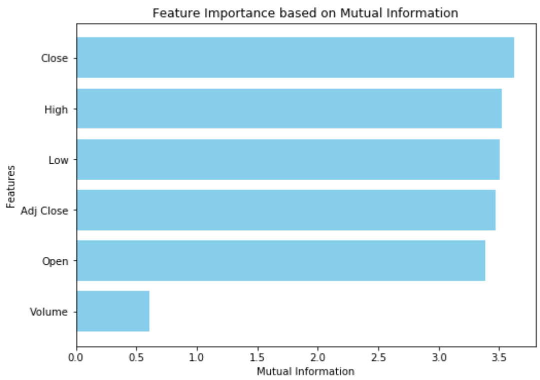

Figure. 6 Feature selection using MI

Table 3. Results of proposed model for univariate stock data prediction

Dataset	MSE	MAE	RMSE
AAPL	8.35	2.33	2.89
Bitcoin (BTC)	446758.61	449.22	668.40
Ethereum (ETH)	4162.94	44.47	64.52
Litecoin (LTC)	7.13	1.86	2.67
GOLD-PRICE	255.11	11.83	15.97

Received: November 22, 2024. Revised: December 17, 2024.

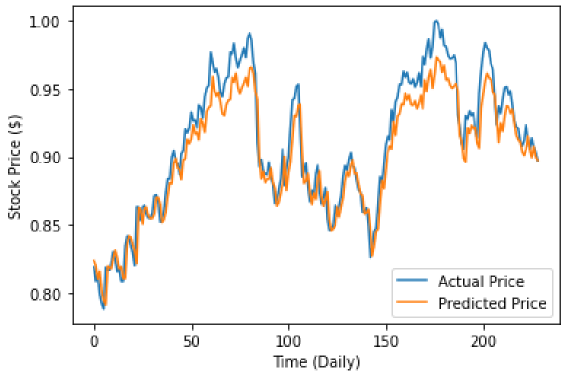

Figure. 6 Comparison of actual price and predicted price of Univariate testing on AAPL dataset

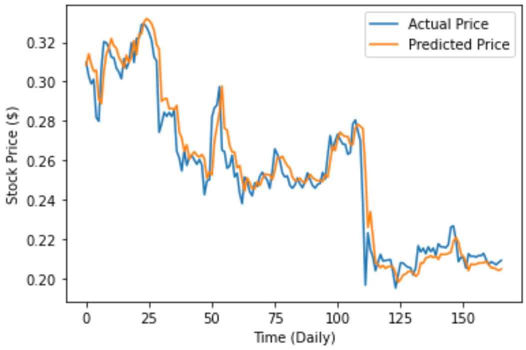

Figure. 7 Comparison of actual price and predicted price of Univariate testing on Bitcoin dataset

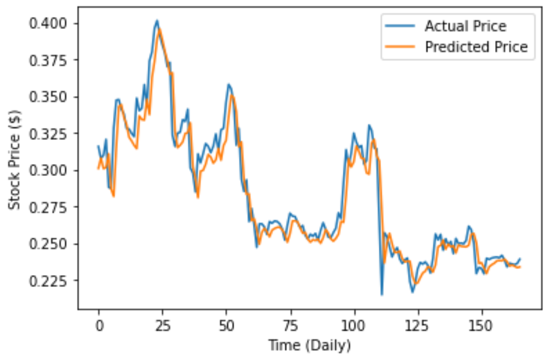

Figure. 8 Comparison of actual price and predicted price of Univariate testing on Ethereum dataset

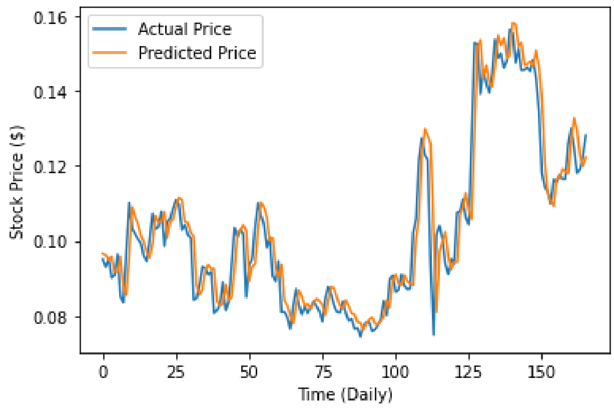

Figure. 9 Comparison of actual price and predicted price of Univariate testing on Litecoin dataset

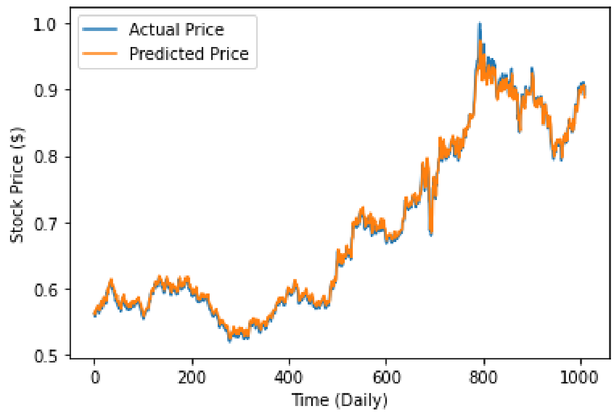

Figure. 10 Comparison of actual price and predicted price of Univariate testing on GOLD-PRICE dataset

Therefore, Close is used for univariate stock data prediction system. Volume is lowest score that it means less relationship for stock price prediction. For multivariate prediction, Close, High, Low, Adj Close, and Open variables are used for analysis of the time series data.

Experimentation results of univariate stock data prediction are shown in Table 3 on five datasets. The errors of prediction proposed model on AAPL are 8.35 of MSE, 2.33 of MAE and 2.89 of RMSE. For Bitcoin dataset, 446758.61 of MSE, 449.22 of MAE and 668.40 of RMSE are achieved respectively. In the test data of Ethereum, the results are MSE=4162.94, MAE=44.47, RMSE=64.52. The Litecoin dataset is used to test the proposed model, obtaining MSE =7.13, MAE =1.86, RMSE = 2.67.

The closing price is most correlated variable. Therefore, this is used to test univariable time series data predictions. The testing results of the proposed model are visualized and compared predicted stock prices with actual prices shown in Figure. 6 on AAPL dataset, Figure. 7 on Bitcoin dataset, Figure. 8 on Ethereum dataset, Figure. 9 for Litecoin dataset, and Figure. 10 for Gold-price dataset.

The developed model is tested on multivariable stock data. The model achieved better performance than univariate testing with less error as described in Table 4.

According to the experimental results, the proposed framework for time series data prediction is stable in training and testing. It overcomes the problem of overfitting for predicting non-linear data.

The model attains the less error rate on both univariate and multivariate time series data. For

Received: November 22, 2024. Revised: December 17, 2024.

Table 4. Results of the proposed model for multivariate stock data prediction

Dataset	MSE	MAE	RMSE
AAPL	5.33	1.79	2.30
Bitcoin (BTC)	396512.35	381.46	629.69
Ethereum (ETH)	3162.75	33.38	56.71
Litecoin (LTC)	11.76	2.25	3.42

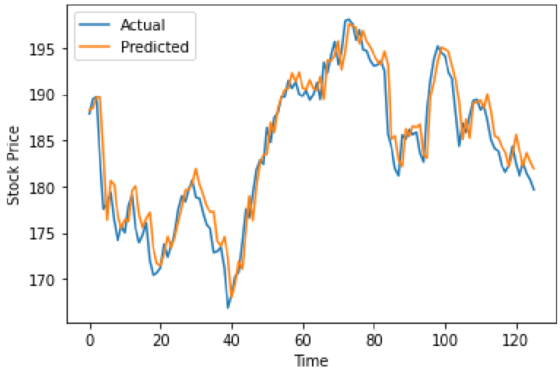

Figure. 11 Comparison of actual price and predicted price of multivariate testing on AAPL dataset

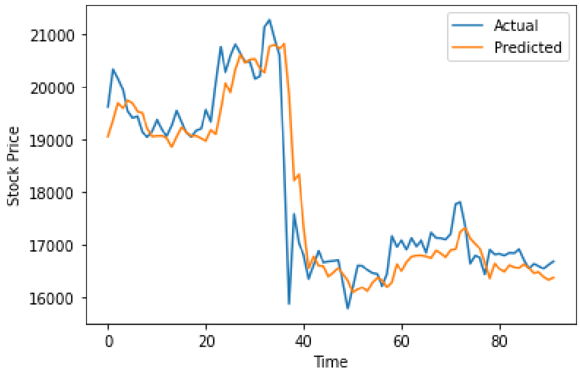

Figure. 12 Comparison of actual price and predicted price of multivariate testing on Bitcoin dataset

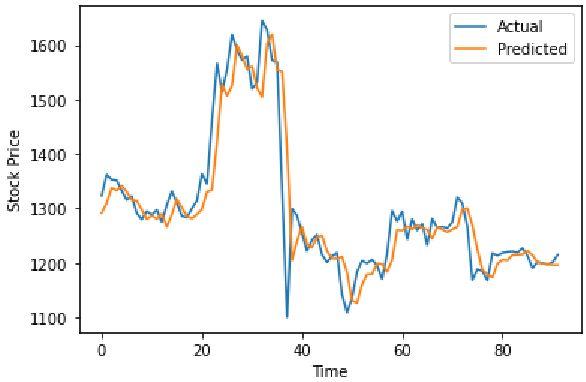

Figure. 13 Comparison of actual price and predicted price of multivariate testing on Ethereum dataset

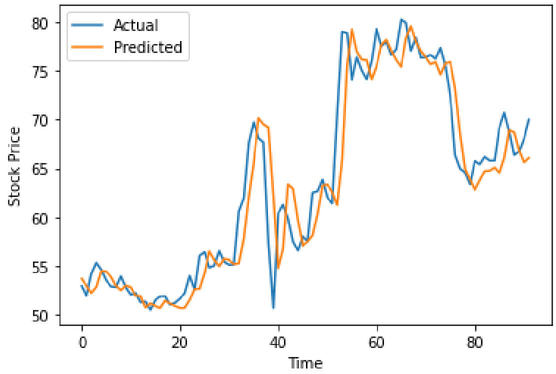

Figure. 14 Comparison of actual price and predicted price of multivariate testing on Litecoin dataset

ulti-variate testing, the actual prices and predicted prices are visualized in Figure. 11, Figure. 12, Figure. 13, and Figure. 14 on AAPL, Bitcoin, Ethereum and Litecoin datasets respectively. Gold-price dataset cannot be tested for multivariate testing because it has only one variable.

4.5 Comparison of the proposed model with existing models

The proposed LSTM model with residual and attention is compared with the three baseline models such as LSTM, Bi-LSTM and LSTM+Bi-LSTM models. The experiments of the models are evaluated on both univariable and multivariable designs.

In the univariate testing, the results demonstrated that the proposed model achieves better performance in term of MSE, MAE and RMSE values on AAPL, Bitcoin, Ethereum and Litecoin datasets as shown in Tables 5-8. For the testing of the models on GOLD-PRICE dataset, the results of Bi-LSTM model and proposed model are comparable performance as shown in Table 9.

Table 5. Comparison of Models on AAPL dataset (univariate)

Model Name	MSE	MAE	RMSE
LSTM	26761.33	163.33	163.58
Bi-LSTM	26761.33	163.33	163.58
LSTM + Bi-LSTM	15.06	3.09	3.88
Proposed Model	**8.35**	**2.33**	**2.89**

Table 6. Comparison of Models on Bitcoin dataset (univariate)

Model Name	MSE	MAE	RMSE
LSTM	745072.62	569.98	863.17
Bi-LSTM	552833.81	481.80	743.52
LSTM + Bi-LSTM	811438.39	586.96	900.79
Proposed Model	**446758.61**	**449.22**	**668.40**

Table 7. Comparison of Models on Ethereum dataset (univariate)

Model Name	MSE	MAE	RMSE
LSTM	1847258.26	1342.91	1359.13
Bi-LSTM	1847258.26	1342.91	1359.13
LSTM + Bi-LSTM	8708.30	64.12	93.31
Proposed Model	**4162.94**	**44.47**	**64.52**

Table 8. Comparison of Models on Litecoin dataset (univariate)

Model Name	MSE	MAE	RMSE
LSTM	13.83	2.52	3.71
Bi-LSTM	1462.98	37.36	38.39
LSTM + Bi-LSTM	13.67	2.50	3.69
Proposed Model	**7.13**	**1.86**	**2.67**

Table 9. Comparison of Models on GOLD-PRICE dataset (univariate)

Model Name	MSE	MAE	RMSE
LSTM	461.64	14.81	21.48
Bi-LSTM	**250.72**	**10.63**	**15.83**
LSTM + Bi-LSTM	448.01	14.54	21.16
Proposed Model	**255.11**	**11.83**	**15.97**

Table 10. Comparison of Models on AAPL dataset (Multivariate)

Model Name	MSE	MAE	RMSE
LSTM	15.62	3.25	3.95
Bi-LSTM	11.13	2.71	3.33
LSTM + Bi-LSTM	17.30	3.38	4.15
Proposed Model	**5.33**	**1.79**	**2.30**

In prediction of models with multivariable, the results illustrated that the proposed model obtains better performance than three baseline models on four datasets as shown in Tables 10-13.

Table 11. Comparison of Models on Bitcoin dataset (Multivariate)

Model Name	MSE	MAE	RMSE
LSTM	716362.20	484.90	846.38
Bi-LSTM	438102.49	359.59	661.89
LSTM + Bi-LSTM	220990154	14780	14865
Proposed Model	**396512.35**	**381.46**	**629.69**

Table 12. Comparison of Models on Ethereum dataset (Multivariate)

Model Name	MSE	MAE	RMSE
LSTM	6654.53	49.93	81.57
Bi-LSTM	4529.13	38.96	67.29
LSTM + Bi-LSTM	6971.82	50.11	83.49
Proposed Model	**3162.75**	**33.38**	**56.71**

Table 13. Comparison of Models on Litecoin dataset (Multivariate)

Model Name	MSE	MAE	RMSE
LSTM	1690.46	39.98	41.11
Bi-LSTM	1690.46	39.98	41.11
LSM + Bi-LSTM	1690.46	39.98	41.11
Proposed Model	**11.76**	**2.25**	**3.42**

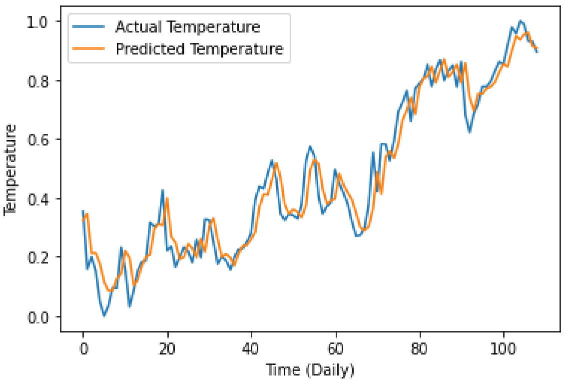

Figure. 15 Result of Proposed Model on Weather dataset

Table 14. Comparison of proposed model with existing state-of-the-art methods in RMSE value

Dataset	**State-of-the-art Models**				**Proposed Model**					
					Univariate Model			**Multivariate Model**		
		MSE	**MAE**	**RMSE**	**MSE**	**MAE**	**RMSE**	**MSE**	**MAE**	**RMSE**
GOLD-PRICE	Ref. [19]	-	27.77	37.94	255.11	**11.83**	**15.97**	**-**	**-**	**-**
AAPL	[24]	11.70	37.96	6.16	8.35	2.33	2.89	**5.33**	**1.79**	**2.30**
BTC	[23]	-	-	1029.36	446758.61	449.22	668.40	396512.35	381.46	**629.69**
ETH	[23]	-	-	83.59	4162.94	44.47	64.52	3162.75	33.38	**56.71**
LTC	[23]	-	-	8.02	7.13	1.86	**2.67**	11.76	2.25	3.42
CSI 300 Index	[13]	2762.82	37.34	52.56	2762.82	33.61	47.55	**764.07**	**20.21**	27.64
SSE Composite Index	[13]	1350.16	28.08	36.74	1042.12	33.61	32.28	**881.32**	**26.88**	**29.68**
SSE Shanghi Composite Index	[25]	1956.03	28.34	44.22	**1431.24**	**27.06**	**37.26**	1261.06	24.60	35.51
China Unicom	[25]	0.028	0.10	0.167	0.09	0.008	0.091	**0.005**	**0.055**	**0.076**

4.6 Results on Weather dataset

The research adopts to develop the optimal model for stock price prediction. The proposed model is tested on Weather dataset downloaded from Kaggle data repository [28]. The model also attains improve performance on time series weather prediction. The loss values are 2.87 in MSE, 1.33 in MAE and 1.69 in RMSE. The actual temperature and predicted temperature are shown in Figure. 15.

4.7 Comparison with existing methods

The proposed LSTM model is compared with existing deep learning-based stock price prediction model. The analysis of the models shows that the proposed system is higher performance for stock price forecasting. The comparison results are tested on AAPL, BTC, ETH, LTC, CSI 300 Index, SSE Composite Index and illustrated in Table 14. Our proposed model achieves lower RMSE error values than the state-of-the-art architecture.

5. Conclusion

This research proposed a model for stock price prediction system using residual unit, attention mechanism, LSTM and Bi-LSTM. The architecture of model focuses on the important features and different dependencies for each time-stamp. The residual unit reduces model overfitting and enhances pattern recognition. Attention unit focus on the most critical features, improving interpretability and robustness. The model is tested on five datasets: AAPL, BTC, ETH, LTC, and GOLD-PRICE. The proposed system is tested on both univariate and multivariate time series data. The proposed univariate model reduces the prediction error 160.69, 194.77, 1294.61, 1.04, 5.51 of RMSE than LSTM model on five datasets respectively. The proposed multivariate model also reduces that of 1.65 on AAPL, 216.69 on BTC, 24.86 on ETH, 37.69 on LTC datasets. Evaluation results shown that the proposed model achieves optimal performance on time series stock data. The proposed attention-based model captures both short-term and long-term dependencies in stock price prediction. The system obtains smooth training and testing progress, lower error rate than LSTM model. The experimental results demonstrate that the proposed network captures both short-term and long-term dependencies for non-linear time series data. The model is also tested on WEATHER dataset; the results show that it improves the performance on weather dataset. This research contributes for stock price prediction by providing more reliable and accurate approach for financial decision making. However, further exploration will be explored by

Received: November 22, 2024. Revised: December 17, 2024.

integrating additional market factors on stock price movement.

Notation List

Symbol	Description
X	The input feature map
W	filter weight
b	Weight Bias vector
μ	Mean value
σ	Standard deviation value
σ^2	Variance
γ	scaling factor
β	shifting factors
T	Total number of time stamps
x_t	Input at time t
i_t	Input gate at time t
C_t	Cell state at time t
f_t	forget gate at time t
h_t	Hidden state at time t
H_t	Concatenated hidden state at time t
O_t	Output gate at time t

Conflicts of Interest

The authors declare that there is no conflict of interest.

Author Contributions

Khin Nyein Myint; Conceptualization, methodology, writing—original draft preparation, software, validation, formal analysis, investigation, resources, Myo Khaing; data curation, review and editing, supervision.

References

[1] Universitas Dian Nuswantoro, A. Syukur, A. Marjuni, and Universitas Dian Nuswantoro, "Stock Price Forecasting Using Univariate Singular Spectral Analysis through Hadamard Transform", *Int. J. Intell. Eng. Syst*, Vol. 13, No. 2, pp. 96–107, Apr. 2020, doi: 10.22266/ijies2020.0430.10.

[2] Annamalai University, L. Maguluri, R. Rengaswamy, and Annamalai University, "An Efficient Stock Market Trend Prediction Using the Real-Time Stock Technical Data and Stock Social Media Data", *Int. J. Intell. Eng. Syst*, Vol. 13, No. 4, pp. 316–332, Aug. 2020, doi: 10.22266/ijies2020.0831.28.

[3] K. N. Myint and M. Khaing, "Time Series Forecasting System for Stock Market Data", In: *Proc. of 2023 IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar: IEEE*, Yangon, Myanmar, pp. 56–61, 2023.

[4] H. M, G. E.A., V. K. Menon, and S. K.P., "NSE Stock Market Prediction Using Deep-Learning Models", *Procedia Comput. Sci*., Vol. 132, pp. 1351–1362, 2018.

[5] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series", In: *Proc. of 2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, pp. 3285–3292, Dec. 2019.

[6] Y. Wen, P. Lin, and X. Nie, "Research of Stock Price Prediction Based on PCA-LSTM Model", *IOP Conf. Ser. Mater. Sci. Eng*., Vol. 790, No. 1, p. 012109, 2020.

[7] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks", arXiv: *arXiv:1910.03151*, 2020.

[8] K. N. Myint and Y. Y. Hlaing, "Predictive Analytics System for Stock Data: methodology, data pre-processing and case studies", In: *Proc. of 2023 IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar*, Yangon, Myanmar, pp. 77–82, 2023.

[9] K. N. Myint, & Y. Y. Hlaing, "Time Series Data Forecasting System for Stock using TA and ARIMA Model", In: *Proc. of 2023 IEEE Conference on Computer Applications (ICCA),* Yangon, Myanmar, pp. 72-76, 2023.

[10] G Pullaiah College of Engineering and Technology et al., "Cross Entropy Based Long Short-Term Memory Recurrent Neural Network Model for Analyzing the Time Series on Stock Market Price", *J. Intell. Eng. Syst*, Vol. 13, No. 2, pp. 259–266, Apr. 2020, doi: 10.22266/ijies2020.0430.25.

[11] CHRIST (Deemed to be University), C. Manjunath, B. Marimuthu, CHRIST (Deemed to be University), B. Ghosh, and RV Institute of Management, "Deep Learning for Stock Market Index Price Movement Forecasting Using Improved Technical Analysis", *Int. J. Intell. Eng. Syst*, Vol. 14, No. 5, pp. 129–141, Oct. 2021, doi: 10.22266/ijies2021.1031.13.

[12] Y. Liu, C. Zhao, and Y. Huang, "A Combined Model for Multivariate Time Series Forecasting Based on MLP-Feedforward Attention-LSTM", *IEEE Access*, Vol. 10, pp. 88644–88654, 2022.

[13] Y. Jia, A. Anaissi, and B. Suleiman, "ResNLS: An improved model for stock price forecasting", *Comput. Intell*, Vol. 40, No. 1, p. e12608, 2024.

[14] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by

Using Bi-Directional LSTM Model", *Front. Bus. Econ. Manag*, Vol. 8, No. 2, pp. 51–54, 2023.
[15] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach", *Fractal Fract*, Vol. 7, No. 2, p. 203, 2023.
[16] S. Seddik, H. Routaib, and A. Elhaddadi, "Multi-Variable Time Series Decoding with Long Short-Term Memory and Mixture Attention", *Acadlore Trans. AI Mach. Learn*, Vol. 2, No. 3, pp. 154–169, 2023.
[17] B. Shaju and V. Narayan, "Prediction Model for Stock Trading using Combined Long Short-Term Memory and Neural Prophet with Regressors," I*nt. J. Intell. Eng. Syst*, Vol. 16, No. 6, pp. 956–964, 2023, doi: 10.22266/ijies2023.1231.79.
[18] Y. Zhang and G. M. Tumibay, "Stock Price Prediction Based on the Bi-GRU-Attention Model", *J. Comput. Commun*, Vol. 12, No. 04, pp. 72–85, 2024.
[19] A. Amini and R. Kalantari, "Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning", *PLOS ONE*, Vol. 19, No. 3, p. e0298426, 2024.
[20] J. Song, Q. Cheng, X. Bai, W. Jiang, and G. Su, "LSTM-Based Deep Learning Model for Financial Market Stock Price Prediction", *Journal of Economic Theory and Business Management,*Vol. 1, No. 2, pp. 43-50, 2024.
[21] B. Gülmez, "Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm", *Expert Syst. Appl*, Vol. 227, p. 120346, 2023.
[22] X. Wang, M. Xia, and W. Deng, "MSRN-Informer: Time Series Prediction Model Based on Multi-Scale Residual Network", *IEEE Access*, Vol. 11, pp. 65059–65065, 2023.
[23] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach", *Fractal Fract*, Vol. 7, No. 2, p. 203, 2023.
[24] D. R. Rizvi and M. Khalid, "Performance Analysis of Stocks using Deep Learning Models", *Procedia Comput. Sci*, Vol. 233, pp. 753–762, 2024.
[25] Y. Chen, R. Fang, T. Liang, Z. Sha, S. Li, Y. Yi, ... & H. Song, "Stock Price Forecast Based on CNN-BiLSTM-ECA Model", *Scientific Programming*, vol. 1, p. 2446543, 2021.
[26] J. Schmidhuber, & S. Hochreiter, "Long short-term memory", *Neural Comput*, Vol. 9, No.8, pp. 1735-1780, 1997.
[27] https://finance.yahoo.com.
[28] https://www.kaggle.com/datasets/sumanthvrao/daily-climate-time-series-data.