Vol. 4 No. (1) (June 2026) 186-196

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

A Hybrid LSTM with GARCH-MIDAS-X Framework for Modelling IDX Composite Volatility

Silviya Indriyani[1], Irhamah*[2], Tintrim Dwi Ary Widhianingsih[3]
[1,2,3]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

DOI: https://doi.org/10.26714/jodi.v4i1.1180

Article Info

Article history:
Received June 15, 2026
Revised June 29, 2026
Accepted June 30, 2026

Keywords:
IDX Composite; GARCH-MIDAS-X; LSTM; Volatility.

Abstract

Stock market volatility forecasting plays a crucial role in supporting investment decision-making and risk management under uncertain market conditions. This study proposes a hybrid LSTM with GARCH-MIDAS-X framework to modelling IDX Composite volatility. The GARCH-MIDAS-X model is first employed to decompose stock return volatility into short-run and long-run components while incorporating multiple low-frequency exogenous variables, including market news sentiment, crude oil prices, and exchange rates. The residual generated by the GARCH-MIDAS-X model is subsequently used as input for the LSTM network to capture complex nonlinear patterns and temporal dependencies that may not be fully explained by the econometric model. Model performance was evaluated using both in-sample criteria (AIC, BIC, and HQIC) and out-of-sample forecasting measures (RMSE and MAE). The empirical results show that, for the Type I specification, the proposed hybrid model slightly improved forecasting accuracy by reducing the RMSE from 5.2846 to 5.2671 and the MAE from 2.8960 to 2.8656 compared with the standalone GARCH-MIDAS-X model. In contrast, for the Type II specification, the hybrid model yielded slightly higher forecasting errors (RMSE = 5.2880 and MAE = 2.9081) than the standalone GARCH-MIDAS-X model (RMSE = 5.2876 and MAE = 2.9070). These findings indicate that the GARCH-MIDAS-X model captures most of the relevant volatility dynamics, while the additional LSTM component provides only limited improvements in forecasting accuracy. Overall, the proposed hybrid model represents a viable alternative for stock market volatility forecasting; however, its superiority over the standalone GARCH-MIDAS-X model is not consistently supported by the empirical results.

✉ Correspondence Address:
E-mail: irhamahr@its.ac.id

e-ISSN: 2988 - 2109

INTRODUCTION

According to the Indonesian Stock Exchange, shares are an investment instrument that many investors choose because shares are able to provide attractive levels of return, namely in the form of capital gain and dividends. The stock market plays an important role so it can be a leading indicator of Indonesia's economic growth [1]. A country's stock price index declines due to the country's economic conditions facing challenges, while a rising stock price index can indicate improvements in the country's economic performance. A significant decline in the JKSE can also signal an early crisis of confidence in a country's economic prospects and, in the long term, could potentially disrupt the stability of the national financial system. Therefore, accurately modeling and forecasting stock market volatility is essential for investors, portfolio managers, and policymakers to improve investment decisions and financial risk management.

Stock price changes can occur in the form of increases or decreases within minutes and seconds; this movement is known as volatility. In the Indonesian context, the IDX Composite (JKSE) exhibits dynamics that are sensitive to global shocks, including changes in international central bank policies and movements in foreign indices. This condition makes predicting JKSE volatility an important aspect in understanding domestic capital market behavior [2]. Therefore, a method is needed to measure the magnitude of JKSE volatility; one method that can be used is Generalized Autoregressive Conditional Heteroscedasticity (GARCH). In the classical linear time series model, the variance value of the residual is assumed to be constant. However, there are often cases where the residual variance becomes non-constant due to high volatility values. The ARCH/GARCH model is a model that includes the possibility of non-constant residual variance [3]. Research on ARCH-based volatility analysis presented by [4] and Generalized ARCH (GARCH) developed by [5]. Although the development of the GARCH model has been widely carried out, GARCH is not capable enough to predict time series influenced by exogenous variables [6]. In the JKSE stock market, [7] stated that interest rates and inflation have a negative effect on the JKSE, while the rupiah exchange rate against the US dollar has a positive effect. [8] also proved that the inflation rate and the rupiah exchange rate against the US dollar have a significant effect on the JKSE.

The GARCH-X model introduced by [9] is a model that adds the cointegration relationship of the stock market with macroeconomic variables into the conditional variance equation. However, all GARCH-type models require the frequency of the data used to be the same. The MIDAS model (Mixed Data Sampling regression) is a regression model introduced by [10] and is a regression reduced form which is strictly parameterized and involves processes observed at different frequencies. MIDAS regression combines variables with different frequencies into one model using a lag weighting function. This concept was later used by [11] into the GARCH model to form GARCH-MIDAS. The GARCH-MIDAS model also allows for direct examination of the relationship between macroeconomic volatility and stock market volatility consisting of short-term and long-term components. Recently, GARCH-MIDAS models have been widely applied to volatility forecasting by incorporating macroeconomic and financial variables, demonstrating their effectiveness in capturing mixed-frequency information and improving forecasting performance [12], [13].

If we expand the short-term component equation by adding exogenous variables that have the same frequency as return stocks, then the appropriate model to use is GARCH-MIDAS-X [14]. However, due to the complex nonlinear correlation structure between variables and the large data size, the prediction results from GARCH-type models are often less than satisfactory. Furthermore, the MIDAS model is only able to capture the linear relationship between the dependent and independent variables, making it difficult to simultaneously capture the influence of macroeconomic variables on volatility, which may be nonlinear. The LSTM method (Long Short-Term Memory) introduced by [15] is a machine learning method which has been proven to provide significant improvements for modeling and forecasting time series data. The use of this method hybrid LSTM with GARCH models has been widely used, one of which is [16] on South Korean stock index data. The results show that combining several GARCH models and deep learning (LSTM) can significantly improve prediction accuracy. However, previous studies have mainly focused on hybrid GARCH-LSTM models and have not explicitly integrated the GARCH-MIDAS-X framework, which incorporates mixed-frequency macroeconomic variables, with the nonlinear learning capability of LSTM. Therefore, there is still a need to develop a hybrid model that combines the advantages of GARCH-MIDAS-X and LSTM for volatility forecasting. Thus, the author would like to propose the use of hybrid LSTM with GARCH-MIDAS-X to model JKSE stock market volatility.

METHOD

2.1 GARCH Model

GARCH model was first proposed by [5] and is a development of the ARCH model introduced by [4]. General form of the GARCH model(p, q)consists of the equation of the mean and variance. Suppose a standard regression model with *error* uncorrelated is defined as follows [17]

$$r_{i,t} = \mu + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, \sigma_{i,t}^2)$$

Where μ assumed constant and $\varepsilon_{i,t}$ is an *error* uncorrelated but have variance that changes over time (heteroscedastic) and

$$\varepsilon_{i,t} = \sigma_{i,t} e_{i,t}, \qquad e_{i,t} \sim N(0,1)$$

Where $\sigma_{i,t}^2$ is the conditional variance and $e_{i,t}$is independent of $\varepsilon_{i-d,t}$ for $d > 0$.

2.2 GARCH-X Model

The GARCH-X model is a development of the classic GARCH model by adding exogenous variables into the GARCH equation. The GARCH-X(1,1) model is as follows:

$$\sigma_{i,t}^2 = \alpha_0 + \alpha_1 \varepsilon_{i-1,t}^2 + \beta_1 \sigma_{i-1,t}^2 + \lambda X_{i-1,t}$$

2.3 MIDAS Regression

MIDAS regression is a time series regression that allows the use of frequency differences in the data. In general, the MIDAS linear regression equation is defined as follows:

$$\tau_t = \gamma_0 + \sum_{c=1}^{C} \gamma_{1c} B(L^{1/m}; \omega)\, x_{tc}^{(m)} + \eta_t^{(m)}$$

$$B\left(L^{\frac{1}{m}}; \omega\right) = \sum_{k=0}^{K} B(k; \omega) L^{k/m}$$

Where τ_t is the observed dependent variable, $x_{tc}^{(m)}$ states the independent variable c which is used to predict τ_t, $\gamma_0$is the intercept, Cshows the number of independent variables used in the MIDAS model, parameters γ_1 states how large the overall effect of the lag variable is $x_{tc}^{(m)}$to τ_t. There are two parameterizations of $B(k; \omega)$, the first one is called as Exponential Almon Lag defined as follows

$$B(k; \omega) = \frac{\exp(\omega_1 k + \omega_2 k^2 + \cdots + \omega_Q k^Q)}{\sum_{a=1}^{K} \exp(\omega_1 a + \omega_2 a^2 + \cdots + \omega_Q a^Q)}$$

Where $\omega_1, \omega_2, \ldots, \omega_Q$ is the MIDAS weighting parameter and $Q$is a polynomial order. The second parameterization is called Beta Lag by using two parameters, namely $\omega = [\omega_1; \omega_2]$ defined as follows

$$B(k; \omega_1, \omega_2) = \frac{f(\frac{a}{K}, \omega_1, \omega_2)}{\sum_{a=1}^{K} f(\frac{a}{K}, \omega_1, \omega_2)}$$

$$f(x, u, v) = \frac{x^{u-1}(1-x)^{v-1}\Gamma(u+v)}{\Gamma(u)\Gamma(v)}$$

$$\Gamma(u) = \int_0^\infty e^{-x} x^{u-1} dx$$

2.4 GARCH-MIDAS-X Model

The GARCH-MIDAS-X model is a development of the GARCH-MIDAS model by including exogenous variables as additional short-term components and is a model that combines short-term daily GARCH-X with MIDAS polynomials applied to monthly, quarterly, semi-annual, and annual macroeconomic variables. For example, *return* on the day i in the th month t modeled as follows:

$$r_{i,t} = \mu + \sigma_{i,t} e_{i,t}, \quad i = 1, \ldots, N_t$$

$$\sigma_{i,t}^2 = \tau_t g_{i,t}$$ ')

Suppose $g_{i,t}$ assumed to follow the daily GARCH-X(1,1) process as follows:

$$g_{i,t} = (1-\alpha-\beta) + \alpha\frac{\left(r_{i-1,t}-\mu\right)^2}{\tau_t} + \beta g_{i-1,t} + \lambda\left(x_{i-1,t} - E(x_{i-1,t})\right) \quad)$$

Where $\alpha, \beta > 0$. There are 2 ways to define τ_t, first can be constructed from realized variance

$$\log\tau_t = m + \theta\sum_{k=1}^{K}\vartheta_k(\omega_1,\omega_2)RV_{t-k} \quad)$$

$$RV_t = \sum_{i=1}^{N_t} r_{i,t}^2 \quad)$$

and the second is directly from macroeconomic variables

$$\log\tau_t = m_l + \theta_l\sum_{k=1}^{K_l}\vartheta_k\left(\omega_{1,l},\omega_{2,l}\right)X_{l,t-k}^{mv} \quad)$$

Where $X_{l,t-k}^{mv}$ is the level value of the macroeconomic variable vm And l is the weight of the macroeconomic variable level. In addition to the macroeconomic variable level, the volatility of the macroeconomic vari able able can also be considered.

$$\log\tau_t = m_v + \theta_v\sum_{k=1}^{K_v}\vartheta_k\left(\omega_{1,v},\omega_{2,v}\right)X_{v,t-k}^{mv} \quad)$$

Where $X_{v,t-k}^{mv}$ represents the volatility of macroeconomic variables.[18] formulate the formation of macroeconomic variable volatility using a model *Vector Autoregressive* (VAR) as follows

$$X_{l,t}^{mv} = \Gamma D_t + \gamma t + \sum_{k=1}^{q} B_i X_{l,t-k}^{mv} + e_t \quad)$$

Where $D_t$is a vector of seasonal *dummy* variables, q is the number of lags based on the smallest AIC value and e_t^2 used as an approximation of the volatility of macroeconomic variables. The weighting function ϑ_k formulated as follows

$$\begin{cases} \vartheta_k(\omega_1,\omega_2) = \dfrac{\left(\frac{k}{K+1}\right)^{\omega_1-1}\left(1-\frac{k}{K+1}\right)^{\omega_2-1}}{\sum_{k=1}^{K}\left(\frac{l}{K+1}\right)^{w_1-1}\left(1-\frac{l}{K+1}\right)^{w_2-1}} & \text{Beta} \\ \vartheta_k(\omega) = \dfrac{\omega^k}{\sum_{k=1}^{K}\omega^k} & \text{Exponential} \end{cases}$$

In the GARCH-MIDAS-X model, parameters are estimated using the *Maximum Likelihood Estimation* (MLE). The parameter space of the GARCH-MIDAS-X model is defined as $\Theta = \{\mu,\alpha,\beta,\lambda,m,\theta,\omega_1,\omega_2\}$. Thus, the log-likelihood function is expressed as follows

$$\mathcal{L}(\Theta) = \ln L(\mu,\alpha,\beta,\lambda,m,\theta,\omega_1,\omega_2)$$

it is assumed that $e_{i,t}$ normally distributed, then the conditional density function is as follows:

$$f\left(r_{i,t}\middle|\Phi_{i-1,t}\right) = \frac{1}{\sqrt{2\pi\tau_t g_{i,t}}}\exp\left(-\frac{\left(r_{i,t}-\mu\right)^2}{2\tau_t g_{i,t}}\right)$$

So the following log-likelihood function maximize

$$\mathcal{L}(\Theta) = \sum_{t=1}^{T}\sum_{i=1}^{N_t} \ln f\left(r_{i,t} \middle| \Phi_{\mathrm{i-1,t}}\right)$$

To lower the standards error, the following variance covariance matrix is used in the context *Quasi Maximum Likelihood* (QMLE)

$$\widehat{Var}\left(\widehat{\Theta}\right) = \boldsymbol{\widehat{H}^{-1}\widehat{OPG}\widehat{H}^{-1}}$$

Where H is the Hessian matrix and OPG represent outer product of gradients.

2.5 LSTM Network

LSTM is a classic type of RNN that can solve problems exploding gradient (gradient that gets bigger and bigger) and vanishing gradient (gradients that get smaller and smaller). LSTM is able to learn tasks deep learning which is very complex and requires long-term memory for an event. LSTM can also handle inputs or signals that have low and high frequency components. The LSTM architecture combines memory cells and gate mechanism that regulates the flow of information, so the network can learn which data needs to be retained or discarded.

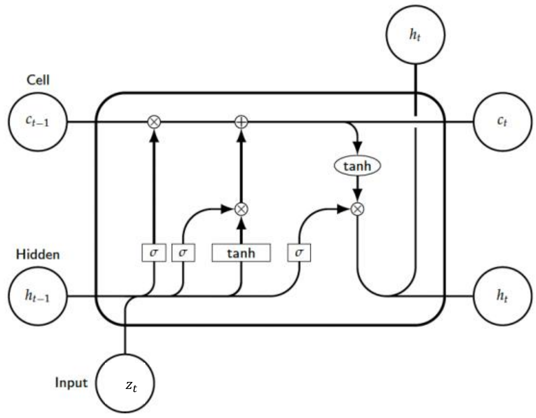

Figure 1. LSTM Architecture

The LSTM gating mechanism consists of a forget gate which is used to determine which information from cell state which must be thrown away. Forget gate evaluate cell state beforehand to determine which information should be discarded, thus preventing cell state filled with irrelevant information. After the forget gate, the input is then processed through the input gate. This gate functions to determine how much new information from the current input will be added to the memory cell state. Next, the input will be processed through the output gate. This gate controls which part of the cell state which will be issued as hidden state. This gate ensures hidden state only reflects the most relevant information from cell state.

$$f_t = \sigma(W_f z_t + W_{fh} h_{t-1} + b_f)$$

$$i_t = \sigma(W_i z_t + W_{ih} h_{t-1} + b_i)$$

$$o_t = \sigma(W_o z_t + W_{oh} h_{t-1} + b_o)$$

$$\tilde{c}_t = \tanh(W_c z_t + W_{ch} h_{t-1} + b_c)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t$$

$$h_t = o_t \odot \tanh(c_t)$$

where σ denotes the sigmoid activation function, which defined as

$$\sigma(Z) = \frac{1}{1 + e^{-Z}}$$

The effectiveness of the model highly depends on the appropriate selection and optimization of hyperparameters, as the LSTM configuration directly affects forecasting performance. The hyperparameters of the LSTM model include the number of layers, the number of unit per layers, epochs, batch size, learning rate, dropout rate and sequence length.

2.6 Data Source

This study uses secondary data obtained from www.investing.com, recorded from February 2015 to May 2026. The data used in this research are IDX Composite stock return prices. The data divided into in-sample and out-sample. The in-sample data start from February 2015 to March 2026, and out-sample data start from April 2026 to May 2026.

The inflation rate used in this study is the year-on-year (YoY) inflation rate obtained from the official website of Statistics Indonesia (BPS). The year-on-year inflation rate for the United States was obtained from the U.S. Bureau of Labor Statistics. The interest rate used in this study is the Bank Indonesia (BI) policy rate, obtained from the official website of Statistics Indonesia. The benchmark interest rate for the United States is determined by the Federal Reserve System. The crude oil price used in this study is the closing price, obtained from the website www.id.investing.com. The market news sentiment used in this study is a news-based sentiment index derived from the GDELT GKG (Global Database of Events, Language, and Tone Global Knowledge Graph). The news data focus on economic topics, particularly the stock market and financial market, and are restricted to events located in Indonesia. The market news sentiment index is constructed following the methodology proposed by [19]. The exchange rate used in this study represents the value of 1,000 Indonesian rupiah expressed in U.S. dollars. The exchange rate is measured using the closing price and was obtained from the website www.id.investing.com.

Table 1. Research Variable

Data	Name of Variable	Frequency
$JKSE_{i,t}$	IDX Composite Closing Price	Daily
$r_{i,t}$	Return of $JKSE_{i,t}$	Daily
$X_{i,t}^{(1)}$	Market news sentiment	Daily
$X_{i,t}^{(2)}$	Exchange rate	Daily
if_t^{INA}	Inflation in Indonesia	Monthly
if_t^{USA}	Inflation in USA	Monthly
br_t^{INA}	Interest rate in Indonesia	Monthly
br_t^{USA}	Interest rate in USA	Monthly
WTI_t	Crude oil price	Monthly

2.7 Analysis Step

The analysis step conducted for modelling JKSE return volatility using a hybrid LSTM with GARCH-MIDAS-X model are outlined as follows:

1. Described the JKSE stock price data and its return
2. Testing stationarity and ARCH effects on JKSE return data
3. Modeling JKSE returns using GARCH-MIDAS-X with predictor variables market news sentiment, crude oil price, exchange rate, inflation and interest rate
4. Conducting a combined estimate of the short-term volatility value using the GARCH-X model with predictor variables of market news sentiment, and exchange rate and the long-term volatility value using MIDAS regression with predictor variables of inflation, interest rate, crude oil price and realized variance.
5. Obtaining residuals from the GARCH-MIDAS-X model
6. Determining the residual lag as an input in an LSTM model
7. Building a hybrid LSTM model with GARCH-MIDAS-X

RESULTS AND DISCUSSION

This subsection presents the results of the exploratory data analysis conducted in this study. First, the IDX Composite (JKSE) and its changes are described. Subsequently, the explanatory variables used in the study, namely the inflation rate, interest rate, market news sentiment, crude oil price, and the exchange rate are presented.

Based on Figure 2, JKSE tends to increase over time. JKSE reached its highest peak in January 2026 but has since declined to date. A sharp decline in value also occurred during the COVID-19 pandemic. During the COVID-19 pandemic, the social and economic conditions shaken by the pandemic affected the JKSE’s value. Furthermore, based on the graph of changes or return The JKSE shows high fluctuations and low fluctuations indicating that the data variance has changed.

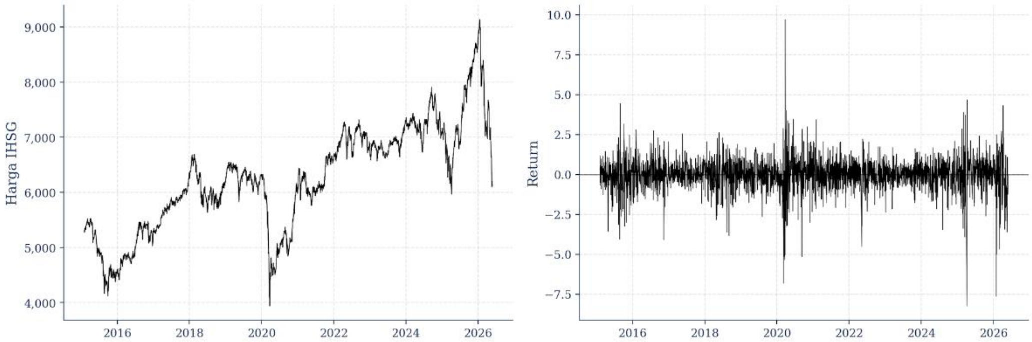

Figure 2. Plot of JKSE Stock Price and Its Return

Descriptive statistics of dataset are shown in Table 2. The JKSE reached its highest value on January 20, 2026, and its lowest on March 3, 2020, due to the COVID-19 pandemic. The largest negative change occurred on April 8, 2025, and the largest positive change occurred on March 26, 2020.

Table 2. Descriptive statistics dataset

Variable	Mean	Minimum	Maximum	Std. Dev.	Skewness	Kurtosis
$JKSE_{i,t}$	6245,96	3937,63	9134,70	959,29	0,0780	−0,2748
$r_{i,t}$	0,0055	−8,2319	9,7042	1,0245	−0,5732	8,9327
$X_{i,t}^{(1)}$	−0,0805	−1,6546	1,6211	0,5930	−0,0728	0,2918
$X_{i,t}^{(2)}$	62,8917	11,5700	119,7800	17,0981	0,3957	0,2243
$X_{i,t}^{(3)}$	0,0688	0,0560	0,0793	0,0051	−0,2931	−0,8667
if_t^{INA}	3,2265	−0,0900	7,2600	1,4487	0,8782	0,7488
if_t^{USA}	2,9023	−0,1995	9,0598	2,1438	1,2184	0,9142
br_t^{INA}	5,2134	3,500	7,7500	1,1572	0,2758	−0,6499
br_t^{USA}	2,2111	0,2500	5,500	1,9072	0,5232	−1,2443

The exploration of the JKSE return continued with several statistical tests and the result are shown in Table 3.. The first test was to test the stationarity of the JKSE return with ADF test, the results show that the return is stationary at the level of significance $\alpha = 0.05$. Then the ARCH effect test was carried out using the Engle LM test, the results showed that there was an ARCH effect on the return so that further modeling can be done with the GARCH-MIDAS-X model. The Terasvirta linearity test indicates that the return exhibits nonlinear characteristic.

Table 3. Diagnostic Test Result of JKSE return

Test	Value	p-value
ADF	−28,9572	0,0000
ARCH Effect	436,1408	$8,26 \times 10^{-86}$
Terasvirta	36,9943	0,0001

First, the return series is subsequently modeled using the GARCH-MIDAS-X Type I model. The details of the model specification are presented below.

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{\left(r_{i-1,t} - \mu\right)^2}{\tau_t} + \beta g_{i-1,t} + \lambda_1\left(X_{i,t}^{(1)} - E(X_{i,t}^{(1)})\right) + \lambda_2\left(X_{i,t}^{(2)} - E(X_{i,t}^{(2)})\right)$$

$$\log \tau_t = m_l + \sum_{k=1}^{K_\pi} \vartheta(\omega_{if,1}, \omega_{if,2})\theta_{if,l}(X_{if^{USA},t-k} - X_{if^{INA},t-k}) + \sum_{k=1}^{K_\pi} \vartheta(\omega_{br,1}, \omega_{br,2})\theta_{br,l}(X_{br^{USA},t-k} - X_{br^{INA},t-k}) + \sum_{k=1}^{K_\pi} \vartheta(\omega_{RV,1}, \omega_{RV,2})\theta_{RV,l} RV_{t-k} + \sum_{k=1}^{K_\pi} \vartheta(\omega_{WTI,1}, \omega_{WTI,2})\theta_{WTI,l}(X_{WTI,t-k} - X_{WTI,t-k-1})$$

The estimation results of the GARCH-MIDAS-X model with the difference values are shown in Table 4. This model uses 3 specifications lag from predictors variables with different lengths, 6 months, 12 months, and 24 months.

Table 4. Estimation Result of GARCH-MIDAS-X Type I

Estimator	GARCH-MIDAS-X Type I		
	$k = 6$ months	$k = 12$ months	$k = 24$ months
$\hat{\mu}$	0,027691	0,027007	0,026656
$\hat{\alpha}$	0,144861	0,141882	0,14668
$\hat{\beta}$	0,748271	0,73354	0,707175
$\hat{\lambda}_1$	-0,010717	-0,03581	-0,031056
$\hat{\lambda}_2$	-4,991892	-8,648642	-12,835896
$\hat{m}$	0,433964	0,776265	0,777802
$\hat{\omega}_{if,2}$	10,0	1,01	10,0
$\hat{\theta}_{if}$	0,43761	0,60186	0,364348
$\hat{\omega}_{br,2}$	1,01	1,01	1,29996
$\hat{\theta}_{br}$	-0,605201	-2,018829	-3,150356
$\hat{\omega}_{WTI,2}$	1,01	1,01	2,71249
$\hat{\theta}_{WTI}$	-0,182683	-0,297599	-0,283153
$\hat{\omega}_{RV,2}$	1,948229	1,01	10,0
$\hat{\theta}_{RV}$	0,008961	0,01154	0,006114

Comparison between GARCH-MIDAS-X Type I models was carried out by looking at several model goodness criteria, AIC, BIC and HQIC, which can be seen in Table 5.

Table 5. Comparison of Model Criteria GARCH-MIDAS-X Type I

Criteria	GARCH-MIDAS-X Type I		
	$k = 6$ months	$k = 12$ months	$k = 24$ months
Loglike	-3493,2140	-3475,3302	-3476,6307
AIC	7014,4280	6978,6603	6981,2613
BIC	7097,0161	7061,2485	7063,8495
HQIC	7044,2971	7008,5295	7011,1305

Based on Table 5, it can be said that the model with the specifications lag $k = 12$ better than the other two specifications because it has a smaller AIC, BIC and HQIC value. Second, the return series is subsequently modeled using the GARCH-MIDAS-X Type II model. The details of the model specification are presented below.

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t} + \lambda_1 (X_{i,t}^{(1)} - E(X_{i,t}^{(1)})) + \lambda_2 (X_{i,t}^{(2)} - E(X_{i,t}^{(2)}))$$

$$\log \tau_t = m_v + \sum_{k=1}^{K_\pi} \vartheta(\omega_{if,1}, \omega_{if,2})\theta_{if,v}\sigma^2_{if,t-k} + \sum_{k=1}^{K_\pi} \vartheta(\omega_{br,1}, \omega_{br,2})\theta_{br,v}\sigma^2_{br,t-k} + \sum_{k=1}^{K_\pi} \vartheta(\omega_{RV,1}, \omega_{RV,2})\theta_{RV,v}RV_{t-k} + \sum_{k=1}^{K_\pi} \vartheta(\omega_{WTI,1}, \omega_{WTI,2})\theta_{WTI,l}\sigma^2_{WTI,t-k}$$
t

The results of the GARCH-MIDAS-X Type II model estimation with volatility values are shown in Table 6. This model also uses 3 specifications lag as in the Type I model.

Table 6. Estimation Result of GARCH-MIDAS-X Type II

Estimator	GARCH-MIDAS-X Type II		
	$k = 6$ months	$k = 12$ months	$k = 24$ months
$\hat{\mu}$	0,029151	0,029153	0,031374
$\hat{\alpha}$	0,142087	0,149038	0,143244
$\hat{\beta}$	0,7494	0,740529	0,708957
$\hat{\lambda}_1$	0,003303	0,00791	-0,00537
$\hat{\lambda}_2$	-2,790874	-1,655009	-7,36355
$\hat{m}$	-0,273979	-0,212816	0,12767
$\hat{\omega}_{if,2}$	1,01	2,386095	3,050414
$\hat{\theta}_{if}$	-1,14684	-1,359679	-4,070311
$\hat{\omega}_{br,2}$	3,959448	1,01	1,843141
$\hat{\theta}_{br}$	0,450402	-5,0	-5,0
$\hat{\omega}_{WTI,2}$	3,854074	1,01	1,01
$\hat{\theta}_{WTI}$	-10,0	-10,0	-10,0
$\hat{\omega}_{RV,2}$	1,830569	1,01	3,824451
$\hat{\theta}_{RV}$	0,017552	0,025101	0,016549

Comparison between GARCH-MIDAS-X Type II models was carried out by looking at several model goodness criteria, AIC and BIC, which can be seen in Table 7.

Table 7. Comparison of Model Criteria GARCH-MIDAS-X Type II

Criteria	GARCH-MIDAS-X Type II		
	$k = 6$ months	$k = 12$ months	$k = 24$ months
Loglike	-3492,8884	-3485,0545	-3474,2982
AIC	7013,7767	6998,1090	6976,5964
BIC	7096,3649	7080,6972	7059,1846
HQIC	7043,6459	7027,9782	7006,4656

Based on Table 7, it can be said that the model with the specifications lag $k = 24$ better than the other two specifications because it has a smaller criteria value.The results of volatility modeling with GARCH-MIDAS-X were then used for modeling *hybrid* GARCH-MIDAS-X and LSTM. Modeling was carried out with the best model, namely GARCH-MIDAS-X type I with the following specifications lag $k = 12$ and GARCH-MIDAS-X type II with specifications lag $k = 24$. The residuals from the model will be the target of the LSTM method. Input which is used for prediction is sequence from the residuals. The selection of LSTM is based on its ability to capture long-term temporal dependencies through a gating mechanism (forget, input, and output). The model parameters were optimized using the Adam optimizer with an initial learning rate of 1×10^{-3}.

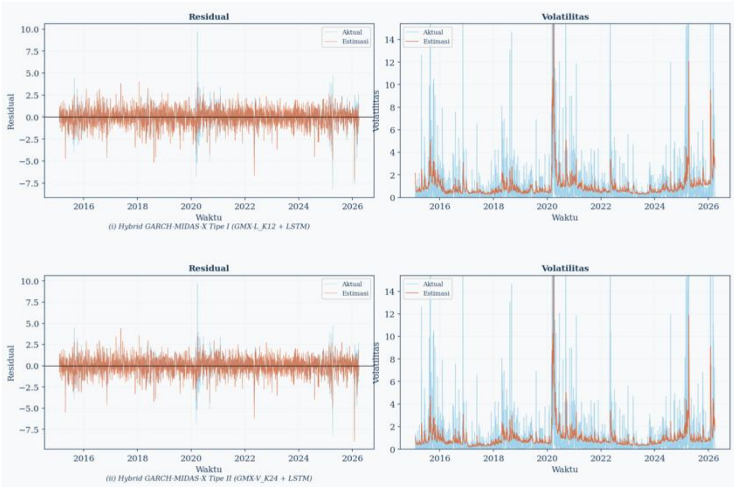

Figure 3. Modelling Result of Hybrid GARCH-MIDAS-X and LSTM

Forecasting and model comparison will be carried out with the best model from each model specification. In the GARCH-MIDAS-X type I model, the best model is the model with lag $k = 12$, while the GARCH-MIDAS-X type II model shows that the best model is the model with lag $k = 24$. Then the model hybrid GARCH-MIDAS-X and LSTM type I and models hybrid GARCH-MIDAS-X and LSTM type II. The four models will be compared in predicting data out-of-sample with RMSE and MAE criteria.

Table 8. Comparison of Prediction Model

Model	*Lag*	RMSE	MAE
GARCH-MIDAS-X Type I	$k = 12$	5,284553	2,896011
Hybrid LSTM with GARCH-MIDAS-X Type I	$k = 12$	5,267061	2,865617
GARCH-MIDAS-X Type II	$k = 24$	5,287606	2,907030
Hybrid LSTM with GARCH-MIDAS-X Type II	$k = 24$	5,287957	2,908127

Based on Table 8, the Hybrid LSTM with GARCH-MIDAS-X Type I model exhibits lower RMSE and MSE values than the GARCH-MIDAS-X Type I model. These results indicate that integrating LSTM with the GARCH-MIDAS-X Type I model improves its forecasting performance. However, the Hybrid LSTM with GARCH-MIDAS-X Type II model produces slightly higher RMSE and MAE values than the GARCH-MIDAS-X Type II model. This suggests that incorporating LSTM into the GARCH-MIDAS-X Type II model does not improve its forecasting performance.

CONCLUSION

Provide This study proposed a hybrid LSTM with GARCH-MIDAS-X for modelling the volatility of the Indonesia Composite Stock Price Index (IDX Composite) by incorporating market news sentiment, and exchange rates as exogenous variables. The GARCH-MIDAS-X framework was utilized to capture both short-run and long-run volatility dynamics, while the LSTM network was employed to model potential nonlinear relationships and temporal dependencies in the estimated volatility series. The empirical results show that the hybrid model achieved forecasting performance comparable to that of the standalone GARCH-MIDAS-X model. For the Type I specification, the hybrid model produced lower forecasting errors than the corresponding GARCH-MIDAS-X model, indicating that the LSTM component provided additional predictive information. In contrast, for the Type II specification, the hybrid model did not outperform the standalone GARCH-MIDAS-X model, suggesting that the inclusion of LSTM did not consistently improve forecasting accuracy across different model specifications. Although the proposed hybrid approach was able to capture the volatility dynamics adequately, its overall improvement in forecasting accuracy was relatively limited, implying that most of the predictive information had already been captured by the GARCH-MIDAS-X framework. Future research may investigate alternative deep learning architectures, incorporate additional exogenous variables, extend the observation period, or examine other financial markets to further assess the potential benefits of hybrid econometric–machine learning models for financial volatility forecasting.

REFERENCES

[1] P. Sitasi, : Aspiansyah, and S. Ayutyas, “Does Stock Market Lead to Economic Growth in Indonesia,” *Jurnal Pasar Modal dan Bisnis*, vol. 4, no. 2, pp. 59–76, 2022, doi: 10.37194/jpmb.v4i2.133.

[2] G. Sitorus, Y. A. L. S. Yolanda, and G. D. S. Gracia, “Perbandingan Kinerja Model GARCH dan LSTM dalam Memprediksi Volatilitas Harian IHSG,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 6, no. 3, pp. 523–529, Dec. 2025, doi: 10.37859/coscitech.v6i3.10741.

[3] V. Ratnasari and M. Nitivijaya, “Pemodelan Inflasi Di Indonesia Menggunakan Pendekatan Model Generalized Autoregressive Conditional Heteroskedasticity (GARCH),” *Inferensi*, vol. 1, no. 2, p. 71, Dec. 2018, doi: 10.12962/j27213862.v1i2.6727.

[4] R. F. Engle, “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, vol. 50, no. 4, p. 987, Jul. 1982, doi: 10.2307/1912773.

[5] T. Bollerslev, “Generalized Autoregressive Conditional Heteroskedasticity,” *J. Econom.*, vol. 31, no. 3, pp. 307–327, Apr. 1986, doi: 10.1016/0304-4076(86)90063-1.

[6] M. Yeasin, K. N. Singh, A. Lama, and R. K. Paul, “Modelling Volatility Influenced byExogenous Factors Using an Improved GARCH-X Model,” *Journal of The Indian Society of Agricultural Statistics*, vol. 74, no. 3, pp. 209–216, Oct. 2020.

[7] E. Endri, Z. Abidin, T. Simanjuntak, and I. Nurhayati, “Indonesian Stock Market Volatility: GARCH Model,” *Montenegrin Journal of Economics*, vol. 16, no. 2, pp. 7–17, Jun. 2020, doi: 10.14254/1800-5845/2020.16-2.1.

[8] M. F. F. Mardianto, E. Pusporani, D. Ulya, I. K. P. K. A. Putra, and R. Ramadhan, “SARIMAX–GARCH Model to Forecast Composite Index with Inflation Rate and Exchange Rate Factors,” *HighTech and Innovation Journal*, vol. 5, no. 3, pp. 743–758, Sep. 2024, doi: 10.28991/HIJ-2024-05-03-014.

[9] T.-H. Lee, “Spread and Volatility in Spot and Forward Exchange Rates,” *J. Int. Money Finance*, vol. 13, no. 3, pp. 375–383, Jun. 1994, doi: 10.1016/0261-5606(94)90034-5.

[10] E. Ghysels, A. Sinko, and R. Valkanov, “MIDAS Regressions: Further Results and New Directions,” *Econom. Rev.*, vol. 26, no. 1, pp. 53–90, Feb. 2007, doi: 10.1080/07474930600972467.

[11] R. F. Engle, E. Ghysels, and B. Sohn, “Stock Market Volatility and Macroeconomic Fundamentals,” *Review of Economics and Statistics*, vol. 95, no. 3, pp. 776–797, Jul. 2013, doi: 10.1162/REST_a_00300.

[12] E. N. N. Nortey, R. Agbeli, G. Debrah, T. Ansah-Narh, and E. F. Agyemang, “A GARCH-MIDAS Approach to Modelling Stock Returns,” *Commun. Stat. Appl. Methods*, vol. 31, no. 5, pp. 535–556, 2024, doi: 10.29220/CSAM.2024.31.5.535.

[13] L. D. Siagian and W. Makaliwe, “Investor Sentiment Dynamics and Market Volatility in Indonesia: Hybrid Approach Using GARCH-Midas and Machine Learning,” *Eduvest-Journal of Universal Studies*, vol. 5, no. 9, 2025, [Online]. Available: http://eduvest.greenvest.co.id

[14] A. Amendola, V. Candila, and G. M. Gallo, “Choosing the Frequency of Volatility Components Within the Double Asymmetric GARCH–MIDAS–X Model,” *Econom. Stat.*, vol. 20, pp. 12–28, Oct. 2021, doi: 10.1016/j.ecosta.2020.11.001.

[15] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[16] H. Y. Kim and C. H. Won, “Forecasting The Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-Type Models,” *Expert Syst. Appl.*, vol. 103, pp. 25–37, Aug. 2018, doi: 10.1016/j.eswa.2018.03.002.

[17] W. W. S. Wei, *Time Series Analysis : Univariate and Multivariate Methods*, Second Edition. Pearson Addison Wesley, 2006.

[18] Y. You and X. Liu, “Forecasting short-run exchange rate volatility with monetary fundamentals: A GARCH-MIDAS approach,” *J. Bank. Financ.*, vol. 116, Jul. 2020, doi: 10.1016/j.jbankfin.2020.105849.

[19] S. Shen, L. Xia, Y. Shuai, and D. Gao, “Measuring News Media Sentiment Using Big Data for Chinese Stock Markets,” *Pacific-Basin Finance Journal*, vol. 74, p. 101810, Sep. 2022, doi: 10.1016/j.pacfin.2022.101810.