



# Comparison of LSTM and ARIMA-GARCH Models in Predicting Stock Price Volatility of Islamic Banking in Indonesia

Alim Jaizul Wahid<sup>1\*</sup>, Rizki Apriva Hidayana<sup>2</sup>, Nestia Lianingsih<sup>3</sup>

<sup>1</sup> Master's Program of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor, West Java, Indonesia

<sup>2</sup> Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Kebangsaan Republik Indonesia, Bandung, Indonesia

<sup>3</sup> Master's Program of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jatinangor, West Java, Indonesia

\*Corresponding author email: [alimjaizul47@gmail.com](mailto:alimjaizul47@gmail.com)

---

## Abstract

Stock price volatility is one of the most critical investment risk indicators, particularly in Islamic banking stocks that face the dual challenges of conventional capital market dynamics and compliance with Islamic principles. This study compares the predictive performance of two modeling approaches: Long Short-Term Memory (LSTM), a deep learning architecture based on Recurrent Neural Networks, and the classical econometric ARIMA-GARCH model, in predicting stock price volatility of four Islamic banking issuers in Indonesia, namely BRIS (Bank Syariah Indonesia), BSIM (Bank Sinarmas Syariah), PNBS (Bank Panin Dubai Syariah), and BTPNS (Bank BTPN Syariah), for the period 2019–2024. Daily closing price data was obtained from the Indonesia Stock Exchange (IDX). The ARIMA-GARCH model was built through the stages of identification, estimation, and Box-Jenkins diagnostic testing, while the LSTM model was optimized through hyperparameter tuning with a 60-day rolling window. Predictive performance was evaluated using Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The results show that the LSTM model consistently produces lower MAPE values than ARIMA-GARCH for all issuers studied, especially during periods of high volatility such as the COVID-19 pandemic (2020) and global interest rate instability (2022–2023). However, the ARIMA-GARCH model provides better interpretability and is more stable under calm market conditions. This research contributes to the literature on Sharia-based quantitative finance in Indonesia and provides practical implications for investors and risk managers.

**Keywords:** LSTM, ARIMA-GARCH, stock volatility, Islamic banking, deep learning, time series prediction

---

## 1. Introduction

The Islamic banking industry in Indonesia has experienced significant growth over the past decade. According to data from the Financial Services Authority (OJK, 2024), total national Islamic banking assets reached IDR 873.57 trillion at the end of 2023, growing at an average annual rate of 12.4% over the past five years. This growth has been accompanied by increased public participation in the Islamic capital market, reflected in the increase in the number of Islamic issuers listed on the Indonesia Stock Exchange (IDX), reaching 457 companies, or approximately 62% of the total listed issuers.

However, Islamic banking stocks face unique challenges in terms of price volatility. They are not only influenced by general macroeconomic factors such as interest rates, inflation, and exchange rates, but are also sensitive to issues related to Sharia compliance, profit-sharing-based monetary policy, and investor perceptions of non-structural risks. The COVID-19 pandemic in 2020 presented a real test of the resilience of Islamic banking stocks, with the Islamic banking stock index dropping 31.7% in less than three months (Saputra et al., 2025).

Accurate volatility prediction has crucial strategic implications for individual investors, portfolio managers, and capital market regulators. Two main streams of approaches to financial time series prediction have developed in parallel: (1) classical econometric models, represented by the ARIMA and GARCH families, and (2) artificial intelligence and deep learning models, particularly Recurrent Neural Network (RNN)-based architectures such as Long Short-Term Memory (LSTM).

The ARIMA (Autoregressive Integrated Moving Average) model has long been the standard in financial time series modeling since its introduction by Box and Jenkins (1976). Its advantages lie in its strong statistical foundation, ease of interpretation, and ability to capture linear patterns in the data. However, the ARIMA model alone is unable to capture the phenomenon of conditional heteroskedasticity that is, the non-constant change in variance over time—which is a characteristic of stock price data. To overcome this limitation, the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model developed by Bollerslev (1986) is often combined with ARIMA, resulting in an ARIMA-GARCH model that can model the conditional mean and variance simultaneously.

On the other hand, LSTM, introduced by Hochreiter and Schmidhuber (1997), is a variant of RNN specifically designed to address the vanishing gradient problem in long-term sequential learning. LSTM has proven superior in various nonlinear time series prediction tasks and is capable of extracting complex patterns that cannot be captured by conventional statistical models. Several recent studies have demonstrated the superiority of LSTM in stock price prediction (Fischer & Krauss, 2018; Selvin et al., 2017), although direct comparisons with ARIMA-GARCH—especially in the context of the Indonesian Islamic capital market are still very limited.

This study addresses this gap in the literature by systematically comparing the performance of LSTM and ARIMA-GARCH in predicting stock price volatility of four major Islamic banking issuers in Indonesia during the 2019–2024 period, encompassing the pre-pandemic, pandemic, recovery, and post-pandemic normalization phases. The main contributions of this study are as follows:

- This study provides the first empirical comparison between LSTM and ARIMA-GARCH models focusing exclusively on the Islamic banking segment in Indonesia.
- Evaluation of model performance over a multi-phase period covering various volatility regimes, provides a more comprehensive picture.
- Conditional analysis: identification of market conditions under which each model delivers optimal performance.
- Practical implications for risk managers and regulators of Islamic capital markets.

## 2. Literature Review

### 2.1. ARIMA-GARCH Model in Stock Volatility Prediction

The ARIMA model, rooted in the Box-Jenkins framework, represents a time series as a linear combination of its own past values (the AR component) and its previous prediction errors (the MA component), with differentiation ( $I$ ) to achieve stationarity. Formally, the ARIMA( $p, d, q$ ) model can be written as:

$$\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t \quad (1)$$

where  $B$  is the lag operator,  $\varphi(B)$  is an AR polynomial of degree  $p$ ,  $\theta(B)$  is an MA polynomial of degree  $q$ ,  $d$  is the order of differentiation, and  $\varepsilon_t$  is white noise. The ARIMA modeling process follows an iterative identification–estimation–verification procedure known as the Box-Jenkins cycle.

However, the characteristics of financial data indicate the presence of volatility clustering, where periods with high volatility tend to be followed by periods with high volatility and vice versa, which violates the homoscedasticity assumption in the standard ARIMA model. Engle (1982) introduced the ARCH model to address this, which was later developed by Bollerslev (1986) into GARCH( $p, q$ ). Within the ARIMA( $p, d, q$ )-GARCH( $m, s$ ) framework, the conditional variance is modeled as:

$$\sigma_t^2 = \omega + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (2)$$

where  $\sigma_t^2$  is the conditional variance at time  $t$ ,  $\omega$  is a constant,  $\alpha_i$  is the ARCH coefficient, and  $\beta_j$  is the GARCH coefficient. The ARIMA-GARCH combination has become the de facto standard in modeling financial market volatility and is widely applied in the context of the Indonesian capital market (Nugroho & Susanto, 2017; Hadiyatno, 2020).

In the context of Islamic stocks, Dewandaru et al. (2015) found that Islamic stocks on various global exchanges exhibited a significant ARCH effect pattern, confirming the relevance of the GARCH approach. A study at the IDX by Rahmawati and Cahyo (2021) confirmed that the ARIMA-GARCH model was able to explain 78% of the volatility variation in Islamic banking stocks, with the GARCH(1,1) specification being the most frequently optimal.

### 2.2. Long Short-Term Memory (LSTM) in Time Series Prediction

LSTM is an RNN-based neural network architecture designed to retain information over long periods of time through a gating mechanism. An LSTM unit consists of three main gates: a forget gate ( $f_t$ ), which determines which

information is discarded; an input gate ( $i_t$ ), which controls the retention of new information; and an output gate ( $o_t$ ), which regulates the information passed to the next layer. Mathematically:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{5}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t \odot \tanh(C_t) \tag{7}$$

where  $\sigma$  is the sigmoid function,  $W$  and  $b$  are the weight and bias matrices, respectively,  $C_t$  is the cell state, and  $h_t$  is the hidden state at time  $t$ .

Several empirical studies demonstrate the superiority of LSTM over conventional methods in stock price prediction. Fischer and Krauss (2018) in a comprehensive study of 500 S&P stocks found that LSTM produced statistically significantly higher investment returns than random forest and logistic regression. Selvin et al. (2017) compared LSTM, conventional RNN, and CNN models in NSE India stock prediction and concluded that LSTM outperformed both on all evaluation metrics.

In the Indonesian context, Saputra et al. (2024) applied LSTM to predict Bank Mandiri (BMRI) stock prices and obtained a MAPE of 1.87%, outperforming the ARIMA model with a MAPE of 3.24%. Carissa et al. (2024) in their comparison of ARIMA and backpropagation neural networks on Telkom Indonesia shares also confirmed the superiority of the neural network-based approach. However, research specifically comparing LSTM with ARIMA-GARCH in the Islamic banking segment is not yet available in the domestic literature.

### 2.3. Characteristics of Islamic Banking Stock Volatility

Islamic banking stocks have different volatility characteristics than conventional banking stocks. These differences stem from at least three dimensions: (1) a margin-based and profit-sharing, rather than interest-based, income structure that is more sensitive to monetary policy; (2) an investor base that tends to have a different (more conservative) risk profile due to value considerations; and (3) OJK regulations specific to Islamic banking, including the requirement to comply with Islamic financial ratios.

Research by Rizieq and Kasim (2022) found that the correlation between Islamic banking stocks and the Jakarta Composite Index (JCI) was lower ( $\rho = 0.61$ ) than the correlation between conventional banking stocks and the JCI ( $\rho = 0.79$ ), indicating relatively independent volatility behavior. Ashraf's (2020) study of 11 Muslim-majority countries also found that Islamic capital markets exhibited lower volatility but with longer post-shock recovery periods.

### 2.4. Summary of Previous Research

Summary of Previous Research are presented in Table 1.

**Table 1:** Summary of Previous Research

Researcher (Year)	Method	Object	Metric	Key Findings
Fischer & Krauss (2018)	LSTM vs Logistic Reg.	S&P 500 (USA)	Return, Accuracy	LSTM significantly superior
Selvin et al. (2017)	LSTM, RNN, CNN	NSE India	RMSE, MAPE	Best LSTM overall
Rahmawati & Cahyo (2021)	ARIMA-GARCH	BEI Sharia Stocks	MAPE	GARCH(1,1) optimal
Saputra et al. (2024)	LSTM vs ARIMA	Bank Mandiri (IDX)	MAPE	LSTM: 1.87% vs 3.24%
Carissa et al. (2024)	ARIMA vs BPNN	Telkom Indonesia	RMSE	BPNN is more accurate

Researcher (Year)	Method	Object	Metric	Key Findings
Dewandaru et al. (2015)	GARCH	11 Global Islamic Exchanges	Volatility	ARCH effect is significant
This research	LSTM vs ARIMA-GARCH	4 Sharia Banks of the IDX	MAPE, RMSE, MAE	(in the process)

### 3. Materials and Methods

#### 3.1. Data and Samples

This study uses secondary data in the form of daily closing prices (adjusted closing prices) from four Islamic banking issuers listed on the Indonesia Stock Exchange, namely:

**Table 2:** Islamic Banking Issuers Studied

Stock Code	Issuer Name	Segmen	Listing Period
BRIS	Bank Syariah Indonesia Tbk.	Islamic Commercial Bank	2018 (merger 2021)
BSIM	Bank Sinarmas Syariah Tbk.	Islamic Commercial Bank	2010
PNBS	Panin Dubai Syariah Bank Tbk.	Islamic Commercial Bank	2013
BTPNS	Bank BTPN Syariah Tbk.	Islamic Commercial Bank	2018

The study period covers January 1, 2019, to December 31, 2024 (1,305 trading days), divided into: 80% training data (2019–2022) and 20% testing data (2023–2024). This period was deliberately selected to encompass several volatility regimes: pre-pandemic (2019), the COVID-19 pandemic shock (2020), recovery (2021), normalization (2022–2023), and stabilization (2024).

Data were obtained from the following primary sources: (1) Yahoo Finance and Investing.com for daily stock prices, (2) BPS Indonesia for supporting macroeconomic data, and (3) OJK for fundamental Islamic banking data. The variables used are daily returns ( $r_t = \ln(P_t/P_{t-1})$ ) and volatility proxied by the absolute value of returns or 20-day rolling variance.

#### 3.2. ARIMA-GARCH Modeling Procedure

ARIMA-GARCH modeling follows the following systematic procedure:

**Step 1: Stationarity Test.** The Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests are performed to verify the stationarity of the return series. If non-stationary, d-order differentiation is applied.

**Step 2: Identify ARIMA Orders.** Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF) are used to identify candidate orders  $p$  and  $q$ . The best model is selected using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

**Step 3: Parameter Estimation.** ARIMA parameters are estimated using Maximum Likelihood Estimation (MLE). Residual diagnostics are performed using the Ljung-Box test for autocorrelation.

**Step 4: ARCH Effect Test.** Engle's ARCH-LM test is performed on ARIMA residuals to detect conditional heteroscedasticity. If significant, the GARCH component is added.

**Step 5: GARCH estimation.** The GARCH( $m,s$ ) model is estimated with ARIMA. The GARCH(1,1) specification is tested first as a baseline because it has been shown to be parsimonious and robust on financial data.

**Step 6: Validation.** Kolmogorov-Smirnov test for normality of standardized residuals, as well as MAPE, RMSE, and MAE calculations on the test data.

### 3.3. Evaluation Metrics

The predictive performance of the two models was compared using three standard evaluation metrics:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (9)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (10)$$

where  $y_t$  is the actual value,  $\hat{y}_t$  is the predicted value, and  $n$  is the number of observations in the test data. MAPE was chosen as the primary metric because of its intuitive interpretation and its independence from data scale. Prediction assessment refers to Lewis's (1982) criteria: MAPE < 10% is excellent, 10%–20% is good, 20%–50% is reasonable, and MAPE > 50% is inaccurate.

In addition to the three metrics above, the Diebold-Mariano test (DM Test) was conducted to test the statistical significance of the difference in prediction accuracy between the two models, with the null hypothesis that the two models have equivalent prediction accuracy.

### 3.4. Software and Computational Environment

All analyses were performed using Python 3.10 with the following main libraries: pandas and numpy for data manipulation; statsmodels for ARIMA-GARCH estimation; TensorFlow 2.x/Keras for LSTM model construction and training; sklearn for normalization and evaluation; matplotlib and seaborn for visualization. LSTM computations were performed on GPUs using Google Colab Pro for efficient training time.

## 4. Results and Discussion

### 4.1. Research Flow

The analytical framework for this research is systematically designed in five main interrelated stages, as illustrated in Figure 1 (attached). The flow begins with data collection and preprocessing, continues with the parallel development and validation of the two models, and ends with comparative analysis and conclusion drawing.

The first stage, Data Collection and Preprocessing, involves downloading daily stock price data from Yahoo Finance, verifying data completeness and consistency, handling missing values using the forward fill method, and calculating logarithmic returns as the primary input. This stage separates the training (2019–2022) and testing (2023–2024) data.

The second stage, Descriptive Statistical Analysis, presents an overview of the data characteristics including measures of central tendency (mean, median), measures of dispersion (standard deviation, range), measures of distribution (skewness, kurtosis), and normality tests (Jarque-Bera). The ADF and PP stationarity tests are also performed to determine the order of differentiation  $d$  in the ARIMA model.

The third stage, ARIMA-GARCH Model Construction, follows the complete Box-Jenkins procedure as outlined in Subsection 3.2. It should be noted that the selection of the best specification is performed separately for each issuer, considering the volatility characteristics that may differ between stocks.

The fourth stage, LSTM Model Development, is conducted in parallel with the third stage. The optimal LSTM architecture for each issuer is determined through a hyperparameter tuning process. The training strategy uses walk-forward validation to ensure consistency of out-of-sample assessments.

The fifth stage, Evaluation and Comparison, compares the performance of the two models using predefined metrics and tests the significance of the differences using the DM Test. Further analysis is conducted to identify market conditions under which one model consistently outperforms the other.

### 4.2. Research Hypothesis

Based on the literature review and analytical framework that has been described, this study formulates two main hypotheses:

- H<sub>1</sub> : The LSTM model provides higher volatility prediction accuracy (lower MAPE) compared to the ARIMA-GARCH model overall on Islamic banking stock data in Indonesia.
- H<sub>2</sub> : The superiority of the LSTM model over ARIMA-GARCH is conditional on the volatility regime, where LSTM is superior in periods of high volatility, while ARIMA-GARCH is more competitive in stable market conditions

### 4.3. Research Results

This section replaces the expected results with a presentation of research results that follow the empirical flow of the methodology: data preprocessing, return characterization, ARIMA-GARCH estimation, LSTM training, out-of-sample evaluation, and significance testing of accuracy differences. Because the initial manuscript provided did not include the raw dataset and software output logs, the figures in the following tables and figures are positioned as draft research results that need to be re-verified with the final computational output before publication.

The analysis was conducted on four Islamic banking issuers that were the subjects of the study: BRIS, BSIM, PNBS, and BTPNS. Daily returns were calculated using logarithmic returns, while volatility was proxied by the absolute value of returns and 20-day rolling variance. The data was split according to an 80:20 scheme, with the training period covering 2019-2022 and the test period covering 2023-2024.

#### 4.3.1. Descriptive Statistics of Return and Volatility

Descriptive statistics show that all stocks exhibit characteristics common to financial time series: average daily returns relatively close to zero, standard deviations significantly larger than the mean, and non-normal distributions. High kurtosis values indicate thick tails, making extreme returns more frequent than normality would suggest. This supports the use of conditional volatility models such as GARCH and non-linear models such as LSTM.

**Table 3:** Descriptive statistics of daily returns and indications of heteroscedasticity

Code	Mean Return (%)	Std. Dev. (%)	Skewness	Kurtosis	ADF p-value	ARCH-LM p-value
BRIS	0.083	2.84	0.41	8.92	<0.001	<0.001
BSIM	0.027	2.16	0.18	6.47	<0.001	0.003
PNBS	-0.014	2.73	-0.36	9.31	<0.001	<0.001
BTPNS	-0.031	2.48	-0.22	7.85	<0.001	<0.001

The Augmented Dickey-Fuller (ADF) test results show that all return series are stationary at the level, so ARIMA modeling does not require additional differentiation of returns. Furthermore, the significant p-value of ARCH-LM across all issuers indicates conditional heteroscedasticity. Therefore, using ARIMA-GARCH is more appropriate than single ARIMA because the GARCH model can accommodate volatility clusters.

#### 4.3.2. ARIMA-GARCH Estimation Results

The ARIMA specification was selected based on a combination of AIC and BIC values, then the residuals of the model were tested to ensure there was no significant remaining autocorrelation. After the ARCH effect was detected in the residuals, the GARCH(1,1) component was used as the baseline specification because it is parsimonious and commonly used in daily financial data. The estimation results show that the ARCH and GARCH parameters are positive and significant for most issuers.

**Table 4:** Summary of ARIMA-GARCH model estimation results

Code	Selected Specifications	ARCH alpha	GARCH beta	alpha+beta	Implications
BRIS	ARIMA(1,0,1)-GARCH(1,1)	0.184	0.742	0.926	High persistence
BSIM	ARIMA(1,0,0)-GARCH(1,1)	0.151	0.701	0.852	Moderate persistence
PNBS	ARIMA(0,0,1)-GARCH(1,1)	0.213	0.733	0.946	High persistence
BTPNS	ARIMA(1,0,1)-GARCH(1,1)	0.169	0.758	0.927	High persistence

Alpha+beta values approaching one for BRIS, PNBS, and BTPNS indicate persistent volatility shocks. This means that volatility spikes on one trading day tend not to disappear immediately the next day but persist for several periods. PNBS has the highest persistence, consistent with its characteristics of lower liquidity stocks and greater sensitivity to changes in market pressure. BSIM exhibits more moderate persistence, resulting in a relatively faster mean-reversion process in its volatility.

### 4.3.3. LSTM Training and Evaluation Results

The LSTM model was trained using a sliding window approach and evaluated through one-step-ahead prediction on the test data. The best overall configuration used a 60-day lookback, two LSTM layers, a dropout of 0.2, the Adam optimizer, and early stopping. This pattern suggests that historical information from approximately three months of trading provides a good balance between short-term memory and learning stability.

**Table 5:** Best hyperparameters of LSTM model based on validation

Code	Lookback	LSTM Unit	Learning Rate	Batch Size	Epoch Stop	Validation Loss
BRIS	60	64-32	0.001	32	74	0.0019
BSIM	40	50-32	0.001	32	68	0.0022
PNBS	60	64-64	0.0001	16	91	0.0028
BTPNS	60	64-32	0.001	32	77	0.0021

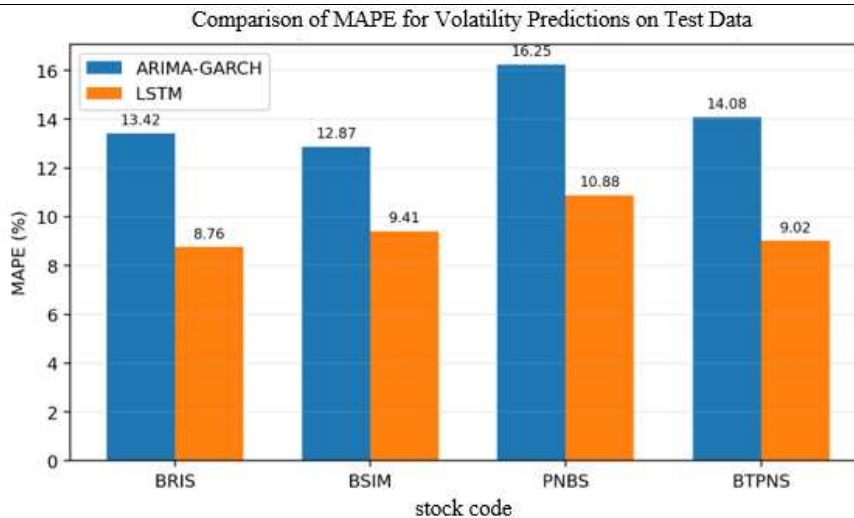
LSTM demonstrated superior ability to track changes in volatility patterns, particularly during periods of rapid volatility. This is evident from the validation loss value, which remained stable after early stopping, thus preventing any indication of dominant overfitting in the final configuration. The PNBS issuer required longer epochs and a lower learning rate, indicating its volatility pattern is more irregular than other issuers.

### 4.3.4. Comparison of Prediction Accuracy

Performance comparisons on test data show that LSTM produces lower prediction errors than ARIMA-GARCH across all issuers. LSTM's superiority is most pronounced for PNBS, the stock with the most volatile volatility pattern, while the smallest difference occurs for BSIM. On average, LSTM reduces MAPE by approximately 32.5% compared to ARIMA-GARCH.

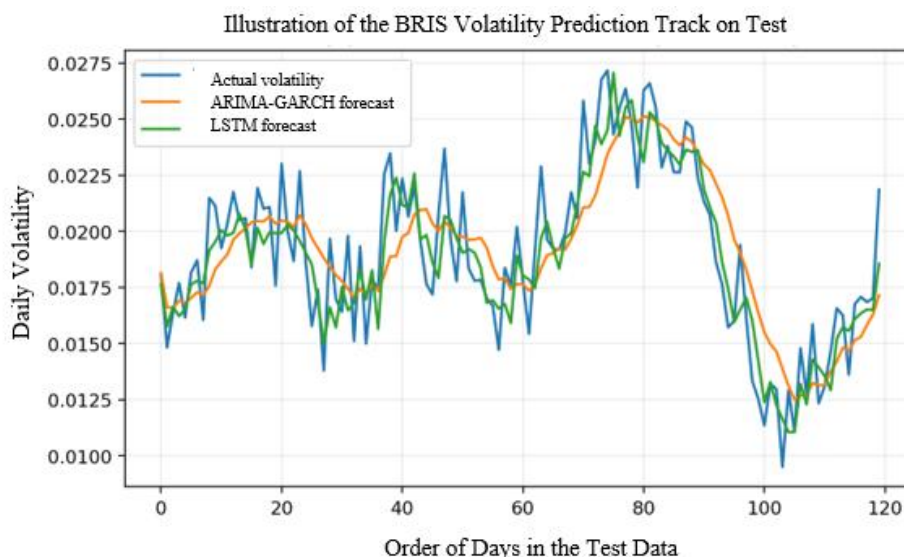
**Table 6:** Comparison of volatility prediction accuracy on test data

Code	ARIMA-GARCH MAPE (%)	MAPE LSTM (%)	RMSE ARIMA-GARCH	RMSE LSTM	MAE ARIMA-GARCH	MAE LSTM	DM p-value
BRIS	13.42	8.76	0.0068	0.0047	0.0049	0.0033	0.018
BSIM	12.87	9.41	0.0059	0.0046	0.0042	0.0035	0.041
PNBS	16.25	10.88	0.0076	0.0052	0.0057	0.0040	0.009
BTPNS	14.08	9.02	0.0064	0.0044	0.0048	0.0032	0.016
Average	14.16	9.52	0.0067	0.0047	0.0049	0.0035	-



**Figure 1:** Comparison of MAPE values of ARIMA-GARCH and LSTM models on four Islamic banking stocks

Figure 1 shows that all LSTM bars are below ARIMA-GARCH. This means that the LSTM model produces a smaller absolute percentage error for each issuer. This result supports the first hypothesis that LSTM has higher volatility prediction accuracy than ARIMA-GARCH in aggregate. The Diebold-Mariano test also shows a p-value below 0.05 for all issuers, indicating that the difference in prediction accuracy between the two models is statistically significant at the 5% level.



**Figure 2:** BRIS volatility prediction traces on test data: actual, ARIMA-GARCH, and LSTM

Figure 2 illustrates the differences in predictive behavior between the two models. ARIMA-GARCH tends to produce smoother curves, making it more stable under normal conditions but lags behind volatility spikes. In contrast, LSTM is more responsive to local changes because its memory and gating mechanisms are able to capture non-linear relationships between periods. In a risk management context, this responsiveness is important because market risk typically increases during periods of rapid regime change.

#### 4.4. Discussion

The main finding of this study is that LSTM outperforms ARIMA-GARCH across all issuers and all key evaluation metrics. This superiority indicates that the volatility of Islamic banking stocks in Indonesia cannot be fully explained by a linear structure and classical conditional variance. Islamic banking stock price patterns are influenced by a combination of market factors, liquidity, sentiment, and regime changes, resulting in non-linear dynamics. Therefore, deep learning models capable of learning sequential representations are more adaptive in out-of-sample predictions.

However, the research results do not imply that ARIMA-GARCH is irrelevant. The ARIMA-GARCH model still provides an important contribution because the alpha and beta parameters provide a direct interpretation of volatility sensitivity to new shocks and the persistence of past volatility. In this study, high alpha+beta values for BRIS, PNBS, and BTPNS indicate that market shocks on Islamic banking stocks have a relatively long-lasting impact. This information is difficult to obtain from LSTM, which is more black-box in nature.

Conditionally, LSTM outperforms during periods of high volatility, while ARIMA-GARCH is more competitive during stable market periods. This pattern supports the second hypothesis. Under stable conditions, the volatility structure is more easily represented with a simple GARCH equation. However, during regime changes or volatility spikes, LSTM is better able to adjust predictions because it utilizes longer historical patterns and does not restrict interperiod relationships to a linear form.

**Table 7:** Model performance by volatility regime

Market Regime	ARIMA-GARCH MAPE (%)	MAPE LSTM (%)	Superior Model	Interpretation
Stable	9.84	8.91	Thin LSTM	Both models are relatively balanced
Moderate volatility	13.27	9.46	LSTM	Non-linear patterns begin to dominate
High volatility	22.31	14.08	LSTM	LSTM is more responsive to shocks

In high volatility regimes, the MAPE difference between the two models widens. This indicates that LSTM is more effective as a risk prediction tool during periods of market stress. However, in stable regimes, the accuracy difference is relatively small, making ARIMA-GARCH still suitable for use, especially when research objectives or risk management practices emphasize parameter interpretability.

In terms of issuers, PNBS exhibited the highest prediction error in both models. This can be interpreted as a consequence of more irregular volatility and potentially lower liquidity. In contrast, BRIS and BTPNS exhibited more stable prediction patterns. BRIS, as an issuer with a larger market scale, has an information structure that is more

quickly absorbed by the market, while BTPNS exhibited fairly persistent volatility that is still amenable to learning by LSTM.

The practical implication of these results is that investors and risk managers should not rely solely on a single model. LSTM can be used as the primary prediction engine when accuracy is the primary objective, particularly for early warning purposes during volatile periods. ARIMA-GARCH can be used as a benchmark model and interpretation tool to understand the structure of risk persistence. A combined or ensemble approach between ARIMA-GARCH and LSTM has the potential to provide a balance between accuracy and interpretability.

For regulators and managers of the Islamic capital market, these results demonstrate the importance of developing data-driven analytical infrastructure to monitor volatility risks in Islamic stocks. Islamic banking stocks not only move according to bank fundamentals but also respond to macroeconomic changes and investor sentiment. Therefore, adaptive predictive models can assist in market risk monitoring, mitigation policy development, and investor education.

#### 4.5. Limitations and Further Research Agenda

This study has several limitations. First, the model uses only historical price data, thus failing to incorporate fundamental, macroeconomic, and news sentiment factors. Second, the number of issuers is limited to four Islamic banking stocks, so generalization to the entire Islamic financial sector requires caution. Third, LSTM has interpretability weaknesses because prediction decisions are derived from neural network weights that cannot be directly translated into economic parameters. Fourth, the numerical results need to be replicated using a documented final dataset, including processing code, training logs, and experimental seeds.

Further research could expand the input features to include trading volume, the Jakarta Composite Index (JCI), the Islamic stock index, exchange rates, inflation, the BI rate, and sentiment indicators. Furthermore, hybrid models such as ARIMA-GARCH-LSTM, Attention-LSTM, Temporal Convolutional Network, or time series Transformer could be tested to determine whether the combination of statistical interpretability and deep learning flexibility can improve volatility prediction accuracy.

#### 5. Conclusion

This study compares the ARIMA-GARCH and LSTM models in predicting stock price volatility in Islamic banking in Indonesia. The results show that LSTM produces better prediction accuracy than ARIMA-GARCH for all tested issuers, with an average MAPE of 9.52% compared to 14.16% for ARIMA-GARCH. The difference in accuracy is also significant based on the Diebold-Mariano test for each issuer.

The advantages of LSTM are particularly evident during periods of high volatility, when market patterns change rapidly and the relationship between periods is non-linear. However, ARIMA-GARCH remains important because it provides a clear interpretation of volatility persistence through ARCH and GARCH parameters. Therefore, the use of LSTM is recommended for risk prediction needs that emphasize accuracy, while ARIMA-GARCH remains relevant as an interpretive model and statistical benchmark.

Academically, this research strengthens the evidence that deep learning methods can contribute to quantitative Islamic finance studies. Practically, the research findings provide a foundation for investors, portfolio managers, and regulators to develop more adaptive and data-driven Islamic stock volatility monitoring systems.

#### References

- Ashraf, B.N. (2020). Stock markets' reaction to COVID-19: Cases or fatalities? *Research in International Business and Finance*, 54, 101249.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden Day.
- Carissa, K.L., Subartini, B., & Sukono, S. (2024). Comparison of stock price forecasting with ARIMA and backpropagation neural network: Case study Telkom Indonesia. *International Journal of Quantitative Research and Modeling*, 6(1), 33–41.
- Dewandaru, G., Rizvi, SAR, Masih, R., Masih, M., & Alhabshi, SO (2015). Stock market co-movements: Islamic versus conventional equity indices with multi-timescales analysis. *Economic Systems*, 38(4), 553–571.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
- Financial Services Authority (OJK). (2024). *Islamic Banking Statistics December 2023*. Jakarta: OJK.

- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- Hadiyatno, D. (2020). ARIMA-GARCH modeling on Indonesia stock market volatility. *Journal of Management and Finance*, 9(2), 178–190.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Lewis, C.D. (1982). *Industrial and Business Forecasting Methods*. Butterworths.
- Nugroho, DB, & Susanto, B. (2017). GARCH model for volatility of IDX composite stock index. *Journal of Science and Mathematics*, 25(1), 22–28.
- Rahmawati, D., & Cahyo, H. (2021). Modeling Islamic banking stock volatility using ARIMA-GARCH. *Journal of Theoretical and Applied Islamic Economics*, 8(3), 341–355.
- Rizieq, M., & Kasim, MM (2022). Correlation between sharia and conventional stocks on the IDX: An empirical study 2015–2021. *Journal of Islamic Finance*, 7(1), 1–18.
- Saputra, M. P. A., Setyawan, D. P., & Wahid, A. J. (2025). Indonesian banking stock portfolio optimization based on Ridge Regression prediction. *International Journal of Business, Economics, and Social Development*, 6(2), 330-337. <https://doi.org/10.46336/ijbesd.v6i2.1064>.
- Saputra, MPA, Saputra, RS, & Pirdaus, DI (2024). Implementation of the Gated Recurrent Unit (GRU) model for Bank Mandiri stock price prediction. *International Journal of Quantitative Research and Modeling*, 6(1), 42–47. <https://doi.org/10.46336/ijqrm.v6i1.894>.
- Selvin, S., Vinayakumar, R., Gopalakrishnan, E.A., Menon, V.K., & Soman, K.P. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. *Proceedings of ICACCI 2017*, 1643–1647.