Prediction of Volatility in Stock Commodities using Deep Learning

Fatimah Abdullah Al Saleem*

Information Systems Department, College of Computer and Information Sciences,
Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia.

Abstract

Investors face risks throughout the investment process, and managing these risks depends on the volatility of stock commodities. Understanding how to calculate and predict the volatility of stock commodities can greatly assist in forecasting future fluctuations. This research aims to assess the effectiveness of deep learning, specifically the long short-term memory (LSTM) model, in predicting volatility in stock commodities. Additionally, the research work aims to determine which model produces the best results for stock investors in forecasting volatility by comparing the GARCH model, ARCH model, and LSTM (deep learning model). The research work utilizes Yahoo Finance datasets for oil, gold, diesel, and the S&P 500 index to conduct the stock market analysis. The results of this research work show that the LSTM model achieved a high accuracy rate of 98.8%, outperforming the GARCH model at 94% and the ARCH model at 89%. Therefore, the predictive power of the LSTM model surpasses that of the ARCH and GARCH models, establishing the effectiveness of deep learning models in predicting volatility. Furthermore, the GARCH model demonstrates superior predictive power compared to the ARCH model.

Keywords: Deep Learning, ARCH, GARCH, LSTM, Forecasting, Stock Commodities, S&P 500, Yahoo Finance.

ARK: https://n2t.net/ark:/47543/JISCOM2023.v4i1.a33
Received August 21, 2023, accepted September 16, 2023, date of publication October 10, 2023

1. Introduction

Investors put most of their focus on monitoring the volatility of market variables before they enter the market or during the investing process. Volatility represents the risk indicator which is used to manage the risk and evaluate the performance [1-2]. Volatility occurs by ongoing social, economic, and political factors that directly affect the stock commodities markets [3]. So, to be able to manage the risk, you must have good forecasting of the volatility of stock commodities [4]. There are multiple methods used to understand and forecast volatility. The generalized autoregressive conditional heteroscedasticity (GARCH) model is widely used to forecast the volatility of stock commodities and other variables. The GARCH model was the result of improving the autoregressive conditional heteroscedasticity (ARCH) model [1-2]. The GARCH model solved some issues and improved some disadvantages of the ARCH model. The GARCH model can deal with different types of financial data by understanding and analyzing them [5-6]. There is another way to forecast the volatility of stock commodities using machine

* Corresponding author.
E-mail address: 442022625@sm.imamu.edu.sa

learning. Machine learning works more on the data-driven, but the GARCH model works more with economic assumptions and statistical logic [2]. Recently deep learning has taken the most attention in the machine learning field. Deep learning performs good work in multiple areas such as image classification and speech recognition [7-8]. However long-short term memory (LSTM) is an effective deep learning architecture. There are different studies, some studies created a hybrid model for forecasting volatility by combining LSTM and GARCH model. Moreover, other researchers combined recurrent neural networks with LSTM and compared them to the GARCH model, the result shows that the combination outperformed the GARCH model [9][2]. Also, some researchers compared the ARCH model with the GARCH model without comparing them with any type of machine learning [10].

The contributions of this research work are as follows:

- The first contribution is identifying the effectiveness of deep learning using the long-short-term memory (LSTM) model to predict volatility in stock commodities.

- The second contribution is determining which model can get the best results for stock investors to forecast the volatility in stock commodities by comparing between the GARCH model, ARCH model, and long-short-term memory (LSTM) (Deep learning model).

For this research work, the Yahoo finance stock/commodity ticker symbol at the dataset for gold is “GOLD”, oil is “CL=F”, diesel is “TR” and S&P 500 is “^GSPC” they are used to make the study of the stock market [35]. The dataset for the research work will be on a daily basis from 23/08/2000 to 01/11/2022 to be studied.

2. Literature Review

Forecasting is useful in daily life since it aids in decision-making. In order for investors to make trading decisions, the volatility forecasts in stock commodities are very crucial. This section shows the related works in the first part. Also, it explains the most important concepts in this research work from the second to the last part.

2.1. Introduction

In this section, comparisons between two or more models and their various offspring from various research are shown. Numerous variables, including stock prices, market indices, and other areas, influence forecasting. As a result, these various LSTM comparisons with various financial model types are shown below.

There are many researchers working in the topic of stock pricing. Similar to this study effort [7], which combined Long-Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN), and Extreme Learning Machines (ELM) to construct system-based solutions with various machine learning models for stock price forecasting. By splitting each machine learning model into two parts—two layers and three layers—they were able to construct fourteen models. They found that using a mixture of CNN's three hidden layers, GRU's three hidden layers, and ELM's two hidden layers, LSTM performs better with two hidden levels. Additionally, the authors of [11] paired deep learning methodologies with multiresolution analysis to investigate how this pairing may increase the precision of financial time-series forecasting. The findings indicated that deep learning models have greater predicting accuracy. Additionally, while forecasting stock prices, M. A. Istiake Sunny, M. M. S. Maswood, and A. G. Alharbi [12] contrasted the LSTM model to the Bi-Directional Long Short-Term Memory (BILSTM) model. The outcomes revealed that the BI-LSTM model performed better than the LSTM model. In time series forecasts for financial data, this research study [13] examined the effectiveness of LSTM with Autoregressive Integrated Moving Average (ARIMA). Due to LSTM's reduced RMSE, the findings indicated that it was superior to ARIMA. To get the greatest outcomes, they were built on an iteration-based foundation. Additionally, the researchers in their study [15] contrasted machine learning and deep learning models for predicting market prices. The outcomes demonstrated that LSTM models could address and resolve issues. On the other hand, multivariate regression and random forest regression were the machine learning models that were the most accurate. However, the study was created by the researchers [20] to demonstrate how well LSTM and convolutional neural networks (CNNs) forecast market values. The researchers discovered that while both models produced correct results, convolutional LSTM was the most accurate. In contrast, the CNN model processed the data more quickly than the convolutional LSTM model.

In other works which depend on the stocks index field, the authors [5] assessed the predictive power of LSTM, GARCH and EWMA in volatility based on the accuracy of the forecasting at the one-month ahead realized volatility. The results showed LSTM was the more accurate between them. On another hand, Jia, F., & Yang, B [2] trained the deep learning (DL) models which were LSTM and DNN using likelihood-based loss function to forecast the volatility of stock index. They chose the likelihood-based loss function for (DL) to compare with economic models fairly, because they discovered that using the estimated volatility causes errors in the forecasting. So, the results showed that the two deep learning models were better than the economic model (ARMA-GARCH) and the better model was LSTM. In another work [21] the researchers compared the performance of LSTM to the performance of Bidirectional LSTM (BiLSTM) and to the ARIMA model. The BiLSTM model had the ability to train the data on both sides from left to right and the opposite direction from right to left. So, the researchers wondered if this ability of BiLSTM can enhance the accuracy of time series forecasting. The researchers found that using both directions in training the data could enhance the accuracy of the time series forecasting. They discovered more features of the BiLSTM model which made it more accurate than the LSTM and ARIMA models.

However, some researchers depend on other fields in their forecasting. So, the researchers made this study [9] to assess the machine learning models were lasso, Random Forest, Gradient Boosting, and Long Short-Term Memory. The assessment will be dependent on their performance in forecasting daily realized volatility of returns. Besides, the researchers targeted the Russian stock market to be studied. After the assessment, the researchers compared these models with the heterogeneous autoregressive realized volatility (HAR-RV) model. The results showed that Lasso and HAR-RV work better in forecasting than LSTM, Gradient Boosting, Random Forest. On the other hand, the authors [14] compared the LSTM model with ARIMA, SARIMA, ARIMAX in short term prediction. The results showed that LSTM had better performance than the others. However, there was a negative relationship between the number of hidden layers and the accuracy of forecasting. While T. Song, J. Jiang, W. Li and D. Xu, [16] applied machine learning in different fields. They proposed Merged LSTM for forecasting the sea surface height anomaly and then compared it to the Stacked-LSTM module, ANN, merged-RNN, TCN, merged-GRU, and 1-D CNN. The results showed that Merged LSTM outperformed the other models on three aspects: performance, stability, and accuracy. Whereas this research work [17] discovered there were problems in photovoltaic (PV) which made the forecast process for PV difficult. So, they decided to use LSTM to create a short-term forecasting model. They compared the LSTM model with three models: ARIMA, SVR and NN models. They found that the LSTM model was the strongest in predicting and had low solar irradiance. Furthermore, the authors made a study about photovoltaic (PV) power here [18] to see which model was better at predicting photovoltaic power. They proposed a long short-term memory recurrent neural network (LSTM-RNN) and compared it to three traditional PV predicting methods. They got the best LSTM model from the suggested five models in predicting the PV power. The workers into [19] focused on intelligent transportation systems. They studied the Traffic flow forecast problem by proposing a LSTM network with combined two domains in the road network. They found that the LSTM network outperformed the SAE, RBF, SVM and ARIMA model in long term forecasting.

Table 1 summarizes the reference number, publication year, forecasting target, models compared, and the most accurate model identified in each study. The comparisons involve various financial models, and in many cases, LSTM models demonstrate better performance and accuracy compared to other models. However, the specific results may vary depending on the forecasting target and dataset used in each study.

Table 1. The summary of the related works.

Reference number	**Publication year**	**Forecasting target**	**Models**	**The most accurate model**
[7]	2018	Stocks prices	Fourteen models for system based on LSTM, GRU, ELM and CNN	LSTM works better with 2 hidden layers with combination of CNN 3 hidden Layers, GRU 3 hidden Layers and ELM 2 hidden layers.
[11]	2021	Stocks prices	DL-EWT and DL-SWT	DL-EWT

[12]	2020	Stocks prices	LSTM and BI-LSTM	BI-LSTM
[13]	2018	Stocks prices	LSTM and ARIMA	LSTM
[15]	2021	Stocks prices	LSTM, Multivariate Regression, Random Forest Regression	Multivariate Regression and Random Forest Regression
[20]	2020	Stocks prices	CNNS, LSTM and Conventional LSTM	Conventional LSTM
[5]	2021	The volatility of stocks indexes	LSTM, GARCH and EWMA	LSTM
[2]	2021	The volatility of stocks indexes	LSTM and DNN using likelihood-based loss function, and ARMA-GARCH	LSTM using likelihood-based loss function
[21]	2020	Indexes prices	LSTM, BI-LSTM and ARIMA	BI-LSTM
[9]	2021	The volatility of returns	Lasso, Random Forest, Gradient Boosting, LSTM and HAR-RV	Lasso and HAR-RV
[14]	2019	Load & temperature	LSTM, ARIMA, SARIMA and ARIMAX	LSTM
[16]	2020	Sea surface height anomaly	Merged-LSTM, Stacked-LSTM, ANN, Merged-ANN, TCN, Merged-GRU and 1-D CNN	Merged-LSTM
[17]	2019	Solar irradiance	LSTM, ARIMA, SVR and NN	LSTM
[18]	2017	Photovoltaic power	LSTM-RNN, MLR, BRT and NN	LSTM-RNN
[19]	2017	Traffic flow	LSTM, SAE, RBF, SVM and ARIMA	LSTM

2.2. The volatility

Volatility in stock commodities means that prices can be up or down depending on the set of returns. The daily returns need to be calculated using the formula:

$$r = \left(\frac{CP-PP}{PP}\right) \times 100 \qquad (1)$$

Then the volatility:

$$V = |r| \qquad (2)$$

Where: r is the daily return, CP is the current close price, PP is the previous close price. Once the daily returns are obtained, the daily volatilities can be calculated using the absolute value of the daily returns:
V = |r|, where V represents the daily volatility [24].

2.3. ARCH model

Engle proposed the ARCH model in 1982 [25], and it was represented by these formulas:

$$\varepsilon_t = v_t \, \sigma_t \, , v_t \sim N(0,1) \qquad (3)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^{2} \qquad (4)$$

The first equation is the mean equation. ε_t represents the residual value which is based on white noise innovations v_t and volatility σ_t. The N (0,1) explains that v_t has unit variance and zero means. The second equation is the equation that helps the model to predict the volatility where the σ_t^2represent the conditional variance. α_0 is a constant factor which must be $\alpha_0 \geq 0$. α_i is the coefficient for the ARCH model and must be $\alpha_i \geq 0$. i defines the time of period t which must be $i \geq 0$. q is the number of lags. q is linked with white noise. One of the ARCH model drawbacks is overpredicting the volatility at most of the time. The reason for overproduction is due to the slow response to separated shocks in returns, and it resulted to be unsuccessful in capturing the leverage impact [34] [25-27].

2.4. GARCH model

The GARCH model is one of the statistical models that is basically used in financial markets to forecast the future price of stocks and their volatility. Bollerslev proposed the GARCH (p, q) model in 1986 which evolved from the ARCH model [28]. GARCH model was represented by these formulas:

$$\varepsilon_t = v_t \, \sigma_t \, , v_t \sim N(0,1) \qquad (5)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^{2} + \sum_{j}^{p} \beta_j \, \sigma_{t-j}^2 \qquad (6)$$

The equations of GARCH model have the same basis as ARCH equations but β_j was adding, it is the coefficient for GARCH model and must be $\beta_j \geq 0$. j defines the time of period t which must be j ≥ 0. p is the number of lags. p is linked with volatility, and it can control the existence of GARCH. The controlling of existence means if p equals zero, the ARCH model will process, and it will ignore the existence of the GARCH. Besides, the difference between the coefficient of the ARCH model and the GARCH model is α_i for the volatility clustering and $\beta_j$is for the persistence of volatility [25 - 26].

2.5. LSTM model

LSTM is a new version of the recurrent neural network. It does the memorization process to the sequences of data, which is done in memory cells even for the long sequences of data. Moreover, LSTM has three layers: an input layer, a hidden layer, and an output layer. And the memory cells exist in the hidden layer. The memory cell is responsible for tracking and saving all the dependencies between inputs [13][29]. Figure 1 shows the memory cell [29].

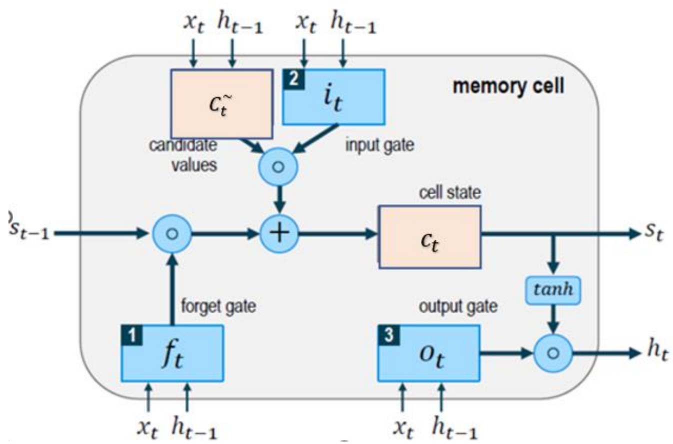

Figure1. The memory cell [29].

From Figure 1, the memory cell has three gates:

- The forget gate: it will do its job by deciding to keep the data or ignore the data.

$$f_t = \sigma(w_f.[h_{t-1}, x_t] + b_f) \qquad (7)$$

So, the forget gate ($\boldsymbol{f_t}$) will be calculated using sigmoid function (σ), bias values ($\boldsymbol{b_f}$), the matrix containing weight ($\boldsymbol{w_f}$) for the input data ($\boldsymbol{x_t}$) and for the result of the previous memory cell from the previous time ($\boldsymbol{h_{t-1}}$). The result of calculation will be either one, zero or between them. If the result is zero or nears to zero that means the forget gate will ignore this data. However, if the result is one or near to one that means the forget gate will keep this data.

- The input gate ($\boldsymbol{i_t}$): it will do its job by choosing the information that must be entering to the memory cell.

$$i_t = \sigma(w_i.[h_{t-1}, x_t] + b_i) \qquad (8)$$

$$\tilde{c_t} = tanh(w_c.[h_{t-1}, x_t] + b_c) \qquad (9)$$

The first equation: the input gate ($\boldsymbol{i_t}$) will be calculated by the same way of forgetting gate calculation. The second equation: the candidate values of the new state ($\boldsymbol{\tilde{c_t}}$) will be calculated by using a hyperbolic tangent function(tanh), bias values ($\boldsymbol{b_c}$), the matrix containing weight ($\boldsymbol{w_c}$) for the input data ($\boldsymbol{x_t}$) and for the result of the previous memory cell from the previous time($\boldsymbol{h_{t-1}}$). The reason of doing the second part is to discover what is the new memory state of the cell.

$$c_t = f_t * c_{t-1} + i_t * \tilde{c_t} \qquad (10)$$

The new state of the memory cell ($\boldsymbol{c_t}$) will be calculated by multiplying two elements are the previous state of the memory cell ($\boldsymbol{c_{t-1}}$) and the forget gate result ($\boldsymbol{f_t}$) then combined with result of multiplying the candidate value of the new state ($\boldsymbol{\tilde{c_t}}$) with the input gate result ($\boldsymbol{i_t}$).

- The output gate ($\boldsymbol{o_t}$): it will do its job by deciding which the result is the best output for the memory cell.

$$o_t = \sigma(w_o.[h_{t-1}, x_t] + b_o) \qquad (11)$$

$$h_t = o_t * \tanh(c_t) \qquad (12)$$

The first equation: the output gate ($\boldsymbol{o_t}$) will be calculated by the same way of forgetting gate and input gate calculation. The second equation: The invisible state output ($\boldsymbol{h_t}$) will be calculated by multiplying two elements are a hyperbolic tangent function(tanh) to the new state of the memory cell ($\boldsymbol{c_t}$) with the output gate ($\boldsymbol{o_t}$) [29-30].

3. Materials and Methods

to construct the models, two elements are needed which are data and methodology. So, this section has two parts, the first part describes the data. The second part describes the methodology that is followed by this research work.

3.1. Data Description

The dataset used for volatility forecasting consists of daily stock commodity prices for gold, oil, diesel, and the S&P 500 index. The data was downloaded from Yahoo Finance. The S&P 500 index is a widely followed index that represents the performance of 500 leading companies in the US stock market. It provides a comprehensive overview of the overall US stock market and includes the largest and most important companies [22]. The ticker symbol used for S&P 500 is "^GSPC", gold is “GOLD”, oil is “CL=F” and diesel is “TR”. The dataset covers the period from 23/08/2000 to 01/11/2022, providing a substantial amount of historical data for analysis. The number of samples available for each asset is as follows: 5583 samples for gold, 5583 samples for diesel, 5653 samples for oil, and 5583 samples for the S&P 500 index.

3.2. Methodology

In this research work, three models have been implemented for volatility forecasting: the ARCH model, the GARCH model, and the LSTM model. Each model serves a different purpose in capturing and predicting volatility patterns. The main four steps on methodology are: data collection, data processing, models implementation and models measuring. Figure 2 explains the whole steps starting from data collection to the models measuring. After that each step has been explained separately.

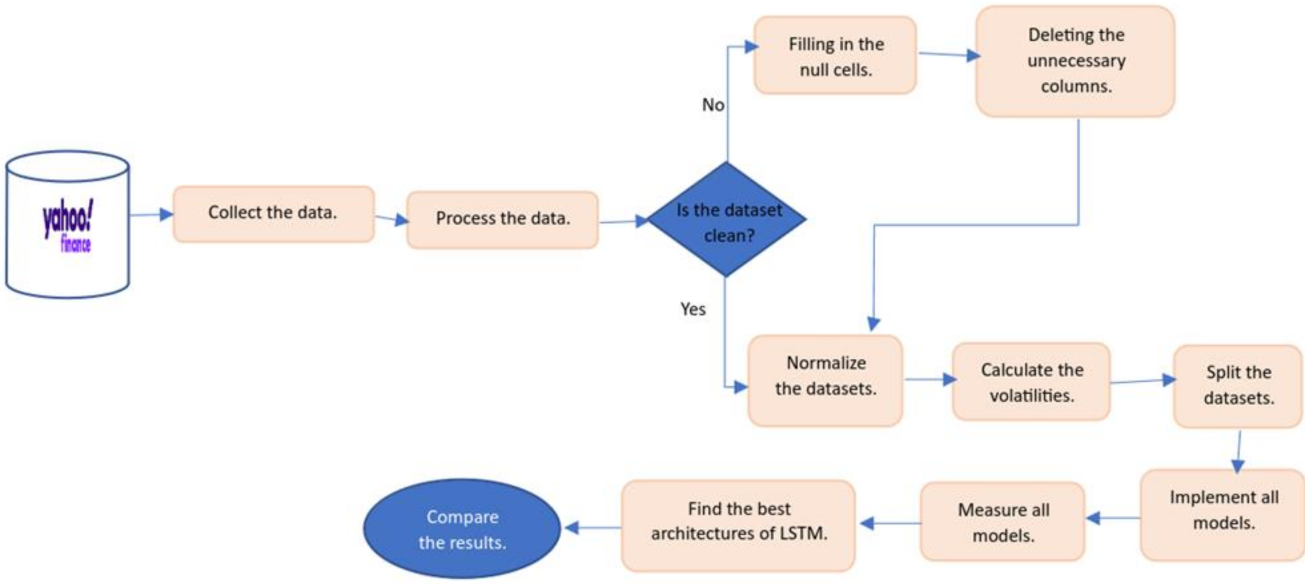

Figure 2. Methodology diagram.

3.2.1. Data collection

The oil, gold, diesel, and S&P 500 datasets were downloaded from Yahoo Finance using Python's Pandas, along with the DateTime and time modules for data manipulation.

3.2.2. Data processing

This part has five steps. If the datasets are clean, there is no need for the first two steps, but all the datasets went through either the first, second step, or both of them.

- Filling in the null cells: the null cells had been filled in by different values to improve the quality of the datasets.
- Deleting the unnecessary columns: the processes of this research will need a close and date column. So, the open, high, low, volume and adj close column had been deleted using drop () function which is one of pandas' functions.
- Data normalization: the normalization was made for the datasets by importing (MinMaxScaler) from (Sklearn.preprocessing) and importing (Scaler) from Panda library.
- The volatilities calculation: for calculating the volatilities, the return must be calculated first. So, the pct_change() function was used from the Pandas library to calculate the daily returns, and then the abs() function was applied to obtain the daily volatilities. By applying these calculations to the stock commodity datasets and the S&P 500 index, the daily volatilities will be obtained for volatility forecasting using the ARCH, GARCH, and LSTM models.

- The datasets splitting: to assess the predictive power for all models, the datasets are needed to split the data to two parts with 8:2 ratio.

Training set: 2000-08-23 to 2018-05-11.
Testing set: 2018-05-14 to 2022-10-31.

3.2.3. Models' implementation

The ARCH (1) model is the first model that had been applied to predict the testing sets. To implement the ARCH model, the ARCH package needs to be installed. So, it had been installed, and the ARCH model was imported from the installed ARCH package. GARCH (1,1) is the second model that had been applied in the training set to predict the testing set. The GARCH (1,1) model was imported from the installed ARCH package.

Hyperparameter tuning of LSTM model.

There are different ways to build LSTM. Each way requires to set the right elements of most hyperparameters. This is a common practice to choose the best architecture of LSTM [5]. For this research work, six architectures of LSTM have been implemented using Keras library. The best performance between them will be chosen. So, the seed has been set to avoid the randomness in the results [5]. Also, 60 days have been defined to input them in the input layer for each sequence of data at each timestep. Besides, 50 memory cells have been determined with return sequences (true) to return all hidden states for all architectures. Also, 50 memory cells have been set in input layer with return sequences(false) for all architectures without return all hidden states. Besides, the state has been set to false for all architectures to avoid resetting the memory cell each time [36]. In addition, optimization has been set to Adam to incorporate a bias correction [24]. Moreover, loss has been set to the mean squared error for all architectures and the size of batch and epoch is one for all architectures. All architectures have one input layer and one output layer.

The next six points show the differences between the different architectures of LSTM:

- LSTM (50,50): the hidden layer had not been specified in this architecture. Only the input layer and output layer had the same number of cells and the input data with other architectures.
- LSTM (50,50) - Dense (25): one hidden layer had been specified with 25 cells in this architecture.
- LSTM (50,50) -Dense (25,10): in this architecture, two hidden layers had been specified with 25 cells in the first layer and 10 cells in the second layer.
- LSTM (50,50)-Dense (25,10,5): in this architecture, three hidden layers had been specified with 25 cells in the first layer, 10 cells in the second layer and 5 cells in the third layer.
- LSTM (50,50)-Dense (50,25,10): in this architecture, three hidden layers had been specified with 50 cells in the first layer, 25 cells in the second layer and 10 cells in the third layer.
- LSTM (50,50)-Dense (60,50,25,10): in this architecture, four hidden layers had been specified with 60 cells in the first layer, 50 cells in the second layer, 25 cells in the third layer and 10 cells in the fourth layer.

3.2.4. Measurements

To compare the performance of all models, two measurements had been used to evaluate each model. They were the mean square error (MSE) and the root mean square error (RMSE). MSE and RMSE were implemented by using sklearn.metrics.

MSE is represented below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y_i})^2 \qquad (13)$$

RMSE is represented below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y_i})^2} \qquad (14)$$

n represent total count for the observations and $\boldsymbol{i}$ represent the counter. While $\boldsymbol{y_i}$ represent forecasting values and $\tilde{\boldsymbol{y_i}}$ represent the actual values [29].

4. Results and Discussion

This section shows and discusses the results. It has four parts starting with descriptive statistics for returns and ending with comparing between GARCH, ARCH and LSTM models. Also, this section is interspersed with an analysis of results from the ARCH (1) model and GARCH (1,1) model and finding the best LSTM architecture.

4.1. Descriptive statistics for returns

Table 2. Descriptive statistics for returns of (oil, gold, diesel and S&P500)

	Oil	**Gold**	**Diesel**	**S&P500**
Mean	0.01	0.12	0.26	0.07
Std	1.96	6.07	8.39	4.51
Skewness	-23.46	4.78	19.74	-0.23
Kurtosis	1193.55	158.81	871.24	126.12
Jarque-Bera (JB) test	335409807.9	5876381.1	176591464.3	3693035.2
P-value (JB test)	0	0	0	0

Table 2 provides descriptive statistics for oil, gold, diesel, and S&P 500 returns. To describe each dataset statistically, this table provides six elements (mean, standard deviation, skewness, kurtosis, Jarque-Bera (JB) test, P-value for the JB test). in table 2, based on the provided information, show the following characteristics of the return distributions:

Standard Deviation: The standard deviation for each dataset is greater than the mean. This indicates that all the datasets exhibit variability and volatility in their returns.

Skewness: The skewness value for the oil and S&P 500 returns is negative, indicating left skewness. This suggests that the returns of these assets have a longer left tail. On the other hand, the skewness value for gold and diesel returns is positive and greater than one, indicating highly right-skewed distributions. This implies that the returns of these assets have a longer right tail and are more concentrated towards lower values.

Kurtosis: The kurtosis value for all datasets is positive and greater than two, indicating that all of them have peaked distributions with heavy tails. This suggests that extreme returns occur more frequently than in a normal distribution.

Jarque-Bera (JB) Test: The JB test is performed to assess the normality assumption of the return distributions. The JB test statistic value for all datasets is greater than one, indicating a deviation from normality. Additionally, the p-value for all datasets is less than 0.05, indicating that the null hypothesis of normality is rejected [25] [31]. This confirms that the return distributions of all datasets do not follow a normal distribution.

In summary, the descriptive statistics indicate that all datasets have variable returns. The return distributions exhibit skewness, with oil and S&P 500 being left-skewed and gold and diesel being highly right-skewed. Furthermore, all datasets have peaked distributions with heavy tails, and they deviate significantly from a normal distribution based on the results of the JB test.

4.2. Analysis of results from ARCH (1) Model and GARCH (1,1) model

Table 3. The result of ARCH (1) and GARCH (1,1) predications

	Mean model:		Constant mean						
	Method:		maximum likelihood						
		Coefficient		**Standard error**		**T-statistic**		**P-value**	
		ARCH	**GARCH**	**ARCH**	**GARCH**	**ARCH**	**GARCH**	**ARCH**	**GARCH**
Oil	α_0	0.6292	0.0114	3.893e-	7.666e-03	16.161	1.486	9.574e-59	0.137

				02					
	α_i	0.2628	0.0502	5.209e-02	2.239e-02	5.046	2.243	4.506e-07	2.492e-02
	β	-	0.9350	-	3.021e-02	-	30.951	-	2.458e-210
	Log-likelihood	-5833.0	-5436.8	**AIC**	10881.7	11672.1	**BIC**	11691.3	10907.4
Gold	α_0	7.8462	0.0514	1.455	2.108e-02	5.393	2.440	6.911e-08	1.469e-02
	α_i	1.0000	0.0715	0.438	1.856e-02	2.281	3.855	2.258e-02	1.156e-04
	β	-	0.9262	-	1.567e-02	-	59.102	-	0
	Log-likelihood	-12155	-10314	**AIC**	20637.3	24316.1	**BIC**	24335.3	20662.9
Diesel	α_0	11.2321	0.1812	1.270	0.146	8.843	1.240	9.329e-19	0.215
	α_i	0.8117	0.1084	0.152	2.996e-02	5.355	3.617	8.554e-08	2.982e-04
	β	-	0.8916	-	3.924e-02	-	22.723	-	2.659e-114
	Log-likelihood	-12732	-11832	**AIC**	23673	25470	**BIC**	25489	23698
S&P500	α_0	3.9006	0.0281	0.464	7.177e-03	8.403	3.914	4.367e-17	9.066e-05
	α_i	1	0.1397	0.231	1.460-e02	4.323	9.568	1.541e-05	1.094e-21
	β	-	0.8603	-	1.242e-02	-	69.284	-	0
	Log-likelihood	-10654	-8267	**AIC**	16542	21315	**BIC**	21334	16568

Table 3 shows all results that belong to the GARCH and the ARCH model for predicting the volatility of the testing set. So, this tables 3 includes (the mean model, method, coefficient, standard error, T-statistic test, P-value, Log-likelihood, AIC, BIC). Here are the key observations and discussions related to the table:

Mean Estimation: Both the ARCH and GARCH models have a constant mean estimation, indicating that the mean value is not a moving average but remains constant.

Maximum Likelihood Method: Both models used the maximum likelihood method to estimate the coefficients. This method is commonly used in ARCH and GARCH models to find the parameters that maximize the likelihood of the observed data.

Log-Likelihood and Information Criteria: Higher values of log-likelihood and lower values of information criteria (such as AIC and BIC) indicate better estimations [30]. According to Table 3, the GARCH model demonstrates higher log-likelihood values and lower AIC and BIC values compared to the ARCH model. This suggests that the GARCH model provides better estimations for the datasets.

Coefficient Significance: To assess the significance of the coefficients, a t-statistic test is performed, and the p-value is examined. Coefficients with a p-value lower than 0.05 are considered statistically significant [30] [32]. From Table 3, it is indicated that all coefficients in the ARCH model for all datasets are statistically significant. However, for the GARCH model, the omega coefficient (representing the constant term) for the oil and diesel datasets fails to pass the t-statistic test since their p-values are higher than 0.05. This implies that the coefficients for the GARCH model in the oil and diesel datasets are not statistically significant and cannot be replaced by zero.

In conclusion, based on the results in Table 3, the GARCH model demonstrates better estimations compared to the ARCH model for the datasets analyzed. However, it should be noted that the coefficients for the GARCH model in the oil and diesel datasets are not statistically significant, indicating that their values are high and cannot be replaced by zero. On the other hand, all coefficients in the ARCH model are statistically significant for all datasets.

4.3. Finding the best architecture of LSTM.

Table 4. Finding the best architecture of LSTM using MSE & RMSE

Stocks	**Oil**		**Gold**		**Diesel**		**S&P500**	
Measurements	**MSE**	**RMSE**	**MSE**	**RMSE**	**MSE**	**RMSE**	**MSE**	**RMSE**
LSTM (50,50)	**10.134**	**3.183**	**6.100**	**2.469**	6.075	2.464	**1.284**	**1.133**
LSTM (50,50) + Dense (25)	10.144	3.185	7.035	2.652	6.139	2.477	1.334	1.155
LSTM (50,50) +Dense (25,10)	10.158	3.187	8.714	2.952	6.039	2.457	1.389	1.178
LSTM (50,50) +Dense (25,10,5)	10.157	3.187	9.162	3.027	6.251	2.500	1.341	1.158
LSTM (50,50) +Dense (50,25,10)	10.177	3.190	9.192	3.031	**6.018**	**2.453**	1.348	1.161
LSTM (50,50) +Dense (60,50,25,10)	10.181	3.190	11.848	3.442	6.207	2.491	1.385	1.176

Table 4 presents the comparison between six architectures of the LSTM model using two error measurements: MSE (Mean Squared Error) and RMSE (Root Mean Squared Error). The lower values of these measurements indicate lower prediction errors. From the results in Table 4, it is observed that the LSTM (50,50) architecture consistently achieves lower MSE and RMSE values compared to the other architectures for all datasets, except for the diesel dataset. For the diesel dataset, the LSTM (50,50) + Dense (50,25,10) architecture performs better in terms of lower MSE and RMSE values. These findings support the idea that the best-performing architecture may vary depending on the specific dataset and its volatility characteristics. As mentioned by Václav, P [26], it is challenging to identify a single best model that works optimally for all periods and all items. However, it is possible to identify the best model for specific periods and specific items. Based on the results in Table 4, the LSTM (50,50) architecture is chosen as the best-performing architecture, as it consistently outperforms the other architectures in terms of lower MSE and RMSE values, except for the diesel dataset where the LSTM (50,50) + Dense (50,25,10) architecture performs better. Therefore, the LSTM (50,50) architecture is selected for further comparison with the GARCH and ARCH models in forecasting volatility.

4.4. Comparing between GARCH, ARCH and LSTM

The three models' performances are shown from Figure 3 to Figure 14. The blue line represents the actual volatility, while the orange line represents the predicted volatility. The first three figures illustrate the three models' performance in forecasting oil volatility. Figures 6 to 8 figure illustrate the three models' performance in forecasting gold volatility. Figures 9 to 11 figure illustrate the three models' performance in forecasting diesel volatility. Figures 12 to 14 figure illustrate the three models' performance in forecasting S&P500 volatility.

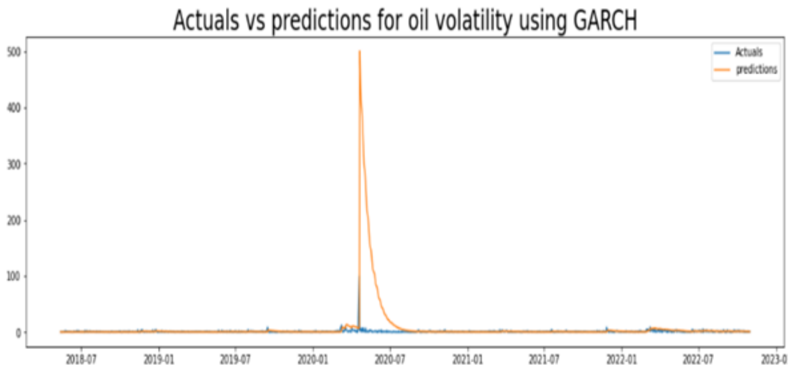

Figure 3. Volatility forecast with GARCH (1,1) for oil.

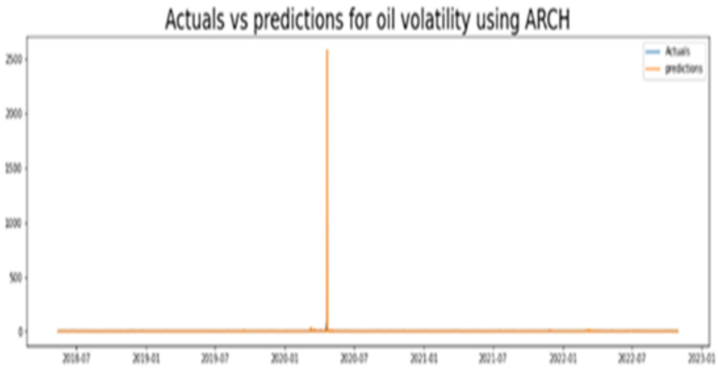

Figure 4. Volatility forecast with ARCH (1) for oil.

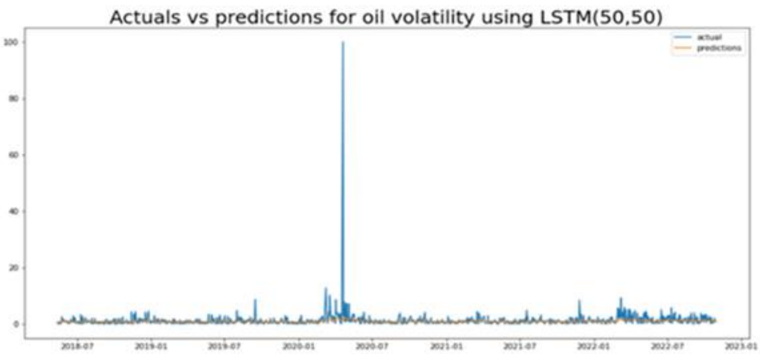

Figure 5. Volatility forecast with LSTM (50,50) for oil.

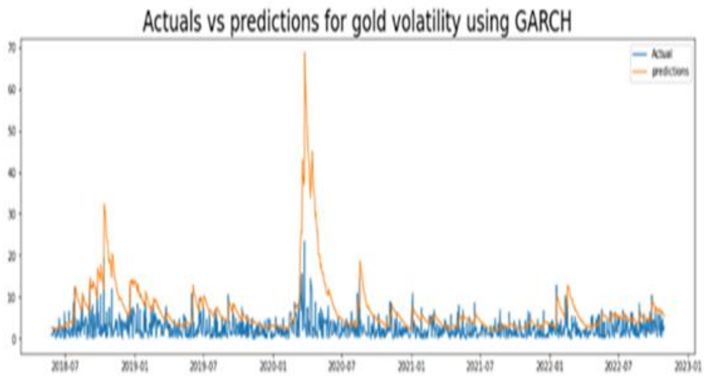

Figure 6. Volatility forecast with GARCH (1,1) for gold.

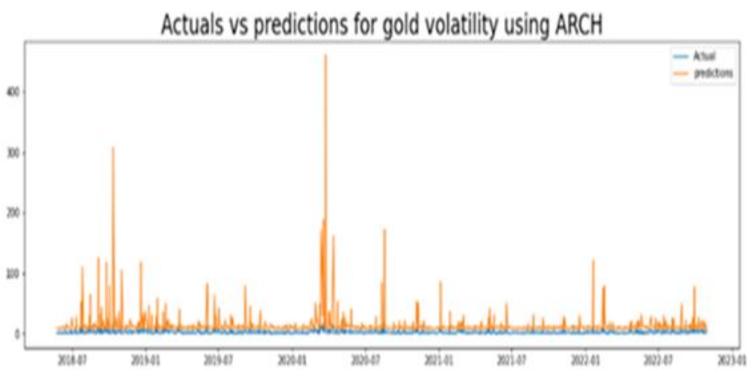

Figure 7. Volatility forecast with ARCH (1) for gold.

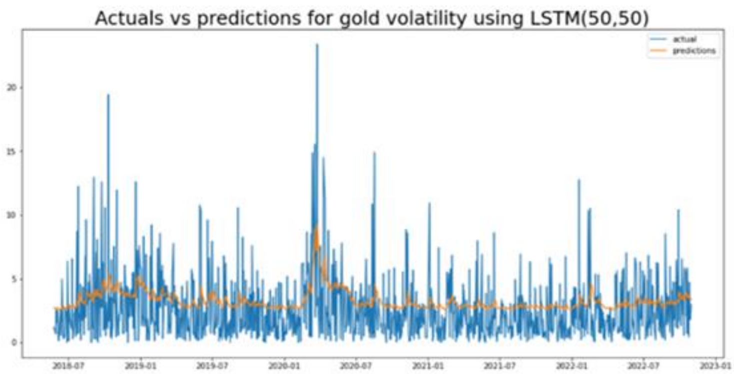

Figure 8. Volatility forecast with LSTM(50,50) for gold.

figures 3 to 5 depict the performance of the three models in forecasting oil volatility. The blue line represents the actual

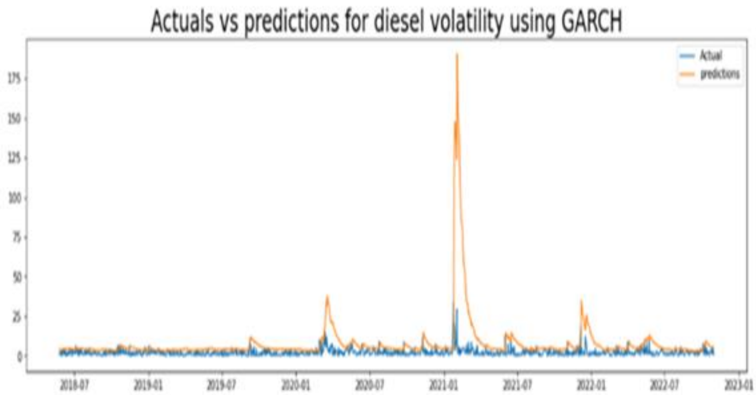

Figure 9. Volatility forecast with GARCH (1,1) for diesel.

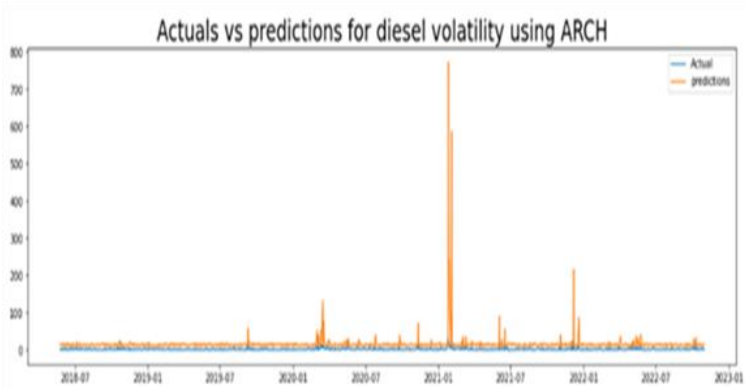

Figure 10. Volatility forecast with ARCH(1) for diesel.

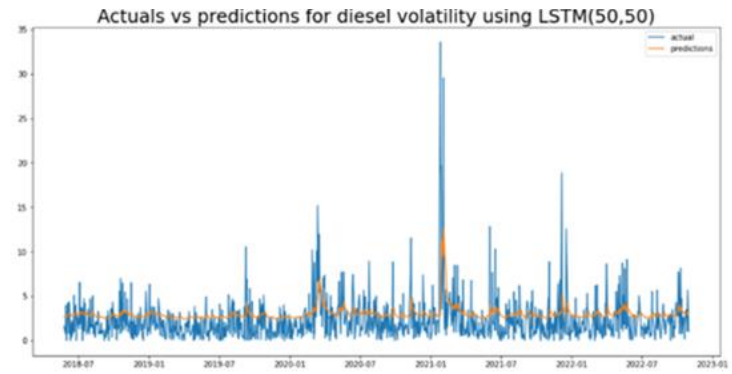

Figure 11. Volatility forecast with LSTM (50,50) for diesel.

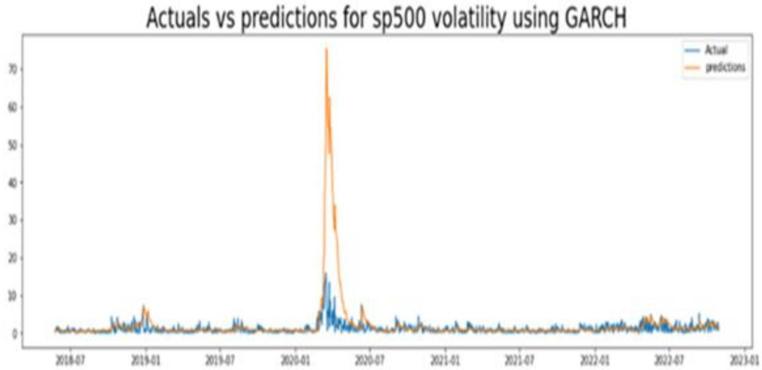

Figure 12. Volatility forecast with GARCH (1,1) for S&P500.

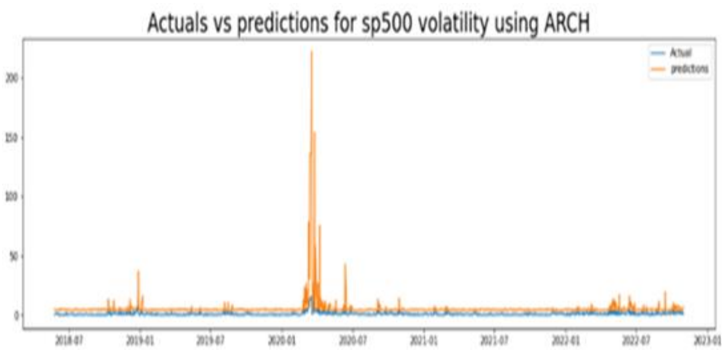

Figure 13. Volatility forecast with ARCH(1) for S&P500.

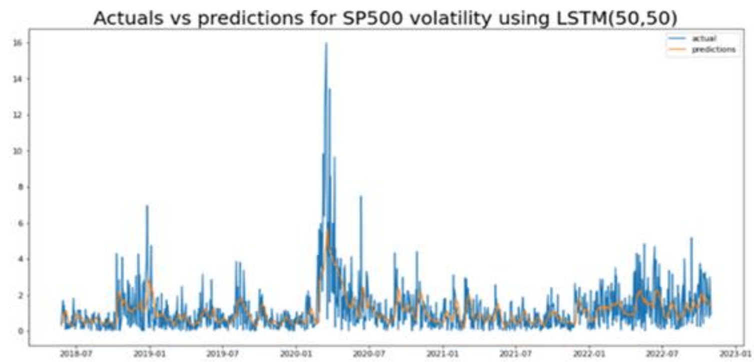

Figure 14. Volatility forecast with LSTM (50,50) for S&P500.

volatility, while the orange line represents the predicted volatility. From these figures, it can be observed that the GARCH model tends to have volatility that deviates from the normal range in previous periods. The estimated volatility for the next day remains highly volatile until it eventually returns to its normal range after a few days. This behavior is attributed to the impact of both the previous shock and previous volatility on the conditional variance in the GARCH model's second equation [27]. Schmidt, L [27] has previously highlighted this as a disadvantage of the GARCH model, and it is noticeable in the figures representing the GARCH model's performance in this research work.

Figures 6 to 8 showcase the performance of the three models in forecasting gold volatility. Similarly, the GARCH model exhibits volatility that deviates from the normal range in previous periods before eventually returning to normal. On the other hand, the ARCH model tends to overpredict the volatility most of the time. This overprediction is attributed to the slow response of the ARCH model in capturing separated shocks in returns and its inability to effectively capture the leverage impact [27]. Schmidt, L [27] has identified this as a weakness of the ARCH model. In contrast, the LSTM model demonstrates smoother volatility predictions by learning the average volatility and applying it to forecasting. This characteristic aligns with the observation made by Romano, S [5], who noted that LSTM can return to the mean faster after periods of high volatility.

Figures 9 to 11 present the performance of the three models in forecasting diesel volatility. The patterns observed in these figures are consistent with the previous findings for oil and gold. The GARCH model shows deviations from the normal range in previous periods, the ARCH model tends to overpredict volatility, and the LSTM model offers smoother volatility predictions.

Lastly, Figures 12 to 14 depict the performance of the three models in forecasting S&P500 volatility. Similar to the previous datasets, the GARCH model exhibits volatility that deviates from the normal range in previous periods before gradually returning to normal. The ARCH model continues to overpredict volatility, while the LSTM model provides smoother volatility predictions. Overall, the LSTM model demonstrates a smoother and more accurate forecasting of volatility compared to the GARCH and ARCH models. It captures the average volatility and adjusts predictions, accordingly, allowing it to respond faster to changes in volatility. This characteristic sets the LSTM model apart and contributes to its superior performance in forecasting volatility.

Table 5 comparing between models using MSE and RMSE

Model	**GARCH**			**LSTM (50,50)**			**ARCH**		
Measurements	**MSE**	**RMSE**	**Accuracy (%)**	**MSE**	**RMSE**	**Accuracy (%)**	**MSE**	**RMSE**	**Accuracy (%)**
Oil	1790	42	58%	**10.134**	**3.183**	**96.8%**	5886	76	24%
Gold	68	8	92%	**6.100**	**2.469**	**97.5%**	656	25	75%
Diesel	260	16	84%	**6.075**	**2.464**	**97.5%**	1153	33	67%
S&P500	38	6	94%	**1.284**	**1.133**	**98.8%**	127	11	89%

Table 5 presents the evaluation of predictions for the GARCH, ARCH, and LSTM models. Lower values in MSE and RMSE indicate lower errors in the predictions. According to the table, the LSTM models achieved an accuracy rate of 98.8%, while the GARCH model reached 94% accuracy, and the ARCH model achieved 89% accuracy. Therefore, the GARCH model outperforms the ARCH model in terms of accuracy. Hence, investors can consider using the GARCH model over the ARCH model when using financial models. However, the LSTM model surpasses both the ARCH and GARCH models in terms of performance. It exhibits the lowest values in MSE and RMSE, indicating better accuracy in volatility predictions. The LSTM model achieves an accuracy rate of 98.8%, which is higher than both the GARCH and ARCH models.

This finding demonstrates that deep learning, represented by the LSTM model, outperforms traditional financial models in volatility forecasting. These results align with the findings of Jia, F., & Yang, B [2], Muzaffar, S., & Afshari, A [14], and Yu, Y., Cao, J., & Zhu, J [17], who obtained similar results in their respective studies. This further supports the conclusion that deep learning models, such as LSTM, offer superior performance in volatility forecasting compared to traditional financial models like ARCH and GARCH.

5. Conclusions

This research aimed to compare financial models (ARCH, GARCH) and deep learning models (LSTM) to forecast volatility in stock commodities. The findings indicate that LSTM outperformed ARCH and GARCH models in terms of predictive power. The research also identified the best architecture among six LSTM architectures based on RMSE and MSE measurements. However, it was observed that the ARCH model had a weakness in overpredicting, while the GARCH model took time to return to the normal range after high volatility. Although GARCH's better predictive performance compared to ARCH, LSTM demonstrated superior predictive power. The effectiveness of deep learning, particularly LSTM, in forecasting volatility in stock commodities was demonstrated in this research. Future work in the field of deep learning can explore and develop different models in various domains to identify the best-performing model.

References

[1] Discla Wenbo Ge, Pooia Lalbakhsh, Leigh Isai, Artem Lenskiy, and Hanna Suominen. 2022. Neural Network–Based Financial Volatility Forecasting: A Systematic Review. ACM Comput. Surv. 55, 1, Arti-cle 14 (January 2022) https://doi.org/10.1145/3483596.
[2] Jia, F., & Yang, B. (2021, February 25). Forecasting volatility of stock index: Deep Learning model with likelihood-based loss function. Complexity. Retrieved March 24, 2022, from https://www.hindawi.com/journals/complexity/2021/5511802/.
[3] Saqware, G. J., & B, I. (2020). Modeling Volatility in the Stock Market for Accuracy in Forecasting. In-ternational Journal of Recent Technology and Engineering (IJRTE), 8 (5), 1–2. https://doi.org/10.35940/ijrte.E5608.018520.
[4] Li, Sophia Zhengzi and Tang, Yushan, Forecasting Realized Volatility: An Automatic System Using Many Features and Many Machine Learning Algorithms (May 23, 2021). Available at SSRN: https://ssrn.com/abstract=3776915 or http://dx.doi.org/10.2139/ssrn.3776915.
[5] Romano, S., Barone, P. E., & Borri, P. N. (2021). (thesis). MACHINE LEARNING FOR VOLATILITY FORECASTING. Retrieved 2022, from http://tesi.luiss.it/29954/1/710741_ROMANO_SIMONE.pdf.
[6] Team, T. I. (2021, June 11). Generalized autoregressive conditional heteroskedasticity (GARCH) defini-tion. Investopedia. Retrieved March 24, 2022, from https://www.investopedia.com/terms/g/garch.asp
[7] Balaji, A. J., Ram, D. S. H., & Nair, B. B. (2018). 8th International Conference on Advances in Computing and Communication (ICACC-2018). In Procedia Computer Science 143 (2018) 947–953. Retrieved 2022, from https://www.sciencedirect.com/science/article/pii/S1877050918320362.
[8] Abe, M., & Nakayama, H. (2018, June 13). Deep learning for forecasting stock returns in the cross-section. arXiv.org. Retrieved March 24, 2022, from https://arxiv.org/abs/1801.01777.
[9] Pyrlik, Vladimir and Leonova, Aleksandra, Forecasting Realized Volatility Using Machine Learning and Mixed-Frequency Data (the Case of the Russian Stock Market) (December 1, 2021). CERGE-EI Working Paper Series No. 713, Available at SSRN: https://ssrn.com/abstract=3996169 or http://dx.doi.org/10.2139/ssrn.3996169.
[10] Kühnert, S. (2020). Functional arch and GARCH models: A Yule-Walker Approach. https://doi.org/10.20944/preprints201912.0163.v2.
[11] K. A. Althelaya, S. A. Mohammed and E. -S. M. El-Alfy, "Combining Deep Learning and Multi-resolution Analysis for Stock Market Forecasting," in IEEE Access, vol. 9, pp. 13099-13111, 2021, doi: 10.1109/ACCESS.2021.3051872 from https://ieeexplore.ieee.org/document/9324831
[12] M. A. Istiake Sunny, M. M. S. Maswood and A. G. Alharbi, "Deep Learning-Based Stock Price Pre-diction Using LSTM and BiDirectional LSTM Model," 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, pp. 87-92, doi: 10.1109/NILES50944.2020.9257950 from https://ieeexplore.ieee.org/document/9257950.
[13] Namin, S. S., & Namin, A. S. (2018, March 16). Forecasting economics and financial time se-ries: Arima vs. LSTM. arXiv.org. Retrieved February 4, 2023, from https://arxiv.org/abs/1803.06386.
[14] Muzaffar, S., & Afshari, A. (2019). 10th International Conference on Applied Energy (ICAE2018), 22-25 August 2018, Hong Kong, China. In Energy Procedia. Retrieved 2022, from https://www.sciencedirect.com/science/article/pii/S1876610219310008.
[15] Sen, Jaydip; Mehtab, Sidra; Dutta, Abhishek (2021): Stock Price Prediction Using Machine Learn-ing and LSTM-Based Deep Learning Models. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.15103602.v1.
[16] T. Song, J. Jiang, W. Li and D. Xu, "A Deep Learning Method With Merged LSTM Neural Networks for SSHA Prediction," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 2853-2860, 2020, doi: 10.1109/JSTARS.2020.2998461 from https://ieeexplore.ieee.org/document/9103934.
[17] Yu, Y., Cao, J., & Zhu, J. (2019). An LSTM short-term solar irradiance forecasting under compli-cated weather conditions. IEEE Access, 7, 145651–145666. https://doi.org/10.1109/access.2019.2946057.
[18] Abdel-Nasser, M., & Mahmoud, K. (2017). Accurate Photovoltaic Power Forecasting models using Deep LSTM-RNN. Neural Computing and Applications, 31(7), 2727–2740. https://doi.org/10.1007/s00521-017-3225-z.
[19] Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: A deep learning approach for short-term Traffic forecast. IET Intelligent Transport Systems, 11(2), 68–75. https://doi.org/10.1049/iet-its.2016.0208.
[20] Mehtab, S., & Sen, J. (2020). Stock price prediction using CNN and LSTM-based deep learning models. 2020 International Conference on Decision Aid Sciences and Application (DASA). https://doi.org/10.1109/dasa51403.2020.9317207.
[21] Namini, S. S., Tavakoli, N., & Namin, A. S. (2020, February 24). The performance of LSTM and BiLSTM in forecasting time series. IEEE Xplore . Retrieved February 5, 2023, from https://ieeexplore.ieee.org/document/9005997.
[22] Kenton, W. (2022, November 3). S&P 500 index: What it's for and why it's important in investing. Investopedia. Retrieved January 31, 2023, from https://www.investopedia.com/terms/s/sp500.asp.
[23] Frankel, M. (2022, September 20). How to invest in the companies in the S&P 500 index? The Mot-ley Fool. Retrieved January 31, 2023, from https://www.fool.com/investing/stock-market/indexes/sp-500/.
[24] Liu, W. K., & So, M. K. (2020). A GARCH model with Artificial Neural Networks. Information, 11(10), 489. https://doi.org/10.3390/info11100489.
[25] Tillgren, V., & Ahmed, S. (2022). Forecasting Time-Varying Volatility In Global Stock Markets (dissertation). (J. Stoklasa, Ed.). Retrieved February 1, 2023, from https://lutpub.lut.fi/handle/10024/164428. https://lutpub.lut.fi/bitstream/handle/10024/164428/Masters%20thesis%20Ville%20Tillgren%20Final%20version.pdf?sequence=1&isAllowed=y.
[26] Václav, P. (2022). Gold as a Stable Asset in Economic Recession: An Econometric Analysis (disser-tation). Zlato jako stabilní aktivum v ekonomické recesi: Ekonometrická analýza. Retrieved February 1, 2023, from https://dspace.cuni.cz/handle/20.500.11956/174020.
[27] Schmidt, L. (2021). Volatility Forecasting Performance of Garch Models: A Study on Nordic Indi-ces During Covid-19 (dissertation). Digitala Vetenskapliga Arkivet. Retrieved February 1, 2023, from https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1566342&dswid=5934.
[28] MUȘETESCU, R.-C., GRIGORE, G.-E., & NICOLAE, S. (2022). The use of GARCH autoregressive models in estimating and forecasting the crude oil volatility. European Journal of Interdisciplinary Studies, 14(1), 13–38. https://doi.org/10.24818/ejis.2022.02.
[29] Vitali, G. (2019). Forecasting stock index volatility: A comparison between Garch and LSTM mod-els. universita' ca' Foscari venezia . Retrieved February 4, 2023, from http://dspace.unive.it/bitstream/handle/10579/15933/868008-1231506.pdf.

[30] Mahajan, V., Thakan, S., & Malik, A. (2022). Modeling and forecasting the volatility of Nifty 50 using GARCH and RNN models. Economies, 10(5), 102. https://doi.org/10.3390/economies10050102.
[31] Gupta, A., & Rajib, P. (2018). Do var exceptions have seasonality? an empirical study on Indian commodity spot prices. IIMB Management Review, 30(4), 369–384. https://doi.org/10.1016/j.iimb.2018.05.008.
[32] Liu, W. K., & So, M. K. (2020). A GARCH model with Artificial Neural Networks. Information, 11(10), 489. https://doi.org/10.3390/info11100489.
[33] Zhang, X., Zhang, L., Zhou, Q., & Jin, X. (2022). A novel bitcoin and gold prices prediction method using an LSTM-P neural network model. Computational Intelligence and Neuroscience, 2022, 1–12. https://doi.org/10.1155/2022/1643413.
[34] Ou, J.; Huang, X.; Zhou, Y.; Zhou, Z.; Nie, Q. Traffic Volatility Forecasting Using an Omnibus Fam-ily GARCH Modeling Framework. Entropy 2022, 24, 1392. https://doi.org/10.3390/e24101392.
[35] Hayes, A. (2022, November 10). What is a stock ticker? definition, how they work, and origins. In-vestopedia. https://www.investopedia.com/ask/answers/12/what-is-a-stock-ticker.asp.
[36] Bernico, M. (2020). Stateful versus stateless LSTMs. O'Reilly Online Learning. Retrieved February 4, 2023, from https://www.oreilly.com/library/view/deep-learning-quick/9781788837996/22117615-1451-4135-8ca6-83fe9c0f6ab6.xhtml.