

Article

Comparative Performance Analysis of Vibration Prediction Using RNN Techniques

Ju-Hyung Lee ¹  and Jun-Ki Hong ^{2,*} ¹ Department of Mobile Software Engineering, Pai Chai University, Daejeon 35345, Korea² Department of Computer Engineering, Pai Chai University, Daejeon 35345, Korea

* Correspondence: jkhong@pcu.ac.kr

Abstract: Drones are increasingly used in several industries, including rescue, firefighting, and agriculture. If the motor connected to a drone's propeller is damaged, there is a risk of a drone crash. Therefore, to prevent such incidents, an accurate and quick prediction tool of the motor vibrations in drones is required. In this study, normal and abnormal vibration data were collected from the motor connected to the propeller of a drone. The period and amplitude of the vibrations are consistent in normal vibrations, whereas they are irregular in abnormal vibrations. The collected vibration data were used to train six recurrent neural network (RNN) techniques: long short-term memory (LSTM), attention-LSTM (Attn.-LSTM), bidirectional-LSTM (Bi-LSTM), gated recurrent unit (GRU), attention-GRU (Attn.-GRU), and bidirectional GRU (Bi-GRU). Then, the simulation runtime it took for each RNN technique to predict the vibrations and the accuracy of the predicted vibrations were analyzed to compare the performances of the RNN model. Based on the simulation results, the Attn.-LSTM and Attn.-GRU techniques, incorporating the attention mechanism, had the best efficiency compared to the conventional LSTM and GRU techniques, respectively. The attention mechanism calculates the similarity between the input value and the to-be-predicted value in advance and reflects the similarity in the prediction.

Keywords: deep learning; recurrent neural network (RNN); time series vibration; vibration prediction; long short-term memory (LSTM); gated recurrent unit (GRU); attention mechanism; bidirectional RNN



Citation: Lee, J.-H.; Hong, J.-K. Comparative Performance Analysis of Vibration Prediction Using RNN Techniques. *Electronics* **2022**, *11*, 3619. <https://doi.org/10.3390/electronics11213619>

Academic Editor: Dionysios Kehagias

Received: 14 October 2022
Accepted: 4 November 2022
Published: 6 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, drone technology has attracted significant attention as one of the pillars of the fourth industrial revolution. Drones are a type of unmanned aerial vehicle that can be flown without a pilot on board. They are used in a variety of industries, such as search and rescue [1,2], shipping and transportation [3], and agriculture [4,5]. According to the drone market forecast report by Drone Industry Insights, the drone market size was 26.3 billion dollars in 2021 and is projected to grow up to 41.3 billion dollars by 2026 [6].

Although drones are widely used in various fields, there is a risk of drone crash due to damage to the motor or propeller driving system, which may result in human casualties and damage or the destruction of the drone. According to the report of the drone safety survey results published by the Korea Consumer Agency in 2017, the number of human casualties caused by drones is increasing each year. The causes of human casualties due to drone-related incidents include drone crashes, battery explosions, and catching fire, as well as people getting hit by a drone's propeller [7]. Therefore, it is essential to conduct studies on predicting vibrations to prevent drone crashes. Previous studies have been conducted to predict time series vibration data using various recurrent neural network (RNN) techniques: long short-term memory (LSTM) [8–16], attention-LSTM (Attn.-LSTM) [17,18], bidirectional-LSTM (Bi-LSTM) [19–22], gated recurrent unit (GRU) [23–27], attention-GRU (Attn.-GRU) [28], and bidirectional GRU (Bi-GRU) [29–31]. Table 1 below

lists the studies in the existing literature that have used two or more RNN techniques to predict time series data and compare the prediction performances.

Table 1. The literature using more than two RNN techniques.

Studies	Purpose	Methods	Evaluation Criteria
Wang et al. [32]	Vibration prediction of turbine	LSTM, GRU	RMSE, MAPE, MSE, and convergence time
Yuan et al. [33]	Fault diagnosis of aircraft engine	RNN, LSTM, GRU	MSE and relative errors
Demidova [34]	Predictive maintenance of aircraft engine	RNN, LSTM, GRU	Train and test accuracies and simulation time
Naren et al. [35]	Predictive maintenance of aircraft engine	RNN, LSTM, GRU	Simulation time and R^2
Chen et al. [36]	Fault detection of drone	LSTM, GRU	RMSE
Hong [37]	Prediction of vibration	LSTM, GRU	RMSE, simulation time, and R^2
Zhang et al. [38]	Prediction of air pollution	LSTM, Bi-LSTM	RMSE, MAE, MAPE and R^2
Tao et al. [39]	Prediction of air pollution	RNN, LSTM, GRU, Bi-GRU	RMSE, MAE and SMAPE
Qin [40]	Prediction of dissolved oxygen	LSTM, Bi-LSTM, GRU, Bi-GRU	RMSE, MAE and R^2
Althelaya et al. [41]	Forecasting of stock market	LSTM, Bi-LSTM	RMSE, MAE and R^2
Siarni-Namini et al. [42]	Forecasting of stock market	LSTM, Bi-LSTM	RMSE
Hollis et al. [43]	Forecasting of stock market	LSTM, Attn.-LSTM	MSE and accuracy
This work	Prediction of vibration	LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, Bi-GRU	Simulation time and R^2

According to Table 1, there have been studies that have incorporated the attention mechanism and bidirectional techniques into the existing LSTM and GRU techniques to predict air pollution [38,39], dissolved oxygen [40], and the stock market [41–43].

However, to our knowledge, no studies have conducted a comparative analysis of the prediction performance of conventional LSTM and GRU techniques and LSTM and GRU techniques incorporating the attention mechanism and bidirectional techniques for the time series vibration data of drones. Therefore, this study compares the prediction performance and simulation runtime of different RNN techniques: LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, and Bi-GRU. Furthermore, the benefit of comparing the performance of vibration prediction by using RNN techniques in this study is that it can verify the most efficient RNN technique among LSTM, Attn.-LSTM, GRU, Attn.-GRU, and Bi-GRU. Furthermore, the best-performing RNN technique validated in this study can be used as a guide for future studies relating to vibration prediction.

The remainder of this paper is organized as follows: Section 2 describes the comparative analysis framework and the collected time series vibration data. Section 3 explains the LSTM, GRU, Attn.-LSTM, Bi-LSTM, Attn.-GRU, and Bi-GRU techniques. Section 4 analyzes the predicted vibration data and compares the prediction performance and runtime efficiency of the LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, and Bi-GRU techniques. Section 5 summarizes the conclusions.

2. Materials and Methods

2.1. Comparative Analysis Framework

This subsection explains the analysis framework used to compare the prediction performance and runtime efficiency of the LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, and Bi-GRU techniques for drone time series vibration data. This subsection also explains

the coefficient of determination used to evaluate the accuracy of the predicted vibrations. Figure 1 below shows the flowchart of the analysis framework.

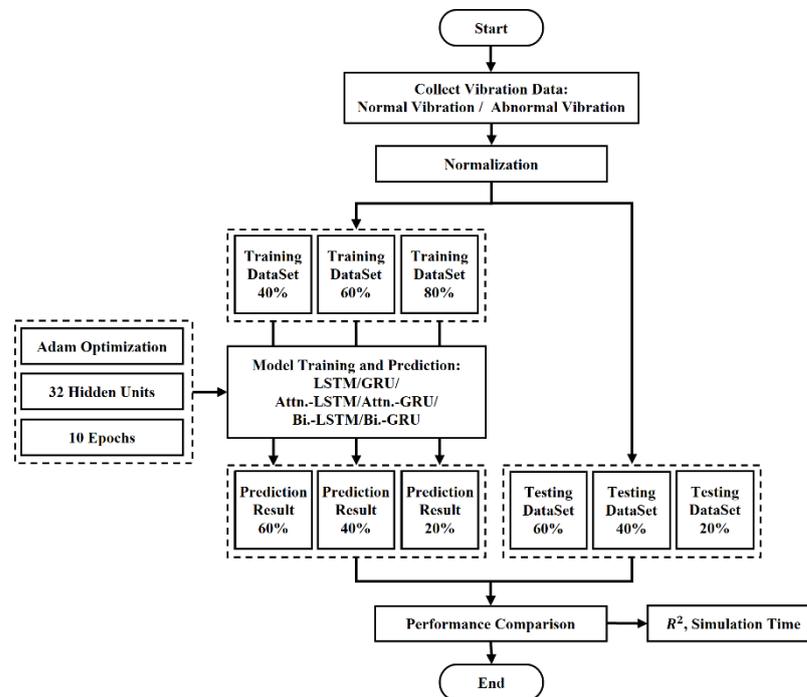


Figure 1. Flowchart of comparative analysis framework.

According to Figure 1, normal and abnormal motor vibration data are collected from the drones. Next, training datasets are built using 40%, 60%, and 80% of the collected normal and abnormal time series data. Then, the LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, and Bi-GRU techniques are used to predict vibrations in the remaining 60%, 40%, and 20% segments. Lastly, the prediction accuracies and runtime efficiencies of the six investigated RNN techniques are compared and analyzed to determine the technique most suitable for predicting the motor vibrations of drones.

2.2. Collection of Vibration Data

Normal and abnormal vibration time series data are collected from the motor connected to the drone's propeller to predict vibrations using different RNN techniques.

Figure 2 demonstrates the configuration of the collection of time series vibration data from the acceleration sensor, which is connected to the motor. Then, the vibration data in the 1 kHz band of 100 ms is collected through the acceleration sensor mounted on the motor of the drone.

Normal vibration data are collected from the motor without damage, as shown in Figure 3a, whereas abnormal vibration data are collected from the motor with a damaged rotor, as shown in Figure 3b. The revolutions per minute (RPM) per volt, the maximum output, and the maximum torque of the motor used to collect the time series vibration data are 180 Kv, 1484.6 W, and 1.992 N m, respectively. In addition, the frequency of the motor is set to 20 Hz to collect vibration data from both normal and abnormal motors. In other words, the number of rotations of the motor is set to 1200 RPM to collect time series vibration data composed of a total of 2000 time steps. The collected time series vibration data values are then normalized to a range from 0 to 1 using the min-max normalization method. Normalization can be expressed as follows.

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where x and x_{norm} denote the original vibration value that was collected and the normalized vibration value, respectively. Figure 3 below shows the waveforms of the collected normal vibrations (N. V.) and abnormal vibrations (Ab. V.) that have undergone the normalization process.

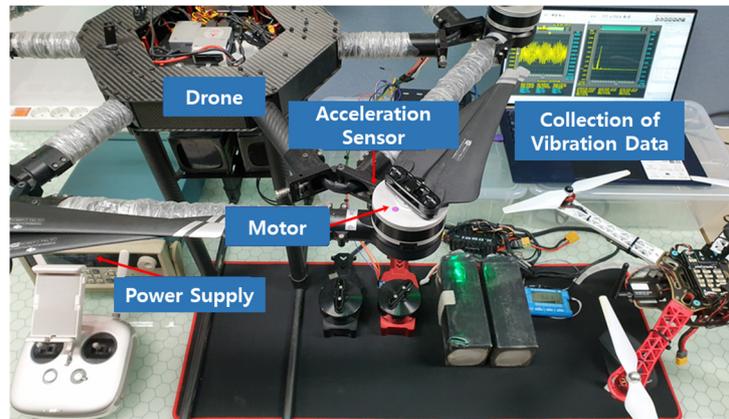


Figure 2. Configuration of collecting time series vibration data.

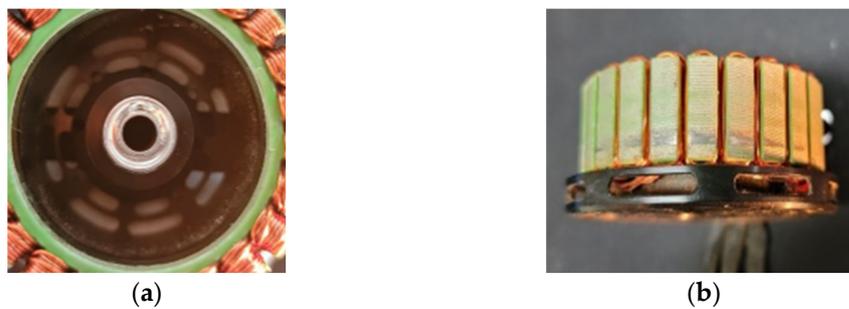


Figure 3. (a) Normal motor; (b) Motor with a damaged rotor.

Figure 4a,b show the waveforms of the normal and abnormal vibration data that were collected, respectively. According to Figure 4, the periods and amplitudes of the normal vibration waveform are consistent. However, the periods and amplitudes of the abnormal vibration waveform are irregular, and more small vibrations occur at the peak of the amplitudes.

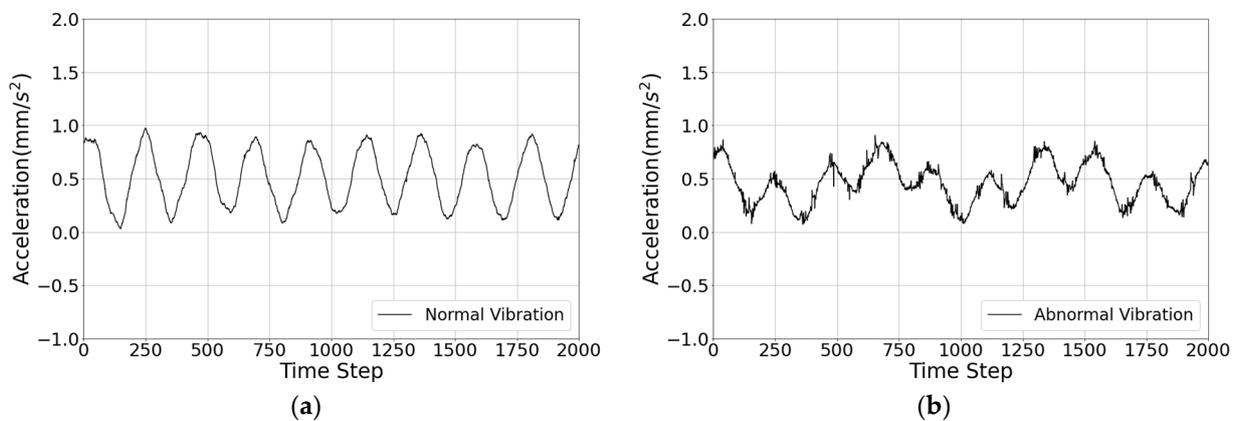


Figure 4. Waveform of the motor's vibrations: (a) a normal motor; (b) an abnormal motor.

3. RNN Techniques

This section explains the LSTM, GRU, Attn.-LSTM, Bi-LSTM, Attn.-GRU, and Bi-GRU techniques as well as the attention mechanism and bidirectional technique.

3.1. LSTM

Existing RNNs use gradient descent, which is a method that adjusts the weights of a neural network to minimize the loss function. RNNs also has a recursive structure in which past outputs are fed back as inputs. Therefore, when an existing RNN is trained with long sequence data, the depth of the neural network increases, and the vanishing gradient problem occurs. Hence, the LSTM technique was proposed to solve this limitation of the RNNs [44].

In Figure 5, C , t , x , and h denote the cell state, time step, input value, and the hidden state corresponding to the output, respectively. In addition, f , i , and o denote the forget, input, and output gates, respectively. LSTM uses a cell state and three gates to solve the existing RNNs' problem of long-term dependencies. Here, the cell state performs the role of passing information to the next LSTM cell without changing the information.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{2}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{3}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{5}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \tag{6}$$

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

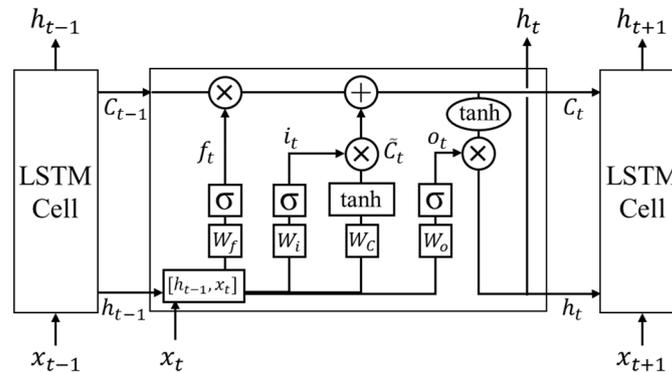


Figure 5. LSTM structure.

Here, W and b in Equations (2)–(5), denote the weight matrix and bias vector, respectively. The first step in LSTM determines which piece of information in the cell state to forget, and this step is expressed as Equation (2), which corresponds to the forget gate. The forget gate takes h_{t-1} , which is the hidden state of the previous time step and x_t , which is the input of the current time step, as the input and passes them through the sigmoid function to generate the output. This output value is then passed to C_{t-1} . Since the output value of the forget gate is generated through the sigmoid function, it has a value between 0 and 1. The larger the output value that the forget gate is, the longer the information is retained. Conversely, the smaller the forget gate's output value is, the faster the information is lost.

The next step in LSTM determines which piece of the information that has newly passed into the cell state will be stored, and this step is expressed as Equation (3), which corresponds to the input gate. As can be seen in Equation (3), \tilde{C}_t , which corresponds to new candidate values to be added to the cell state, this is generated through the \tanh function: an activation function. The input gate determines which of the candidate values will be added to the cell state.

The last step in LSTM updates the information in the cell state and determines which value will be output, and this step corresponds to Equations (6) and (7). The update of the cell state corresponds to Equation (6) and is performed by adding the element-wise product of f_t , the output of the forget gate, and C_{t-1} , the cell state of the previous time step, to the element-wise product of i_t , the output of the input gate, and \tilde{C}_t , the vector of the new candidate values. Lastly, as shown in Equation (7), the output value, h_t , is generated by taking the element-wise product of o_t , the value of the output gate, and the value obtained by applying the tanh function to C_t , to obtain the updated cell state value.

3.2. GRU

GRU is a modification to the LSTM technique. It has fewer parameters and requires lower amounts of computation than LSTM [45].

In Figure 6, t , x , and h , respectively, denote the time step, input, and the hidden state corresponding to the output. In addition, z and r represent the update and reset gates, respectively.

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \tag{8}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \tag{9}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) \tag{10}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{11}$$

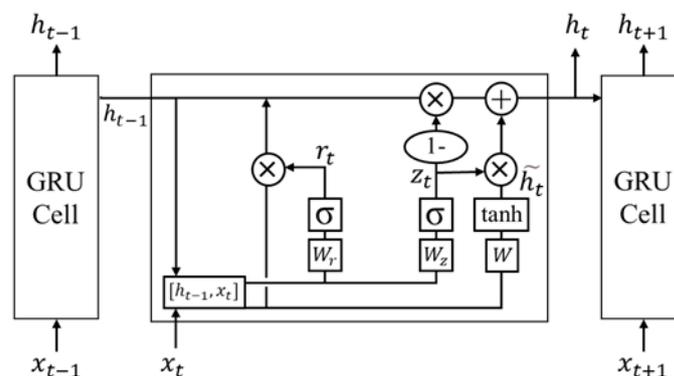


Figure 6. GRU structure.

W in Equations (8)–(11) denotes the weight matrix. Unlike LSTM, GRU passes information with only one hidden state without a separate cell state. The first step in GRU determines how much of the information from the output of the previous time step will be utilized by applying the sigmoid function to h_{t-1} , the output of the previous time step, and x_t , the input to the current time step. The first step is expressed as Equation (8), which corresponds to the reset gate. The result value of the reset gate is not used as it is. Instead, it is multiplied by the output of the previous time step, as shown in Equation (10).

The last step in GRU determines the proportion of the information from the previous and current time steps to be used. This step is expressed as Equation (9), which corresponds to the update gate. The result value of the update gate, z , corresponds to the importance of the current information and performs the role of the input gate of LSTM. In addition, $(1 - z_t)$ it corresponds to the importance of past information. The proportion of the current and past information to be used is determined using two values. Moreover, Equation (9) also expresses the forget gate of LSTM.

Finally, Equation (11) represents the element-wise product of $(1 - z_t)$, the importance of the past information, and h_{t-1} , the output of the previous time step, as well as the importance of the current information. \tilde{h}_t is obtained by applying the tanh function to the product of h_{t-1} , the output of the previous time step, and r_t , the result value of the

reset gate. Then, h_t , the output of the current time step, is obtained by performing the element-wise multiplication on \tilde{h}_t and z , which is the result value of the update gate.

3.3. Attention Mechanism

This subsection explains the attention mechanism applied to the existing LSTM and GRU. The attention mechanism is a mechanism that performs well in sequence-to-sequence problems. It considers the entire input data, calculates the similarity between the input data and the to-be-predicted value in advance, and reflects the calculated similarity in the output layer to generate the predicted value [46].

Figure 7 shows the architecture of LSTM and GRU with the attention mechanism. Here, $\langle \text{sos} \rangle$ denotes the start of the sequence, and x and M denote the vibration value of a single time step entered for prediction and the total number of vibration data values, respectively. The decoder block can have a hidden state since it is composed of LSTM/GRU cells similar to the encoder. First, the decoder block passes the hidden state of the current time, $\langle \text{sos} \rangle$, to the encoder. Then, the hidden state of the decoder block, which has been passed to the encoder, is used to calculate the attention score.

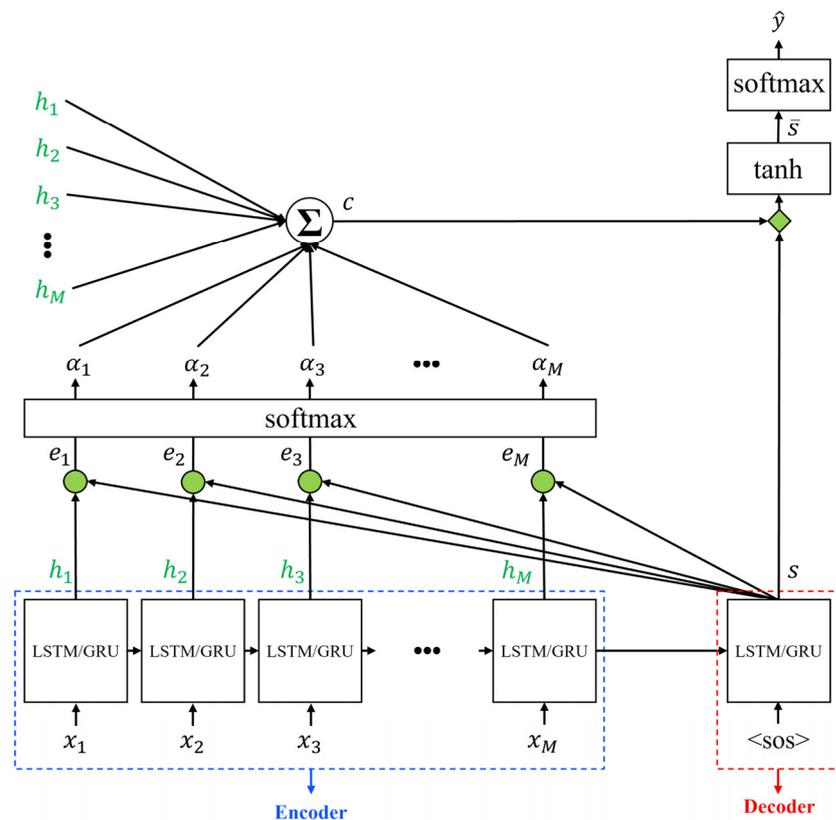


Figure 7. Architecture of LSTM and GRU with the attention mechanism.

The attention value obtained from the attention score and the hidden state of the decoder block are concatenated to create a vector, and the attention vector is calculated by a neural network operation in which tanh is the active function. Then, the final output value, \hat{y} , is calculated by using the generated attention vector as the input of the output layer. Equations for applying the attention mechanism are shown below.

$$e_i = s^T W_c h_i \tag{12}$$

$$\alpha_i = \text{softmax}(e_i) \tag{13}$$

$$c = \sum_{i=1}^M \alpha_i h_i \tag{14}$$

$$\bar{s} = \tanh(W_c [c; s]) \tag{15}$$

$$\hat{y} = \text{softmax}(W_y \bar{s} + b_y) \tag{16}$$

In Equation (12), e_i , s^T , W_c , and h_i represent the i th attention score of the encoder, the transpose of the hidden state of the decoder, the trainable weight matrix, and the i th hidden state of the encoder, respectively.

Equation (12) is the first step in the attention mechanism for generating the output value. It represents the attention score, which indicates the similarity between the input vibration value and the predicted vibration value.

Equation (13) is used to calculate the attention weight from the attention score. α_i and e_i in Equation (13) denote the attention weight calculated from the i th attention score of the encoder and the i th attention score of the encoder, respectively. For α_i , the softmax function is applied to the attention score to generate a probability distribution in which the sum of M attention weights are 1. These probability distribution values indicate the similarity between the input vibration value and the to-be-predicted vibration value.

Equation (14) is the third step of the attention mechanism for generating the output value, and it represents the attention value. c , M , α_i , and h_i in Equation (14) denote the attention value, the total number of input vibration values, the i th attention weight, and the i th hidden state of the encoder, respectively. Here, the attention value is the vector generated by the weighted sum of the attention weight and the hidden state of the encoder.

Equation (15) is the last step of the attention mechanism for generating the output value, and it represents the attention vector. Here, \bar{s} , W_c , c , and s denote the attention vector, the trainable weight matrix, the attention value, and the hidden state of the decoder, respectively. Then, to calculate the attention vector that will be used as the input to the output layer (W_y), the attention value and the hidden state of the decoder are combined to generate a single vector. This vector is then input to the trainable weight matrix to compute \bar{s} of the attention vector, which is the final output value of the attention mechanism.

Further, \hat{y} , W_y , \bar{s} , and b_y in Equation (16) denote the predicted value, the weight matrix of the output layer, the final output value of the attention mechanism, and the bias of the output layer, respectively. The final output value of the attention mechanism is input to the output layer to calculate the predicted value \hat{y} . Since the attention mechanism considers the similarity between the input vibration value and the to-be-predicted vibration value, and the final output value of the attention mechanism is passed as the input to the output layer, more accurate predicted values can be obtained.

3.4. Bidirectional RNN Techniques

Regular unidirectional RNNs can be transformed into bidirectional RNNs (Bi-RNNs) by connecting two independent hidden layers in opposite directions [47]. Figure 8 below shows the process of how Bi-LSTM and Bi-GRU generate predicted values.

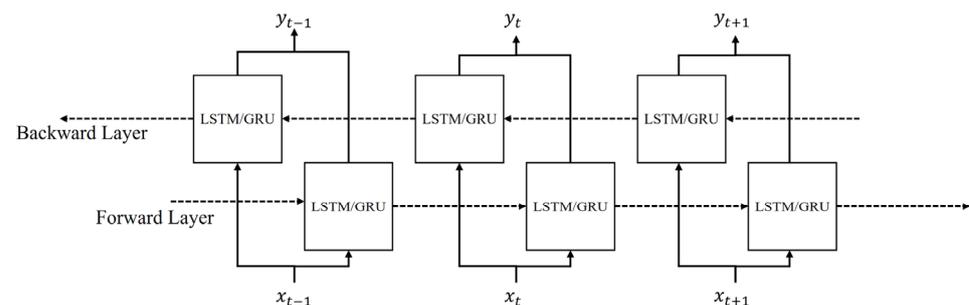


Figure 8. Architecture of Bi-LSTM and Bi-GRU.

In Figure 8, x , y , and t denote the input, output, and time step, respectively. As can be seen in Figure 8, Bi-RNNs generate predicted vibration values after performing the forward and backward operations. Therefore, the hidden state of a Bi-LSTM and Bi-GRU cell can have information in both forward and backward directions. As a result, Bi-LSTMs and Bi-GRUs typically outperform regular LSTMs and GRUs; however, they have a disadvantage in that the simulation runtime is increased.

4. Simulation Results and Discussion

The six investigated RNN (LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, and Bi-GRU) models were trained with 40%, 60%, and 80% segments of the collected normal and abnormal vibration data in this study. Then, the prediction performance of each model was evaluated on the validation sets. In this section, the predicted vibration results are explained. To compare and analyze the simulation results in detail, the efficiency of each model was comparatively analyzed based on the waveforms of the vibrations, scatter plots, the coefficient of determinations, and simulation runtime.

4.1. Simulation Environment and Parameters

This subsection describes the simulation environment and parameters used to predict the vibrations of the motor using the six investigated techniques. Table 2 below describes the simulation environment.

Table 2. Simulation environment.

Parameter	Specification
CPU	Intel(R) Xeon(R) CPU @ 2.20 GHz
GPU	NVIDIA Tesla T4
Memory	26 GB
Simulation Tools	Google Colaboratory Pro/Python 3.7.13/Tensorflow 2.8.2

In addition, Table 3 below lists the parameters configured for the simulation.

The values of ‘No. of epoch’, ‘No. of hidden units’, and ‘Initial learning rate’ in Table 2 are set to prevent overfitting during the simulation process and to compare, with more precision, the differences in the prediction accuracy using RNN techniques. The epoch was set to 10 to verify clear performance differences according to the size of the training data. In addition, the batch size was set to 32 since a batch size of between 32 and 128 is recommended [48].

Table 3. Simulation parameters.

Parameter	Value
No. of hidden units	32
Initial learning rate	0.0002
No. of epochs	10
Minimum batch size	32
Optimizer	Adam [49]
Iterations	10

The adaptive moment estimation (Adam) algorithm was used as the optimization algorithm to minimize the training error in the simulation and prevent the local minimum problem. The Adam algorithm is an improved technique compared to the stochastic gradient descent (SGD), in which there is a higher probability that the local minima problem will occur [49]. Furthermore, 10 iterations of the simulation have been conducted in total in order to compare the average coefficient of the determination and simulation runtimes for a more accurate performance comparison of the six RNN techniques.

4.2. Waveform of the Predicted Vibrations and Comparison of Scatter Plots

This subsection compares and analyzes the waveforms and scatter plots of the vibrations predicted from the normal and abnormal motor vibration data for the six investigated models.

Figures 9 and 10 show the waveforms of the vibrations predicted from the normal and abnormal vibration data, respectively, using the six investigated techniques. For the prediction results of both the normal and abnormal vibration data, the vibration prediction accuracy increased for all six RNN models as the training set increased from 40% to 80%.

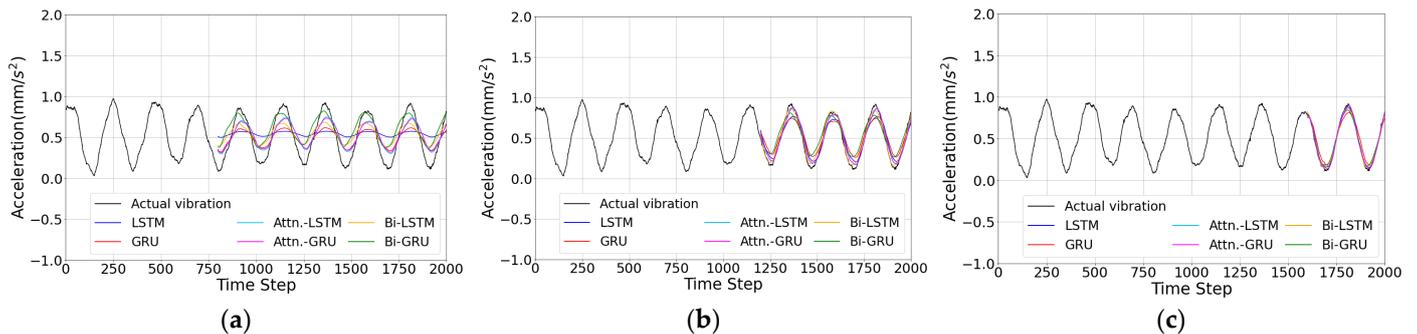


Figure 9. Waveforms of the vibrations predicted from the normal vibration: (a) 40%, (b) 60% (c) 80% segments of training data.

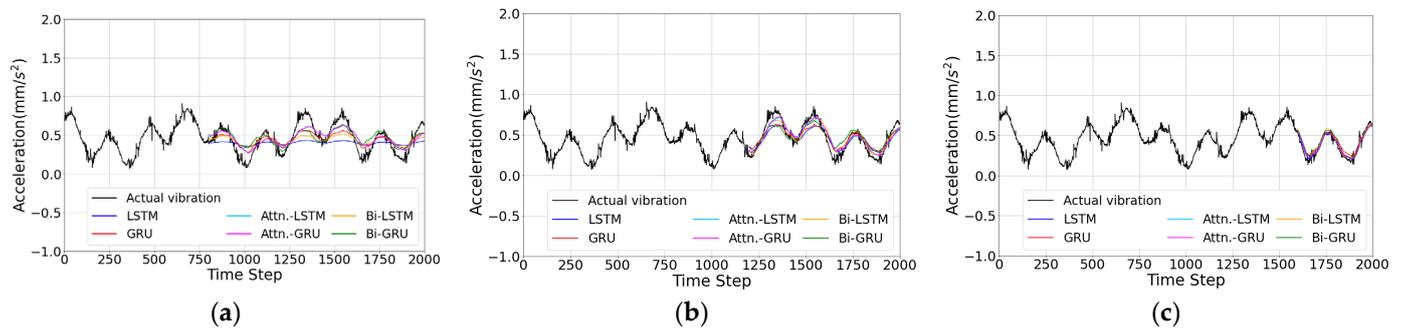


Figure 10. Waveforms of the vibrations predicted from the abnormal vibrations: (a) 40%, (b) 60% (c) 80% segments of training data.

In addition, according to the vibration prediction results in Figures 9a and 10a, the vibration prediction accuracy of LSTM was the lowest among the six RNN models when the size of the training dataset was 40%. As shown in Figures 9b and 10b, when the size of the training dataset was 60%, Attn.-LSTM and Attn.-GRU had the highest prediction accuracy, whereas LSTM and GRU exhibited the lowest prediction accuracy. Moreover, the results in Figures 9c and 10c show that when the size of the training dataset was 80%, the RNN techniques with the attention mechanism and bidirectional method achieved vibration prediction accuracies that were very similar to the actual vibrations.

Figures 11 and 12 below show the scatter plots of the similarities between the actual vibration values and the vibration values predicted from the normal and abnormal vibration data, respectively, for the six investigated models.

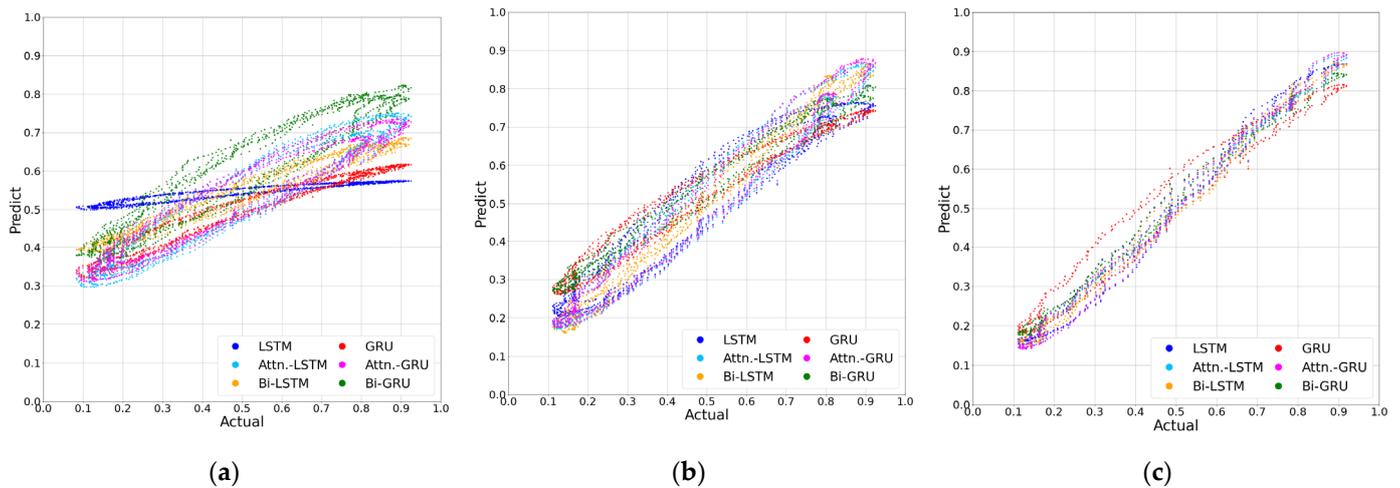


Figure 11. Scatter plots of normal vibrations: (a) 40%, (b) 60%, (c) 80% segment of training data.

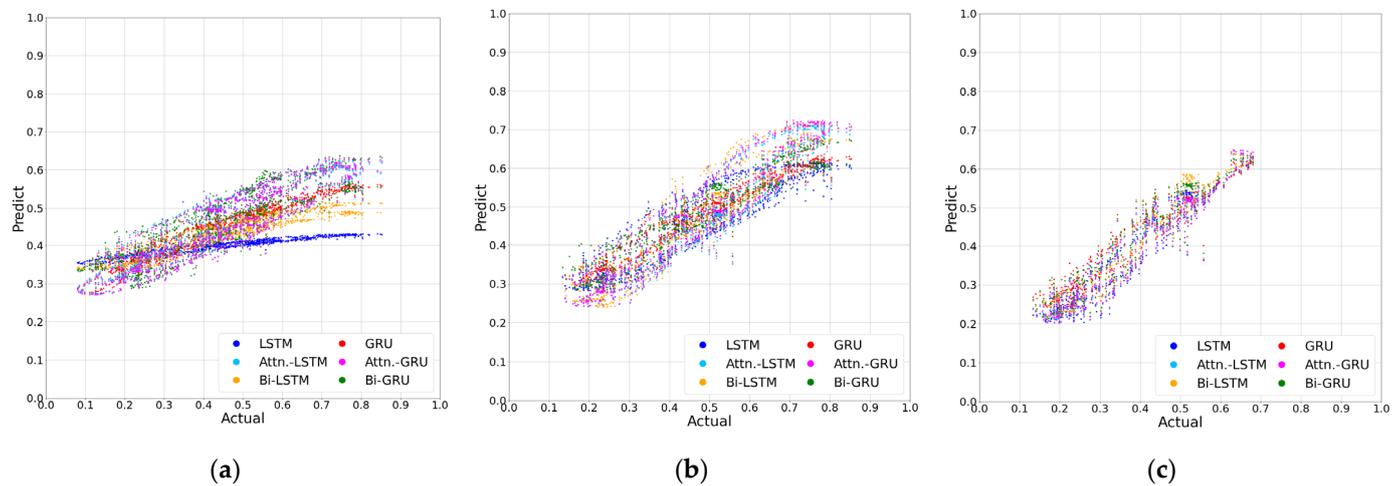


Figure 12. Scatter plots of abnormal vibrations: (a) 40%, (b) 60%, (c) 80% segment of training data.

According to Figures 11 and 12, which show the scatter plots of the results predicted from the normal and abnormal vibration data, respectively, the prediction accuracy of all six RNN models increased with the increase in the training segment.

In particular, according to Figures 11a and 12a, the accuracy of the vibrations predicted using LSTM was much lower than the accuracy of the vibrations predicted using the other five RNN models when the size of the training data was 40%. Additionally, according to Figures 11c and 12c, the vibrations predicted from the normal vibrations were more accurate than the vibrations predicted from the abnormal vibrations for all six RNN models.

This subsection comparatively analyzed, through the waveforms and scatter plots, the vibration prediction accuracies of the six RNN models according to the increase in the segment of the training data. To comparatively analyze the accuracies of the vibrations predicted in more detail, the changes in the coefficients of determination of the six RNN models according to the increase in the segment of the training data are analyzed in Section 4.3.

4.3. Comparative Analysis of the Accuracy of the Predicted Vibrations

This subsection comparatively analyzes the accuracy of the predicted values for each RNN model based on the changes in their coefficients of determination according to the

increase in the segment of the training data. To compare the accuracies of the vibrations predicted for the six RNN techniques, the coefficient of determination is defined as follows:

$$R^2 = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \tag{17}$$

where N , y_i , \hat{y}_i , and \bar{y} denote the total number of vibration data values, the actual vibration values, the predicted vibration values, and the average of the vibration value, respectively.

Figure 13 below shows the average value of the coefficient of determination according to the change in the size of the training data. Here, the value of the coefficient of determination was additionally calculated for the 50%, 70%, and 90% segments of the training data to analyze in more detail the changes in the prediction accuracies of the LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, and Bi-GRU models with the increase in the size of the training data.

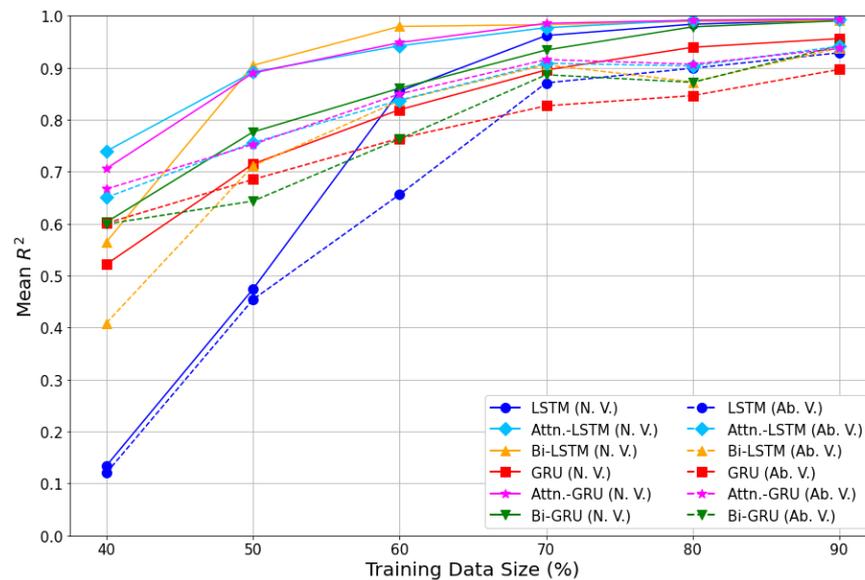


Figure 13. Average value of the coefficient of determination according to the size of the training data.

The solid and dotted lines in Figure 13 represent the average value of the coefficient of determination for the results predicted from normal and abnormal vibration data, respectively. According to Figure 13, the average value of the coefficient of determination for the normal vibration data was higher than that of the abnormal vibration data in most segments of the training data for the RNN techniques. However, as an exception, the average value of the coefficient of determination for GRU was lower for the normal vibration data than for the abnormal vibration data when the size of the training data was 40%.

Table 4 below shows the average value of the coefficient of determination for each segment of the training data, as shown in Figure 13.

According to Table 4, the average value of the coefficient of determination for the vibrations predicted from all six RNN models increased with the increase in the size of the training data. In addition, the average value of the coefficient of determination for the normal predicted vibrations converged to one for all six RNN models. However, the average value of the coefficient of determination for the results predicted from abnormal vibration data converged to 0.9, which is less than one. This subsection compared the average values of the coefficient of determination according to the increase in the segment of the training data. Next, in Section 4.4, the time taken to predict the vibrations is also considered to analyze the runtime efficiency of the investigated models.

Table 4. The average value of the coefficient of determination for each segment of the training data.

RNN	Vibration	Avg. R^2					
		40%	50%	60%	70%	80%	90%
LSTM	N. V.	0.134	0.474	0.854	0.961	0.983	0.991
	Ab. V	0.120	0.453	0.655	0.870	0.898	0.928
Attn.-LSTM	N. V.	0.739	0.891	0.942	0.976	0.991	0.993
	Ab. V	0.600	0.684	0.763	0.826	0.846	0.896
Bi-LSTM	N. V.	0.565	0.904	0.979	0.982	0.989	0.990
	Ab. V	0.408	0.711	0.836	0.905	0.872	0.937
GRU	N. V.	0.521	0.714	0.819	0.895	0.939	0.955
	Ab. V	0.599	0.643	0.762	0.886	0.871	0.942
Attn.-GRU	N. V.	0.705	0.889	0.948	0.985	0.991	0.993
	Ab. V	0.650	0.755	0.837	0.908	0.903	0.941
Bi-GRU	N. V.	0.602	0.776	0.860	0.933	0.978	0.990
	Ab. V	0.666	0.752	0.849	0.915	0.906	0.936

4.4. Comparative Analysis of Runtime Efficiency

Sections 4.2 and 4.3 verified that the vibration prediction accuracies for the LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, and Bi-GRU models all increased with the increase in the size of the training data. Here, this subsection comparatively analyzes the efficiency of each RNN model by comparing the simulation runtimes and the coefficients of determination.

Table 5 shows the average simulation runtime of each model when the size of the training data was 40%, 60%, and 80%. According to Table 5, the simulation runtime required to predict the vibrations increased with the increase in the size of the training data for all six RNN models. Therefore, the model with the highest coefficient of determination compared to the simulation runtime required to predict the vibrations was the most efficient technique for predicting vibrations.

Table 5. Average simulation runtime for each RNN model for different segments of training data (60%, 40%, 80%).

RNN	Vibration	Avg. Simulation Runtime (s)		
		40%	60%	80%
LSTM	N. V.	6.974	6.974	6.974
	Ab. V	7.329	7.329	7.329
Attn.-LSTM	N. V.	8.499	8.499	8.499
	Ab. V	8.603	8.603	8.603
Bi-LSTM	N. V.	11.081	11.081	11.081
	Ab. V	11.299	11.299	11.299
GRU	N. V.	7.256	7.256	7.256
	Ab. V	7.518	7.518	7.518
Attn.-GRU	N. V.	10.967	10.967	10.967
	Ab. V	10.905	10.905	10.905
Bi-GRU	N. V.	8.475	8.475	8.475
	Ab. V	8.496	8.496	8.496

Figure 14 below compares both the average simulation runtimes and average values of the coefficient of determination for the six RNN models to compare the efficiency of each model.

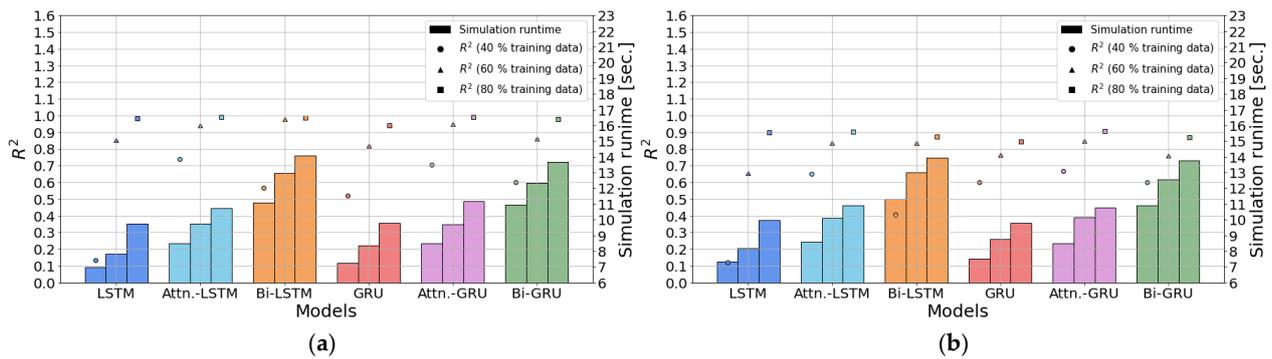


Figure 14. Comparison of the average value of the coefficient of determination for each RNN model according to the average simulation runtime: (a) Normal vibrations; (b) Abnormal vibrations.

In Figure 14, the circle, triangle, and square represent the average value of the coefficient of determination when the size of the training data was 40%, 60%, and 80%, respectively. In addition, the bar graph shows the simulation runtimes of the RNN models. According to Figure 14a,b, the average values of the coefficient of determination of Attn.-LSTM and Attn.-GRU, with the attention mechanism, were higher than or similar to those of the other RNN models. Therefore, in this subsection, the simulation runtime required for the prediction and the rate of change for the coefficient of determination is analyzed when the attention mechanism and bidirectional technique were incorporated into the existing LSTM and GRU techniques.

Table 6 below compares the rate of change for the coefficient of determination and the simulation runtime when the attention mechanism and bidirectional technique were incorporated into the LSTM technique.

Table 6. Comparison of the rate of change between Attn.-LSTM and Bi-LSTM.

RNN	Vibration	Efficiency (%)					
		40%		60%		80%	
		R^2	Sim. Runtime	R^2	Sim. Runtime	R^2	Sim. Runtime
Attn.-LSTM	N. V.	+451.5	+21.9	+10.3	+25.1	+0.8	+10.3
	Ab. V	+441.7	+17.4	+27.8	+23.9	+0.6	+9.2
	Avg.	+446.6	+19.7	+19.1	+24.5	+0.7	+9.8
Bi-LSTM	N. V.	+359.2	+56.0	+19.1	+63.1	+0.7	+43.5
	Ab. V	+240.0	+54.2	+27.6	+59.3	−2.9	+39.7
	Avg.	+299.6	+55.1	+23.4	+61.2	−1.1	+41.6

According to Table 6, the accuracy compared to the rate of increase in the simulation runtime was higher in Attn.-LSTM, with the attention mechanism, than in Bi-LSTM, with the bidirectional technique, for both normal and abnormal vibration prediction results in all segments of the training data. For example, when the size of the training data was 40%, the average value of the coefficient of determination for Attn.-LSTM increased by about 446.6% compared to that for LSTM, whereas the simulation runtime increased by only 19.7%.

However, the average value of the coefficient of determination increased by about 299.6% when predicting vibrations using Bi-LSTM compared to when LSTM was used, while the simulation time increased by 55.1%, which was greater than the increase when Attn.-LSTM was used. When the segment of the training data was 40%, the rate of increase in the simulation runtime was lower for Attn.-LSTM than for Bi-LSTM, while the vibration prediction accuracy increased more for Attn.-LSTM. Hence, the comparative analysis verified that Attn.-LSTM was more efficient than Bi-LSTM.

Moreover, when the segment of training data was 60%, the increase in the average value of the coefficient of determination for Bi-LSTM was about 4.3% higher than that for

Attn.-LSTM; however, the simulation runtime increased by about 36.7% more for Bi-LSTM than for Attn.-LSTM. Hence, Attn.-LSTM had better efficiency than Bi-LSTM. In particular, when the abnormal vibration data predicted from the segment of the training data was 80%, the average value of the coefficient of determination decreased slightly by about 1.1%, while the simulation runtime increased by about 41.6%. Therefore, the efficiency of Attn.-LSTM was superior to that of Bi-LSTM. Therefore, the rate of change in the simulation runtime and the average value of the coefficient of determination were compared between the Attn.-LSTM and Bi-LSTM techniques and analyzed. This comparative analysis confirmed that Attn.-LSTM had the best efficiency.

Table 7 below shows the change rate in the simulation runtime and the average value of the coefficient of determination when the attention mechanism and bidirectional techniques were incorporated into the GRU technique.

Table 7. Comparison of the rate of change between Attn.-GRU and Bi-GRU.

RNN	Vibration	Efficiency (%)					
		40 %		60 %		80 %	
		R^2	Sim. Runtime	R^2	Sim. Runtime	R^2	Sim. Runtime
Attn.-GRU	N. V.	+35.3	+16.8	+15.8	+16.4	+5.5	+14.2
	Ab. V	+11.0	+13.0	+11.3	+15.8	+7.1	+9.80
	Avg.	+23.2	+14.9	+13.6	+16.1	+6.3	+12.0
Bi-GRU	N. V.	+13.5	+49.4	+5.40	+45.8	+4.6	+39.5
	Ab. V	−0.20	+45.1	−0.10	+43.4	+3.0	+40.6
	Avg.	+6.70	+47.3	+2.70	+44.6	+3.8	+40.1

According to Table 7, both Attn.-GRU and Bi-GRU have an increased simulation runtime compared to general GRUs in all segments of the training data. However, the increase in the average value of the coefficient of determination was greater for Attn.-GRU than for Bi-GRU. For example, when the size of the training data was 40%, the average value of the coefficient of determination for Attn.-GRU increased by about 23.2% compared to that for GRU, and the simulation runtime increased by only about 14.9%. Meanwhile, the average value of the coefficient of determination for Bi-GRU increased slightly by about 6.7% compared to that for GRU, and the average simulation runtime increased by about 47.3%.

Similarly, when the size of the training data was 60%, the average value of the coefficient of determination for Attn.-GRU increased by about 6% compared to that for GRU, and the simulation runtime increased by about 16.1%. Meanwhile, the average value of the coefficient of determination for Bi-GRU increased by about 2.7% compared to that for GRU, and the average simulation runtime increased by about 44.6%.

In addition, when the size of the training data was 80%, the average value of the coefficient of determination for Attn.-GRU increased by about 6.3% compared to that for GRU, whereas the average value of the coefficient of determination for Bi-GRU increased by about 3.8%. As for the average simulation runtime, the average simulation runtime for Bi-GRU increased by 40.1%, which was 28.1% higher than the increase in the average simulation runtime for Attn.-GRU.

Therefore, the comparison of the rate of change in the average simulation runtime and the average coefficient of determination confirmed that Attn.-GRU had the best vibration prediction efficiency. Section 4.5 explains why the RNN techniques with the attention mechanism had the best efficiency.

4.5. Analysis of Vibration Prediction of Attention Mechanism

This subsection explains why the best performance was achieved when the attention mechanism was incorporated into the LSTM and GRU techniques.

In Figure 15, the values in the blue boxes represent the input vibration values used for the prediction, and the values in the red boxes represent the predicted vibration values. In addition, v , x , and y in Figure 15 denote the collected vibration data values, the input vibration values used for the prediction, and the predicted output vibration values, respectively.

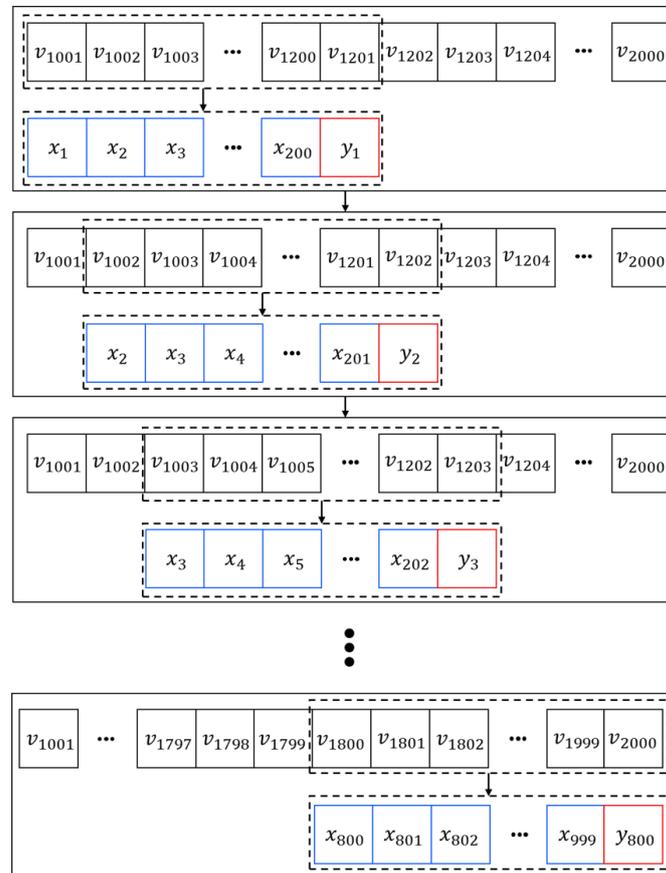


Figure 15. Process of predicting vibration values in the 40% prediction segment.

As shown in Figure 15, the number of input vibration values was set to 200, the same size as one cycle of the waveform of the vibration data. Then, a total of 800 vibration values (y_t) were predicted sequentially from these input vibration values. The 200 vibration values were used as the input to the encoder of the attention mechanism. The attention value, a vector representing the correlation with the predicted vibration values as in Equation (16), was calculated using these input values and used as the input to the output layer. Hence, the correlation with the to-be-predicted vibration values was passed to the output layer (W_y) of the neural network. In other words, Attn.-LSTM and Attn.-GRU calculated the correlation between the input vibration values and the to-be-predicted vibration values before predicting the vibrations. Conversely, the general LSTM and GRU, Bi-LSTM, and Bi-GRU predicted the vibration values without calculating the correlation between the input vibration values and the to-be-predicted vibration values. Therefore, Attn.-LSTM and Attn.-GRU have a higher prediction accuracy than other models.

On the one hand, Bi-LSTM and Bi-GRU collected information in both the backward and forward directions to increase the amount of information to be used for prediction. As a result, the vibration prediction accuracy was improved. However, the simulation runtime also increased rapidly. On the other hand, Attn.-LSTM and Attn.-GRU obtained information on the correlation between the input values and the to-be-predicted values only in the forward direction without unnecessary repetition. Hence, their vibration prediction accuracy is similar to or higher than that of Bi-LSTM and Bi-GRU, but their simulation runtime is shorter.

5. Conclusions

This paper conducts a comparative analysis of the prediction accuracy and runtime efficiency of six RNN techniques (i.e., LSTM, Attn.-LSTM, Bi-LSTM, GRU, Attn.-GRU, and Bi-GRU) using the motor vibration data of drones. The six RNN models were trained on 40%, 60%, and 80% segments of the collected normal and abnormal vibration data. Then, the simulation runtime required to predict the remaining 60%, 40%, and 20% segments and the accuracy of the prediction results were comparatively analyzed.

Based on the simulation results, it was verified that the simulation runtime and prediction accuracy increased with the increase in the size of the training data for all six investigated RNN models. However, the results of comparing the efficiencies by analyzing both the simulation runtime and prediction accuracy showed that the Attn.-LSTM and Attn.-GRU techniques achieved the best efficiency.

When the attention mechanism was incorporated into the regular LSTM and GRU models, the attention value, which represents the correlation between the input vibration values for training and the to-be-predicted vibration values, was calculated and used to predict the vibrations. The simulation results verified that Attn.-LSTM and Attn.-GRU could predict accurate vibration values more quickly than the LSTM, GRU, Bi-LSTM, and Bi-GRU models, which did not calculate the correlation between the input vibration values and the to-be-predicted vibration values.

Therefore, based on the simulation results, it was confirmed that RNN models with an attention mechanism are most suitable for predicting time series vibration data. In the future, we plan to conduct research on predicting vibrations when the normal and abnormal vibrations coexist in near real-time using RNN techniques with attention mechanisms.

Author Contributions: Conceptualization, J.-K.H.; methodology, J.-K.H.; software, J.-H.L. and J.-K.H.; validation, J.-H.L. and J.-K.H.; formal analysis, J.-K.H.; writing—original draft preparation, J.-H.L. and J.-K.H.; writing—review and editing, J.-K.H.; visualization, J.-H.L.; supervision, J.-K.H.; funding acquisition, J.-K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-004).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gupta, L.; Jain, R.; Vaszkun, G. Survey of Important Issues in UAV Communication Networks. *IEEE Commun. Surv. Tuts.* **2016**, *18*, 1123–1152. [[CrossRef](#)]
2. Yao, P.; Xie, Z.; Ren, P. Optimal UAV Route Planning for Coverage Search of Stationary Target in River. *IEEE Trans. Control Syst. Technol.* **2019**, *27*, 822–829. [[CrossRef](#)]
3. Agha-mohammadi, A.; Ure, N.K.; How, J.P.; Vian, J. Health Aware Stochastic Planning for Persistent Package Delivery Missions using Quadrotors. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 3389–3396.
4. Guo, Y.; Guo, J.; Liu, C.; Xiong, H.; Chai, L.; He, D. Precision Landing Test and Simulation of the Agricultural UAV on Apron. *Sensors* **2020**, *20*, 3369. [[CrossRef](#)] [[PubMed](#)]
5. Mazzia, V.; Comba, L.; Khaliq, A.; Chiaberge, M.; Gay, P. UAV and Machine Learning Based Refinement of a Satellite-Driven Vegetation Index for Precision Agriculture. *Sensors* **2020**, *20*, 2530. [[CrossRef](#)]
6. Schroth, L. *The Drone Market Size 2020–2025: 5 Key Takeaways*; Drone Industry Insights UG: Hamburg, Germany, 2020.
7. Product Safety Team. *Drone Safety Status Survey Results*; Korea Consumer Agency: Chungbuk Innovation City, Korea, 2017.
8. Khan, T.; Alekhya, P.; Seshadrinath, J. Incipient Inter-turn Fault Diagnosis in Induction Motors using CNN and LSTM based Methods. In Proceedings of the 2018 IEEE Industry Applications Society Annual Meeting, Portland, OR, USA, 23–27 September 2018; pp. 1–6.
9. Tian, H.; Ren, D.; Li, K.; Zhao, Z. An Adaptive Update Model based on improved Long Short Term Memory for Online Prediction of Vibration Signal. *J. Intell. Manuf.* **2020**, *32*, 37–49. [[CrossRef](#)]
10. Xiao, D.; Huang, Y.; Zhang, X.; Shi, H.; Liu, C.; Li, Y. Fault Diagnosis of Asynchronous Motors Based on LSTM Neural Network. In Proceedings of the 2018 Prognostics and System Health Management Conference, Chongqing, China, 26–28 October 2018; pp. 540–545.

11. Hong, J.-K.; Lee, Y.-K. LSTM-based Anomal Motor Vibration Detection. In Proceedings of the 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Kanazawa, Japan, 20–22 June 2021; pp. 98–99.
12. ElSaid, A.; Wild, B.; Higgins, J.; Desell, T. Using LSTM Recurrent Neural Networks to Predict Excess Vibration Events in Aircraft Engines. In Proceedings of the 2016 IEEE 12th International Conference on e-Science, Baltimore, MD, USA, 23–27 October 2016; pp. 260–269.
13. Wang, R.; Feng, Z.; Huang, S.; Fang, X.; Wang, J. Research on Voltage Waveform Fault Detection of Miniature Vibration Motor Based on Improved WP-LSTM. *Micromachines* **2020**, *11*, 753. [[CrossRef](#)]
14. ElSaid, A.; Desell, T.; Jamiy, E.F.; Higgins, J.; Wild, B. Optimizing Long Short-Term Memory Recurrent Neural Networks Using Ant Colony Optimization to Predict Turbine Engine Vibration. *Appl. Soft Comput.* **2018**, *73*, 969–991. [[CrossRef](#)]
15. Yang, Y.; Qin, N.; Huang, D.; Fu, Y. Fault Diagnosis of High-Speed Railway Bogies Based on LSTM. In Proceedings of the 2018 5th International Conference on Information, Cybernetics, and Computational Social Systems, Hangzhou, China, 16–19 August 2018; pp. 393–398.
16. Xiao, D.; Huang, Y.; Qin, C.; Shi, H.; Li, Y. Fault Diagnosis of Induction Motors Using Recurrence Quantification Analysis and LSTM with Weighted BN. *Shock Vib.* **2019**, *2019*, 8325218. [[CrossRef](#)]
17. Jiang, J.R.; Lee, J.E.; Zeng, Y.M. Time Series Multiple Channel Convolutional Neural Network with Attention-Based Long Short-Term Memory for Predicting Bearing Remaining Useful Life. *Sensors* **2020**, *20*, 166. [[CrossRef](#)]
18. Xu, H.; Ma, R.; Yan, L.; Ma, Z. Two-stage Prediction of Machinery Fault Trend based on Deep Learning for Time Series Analysis. *Digit. Signal Process.* **2021**, *117*, 103150. [[CrossRef](#)]
19. Huang, C.G.; Huang, H.Z.; Li, Y.F. A Bidirectional LSTM Prognostics Method Under Multiple Operational Conditions. *IEEE Trans. On Indust. Electro.* **2019**, *66*, 8792–8802. [[CrossRef](#)]
20. Zhang, Y. Aeroengine Fault Prediction Based on Bidirectional LSTM Neural Network. In Proceedings of the 2020 IEEE International Conference on Big Data, Artificial Intelligence and Internet of Things, Fuzhou, China, 12–14 June 2020; pp. 317–320.
21. Lee, K.; Kim, J.K.; Kim, J.; Hur, K.; Kim, H. Stacked Convolutional Bidirectional LSTM Recurrent Neural Network for Bearing Anomaly Detection in Rotating Machinery Diagnostics. In Proceedings of the 2018 1st IEEE Inter. Confer. On Know. Innov. And Inven, Jeju Island, South Korea, 23–27 July 2018; pp. 98–101.
22. Liang, J.; Wang, L.; Wu, J.; Liu, Z.; Yu, G. Prediction of Spindle Rotation Error through Vibration Signal based on Bi-LSTM Classification Network. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1043*, 042033. [[CrossRef](#)]
23. Zhu, J.; Ma, H.; Ji, L.; Zhuang, J.; Wang, J.; Liu, B. Vibration Trend Prediction of Pumped Storage Units based on VMD and GRU. In Proceedings of the 2020 5th International Conference on Mechanical, Control and Computer Engineering 2020, Harbin, China, 25–27 December 2020; pp. 180–183.
24. Lee, K.; Kim, J.-K.; Kim, J.; Hur, K.; Kim, H. CNN and GRU Combination Scheme for Bearing Anomaly Detection in Rotating Machinery Health Monitoring. In Proceedings of the 2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII), Jeju Island, Korea, 23–27 July 2018; pp. 102–105.
25. Ma, M.; Mao, Z. Deep Wavelet Sequence-Based Gated Recurrent Units for the Prognosis of Rotating Machinery. *Struct. Health Monit.* **2021**, *4*, 1794–1804. [[CrossRef](#)]
26. Zhao, K.; Shao, H. Intelligent Fault Diagnosis of Rolling Bearing Using Adaptive Deep Gated Recurrent Unit. *Neural Processing Lett.* **2020**, *51*, 1165–1184. [[CrossRef](#)]
27. Zhao, K.; Jiang, H.; Li, X.; Wang, R. An Optimal Deep Sparse Autoencoder with Gated Recurrent Unit for Rolling Bearing Fault Diagnosis. *Meas. Sci. Technol.* **2019**, *31*, 015005. [[CrossRef](#)]
28. Zhang, X.; Cong, Y.; Yuan, Z.; Zhang, T.; Bai, X. Early Fault Detection Method of Rolling Bearing Based on MCNN and GRU Network with an Attention Mechanism. *Shock. Vib.* **2021**, *2021*, 6660243. [[CrossRef](#)]
29. Cao, Y.; Jia, M.; Ding, P.; Ding, Y. Transfer Learning for Remaining Useful Life Prediction of Multi-conditions Bearings based on Bidirectional-GRU Network. *Measurement* **2021**, *178*, 109287. [[CrossRef](#)]
30. She, D.; Jia, M. A BiGRU Method for Remaining Useful Life Prediction of Machinery. *Measurement* **2021**, *167*, 108277. [[CrossRef](#)]
31. Shang, Y.; Tang, X.; Zhao, G.; Jiang, P.; Lin, T.R. A remaining Life Prediction of Rolling Element Bearings based on a Bidirectional Gate Recurrent Unit and Convolution Neural Network. *Measurement* **2022**, *202*, 11893. [[CrossRef](#)]
32. Wang, Z.; Qian, H.; Zhang, D.; Wei, Y. Prediction Model Design for Vibration Severity of Rotating Machine Based on Sequence-to-Sequence Neural Network. *Math. Probl. Eng.* **2019**, *2019*, 4670982. [[CrossRef](#)]
33. Yuan, M.; Wu, Y.; Lin, L. Fault Diagnosis and Remaining Useful Life Estimation of Aero Engine using LSTM Neural Network. In Proceedings of the 2016 IEEE international conference on aircraft utility systems (AUS), Beijing, China, 10–12 October 2016; pp. 135–140.
34. Demidova, L.A. Recurrent Neural Networks' Configurations in the Predictive Maintenance Problems. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *714*, 012005. [[CrossRef](#)]
35. Naren, R.; Subhashini, J. Comparison of Deep Learning Models for Predictive Maintenance. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *912*, 022029. [[CrossRef](#)]
36. Chen, B.; Peng, Y.; Gu, B.; Luo, Y.; Liu, D. A Fault Detection Method Based on Enhanced GRU. In Proceedings of the 2021 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), Nanjing, China, 21–23 October 2021; pp. 1–4.

37. Hong, J.-K. Vibration Prediction of Flying IoT Based on LSTM and GRU. *Electronics* **2022**, *11*, 1052. [[CrossRef](#)]
38. Zhang, L.; Liu, P.; Zhao, L.; Wang, G.; Zhang, W.; Liu, J. Air Quality Predictions with a Semi-supervised Bidirectional LSTM Neural Network. *Atmos. Pollut. Res.* **2021**, *12*, 328–339. [[CrossRef](#)]
39. Tao, Q.; Liu, F.; Li, Y.; Sidorov, D. Air Pollution Forecasting using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU. *IEEE Access* **2019**, *7*, 76690–76698. [[CrossRef](#)]
40. Qin, H. Comparison of Deep Learning Models on Time Series Forecasting: A Case Study of Dissolved Oxygen Prediction. *arXiv* **2019**, arXiv:1911.08414.
41. Althelaya, K.A.; El-Alfy, E.M.; Mohammed, S. Evaluation of Bidirectional LSTM for Short- and Long-term Stock Market Prediction. In Proceedings of the 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 3–5 April 2018; pp. 151–156.
42. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
43. Hollis, T.; Viscardi, A.; Yi, S.E. A Comparison of LSTMs and Attention Mechanisms for Forecasting Financial Time Series. *arXiv* **2018**, arXiv:1812.07699.
44. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
45. Chung, J.; Gülçehre, Ç.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.
46. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
47. Schuster, M.; Paliwal, K. Bidirectional Recurrent Neural Networks. *IEEE Trans. Signal Processing* **1997**, *45*, 2673–2681. [[CrossRef](#)]
48. Wu, Y.; Johnson, J. Rethinking “Batch” in BatchNorm. *arXiv* **2021**, arXiv:2105.07576.
49. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.