

Unidirectional and bidirectional LSTM for nonlinear synthetic time series forecasting

Nur Haizum Abd Rahman¹, Hani Syahida Zulkafli²

¹Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebu Persiaran Tun Khalil Yaakob, Gambang, Malaysia

²Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, Serdang, Malaysia

Article Info

Article history:

Received Feb 7, 2026

Revised Jul 24, 2026

Accepted Aug 28, 2026

Keywords:

Threshold autoregressive model

Self-exciting threshold autoregressive model

Time series

Hyperparameter tuning

Grid search

ABSTRACT

Deep learning models are sensitive to architecture choice and hyperparameter tuning. This study evaluates the forecasting performance of unidirectional long short-term memory (LSTM) and bidirectional LSTM (BiLSTM) networks using synthetic nonlinear time series data. Synthetic data were used to provide a controlled experimental setting in which the nonlinear data-generating process, sample size, and hyperparameter effects could be systematically examined. Threshold autoregressive (TAR) and self-exciting threshold autoregressive (SETAR) models were simulated with sample sizes, 1000 and 3000. Each series was partitioned into training and testing sets using an 80:20 split. Both networks were trained using grid search by varying the number of hidden neurons (32, 64, 128) and training epochs (20, 50, 100). Forecast accuracy was evaluated using mean squared error (MSE) and mean absolute error (MAE), averaged across 30 replications. The results show that increasing epochs and hidden neurons improves accuracy, but the improvements decline at higher settings. BiLSTM achieves lower forecasting errors than LSTM for both TAR and SETAR datasets, and the Wilcoxon test confirms that the differences are statistically significant. The main contribution of this study is a controlled empirical framework that links the architecture, sample size, and hyperparameter settings to nonlinear regime-switching forecasting performance. The findings suggest that BiLSTM is preferable when higher accuracy is required, while LSTM remains a practical alternative for smoother nonlinear series or resource-constrained applications.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nur Haizum Abd Rahman

Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah

Lebu Persiaran Tun Khalil Yaakob, Gambang, Malaysia

Email: haizum@ump.edu.my

1. INTRODUCTION

Accurate time series forecasting is crucial for decision-making in sectors such as banking, transportation, energy, and environmental science, where predicting future trends and anomalies is essential for operational efficiency and risk mitigation [1]-[3]. The increasing complexity of temporal connections in real-world datasets has emphasised the shortcomings of conventional statistical models, such as autoregressive integrated moving average (ARIMA) and exponential smoothing. These models often struggle to capture nonlinear patterns and long-range correlations in complex forecasting tasks. As a result, the research community has transitioned to deep learning techniques, particularly recurrent neural networks (RNNs), to address these issues and achieve enhanced accuracy in time-series forecasting [4].

Among RNN variants, long short-term memory (LSTM) networks have emerged as a robust solution for modelling sequential data, due to the LSTM characteristic of a gated architecture that facilitates memory retention and selective exclusion of information over extended time periods [5]. LSTM models excel at capturing short-term variations and long-term dependencies, making the LSTM highly suitable for complicated data forecasting applications. Two principal LSTM architectures have gained prominence, which are unidirectional LSTM, which processes input sequences in a single forward direction, and bidirectional LSTM (BiLSTM), which analyses data in both forward and backward directions [6]. This bidirectional mechanism allows BiLSTM models to leverage contextual information from both past and future states, potentially enhancing forecasting performance in applications where comprehensive sequence context is essential.

In recent years, the literature on LSTM-based forecasting has rapidly evolved to tackle increasingly complicated temporal patterns across several application areas. Simulation studies consistently show that BiLSTM models outperform unidirectional models in capturing intricate temporal dependencies, especially in short- and medium-term forecasting contexts such as traffic flow, stock market prediction, and hydrological modelling. The bidirectional architecture allows the model to analyse input sequences in both directions, utilising information from both past and future states to improve forecast accuracy. Multi-layer BiLSTM models have demonstrated considerable improvements in forecast accuracy relative to stacked or hybrid Uni-LSTM models, particularly for extended forecast horizons [7]-[9].

In comparative analysis, the performance evaluation relies on a variety of error metrics, such as root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R²). Besides the inferential statistical tests, such as the Wilcoxon Signed-Rank Test, is also used [10]. BiLSTM models, which produce lower error rates and greater correlation scores across many applications and forecast horizons, are frequently favoured by empirical results. These improvements, however, come at the cost of longer training times and greater computational complexity, prompting researchers to explore hardware acceleration and optimisation techniques to reduce resource requirements. While ensemble approaches that integrate LSTM, BiLSTM, and other deep learning models offer promising avenues for enhanced resilience, applying cross-validation, multi-horizon forecasting, and domain-specific measures further improves the reliability of these comparative evaluations [11]-[14].

In addition, although BiLSTM models offer clear benefits in accuracy and capturing context, their practical application is often limited by computational trade-offs. Research indicates that BiLSTM models require more training time and computational resources and may converge more slowly than LSTM models. This has sparked interest in optimisation techniques, such as hyperparameter tuning and hardware acceleration, to balance predictive performance with efficiency. Additionally, the interpretability of deep learning models remains a significant challenge, with attention mechanisms and feature selection emerging as important strategies for improving transparency and trust. The literature also emphasises the importance of better integrating LSTM models with complementary methods, such as autoregressive integrated moving average with exogenous variable (ARIMAX) and transformer-based networks, to further enhance forecasting accuracy and generalisability [15], [16].

A notable research gap persists in the reliable application of LSTM architectures for long-term forecasting horizons. Although BiLSTM models perform exceptionally well in short- and medium-term forecasts, their performance tends to deteriorate over longer time horizons, and there are no established recommendations for optimising these models in long-term scenarios. More thorough research on architectural tuning, interpretability, and methodological rigour is needed, according to the literature, particularly in understudied fields such as financial risk assessment and environmental monitoring. To address the issues of scalability, transparency, and long-term predictive dependability, future research aims to build effective training methodologies, improved hybrid structures, and standardised evaluation frameworks [9], [17], [18].

While the superiority of advanced LSTM and BiLSTM architectures is well documented, direct, head-to-head comparisons between unidirectional and BiLSTM models on both simulated and real stock price datasets remain limited. Recent studies that provide such comparisons utilise standardised metrics, such as MAE, and report that BiLSTM models generally achieve higher accuracy [19]. However, limited examination of how these performance differences depend jointly on data length and key hyperparameters. In addition, although recent work has explored grid search and adaptive gradient-based strategies for hyperparameter optimisation in deep learning models [3], [20], systematic evaluations that link these choices to nonlinear data-generating mechanisms and varying sample sizes are still insufficient.

This study investigated the effects of model architecture, sample size, and hyperparameter settings on the forecasting performance of LSTM and BiLSTM models for nonlinear synthetic time series. While earlier studies have explored the performance of LSTM and BiLSTM models in various forecasting applications, they have not explicitly addressed how their performance differences are influenced by controlled nonlinear data-generating processes, different sample sizes, and key hyperparameter settings.

Therefore, this study aims to design a unified experimental framework that combines nonlinear TAR and SETAR simulations and conduct a systematic grid search over hidden neurons and epochs for both LSTM and BiLSTM, allowing us to quantify how model architecture and hyperparameters interact with data length. This integrated approach provides practical, evidence-based guidelines for selecting both the model type and its hyperparameters in time series forecasting applications.

2. RESEARCH METHOD

2.1. Unidirectional LSTM

The unidirectional LSTM architecture aims to solve the vanishing gradient problem often seen in traditional RNNs [21]. The LSTM unit includes a cell state, input gate, output gate, and forget gate, which work together to control the flow of information. The cell state acts as a memory that carries data across time steps, while the gates manage the addition and removal of information. The forget gate determines which information should be removed or discarded from the previous cell states. It evaluates the current input alongside the output and memory cell state from the previous time step to make this decision. The forget gate was not part of the original LSTM architecture but was added later to allow the network to reset its state. Next gate, input gate decides what new information should be retained and added to the network's cell states. It determines the activation values and helps select the candidate values that could potentially be incorporated into the memory cell. Finally, the output gate calculates the final output of the LSTM block for the current iteration. It does this by combining the current input and the previous unit output with the newly updated cell value. This design enables LSTMs to retain relevant information over long sequences, making them especially effective for time series forecasting tasks. Figure 1 shows the architecture of the LSTM.

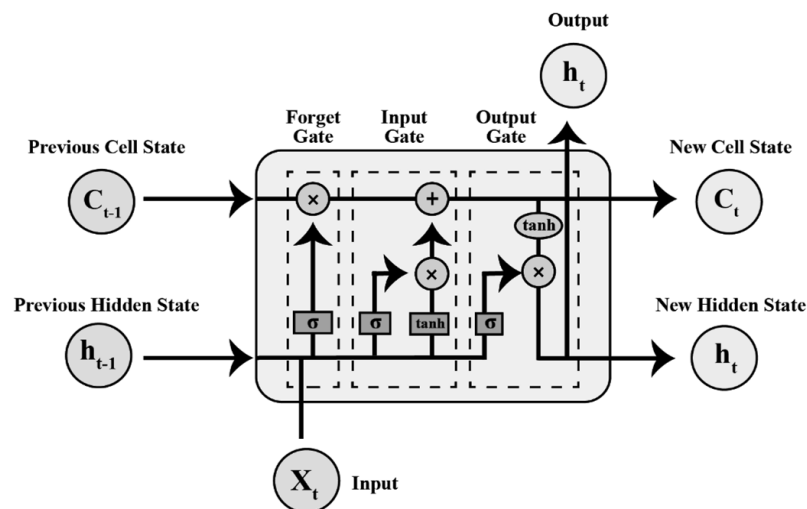


Figure 1. The architecture of unidirectional LSTM

In this study, the LSTM model is implemented using a single layer of LSTM units, followed by a fully connected layer that outputs the forecasted values. The model is trained using a mean squared error (MSE) loss function, which quantifies the difference between the predicted and actual values. The training process involves backpropagation through time (BPTT), where the gradients are computed and updated using an optimisation algorithm, the Adam optimiser.

2.2. Bidirectional LSTM

The BiLSTM architecture extends the capabilities of the standard LSTM by processing input sequences in both forward and backward directions [22]. This dual processing enables the model to capture contextual information from both past and future time steps, which can be especially useful in time series forecasting, where future context may offer valuable insights into the current state. The BiLSTM consists of two separate LSTM layers, where one processes the input sequence from start to finish (forward LSTM), and another processes it from end to start (backward LSTM). The outputs from both layers are then concatenated to maintain temporal relationships. Figure 2 illustrates the architecture of the BiLSTM.

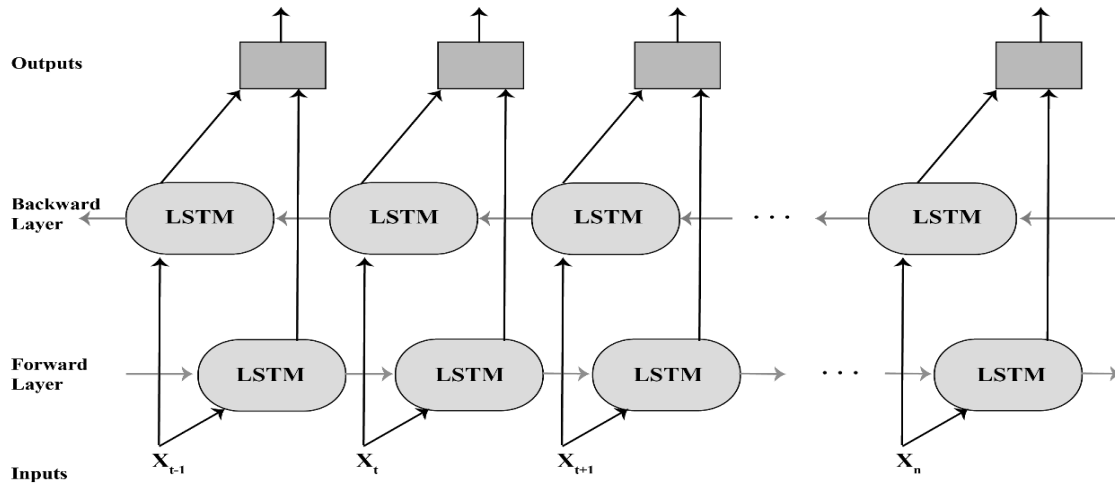


Figure 2. The architecture of BiLSTM

Similar to the LSTM model, the BiLSTM is trained using a MSE loss function and optimised through backpropagation. The architecture may include additional layers, such as dropout layers, to prevent overfitting and improve generalisation. The choice of hyperparameters used in the BiLSTM in this paper is the same as those used in the LSTM model.

2.3. Forecast accuracy measure

Many metrics can be used to assess the accuracy of prediction models. MSE is often utilised in studies to measure accuracy. MSE measures the differences between the actual values and predicted values. MSE measures the errors' variability. Hence, the smaller the MSE, the better the prediction performance. One drawback of MSE is that it is sensitive to data transformations or changes in scale, and it penalises extreme errors. Therefore, another measure used in this study is the MAE. MAE measures the average magnitude of the errors in the predicted values. The formulas to compute MSE and MAE are as (1), (2):

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

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

where y_t is the actual value at timestep t , and $\hat{y}_t$ is the predicted value at timestep t .

2.4. Simulation study

Two regimes-switching non-linear models, the threshold autoregressive (TAR) and self-exciting threshold autoregressive (SETAR) models, are used in this simulation study to generate non-linear time series data. These classical regime-switching models are widely used to model financial series with different dynamics in "low" and "high" regimes. Stock prices often exhibit nonlinear behaviour, asymmetric adjustment and regime, which can be captured by allowing the autoregressive parameters to change when a threshold is crossed [23].

Specifically, this study consider a two-regime TAR(2;1,1) model as (3):

$$y_t = \begin{cases} -0.2 y_{t-1} + \varepsilon_t, & |y_{t-1}| \leq 1, \\ 1.0 y_{t-1} + \varepsilon_t, & |y_{t-1}| > 1, \end{cases} \quad (3)$$

and a two-regime SETAR (2;1,1) model as (4):

$$y_t = \begin{cases} 0.19 + 0.92 y_{t-1} + \varepsilon_t, & |y_{t-1}| \leq 2.5, \\ 0.78 + 0.72 y_{t-1} + \varepsilon_t, & |y_{t-1}| > 2.5, \end{cases} \quad (4)$$

where $\varepsilon_t \sim N(0,1)$. The autoregressive coefficients are chosen in the interval $[-1,1]$ to induce nonlinear but stationary dynamics that resemble stock price behaviour [23].

3. RESULTS AND DISCUSSION

Each non-linear model simulated the dataset with two different sample sizes, 1000 and 3000, respectively. For the network hyperparameters, hidden neurons and epochs, three different numbers were used for each model. Consequently, the total number of combinations for simulated data and hyperparameter configurations is 18 for both the LSTM and BiLSTM models. The TAR and SETAR simulated series were replicated 30 times using independent random draws from the default pseudo-random number generator in R, and the reported MSE and MAE are averages over these 30 replications. The error term is assumed to be independent and identically distributed with $N(0,1)$. The datasets are split at 80:20, with 80% used for training and 20% for testing. Hence, the testing datasets for each sample size will be 200 and 600. The training dataset is used to train and run through LSTM and BiLSTM models. The trained models were then used to predict stock prices, and their prediction accuracy was evaluated. Table 1 summarises the hyperparameter values used for LSTM and BiLSTM in this study.

Table 1. Hyperparameter grid search values of LSTM and BiLSTM

Hyperparameters	Values
Hidden neuron	32,64,128
Hidden layer	1
Output neuron	1
Epoch	20,50,100
Batch size	32
Optimizer	Adaptive moment estimation (Adam)
Loss function	MSE

Grid search was selected for hyperparameter optimisation because this study involved a limited and predefined set of hyperparameter values. This approach allows each combination of hidden neurons and epochs to be evaluated systematically for both LSTM and BiLSTM models. The main advantage of grid search in this study is its transparency and reproducibility, as it enables a fair comparison of model performance across the same hyperparameter settings. Although more adaptive methods, such as Bayesian optimization and metaheuristic algorithms, may be more efficient for larger and more complex search spaces, grid search and its variants remain relevant for LSTM tuning when the hyperparameter space is small and well-defined [24]. Therefore, grid search is appropriate in this controlled simulation setting because the study aims to examine the effect of each selected configuration on forecasting accuracy.

Data transformation was performed to enhance the learning of the LSTM and BiLSTM models. Min-max normalisation was conducted first in 80% training, ensuring the data ranged from 0 to 1, and the same scaling parameters were applied to transform the test set. In addition, the LSTM and BiLSTM neural networks for each hyperparameter combination were replicated to account for the stochastic nature of the neural network algorithm. The average MSE and MAE were then computed and shown in Tables 2 and 3.

Table 2. MSE and MAE of LSTM and BiLSTM for TAR simulated data

Model	Sample size	Hidden neurons	MSE			MAE		
			20	50	100	20	50	100
LSTM	1000	32	5.8500	3.5955	2.5102	1.8985	1.4811	1.2177
		64	5.0613	2.9761	2.0308	1.7590	1.3492	1.0351
		128	4.2507	2.5305	1.9004	1.6184	1.2514	1.0726
	3000	32	3.7990	2.2055	1.7798	1.5259	1.1670	1.0267
		64	3.2514	1.9021	1.5244	1.4168	1.0799	0.9740
		128	2.8635	1.6384	1.4872	1.3317	1.0134	0.9657
BiLSTM	1000	32	3.7852	2.4445	1.9111	1.5201	1.2211	1.0828
		64	3.1001	2.1213	1.8122	1.3798	1.1399	1.0501
		128	2.6926	1.9390	1.7200	1.2967	1.0888	1.0363
	3000	32	2.7284	1.8797	1.6979	1.2973	1.0853	1.0245
		64	2.3327	1.8095	1.5039	1.2037	1.0582	0.9711
		128	2.2675	1.5626	1.4585	1.1927	0.9910	0.9601

For the TAR simulations result (refer to Table 2), the best-performing LSTM configuration for both sample sizes is the model with 128 hidden neurons and 100 epochs, yielding MSE of 1.9004 and MAE of 1.0726 for $n = 1000$ and MSE of 1.4872 and MAE of 0.9657 for $n = 3000$ as shown in Figure 3. The BiLSTM model shows the same optimal configuration, with slightly lower errors. The MSE showed the

value of 1.7200 and MAE of 1.0363 for $n = 1000$, and MSE of 1.4585 and MAE of 0.9601 for $n = 3000$. Figures 3(a) and 3(b) show that increasing the number of hidden neurons and epochs generally decreases the MSE for both architectures.

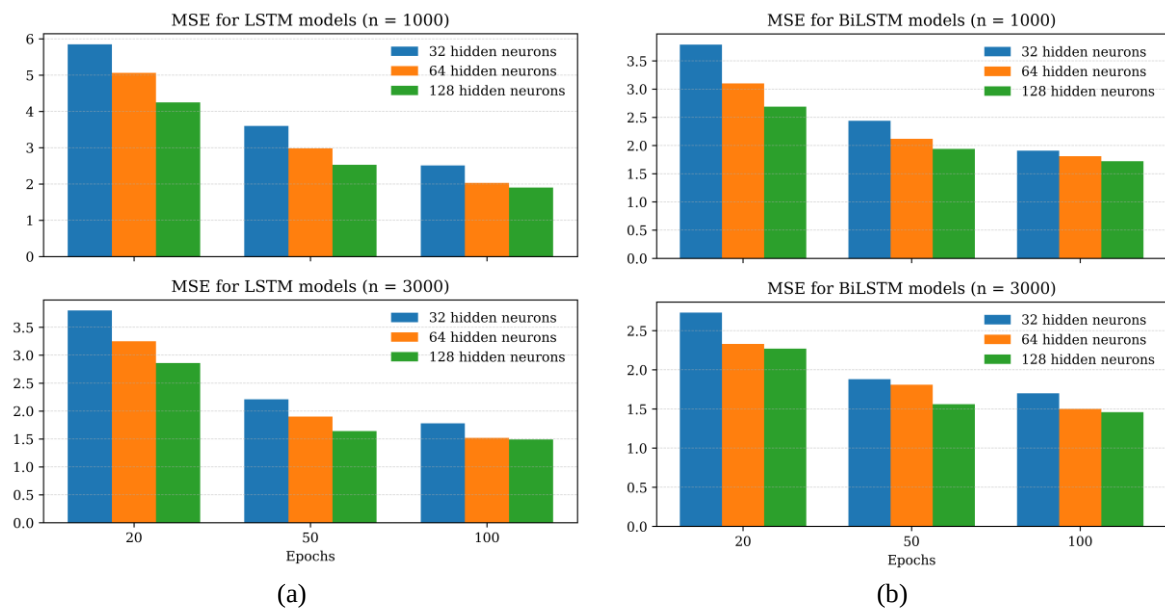


Figure 3. Effect of hidden neurons and epochs on (a) LSTM and (b) BiLSTM performance for TAR-simulated data

Table 3. MSE and MAE of LSTM and BiLSTM for SETAR simulated data

Model	Sample size	Hidden neurons	MSE			MAE		
			Epochs	20	50	Epochs	20	50
LSTM	1000	32	2.0317	1.3554	1.0693	1.1279	0.9233	0.8256
		64	1.8241	1.2050	1.0498	1.0693	0.8727	0.8231
		128	1.6331	1.1369	1.0492	1.0150	0.8504	0.8227
	3000	32	1.2897	1.0685	1.0505	0.9084	0.8287	0.8224
		64	1.1526	1.0672	1.0496	0.8596	0.8285	0.8222
		128	1.1162	1.0659	1.0485	0.8459	0.8283	0.8218
BiLSTM	1000	32	1.4529	1.1702	1.0640	0.9581	0.8609	0.8233
		64	1.3294	1.1307	1.0485	0.9177	0.8486	0.8188
		128	1.2665	1.1076	1.0480	0.8973	0.8403	0.8186
	3000	32	1.1688	1.0648	1.0503	0.8653	0.8277	0.8223
		64	1.1348	1.0646	1.0474	0.8527	0.8274	0.8204
		128	1.1037	1.0622	1.0441	0.8416	0.8264	0.8199

A similar pattern is observed for the SETAR simulations (Table 3). For both sample sizes, the LSTM model with 128 hidden neurons and 100 epochs again gives the lowest errors, with MSE = 1.0492 and MAE = 0.8227 for $n = 1000$, and MSE = 1.0485 and MAE = 0.8218 for $n = 3000$. The corresponding BiLSTM configuration yields slightly smaller errors, with MSE = 1.0480 and MAE = 0.8186 for $n = 1000$, and MSE = 1.0441 and MAE = 0.8199 for $n = 3000$ as shown in Figure 4. Figures 4(a) and 4(b) confirm that, across both TAR and SETAR models, MSE and MAE decrease as the number of hidden neurons increases from 32 to 64 to 128 and as the number of epochs increases from 20 to 50 to 100. In addition, the larger sample size ($n = 3000$) consistently yields lower errors than $n = 1000$, and for $n = 3000$ the performance differences between neighbouring hyperparameter settings become relatively small for both LSTM and BiLSTM.

Across both TAR and SETAR, increasing the sample size from $n = 1000$ to $n = 3000$ reduces MSE and MAE for both architectures, consistent with improved parameter estimation and reduced generalisation error under larger training sets. However, the extent of improvement depends on the data-generating process. For TAR simulation, the benefits strongly from increased capacity and training more neurons and epochs, and shows a clearer advantage for BiLSTM especially at smaller sample size. In the

SETAR simulation, the results tend to converge more quickly and form a more stable pattern with few changes. This means that once the number of epochs and neurons reaches moderate-to-high levels, differences between architectures become less significant, indicating a smoother, more efficient process. To further support this, a Wilcoxon signed-rank test was conducted to verify that the BiLSTM model outperformed the LSTM model in forecasting the time series data. The results, presented in Table 4, indicate that BiLSTM significantly outperforms LSTM for both TAR and SETAR simulated datasets at sample sizes of 1000 and 3000. All p -values are below 0.05, confirming that the differences in forecasting errors between LSTM and BiLSTM are statistically significant.

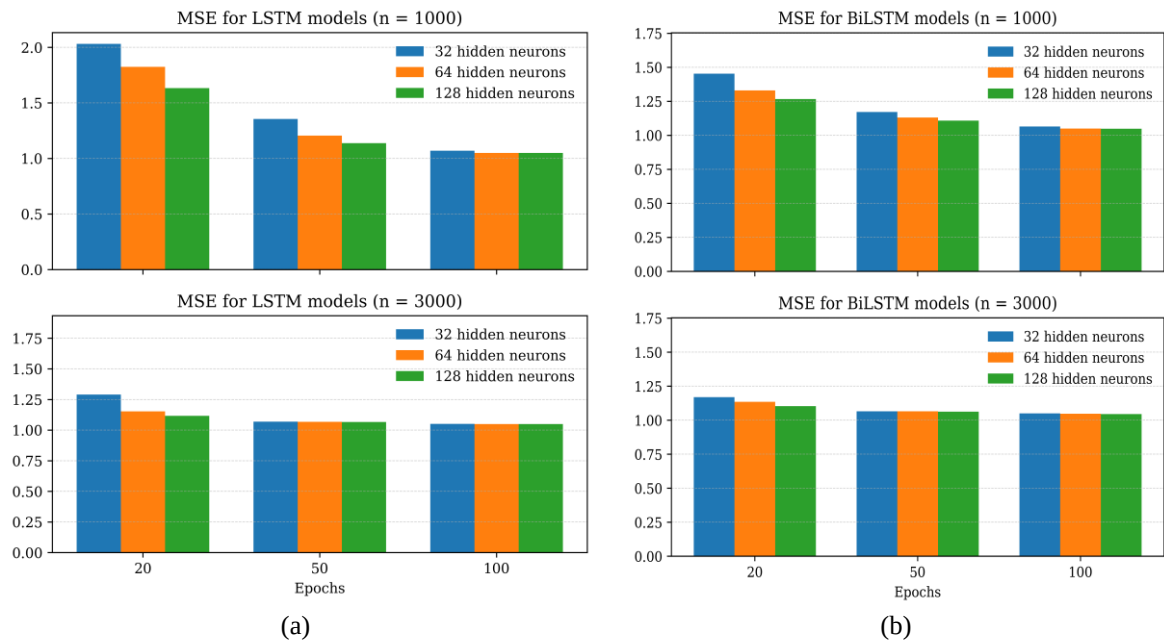


Figure 4. Effect of hidden neurons and epochs on (a) LSTM and (b) BiLSTM performance for SETAR-simulated data

Table 4. Wilcoxon signed-rank test results between LSTM and BiLSTM

Sample size	Model	Measurement used	p -value
1000	TAR	MSE	0.0020
		MAE	0.0039
3000	TAR	MSE	0.0020
		MAE	0.0020
1000	SETAR	MSE	0.0020
		MAE	0.0020
3000	SETAR	MSE	0.0020
		MAE	0.0020

From a tuning perspective, two practical patterns emerge where epochs from 20 to 50, and 50 to 100 generally reduce error, but gains from 50 to 100 epochs are smaller than from 20 to 50 in several settings, suggesting potential overtraining risk or convergence saturation. On the other hand, hidden neurons from 32 to 64 and 64 to 128 improve performance consistently, with the largest gains often occurring when moving from 32 to 64, and smaller incremental gains from 64 to 128, particularly under $n = 3000$. Thus, the results indicate that BiLSTM is most beneficial when the data are limited and the nonlinear regime-switching behaviour is pronounced TAR, while standard LSTM can be nearly as effective for SETAR-like series, especially at larger sample sizes where error plateaus are observed. This finding is consistent with previous studies reporting that BiLSTM can outperform standard LSTM in time series forecasting because it learns temporal information from both forward and backward directions [4], [6], [25]. However, the present results further show that this advantage is condition-dependent, where BiLSTM provides clearer improvement for TAR simulated data but only marginal improvement for smoother SETAR simulated data, especially when the sample size is large [26].

4. CONCLUSION

This research conducted experimental studies on the application of machine learning to time series data. Nonlinear time series data were generated from the TAR and SETAR models to capture the nonlinearity in stock prices. Simulated non-linear time series data from each model consisted of two sample sizes, 1000 and 3000. The experiment's results were reported, and the accuracy and performance of unidirectional LSTMs and BiLSTMs were compared and analysed. Grid search was used to tune the hyperparameters of LSTM and BiLSTM models to optimise the neural network's performance. The main hyperparameter tuned in the experiment is the number of hidden neurons, with values 32, 64, and 128. The results of the experiment showed that the MSE and MAE values decrease as the number of hidden neurons increases. The optimal number of hidden neurons is 128, as it provides the highest prediction accuracy. Hidden neurons are artificial neurons that receive weighted inputs and generate an output via an activation function. Although a small number of hidden neurons may improve generalisation, it may cause the neural network to lose the capacity to solve the problem. Hence, it is essential to determine the optimal number of hidden neurons by analysing different datasets, as the network architecture, the degree of noise, and the complexity of the function should be considered. In this study, results showed that a higher number of hidden neurons (128) is preferred for forecasting stock price time series data, as it achieves the highest prediction accuracy.

The results showed that sample size affected the performance of LSTM and BiLSTM models. Both models performed better at forecasting simulated data with a 3000-sample size than with a 1000-sample size. BiLSTM models are neural networks that are well known for their ability to learn long-term dependencies in data. Additionally, the comparison between the LSTM and BiLSTM models showed that the BiLSTM model achieved higher prediction accuracy and lower error values overall. This result indicates that learning temporal information from both forward and backward directions positively affects the precision of time series forecasting. Even though BiLSTM provides more accurate results, it requires more computational power. Therefore, in situations where resources are limited or time is tight, a well-suited unidirectional LSTM could be a better, more practical choice, balancing performance and efficiency.

However, since this study is based on synthetic TAR and SETAR data with selected hyperparameter settings, the findings should be interpreted as evidence from a controlled simulation setting rather than as a universal conclusion for all nonlinear forecasting problems. Future studies may extend this work by using real-world nonlinear time series data, broader hyperparameter spaces, multi-step forecasting horizons, computational time analysis, and hybrid modelling approaches to provide a more complete evaluation of forecasting accuracy and model complexity.

ACKNOWLEDGMENTS

We are grateful to the anonymous reviewers for their valuable feedback. We appreciate their input and support.

FUNDING INFORMATION

The authors would like to thank the Ministry of Higher Education for providing financial support under the Fundamental Research Grants Scheme - Early Career Research (FRGS-EC) No. FRGS-EC/1/2024/STG06/UMPSA/02/1 (University reference RDU243711).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nur Haizum Abd Rahman	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Hani Syahida Zulkifli		✓			✓	✓		✓	✓	✓	✓			

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The data supporting the findings of this study were generated through simulation using the equations and settings described in the paper.

REFERENCES

- [1] F. Hamami and I. A. Dahlan, "Univariate time series data forecasting of air pollution using LSTM neural network," in *International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, Oct. 2020, pp. 55–59, doi: 10.1109/ICADEIS49811.2020.9277393.
- [2] A. I. Sunny, M. M. S. Maswood, and A. Alharbi, "Deep learning-based stock price prediction using LSTM and bi-directional LSTM model," in *International Conference on Novel Intelligent and Leading Emerging Sciences (NILES)*, Oct. 2020, pp. 87–92, doi: 10.1109/NILES50944.2020.9257950.
- [3] I. A. Ibrahim, "Improved grid search algorithm for optimal LSTM performance: a case study on australian electricity price forecasting," in *2024 IEEE Power Energy Society General Meeting (PESGM)*, Jul. 2024, pp. 1–5, doi: 10.1109/PESGM51994.2024.10688849.
- [4] R. L. Abduljabbar, H. Dia, and P.-W. Tsai, "Unidirectional and bidirectional LSTM models for short-term traffic prediction," *Journal of Advanced Transportation*, vol. 2021, 2021, pp. 1–16, doi: 10.1155/2021/5589075.
- [5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [6] S. Siarni-Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 2019, pp. 3285–3292, doi: 10.1109/BigData47090.2019.9005997.
- [7] P. N. Hai *et al.*, "An empirical research on the effectiveness of different LSTM architectures on vietnamese stock market," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, 2020, pp. 144–149, doi: 10.1145/3437802.3437827.
- [8] M. Pirani, P. Thakkar, P. Jivani, M. H. Bohara, and D. Garg, "A comparative analysis of ARIMA, GRU, LSTM and BiLSTM on financial time series forecasting," in *Proceedings of the IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Apr. 2022, pp. 1–6, doi: 10.1109/ICDCECE53908.2022.9793213.
- [9] A. Bhambu, "Stock market prediction using deep learning techniques for short and long horizon," in *Soft Computing for Problem Solving*, M. Thakur, S. Agnihotri, B. S. Rajpurohit, M. Pant, K. Deep, and A. K. Nagar, Eds. Singapore: Springer, 2023, pp. 121–135, doi: 10.1007/978-981-19-6525-8_11.
- [10] N. H. A. Rahman, Q. P. Yin, and H. S. Zulkafli, "Optimising LSTM and BiLSTM models for time series forecasting through hyperparameter tuning," *Journal of Quality Measurement and Analysis*, vol. 21, no. 3, pp. 191–205, 2025, doi: 10.17576/jqma.2103.2025.12.
- [11] M. Jain, T. Alskaf, and S. Dev, "Are deep learning models more effective against traditional models for load demand forecasting?," in *2022 International Conference on Smart Energy Systems and Technologies (SEST)*, Eindhoven, Netherlands, 2022, pp. 1–6, doi: 10.1109/SEST53650.2022.9898424.
- [12] S. M. Al-Selwi *et al.*, "RNN-LSTM: from applications to modeling techniques and beyond—systematic review," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 5, 2024, doi: 10.1016/j.jksuci.2024.102068.
- [13] M. Rhif, A. B. Abbas, B. Martinez, and I. R. Farah, "Deep learning models performance for NDVI time series prediction: a case study on North West Tunisia," in *2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, 2020, pp. 9–12, doi: 10.1109/M2GARSS47143.2020.9105149.
- [14] F. R. Alharbi and D. Csala, "Short-term solar irradiance forecasting model based on bidirectional long short-term memory deep learning," in *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 2021, pp. 1–6, doi: 10.1109/ICECCE52056.2021.9514233.
- [15] C. Chen, Y. Li, K. Zhou, X. Chen, and B. Tan, "Air quality prediction model utilizing Bi-LSTM-MA," *Journal of Safety and Environment*, vol. 25, no. 8, pp. 3240–3251, 2025, doi: 10.13637/j.issn.1009-6094.2024.2216.
- [16] S. Osman, N. S. Gheriyani, S. Salem, A. Keshlaf, and J. Ibrahim, "Dual-LSTM model for real-time traffic prediction and anomaly detection in IoT environment," in *Studies in Computational Intelligence*, A. O. Albaji, Ed., Springer Science and Business Media Deutschland GmbH, 2026, pp. 751–766, doi: 10.1007/978-3-032-00232-7_49.
- [17] M. T. Pawitra, H. Fakhruroja, and L. Abdurrahman, "Predicting stock market using CNN and BiLSTM model," in *International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, 2024, pp. 267–272, doi: 10.1109/IC3INA64086.2024.10732245.
- [18] N. I. A. Kader, U. K. Yusof, M. N. A. Khalid, and N. R. N. Husain, "A review of long short-term memory approach for time series analysis and forecasting," in *Lecture Notes in Networks and Systems*, M. A. Al-Sharafi, M. Al-Emran, K. Shaalan, and M. N. Al-Kabi, Eds., Springer Science and Business Media Deutschland GmbH, 2023, pp. 12–21, doi: 10.1007/978-3-031-20429-6_2.
- [19] N. Q. Viet, N. N. Quang, N. King, D. T. Huu, N. D. Toan, and D. N. H. Thanh, "An exploratory comparison of LSTM and BiLSTM in stock price prediction," in *Inventive Communication and Computational Technologies*, 2023, pp. 513–524, doi: 10.1007/978-981-99-5166-6_35.
- [20] A. E. Ouardi, B. Er-Raha, and K. Tatane, "LSTM adaptive hyperparameter tuning for financial time series forecasting using custom gradient-based methods: a comparative study with Bayesian optimization," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 24, pp. 8913–8936, 2024.
- [21] M. Krichen and A. Mihoub, "Long short-term memory networks: a comprehensive survey," *AI*, vol. 6, p. 215, Sep. 2025, doi: 10.3390/ai6090215.
- [22] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Computers in Biology and Medicine*, vol. 96, pp. 189–202, 2018, doi: 10.1016/j.combiomed.2018.03.016.

- [23] E. Firat, "SETAR (Self-exciting threshold autoregressive) non-linear currency modelling in EUR/USD, EUR/TRY and USD/TRY parities," *Mathematics and Statistics*, vol. 5, pp. 33–55, Jan. 2017, doi: 10.13189/ms.2017.050105.
- [24] A. Loussaief, R. Ktari, and Y. Hadj Kacem, "HGS-At-LSTM: attention-based long short-term memory model combined with halving grid search optimizer for harmful algal bloom forecasting," *International Journal of Data Science and Analytics*, vol. 20, no. 6, pp. 5457–5483, 2025, doi: 10.1007/s41060-025-00783-2.
- [25] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," *Transportation Research Part C: Emerging Technologies*, vol. 118, p. 102674, 2020, doi: 10.1016/j.trc.2020.102674.
- [26] A. Thomas, M. Petridis, S. Walters, S. M. Gheytaasi, and R. Morgan, "On predicting the optimal number of hidden nodes," in *International Conference on Computational Science and Computational Intelligence (CSCI)*, 2015, pp. 565–570, doi: 10.1109/CSCI.2015.33.

BIOGRAPHIES OF AUTHORS



Nur Haizum Abd Rahman    received her first degree and Ph.D. degree from Universiti Teknologi Malaysia, in 2010 and 2017, respectively. She is currently a senior lecturer at Universiti Malaysia Pahang Al-Sultan Abdullah, with her main research interests focusing on statistics, particularly in time series analysis. Previously, she worked as a senior lecturer at Universiti Putra Malaysia. She has a total of nine years' teaching experience. Her research publications are regularly published in leading journals, and she is actively involved in the academic community, attending conferences, symposiums, and seminars. Beyond her university academic responsibilities, she holds a lifelong membership in Institut Statistik Malaysia (ISMy) and currently serves as a council member, fulfilling the roles of Honourary Secretary from 2022 to 2024 and from 2024 to 2026. She can be contacted at email: haizum@umpsa.edu.my.



Hani Syahida Zulkafli    received her Master in Applied Statistics and B.Sc (Hons.) in Statistics from Universiti Putra Malaysia in year 2011 and 2009, respectively. Her Ph.D. degree from Heriot-Watt University. She has been working as a Senior Lecturer in the Department of Mathematics and Statistics at Universiti Putra Malaysia since 2017. She has 10 years of teaching and research experience at various reputed institutions. Her research interests include applications of Bayesian networks, Bayesian modeling, and inference in chronic diseases, such as diabetes. She can be contacted at email: hsyahida@upm.edu.my.