Climate Forecasting by Bidirectional Recurrent Neural Networks

Sepehr Abbaspour* Hedieh Sajedi†

Abstract

Regression problems have been extensively studied and addressed using a variety of algorithms and models, including both statistical and machine learning-based approaches. One notable application of regression tasks is in predicting weather conditions, which has significant implications for various sectors. To enhance the accuracy of real-valued predictions in time series or sequential data, memory-based models are particularly effective. Among these, Bidirectional Recurrent Neural Networks stand out because they learn from both past and future data points. Bidirectional learning approach allows for more precise parameter tuning and improved predictive performance. This study focuses on three specific types of Bidirectional Recurrent Neural Networks: Bidirectional Recurrent Neural Network, Bidirectional Long Short-Term Memory, and Bidirectional Gated Recurrent Unit. The primary objective is to investigate and compare their performance in regression tasks. Through a comprehensive analysis, the models are trained on a relevant dataset and evaluated based on their ability to adapt and fit the data and predict unseen values. The findings of this study provide worthwhile intelligence into the efficiency of each model, regarding the advancement of memory-based approaches in regression tasks. The Bidirectional Gated Recurrent Unit model has demonstrated desirable performance, achieving a high R^2 score of 0.93233. This indicates a highly acceptable level of modeling accuracy.

Keywords: Bidirectional RNN, Climate Forecast, and Regression

1 Introduction

Have you ever planned a camping trip only to be disappointed by unsuitable weather conditions? This is one of the many reasons why regression analysis is so valuable. Climate forecasting has a rich history, dating back to ancient times when people used wind patterns and clocks to predict weather. The first known instance of this was at ”The Tower of the Winds”, a historical landmark in Athens, Greece. Today, meteorologists use time series data to make predictions. But what if a system could make these predictions autonomously, just like a human? Such a system would need to learn from past events and identify influential features, much like humans do.

Climate forecasting lacks precise mathematical formulas, but it involves patterns that can be analyzed using Machine Learning (ML). With sufficient training data, deep learning techniques excel at making predictions by understanding complex nonlinear relationships between input attributes and their corresponding outputs. Due to their impressive predictive accuracy and versatility, ML techniques are increasingly used for climate forecasting and other time series tasks. Numerous studies have evaluated the efficiency of different models in forecasting climate and predicting related dataset features.

The goal of this study is to comprehensively evaluate and compare three Bidirectional Recurrent Neural Networks (BiRNNs): Bidirectional Recurrent Neural Network (BiRNN) [1], Bidirectional Long Short-Term Memory (BiLSTM) [2], and Bidirectional Gated Recurrent Unit (BiGRU) [3]. This comparative analysis involves training and evaluating these models to determine their effectiveness in fitting data and predicting unseen values. By critically assessing each BiRNN, this research aims to identify the most robust and reliable model for climate forecasting, contributing to the ongoing discourse on machine learning applications in this field.

In this study, we delve into the intricacies of each model, examining their architectures, training processes, and performance metrics. We also explore the theoretical underpinnings of BiRNNs and their advantages over traditional unidirectional models. By leveraging bidirectional learning, these models can capture dependencies in both forward and backward directions to conclude more insightful predictions. Furthermore, we discuss the practical implications of our findings, highlighting how these models can be applied to real-world climate forecasting scenarios and other time series prediction tasks.

Ultimately, this research aims to advance the understanding of BiRNNs and their potential to revolutionize climate forecasting. By providing a thorough comparative analysis, we hope to offer valuable insights that

*School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran. Tehran, Iran, sepehr.abbaspour@ut.ac.ir

†School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran. Tehran, Iran, hhsajedi@ut.ac.ir

can guide future research and development in this area, detecting ways for more accurate predictive models.

2 Related work

Climate forecasting can be viewed as both sequence regression and sequence classification problem. Outcomes in study [4] concluded that ML models could have an encouraging future in arithmetic weather forecasting. Good result from Random Forest-based regressor trained on ERA5 Monthly Aggregates was achieved in study [5]. Study [6] suggests a big view into the changing role of ML in climatological prediction. It exclusively merges short-run weather prediction with middling and long-run climate forecastings, covering twenty models and producing a preface of eight select models that abide in the vanguard of the production. In study [7], the mean temperature was predicted using the Prophet model, which yielded sufficiently accurate results. Another study in [8] examined the performance of a applied model named Wavelet Decomposition-Seasonal Auto-Regressive Integrated Moving Average with Exogenous Variables (WD-SARIMAX) and several ML regression models. The tentative outcomes concluded that the WD-SARIMAX provided the best results, while the DR model yielded the worst performance.

As this study aims to utilize recurrent neural networks, it is essential to consider previous research comparing these models. One article concluded that the BiLSTM model outperformed traditional LSTM and GRU models in predicting cryptocurrency prices. The study highlighted that BiLSTM's ability to catch long-term relations and bidirectional context contributed to its superior performance [9]. Additionally, a study on traffic stream prediction reported that the BiGRU model showed finer prediction outcomes during peak times compared to low apex times. The study also noted that the suggested model had a definite lag, and recommended using the GRU model for predicting traffic stream in big route lattice outlines, balancing computational proficiency and prediction reliability [10].

Recurrent neural networks are utilized for other time series tasks. One is to do predictions on Satellite Image Time Series (SITS) which is a series of satellite images that write down a given zone at a few successive times. In study [11], SITS has been modeled by numerous Deep Learning (DL) methods, with 1D Convolutional Neural Networks (1D-CNNs) and RNNs wielded to model time-related data, and more complicated methods such as RNNs, hybrid CNN and 3D-CNNs wielded for spatial temporal modelling. Additionally, DL methods utilized for SITS forecasting are put into three principal batches, namely feed-forward-based models, hybrid models and RNN-based models in study [12]. Study [13] offered a novel BiGRU named KT-Bi-GRU student efficiency forecasting task. This model presents a adjusted construction with two subnetwork parts. The first subnetwork is for approximating the student lore state based on her/his link history utilizing a RNN and the second forecasts the student efficiency utilizing Multi-Layer Perceptron (MLP).

Study [14] states that time series predictions can also be done by Genetic Algorithm (GA), Support Vector Machine (SVM) as a ML method. SVM is also applied in monetary time series prediction. According to study [15], SVM provided a more reliable and more effectual prediction method for such economic data than the Box and Jenkins' Autoregressive Integrated Moving Averages (ARIMA) [16] and Artificial Neural Network (ANN) models did. Additionally, study [17] concluded that SVM has smaller number of free parameters, does prediction better in quality and has faster training process in comparison with a MLP trained by the Back-Propagation (BP) algorithm. Cuffless blood pressure estimation, a physiological time series, has been modeled using Physics-Informed Neural Network (PINN) [18] in study [19].

This study aims to contribute to the ongoing advancements in ML and its applications in climate forecasting as a time series prediction tasks by leveraging the strengths of BiRNN models.

3 Dataset

The intended dataset spans from January 1, 2013, to April 24, 2017, and pertains to the city of Delhi, India. It comprises 1,463 training samples and 115 test samples. Each record includes five features: date, mean temperature, humidity, wind speed, and mean pressure [20]. The 'date' feature is excluded from the dataset as it serves as a unique primary key for each record and is not useful for modeling purposes.

A correlation analysis of the data (illustrated in Figure 1) reveals that the 'mean pressure' feature has the least correlation with the target variable, 'mean temperature,' indicating its minimal impact on the modeling process. Additionally, the 'wind speed' feature contains

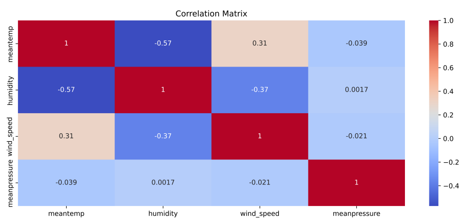

Figure 1: Feature Correlation Matrix

some outliers (illustrated in Figure 2), which are addressed using the Z-Score method. This method helps in identifying and mitigating the influence of extreme values (illustrated in Figure 3), ensuring that the data is more representative of typical conditions.

Following the preprocessing phase, the data is normalized to ensure consistency and improve model performance. Normalization scales the features to a standard range, typically between 0 and 1, which helps in accelerating the convergence of gradient-based learning algorithms and enhances the overall stability of the model training process. The dataset provides a comprehensive overview of the climatic conditions in Delhi over the specified period, capturing variations in temperature, humidity, wind speed, and pressure. This rich dataset serves as a robust foundation for training and evaluating the BiRNNs explored in this study. By leveraging this data, the study aims to develop accurate and reliable models for climate forecasting, contributing to the broader field of ML applications in meteorology.

4 Methods and Models

A BiRNN has of two easily distinguishable recurrent hidden layers. One layer processes the sequence-type input in the forward side, while the other processes the input in the backward side.

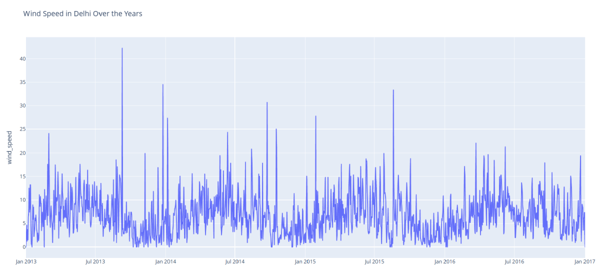

Figure 2: 'wind speed' in Delhi over the years including outliers

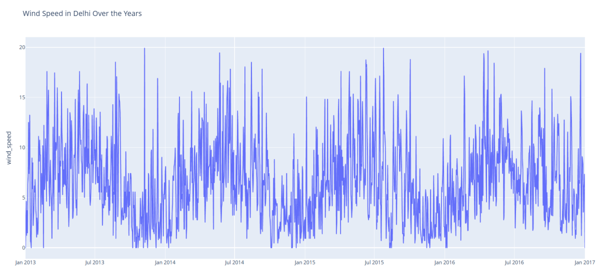

Figure 3: 'wind speed' in Delhi over the years (outliers are dealt with)

The results from the processes are then combined and fed into a layer for final forecasting. In the forward side, the BiRNN functions similarly to conventional recurrent neural networks, modernizing the hidden state based on the present input and the previous hidden state step by step. Conversely, the backward hidden layer processes the input in reverse, modernizing the hidden state based on the present input and the hidden state of the next step. This bidirectional processing improves the reliability of the BiRNN by considering both directions. The two hidden layers complement each other, providing the final forecasting layer with more comprehensive data, which also serves as a form of model regularization. During training, gradients are calculated for forward and backward passes using the BackPropagation Through Time (BPTT) technique. The BiRNN architecture is depicted in Figure 4.

BiLSTM is a recurrent neural network which is applied in Natural Langauge Processing (NLP). Like BiRNN, it processes input in both directions, utilizing information from both sides. BiLSTM has an extra LSTM layer that drives backward the direction of knowledge stream, meaning the sequence-type input passes backward in this layer. The outputs from both LSTM layers are then combined using methods such as averaging, summation, multiplication, or concatenation. This architecture has significant advantages in addressing real-world problems, as each part of an input contains knowledge from the past and current. Consequently, BiLSTM provide more significant outputs by merging LSTM layers from both sides. It is worth noting that BiLSTM requires more time for training and hence it's a slower model. The BiLSTM architecture is depicted in Figure 5.

BiGRU is a bidirectional recurrent neural network that has just forget and input gates. It also has two GRUs: one processing the input in a forward side (orig-

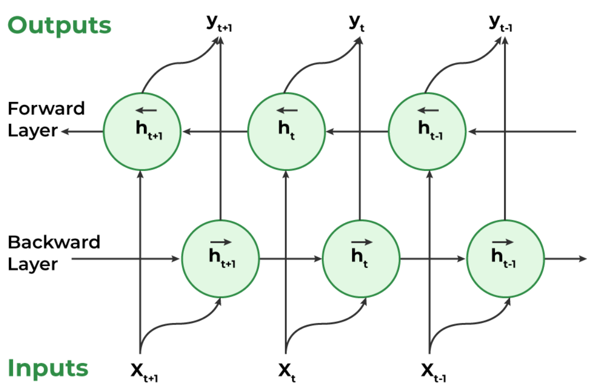

Figure 4: An overview of BiRNN architecture [21]

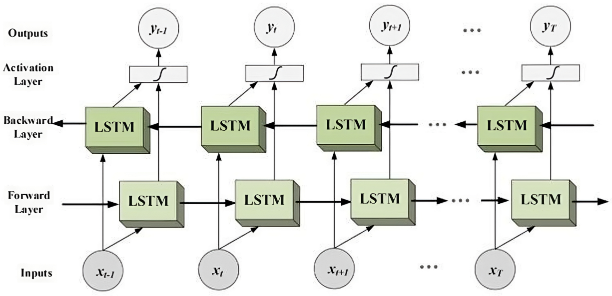

Figure 5: An overview of BiLSTM architecture [22]

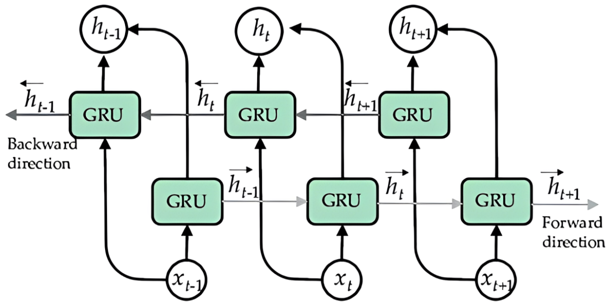

Figure 6: An overview of BiGRU architecture [23]

inal order) and the other in a backward side (the reverse order). This bidirectional approach allows BiGRU to efficiently catch underlying knowledge from both foregoing and succeeding inputs. The BiGRU architecture is depicted in Figure 6.

In summary, bidirectional recurrent neural networks are exceedingly effectual for time series and sequential data because they can catch serial correlations through memory mechanisms.

5 Results

To evaluate the effectiveness and accuracy of the computations described in this work, various performance indicators are utilized. We begin by explaining and elaborating on these assessment criteria to provide a comprehensive overview.

1. MAE (Mean Absolute Error): MAE estimates the mean enormity of errors in a set of forecastings, without pondering their side. It is computed as the mean of the absolute disparities between real and estimated values:

$$\text{MAE} = \frac{1}{n}\sum_{k=1}^{n}|z_k - \hat{z}_k| \tag{1}$$

while n is the number of observations, $\hat{z}_k$ is the estimated value and z_k is the real value. It supplies a straightforward explanation of the mean error enormity, making it a helpful metric for realizing the all-embracing reliability of the model's predictions.

2. MSE (Mean Squared Error): MSE estimates the mean of the squares of the errors. It gives more attention to bigger errors, making it practical for recognizing important errors in predictions. The formula is:

$$\text{MSE} = \frac{1}{n}\sum_{k=1}^{n}(z_i - \hat{z}_k)^2 \tag{2}$$

while n is the number of observations, $\hat{z}_k$ is the estimated value and z_k is the real value. It is especially sensitive to outliers, as larger errors are squared, thus having a greater impact on the overall error metric. This makes MSE a valuable tool for detecting and addressing significant deviations in the model's predictions.

3. R² Score (R-squared Score): R² which is acknowledged as the coefficient of determination, implies the part of the variance in the dependent variable that is foreseeable from the self-reliant variables. It ranges from 0 to 1, where 0 concludes the model interprets none of the variability and 1 concludes it interprets all the variability. A negative value indicates that using the average of the data would be more reasonable than modeling for prediction. The formula is:

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(z_k - \hat{z}_k)^2}{\sum_{k=1}^{n}(z_k - \bar{z})^2} \tag{3}$$

while n is the number of observations, $\hat{z}_k$ is the estimated value, z_k is the real value and $\bar{z}$ is the mean of the real values. It supplies an estimation of how good the model's predictions match with the actual data, giving perspicacities into the model's explanatory power or overall performance.

To predict mean temperature values, we conducted a comparative analysis. The output layer was configured with a single node utilizing a linear activation function, appropriate for regression tasks. The study utilized key hyperparameters, including Number of Hidden Nodes (HN), Number of Hidden Layers (HL), Activation Function for Hidden Layer (AF), Loss Function (LF), Optimization Algorithm (OA), Batch Size (BS), Number of Epochs (NE) and Validation Split (VS) which is a percentage of training set and none of the study models requires particular hyperparameter. Each hyperparameter was assigned a specific value for all modelings, as detailed in Table 1.

An additional evaluation was performed using the same hyperparameters but with 150 training epochs instead of 75. This led to overfitting across all models. Consequently, we decided to proceed with 75 training epochs as a more effective strategy. We also examined the performance of unidirectional recurrent neural networks, which process sequential data in one direction, capturing dependencies based on past context, while bidirectional recurrent neural networks process the data in both directions, capturing dependencies from both past and future context and pull in future data to enhance the reliability of forecasting-in other words, bidirectional recurrent neural networks have outperformed their unidirectional counterparts, as mathematically demonstrated in [24]. The performance results of each model are presented in Table 2, offering a detailed comparison of metrics such as MAE, MSE and R^2 scores across training epochs. By probing the outcomes, we can recognize the frailties and powers of each model and their suitability for the regression task.

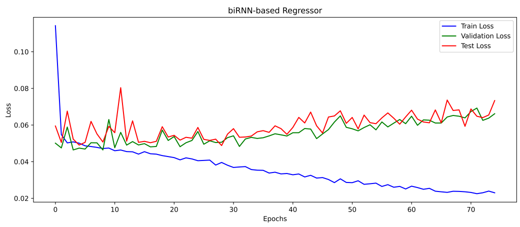

Figure 7: BiRNN-based regressor.

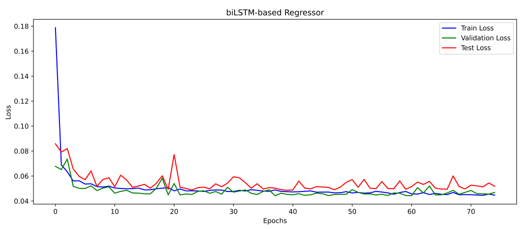

Figure 8: BiLSTM-based regressor.

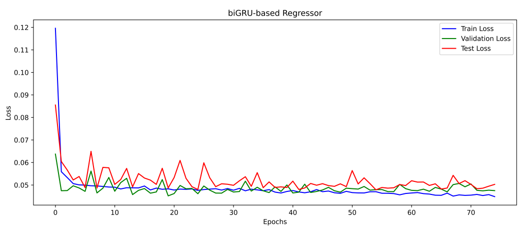

Figure 9: BiGRU-based regressor.

Table 2: Model performances after learning in 75 Epochs.

Model \ Metric	MAE	MSE	R^2 Score
GRU	0.17311	0.05035	0.93223
LSTM	0.17899	0.05209	0.92988
RNN	0.19497	0.06550	0.91185
BiRNN	0.20987	0.07334	0.90129
BiLSTM	0.17867	0.05180	0.93028
BiGRU	**0.17429**	**0.05028**	**0.93233**

6 Comparison

This paper seeks to identify and discuss some attributes of the models. Let's first examine their generalization capabilities.

RNN and BiRNN became overfitted and failed to generalize effectively, as evidenced by higher error values and lower R^2 score. In contrast, LSTM, GRU, BiLSTM and BiGRU maintained nearly the same performance quality, with acceptable high R^2 scores (greater than 0.92) and low error values for MAE and MSE (less than 0.18 and 0.06, respectively). RNN and BiRNN's tendency to overfit highlights the need for careful regularization and early stopping techniques to prevent performance degradation. On the other hand, other studied models' ability to catch long-term dependencies makes them particularly effective for tasks involving complex sequential data. By the way, study [8] reported that the WD-SARIMAX model gave the best outcomes in this Delhi climate temperature forecasting with R^2, MAE, and MSE are 0.91, 1.13, and 2.8, respectively, which is definitely worst in comparison with the performance results reported here.

According to the outcomes, bidirectional recurrent neural network did better performance in comparison with the correspondent unidirectional recurrent neural network, except that RNN did better than BiRNN probably because it has more complex architecture and becomes overfitted sooner than RNN did in the 75 epochs, or due to applying shuffling the training dataset for making batches in training phase. Nevertheless, the outcomes state that bidirectional recurrent neural networks does not enhance reliability and accuracy so much, probably because the number of train samples wasn't great enough to demonstrate the superiority of them as well.

Table 1: Hyperparameters Initialization

Hyperparameter	HL	HN	AF	LF	OA	BS	NE	VS
Assigned Value	1	50	ReLU	MSE	Adam	5	75	20%

7 Conclusion

In this comparative examination on bidirectional recurrent neural networks' performances, none of the studied models failed to model the data effectively, except for BiRNN, which experienced overfitting. However, increasing the number of training epochs can lead to more accurate models. Both BiLSTM and BiGRU achieved good results, with performance metrics that were nearly identical. According to Occam's razor, which favors simpler models when performance is comparable, we can conclude that the BiGRU model is the best choice due to its simpler architecture and efficient performance. The results underline the significance of working with the appropriate model and training duration to achieve optimal results. While BiRNN showed a tendency to overfit, BiLSTM and BiGRU demonstrated strong generalization capabilities. BiGRU's resilience to overfitting and its ability to maintain high performance with fewer parameters make it a particularly attractive option for practical applications.

The study doesn't underscore on hyperparameter tuning which might affect the results and conclusion of this study and change them. By leveraging the strengths of each model, researchers can seek more reliable predictive approaches, contributing to advancements in machine learning and its applications.

In summary, the BiGRU model stands out as the most effective and efficient choice for predicting mean temperature, offering a balance between simplicity and performance. These findings contribute to the broader understanding of bidirectional recurrent neural networks and their applications.

References

[1] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673–2681, 1997.

[2] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," IEEE Xplore, 2005.

[3] R. Zhang, T. Chen, F. Xiao, and J. Luo, "Bidirectional gated recurrent unit recurrent neural networks for failure prognosis of proton exchange membrane fuel cells," International Journal of Hydrogen Energy, vol. 47, no. 77, pp. 33027–33038, Sep. 2022.

[4] Z. Ben-Bouallegue et al., "The rise of data-driven weather forecasting," arXiv.org, Nov. 03, 2023. https://arxiv.org/abs/2307.10128 (accessed Nov. 16, 2023).

[5] Ahmed Almeslemani, Saad Aldoihi. A Machine Learning Approach to Climate Forecasting. GRI Enrichment Program, Aug 2023, Riyadh, Saudi Arabia. hal-04409406

[6] L. Chen, B. Han, X. Wang, J. Zhao, W. Yang, and Z. Yang, "Machine Learning Methods in Climate Prediction: A Survey," Sep. 2023.

[7] Aniruddha, A. Vats, and S. Srivastava, "Weather Forecasting using Python (Prophet model)," DTC Journal of Computational Intelligence, vol. 2, no. 2, p. 8, 2023.

[8] A. M. Elshewey et al., "A Novel WD-SARIMAX Model for Temperature Forecasting Using Daily Delhi Climate Dataset," Sustainability, vol. 15, no. 1, p. 757, Dec. 2022.

[9] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach," Fractal and Fractional, vol. 7, no. 2, p. 203, Feb. 2023.

[10] S. Wang, C. Shao, J. Zhang, Y. Zheng, and M. Meng, "Traffic flow prediction using bi-directional gated recurrent unit method," Urban Informatics, vol. 1, no. 1, Dec. 2022.

[11] L. Miller, C. Pelletier, and G. I. Webb, "Deep Learning for Satellite Image Time Series Analysis: A Review," arXiv (Cornell University), Apr. 2024.

[12] W. R. Moskolaï, W. Abdou, A. Dipanda, and Kolyang, "Application of Deep Learning Architectures for Satellite Image Time Series Prediction: A Review," Remote Sensing, vol. 13, no. 23, p. 4822, Nov. 2021.

[13] M. Delianidi and K. Diamantaras, "KT-Bi-GRU: Student Performance Prediction with a Bi-Directional Recurrent Knowledge Tracing Neural Network," Journal of Educational Data Mining, vol. 15, no. 2, pp. 1–21, Jun. 2023.

[14] W. Wang, "Time Series Prediction Based on Support Vector Machines Experts and Genetic Algorithm," Jan. 2007.

[15] M. K. Okasha, "Using Support Vector Machines in Financial Time Series Forecasting," International journal of statistics and applications, vol. 4, no. 1, pp. 28–39, Jan. 2014.

[16] G. E. P. Box and E. Al, Time series analysis : forecasting and control. Hoboken, New Jersey: John Wiley Sons, 2015.

[17] L. Cao and F. E. H. Tay, "Financial Forecasting Using Support Vector Machines," Neural Computing Applications, vol. 10, no. 2, pp. 184–192, May 2001.

[18] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, vol. 378, pp. 686–707, Feb. 2019.

[19] K. Sel, A. Mohammadi, R. I. Pettigrew, and R. Jafari, "Physics-informed neural networks for modeling physiological time series for cuffless blood pressure estimation," npj Digital Medicine, vol. 6, no. 1, pp. 1–15, Jun. 2023.

[20] Gamze Ç, "Weather Data Time Series Modelling Delhi-Kaggle Notebook," Medium, Sep. 03, 2023. https://medium.com/@mervegamzenar/weather-forecasting-delhi-kaggle-notebook-2240f49988ef (accessed Aug. 14, 2024).

[21] "Bidirectional Recurrent Neural Network," GeeksforGeeks, Jun. 08, 2023. https://www.geeksforgeeks.org/bidirectional-recurrent-neural-network (accessed Jul. 24, 2024).

[22] Anishnama, "Understanding Bidirectional LSTM for Sequential Data Processing," Medium, May 18, 2023. https://medium.com/@anishnama20/understanding-bidirectional-lstm-for-sequential-data-processing-b83d6283befc

[23] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "A Hybrid Deep Residual Network for Efficient Transitional Activity Recognition Based on Wearable Sensors," Applied Sciences, vol. 12, no. 10, p. 4988, May 2022.

[24] better, "DeepwizAI," DeepwizAI, Sep. 14, 2021. https://www.deepwizai.com/simply-deep/a-mathematical-proof-of-why-bidirectional-rnns-are-better-than-unidirectional-rnns (accessed Aug. 15, 2024).