

Applications of Long-Short-Term-Memory Architecture and Gated-Recurrent-Unit Model Based on Neural Network for Predictive Analytics

Suchada Sitjongsataporn* and Panavy Pookaiyaudom

Department of Electronic Engineering, School of Electrical and Electronic Engineering (SEE)
Mahanakorn University of Technology, Nongchok, Bangkok, Thailand

Email: ssuchada@mut.ac.th and panavy@mut.ac.th

Abstract

This paper presents the application of Long Short-Term Memory (LSTM) and Gated-Recurrent-Unit (GRU) models in order to investigate the effectiveness of LSTM-based, GRU architecture. Bi-LSTM model is an extended LSTM that processes the input data in the forward and backward directions for improving the prediction accuracy by utilizing the complete context of sequential data. Using a dataset enriched with temporal features like temperature, humidity, and wind speed, the model architecture and hyper-parameters were optimized for enhanced accuracy. The dataset comprises global time-series data, with preprocessing focused on extracting country-specific trends. Experiments demonstrate for analyzing and forecasting in weather patterns over a 30-day period. Experimental results show that the Bi-LSTM model outperforms traditional models, achieving high accuracy in temperature forecasting.

Keywords: Keras library, TensorFlow, LSTM, GRU, BiLSTM, predictive analytics.

1. Introduction

Based on the deep learning improvement, classification presents efficient ability in many field as text feature and feature extraction. With the multi-dimensional dataset, it can help to extract more features and suitable for the text defection.

Main classification models based on deep learning [1] that comprise with Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) [2]. LSTM networks is a modification of recurrent neural network (RNN) that is particularly suited for sequential data analysis due to their ability to handle long-term dependencies.

As stated in text extraction, the authors in [3] have used LSTM to extract the text information of fault defects in power system by considering only the information before the processed data. For example, the temperature forecasting plays an essential role in various industries, including agriculture, energy, and climate studies.

Traditional time-series prediction models as an unidirectional LSTM has been used for temperature prediction tasks. However, these models often struggle with capturing long-term dependencies in sequential data. In [4], LSTM-based model is designed to analyze and forecast time-series data, specifically focusing on

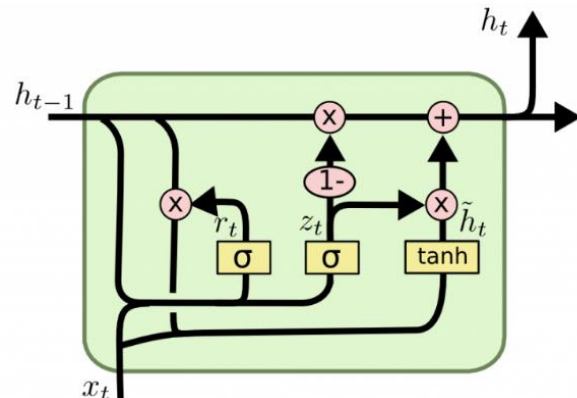


Fig. 1 GRU model [7]

COVID-19 confirmed case trends. It employs two advanced recurrent neural network architectures: LSTM and GRU. These architectures are particularly well-suited for time-series prediction in terms of long-term dependencies and temporal patterns in sequential data.

As stated about LSTM extended model, Bi-LSTM takes two parts into account with a bidirectional mechanism which processes the sequence both forwards and backwards. It can improve predictions by leveraging additional context. In [5], the authors have presented to extract deep model used for a multitask classifier based on a Bi-LSTM model to perform multitasking prediction based on the decision fusion of the deep representations extracted from time-series, demographics, and MRI dataset. This model has been optimized to predict the three tasks of Parkinson's disease. In [6], Bi-LSTM mechanism has been enhanced by learning from past and future data for increasing diagnostic accuracy in sequential clinical examinations in medicine.

As presented in the field of renewable energy, the GRU model in [7] has been introduced for fault level forecasting in grid system operators of renewable energy source. This GRU-based model is modified the relationship between system strength input level and forecasting output. Results show that this GRU-based model can show with higher accuracy in forecasting. In [8], the GRU model has been used to forecast the electrical load with less error than RNN and LSTM. In [9], GRU with the particle swarm optimization learning method has been presented to predict the short-term wind power by updating its parameters.

*Corresponding Author

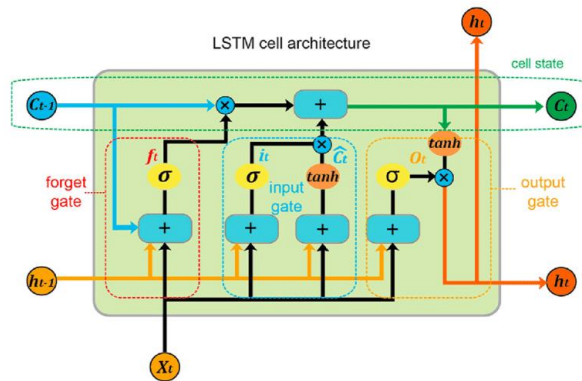


Fig. 2 LSTM cell architecture [10]

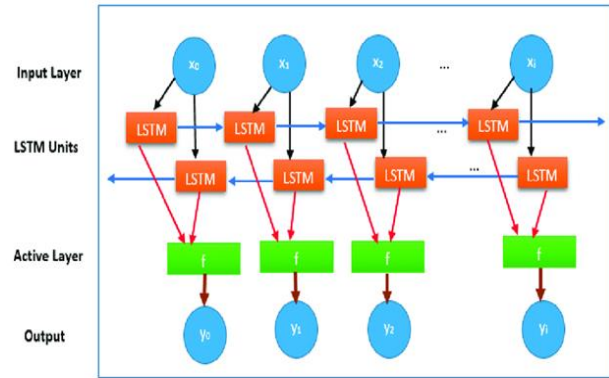


Fig. 4 Bi-LSTM architecture [13]

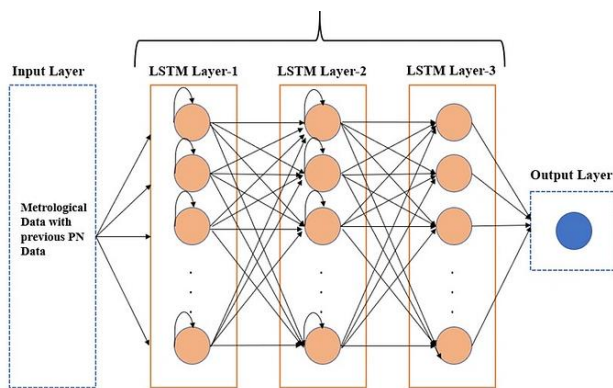


Fig. 3 LSTM structure [11]

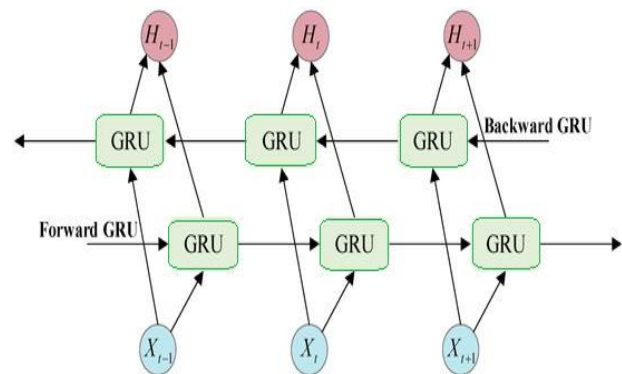


Fig. 5 Bi-GRU architecture [14]

In this paper, the applications of a Bi-LSTM network are explored for predicting mean temperature based on historical data. The objective is to focus on applications of LSTM-based architecture as LSTM, Bi-LSTM and GRU for predictive analytics. The goal is to leverage these models to predict future case trends using historical data, allowing for insights into pandemic dynamics. The workflow includes data preprocessing, visualization, and model training using the Keras library, with configurations optimized for accurate predictions.

This paper is organized as follows. First, the deep learning models with LSTM, Bi-LSTM are provided briefly in Section 2. Procedures of models are explained in Section 3. Case studies and simulation results of each case are presented in Section 4. Finally, Section 5 is conclusion of this work.

2. Models

2.1 GRU model

Following [7], the GRU cell consists of the update gate and reset gate. The data stream is processed using two individual gates as an update gate (u_t) and reset gate (r_t) as shown in Fig. 1, where σ denotes as sigmoid function and $\tanh$ is tan hyperbolic function, respectively. During this process, the hidden state (h_t) is updated using the update gate and reset gate considers the previous data.

2.2 LSTM model

Following [11], an LSTM cell consists of three main components as a cell state, hidden state, and gates. These LSTM gates are including with an input, forget, and output gates, which regulate the information flow in and out of the cell. A forget gate can decide what kind of information from the previous cell state to ignore. While, an input gate can determine which new information should be stored in the cell state. Finally, an output gate can control information is passed to next hidden state.

This architecture allows LSTMs to keep information along periods suitable for sequential data. Fig. 2 shows a core of LSTM that is the cell inside the blocks. This LSTM cell has both feedback connections and back route to the previous output [12].

LSTM structure is shown in Fig. 3 that consists of three types of layer as an input layer, an output layer and multiple LSTM hidden layers. Inside LSTM layer comprises with LSTM cell shown in Fig. 2. As compared with RNN architecture, the hidden layer of LSTM is a gated cell which consists of three layers that can interact with one another to produce the output along with the cell state. These two things are then passed onto the next hidden layer.

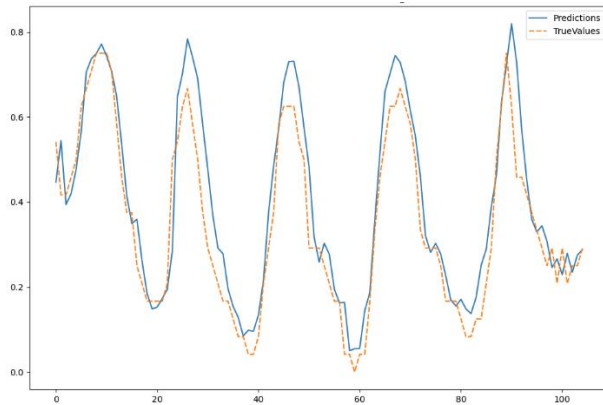


Fig. 6 Prediction with un-optimized LSTM model.

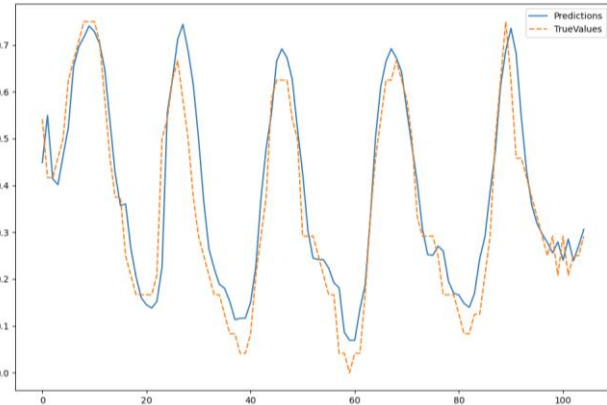


Fig. 7 Prediction with optimized LSTM model.

2.3 Bi-LSTM model

Following [15], Bi-LSTM network is a type of RNN that addresses the limitations of traditional neural networks. It is tailored to effectively control the gradient problem and capture long-term dependencies in sequences by tracking data flow from previous and current steps in the bidirectional of forward and backward signal. Fig. 4 shows the Bi-LSTM architecture within LSTM cell presented in Fig. 2.

2.4 Bi-GRU model

Bidirectional-GRU (Bi-GRU) network is an extended of GRU that consists of forward GRU layer and reverse GRU layer. It focuses to capture the dynamic features in time-series data in order to improve the feature extraction. Fig. 5 shows a diagram of Bi-GRU model within GRU cell shown in Fig. 1.

3. Procedure and Experimental Results

3.1 Procedure of LSTM model

Weather prediction is a vital application of machine learning, crucial for agriculture, disaster management, and daily planning. LSTM network is generally suited for sequential data analysis due to their ability to handle long-term dependencies.

The experiment focuses on applying LSTM to predict weather metrics over 30 days by historical data. Using a dataset enriched with temporal features like temperature, humidity, and wind speed, the model architecture and hyper-parameters were optimized for enhanced accuracy. The results demonstrate the utility of LSTM in capturing temporal dependencies critical for weather forecasting, albeit with challenges in long-term prediction reliability.

The LSTM model consisted of input layer with processing features like temperature, humidity, and wind speed. For hidden layers, three LSTM layers are each with 64 units followed by dropout for regularization. At output layer, a dense layer is providing predictions for temperature and other weather metrics. Hyper-parameters were optimized to balance accuracy and computational efficiency.

3.1.1 Procedure of LSTM architecture

- 1) Data preparation: historical weather data was normalized and split into training, validation, and testing sets (70:15:15 ratio). Temporal sequences were generated for LSTM training.
- 2) Model Training: The model was trained over 100 epochs with a batch size of 32, using the Adam optimizer and a mean squared error (MSE) loss function.
- 3) Metric Evaluation: Root mean squared error (RMSE), mean absolute error (MAE), and R-squared were used to assess model performance.

3.1.2 Experimental Results

This experiment is separated to test with two models as an un-optimized model compared with an optimized model. Fig. 6 shows results of prediction compared with true value with un-optimized LSTM model. It is achieved an RMSE of 4.2°C and MAE of 3.5°C for temperature prediction. Fig. 7 shows results of prediction compared with true value with optimized LSTM model. While, an optimized model can improve RMSE to 3.1°C and MAE to 2.7°C by adjusting the learning rate and model architecture.

For limitations of this experiment, while short-term forecasts between 1-7 days were highly accurate, errors increased for longer-term predictions.

3.1.3 Comparison

- 1) Alignment with True Values:

- a. In the un-optimized model shown in Fig. 6, there are noticeable deviations where the predictions lag or overshoot the true values, particularly around the peaks and troughs of the cycles.
- b. In the optimized model shown in Fig. 7, the predictions align more closely with the true values, showing a significant reduction in both the amplitude of errors and timing mismatches

- 2) Improved Consistency

Fig. 7 demonstrates smoother and more consistent trends that better capture the shape of the true values. This indicates that the optimized model can more accurately learn and predict temporal dependencies.

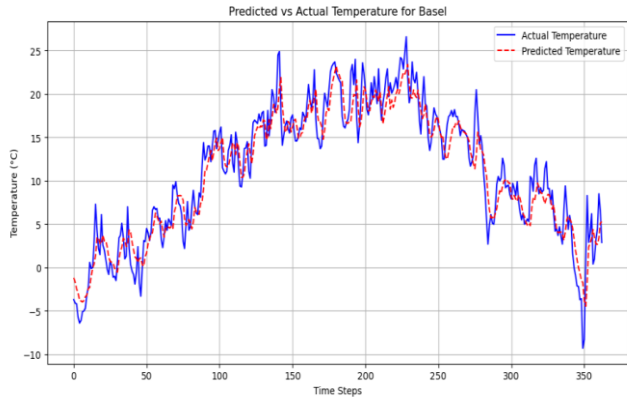


Fig. 8 Predicted compared with actual temperature by Bi-LSTM model

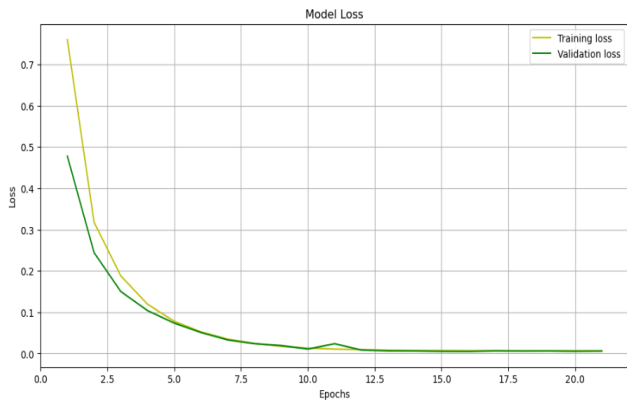


Fig. 9 Model loss by Bi-LSTM model

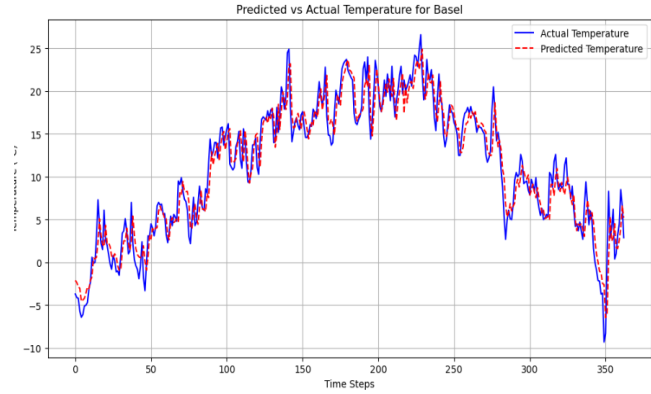


Fig. 10 Predicted compared with actual temperature by Bi-GRU model

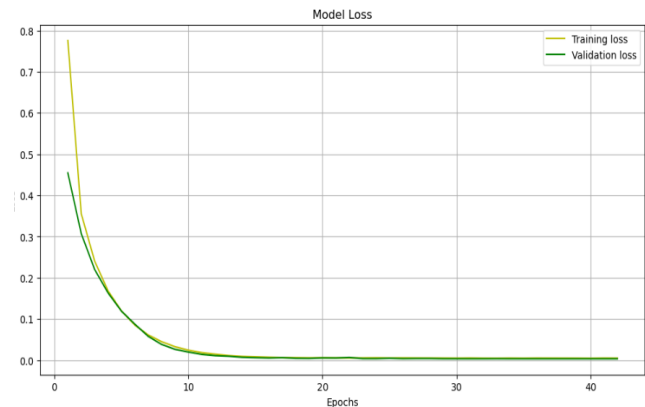


Fig. 11 Model loss by Bi-GRU model

3) Error Reduction

Comparing the two graphs visually, Fig. 7 shows fewer large gaps between predictions and true values, reflecting the improved RMSE from 4.2°C to 3.1°C and MAE from 3.5°C to 2.7°C reported in the experimental results.

3.1.4 Summary

In summary, the modified graph reflects a significant improvement in the model's accuracy and predictive reliability, making it the better representation of the optimized LSTM model's performance. This directly supports the claims made in the report about the benefits of hyper-parameter optimization.

3.2 Procedure of Bi-LSTM model

The Bi-LSTM architecture includes several key components that work together to improve performance in predicting sequential data. First, at Bi-LSTM layers, a Bi-LSTM processes the data in both directions of forward and backward to capture more comprehensive patterns in the time-series data. Then, a dropout layer with a rate of 0.2 is added to prevent overfitting by randomly dropping neurons during training.

A dense layer is used with 25 units and ReLU activation functions as the fully connected layer between the LSTM output and the final prediction. For batch normalization, this layer helps stabilize the learning process and reduce overfitting by normalizing activations.

At the final layer, the output layer uses a linear activation function to output the predicted temperature value.

3.2.1 Data pre-processing

- 1) Data Collection: Historical temperature data was collected for the training process.
- 2) Normalization: The data was normalized using MinMaxScaler to ensure consistent scaling and faster convergence during training.
- 3) Data Splitting: The data was divided into training (80%) and testing (20%) datasets.

3.2.2 Model Building

- 1) The model was constructed using TensorFlow and Keras with the following layers:
- 2) Bidirectional LSTM Layer: 128 units with L2 regularization.
- 3) Dropout Layer: 20% dropout to mitigate overfitting.
- 4) Dense Layer: A fully connected layer with 25 units and ReLU activation.
- 5) Batch Normalization Layer: To standardize outputs before passing them to the next layer.
- 6) Output Layer: A dense layer with a single unit and linear activation to predict the temperature value.

3.2.3 Training the Model

- 1) Optimization Algorithm: Adam optimizer was used with a learning rate of 0.001.
- 2) Epochs: 50 epochs were set for training, with early stopping used to prevent overfitting.
- 3) Loss Function: Mean Squared Error (MSE) was chosen as the loss function to optimize for the temperature prediction task.

3.2.4 Model Evaluation

The performance of model was evaluated using MAE, MSE, RMSE, and R-squared (R^2).

3.2.5 Experimental Results of Bi-LSTM model

- 1) Model training (1-50 epochs)

During training, the model exhibited a significant decrease in loss from epoch 1 to epoch 16, achieving a low validation loss by the end of training. Notably, early stopping was triggered after epoch 21 to prevent overfitting.

- 2) Performance Metrics:

MAE is of 1.8788, MSE is of 5.7632, RMSE is of 2.4007 and R^2 is of 0.9032.

Fig. 8 shows these results indicate the model's effectiveness in predicting temperature with high accuracy using Bi-LSTM. Fig. 9 shows the Bi-LSTM model loss of training and validation loss.

3.2.6 Experimental Results of Bi-GRU model

In this experiment, modifications were made to the original LSTM architecture by introducing a Bidirectional GRU (Bi-GRU) layer instead of an LSTM layer. The data input size was also adjusted to evaluate how changes in the data and model architecture affect performance.

The modified model architecture is as follows.

- 1) A Bidirectional GRU Layer with 128 units.
- 2) A Dense Layer with 25 units and ReLU activation.
- 3) A Dropout Layer to prevent overfitting.

Fig. 10 shows these results indicate the model's effectiveness in predicting temperature with high accuracy by Bi-GRU model. Fig. 11 shows a graphical representation of the training and validation losses across the epochs. It can be seen that both the training and validation losses decreased in parallel, with no significant divergence, suggesting good model convergence and the effectiveness of the regularization techniques used.

4. Discussions

The LSTM model successfully captured temporal trends, particularly for short-term forecasts. However, diminishing accuracy for extended predictions suggests room for improvement through techniques such as attention mechanisms or ensemble models.

5. Conclusion

This paper demonstrated the viability of LSTM for weather forecasting, highlighting its strengths in short-

term predictions. Future work could incorporate external variables like satellite imagery or employ hybrid models combining LSTM with convolutional networks. For Bi-LSTM, the predictive performance of the original LSTM model was compared with a modified version that incorporated a Bi-GRU layer and additional regularization. The results indicate that the modified model provided improved accuracy in predicting mean temperature, as shown by the lower error metrics of MAE, MSE, RMSE and the higher R-squared value. Although the training and validation losses were slightly higher in the initial epochs, the modified model exhibited better overall performance, with a noticeable increase in the model's ability to generalize. This suggests that the added complexity of the Bi-GRU layer and regularization techniques enhanced the model's robustness.

References

- [1] P. T. Jia, W. Sun, "A Survey of Text Classification Based on Deep Learning," *Computer and Modernization*, no. 7, pp. 29-37, 2021.
- [2] F. Qu, Y. Shi, C. Xie, J. Zhang, W. Li and C. Lin, "Improvement of power equipment defect text quality based on improved BI-LSTM," *CSEE Journal of Power and Energy Systems*, pp. 1-9, Feb. 2024,
- [3] W. Daqian, W. Bo, and L. Gang, "Research on Unstructured Text Data Mining and Fault Classification Based on RNN-LSTM with Malfunction Inspection Report," *Energies*, vol. 10, no. 3, pp. 406-416, 2017.
- [4] Y. Shao, T.K. Wan and K.H.K. Chan, "Prediction of COVID-19 cases by multifactor driven long short-term memory (LSTM) model", *Scientific Reports*, vol. 15, 4935, 2025.
- [5] M. Junaid, M. Ghergherehchi and S. Lee, "Multitask Deep Learning for Predicting Parkinson's Progression and Depression from Multimodal Time Series Data," *IEEE Access*, pp. 1-25, July 2025.
- [6] A. Sagheer, M. Kotb, "Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems," *Scientific Reports*, vol. 9, no. 1, pp. 1-16, 2019.
- [7] M. O. Qays, I. Ahmad, D. Habibi and M. A. S. Masoum, "System strength forecasting using hedge feedforward feedback-based gated recurrent unit algorithm with drift detection analysis," *CSEE Journal of Power and Energy Systems*, pp. 1-10, Jan. 2025.
- [8] M. Abumohsen, A. Y. Owda, and M. Owda, "Electrical Load Forecasting Using LSTM, GRU, and RNN Algorithms," *Energies*, vol. 16, no. 5, p. 2283, Feb. 2023.
- [9] Y. Xiao, C. Zou, H. Chi, and R. Fang, "Boosted GRU model for short-term forecasting of wind power with feature-weighted principal component analysis," *Energy*, vol. 267, p. 126503, Mar. 2023.
- [10] G. Li, W. Wang, and Y. Qi, "Defect Text Analysis Method of Electric Power Equipment Based on



- Double-Layer Bidirectional LSTM Model,” in *Proc. IEEE 3rd International Electrical and Energy Conference (CIEEC)*, pp. 1318-1324, 2019.
- [11] Medium, “Complete Guide to Learn LSTM Models: Types, Applications, and When to Use Which Model”, Aug. 11, 2024. Available at: <https://blog.stackademic.com/complete-guide-to-learn-lstm-models-types-applications-and-when-to-use-which-model-f9b779f31714>. Access on 14 Aug. 2025.
- [12] R. Kang, B. Park, Q. Ouyang, N. Ren, “Rapid identification of foodborne bacteria with hyperspectral microscopic imaging and artificial intelligence classification algorithms”, *Food Control*, vol. 130, 2021.
- and Applications (IJACSA), vol. 14, no. 8, pp. 338-347, 2023
- [14] X. Wang, G. Xie, Y. Zhang, H. Liu, L. Zhou, W. Liu, Y. Gao, “The Application of a BiGRU Model with Transformer-Based Error Correction in Deformation Prediction for Bridge SHM”, *Buildings*, vol. 15, no. 4, 542, 2025.
- [15] S. Ghosh, A. Ekbal, P. Bhattacharyya, “Chapter 2 - Natural language processing and sentiment analysis: perspectives from computational intelligence” , Editor(s): D. Das, A. Kumar Kolya, A. Basu, S. Sarkar, In “Hybrid Computational Intelligence for Pattern Analysis and Understanding, Computational Intelligence Applications for Text and Sentiment Data Analysis”, *Academic Press.*, pp. 17-47, 2023.
- [13] M. Nair, M. I. Marie , L. A. Abd-Elmegid, “Prediction of Cryptocurrency Price using Time Series Data and Deep Learning Algorithms”, *International Journal of Advanced Computer Science*