International Journal on Advanced Science Engineering Information Technology

Vol.15 (2025) No. 1
ISSN: 2088-5334

Multivariate Variable-Based LSTM-AE Model for Solar Power Prediction

GunHa Park [a], JongChan Kim [a,*]

[a] Department of Computer Engineering, Sunchon National University, 255, Jungang-ro, Suncheon-si, Republic of Korea
Corresponding author: *seaghost@scnu.ac.kr

Abstract—**This study proposes a multivariate-based LSTM-Autoencoder (LSTM-AE) model for short-term photovoltaic power generation prediction. The LSTM-based encoder-decoder structure effectively learns multivariate relationships and time series dependencies between significant environmental and power-related variables. Input variables include DC voltage, DC current, DC power, ambient temperature, solar radiation, and environmental factors, and they are preprocessed through scaling to increase learning efficiency. The encoder compresses multivariate time series data to the latent space, and the decoder restores the corresponding sequence to learn the complex time series patterns of the data. The normalization technique was applied to the algorithm to prevent overfitting and improve the model's generalization performance. The prediction accuracy evaluation was made through mean absolute percentage error (MAPE), mean square root error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). As a result of the experiment, the proposed LSTM-AE model outperformed the existing model in capturing nonlinear and long-term dependence. The results of this study suggest that the LSTM-AE architecture can contribute to the development of renewable energy prediction fields, contributing to the development of more accurate and reliable photovoltaic prediction systems.**

Keywords— **Solar; power; LSTM; autoencoder; ARIMA; normalization.**

Manuscript received 7 Oct. 2024; revised 19 Jan. 2025; accepted 12 Feb. 2025. Date of publication 28 Feb. 2025.

I. INTRODUCTION

Forecasting is applied prominently in various areas such as finance, transportation, and the environment by grasping future outcomes and supporting rational decision-making. Forecasting goes beyond predicting the future and performs key functions such as supporting strategic decisions, improving efficiency, and managing risk. The higher the accuracy and reliability of predictions, the better the quality of decision-making, creating continuous value across society and industry. Compared to current forecasting techniques based on statistics and regression, deep learning-based prediction technology has led to rapid development, learning vast amounts of data to recognize complex patterns, and based on this, it has excellent predictive performance.

Environmental interest is increasing worldwide, and many companies are adopting eco-friendly management policies such as ESG (Environmental Social Coverage). The use of renewable energy has become an essential factor to implement eco-friendly management policies [1], [2], [3], [4], [5]. Renewable energy can be obtained from nature through solar and wind power. Since it is a pollution-free energy that does not emit carbon in the energy production process, and there is no fear of depletion, many studies have been conducted worldwide. Solar power generation has a simple structure, so it has been frequently used because it is easy to maintain, and installation cost is relatively low. Solar power generation can hold a longer lifespan than other renewable energy sources after installation, so electricity can be produced at no added cost. However, when establishing a power consumption plan, solar power generation can be limited to the output. The output limit means that the power generation output is intentionally limited to maintain the safety and stability of the power grid when there is much power caused by intermittent renewable energy such as solar power generation. Several factors may exist for the output limitation of renewable energy, such as the power grid's capacity, an imbalance between supply and demand, a power quality problem, or a technical limitation of the power grid.

Statistics-based prediction algorithms and deep learning-based algorithms exist as methodologies for prediction. Among statistical-based prediction algorithms, regression analysis is a technique that models and analyzes the relationship between two or more variables. Time series data have characteristics such as a trend of increasing or decreasing over time, seasonality, a pattern that repeats in a specific cycle, and autocorrelation, in which data from the previous time affect the current data. Regression analysis is mainly used to understand and predict the relationship between the dependent variable and one or more independent variables [6], [7], [8], [9], [10].

Regression analysis techniques using time series data include simple regression, multiple regression, and ARIMA (AutoRegressive Integrated Moving Average) model. Simple regression is a method of linearly modeling the relationship between time and observed values, but there is a limit to sufficiently reflecting the complex characteristics of time series data. In contrast, multiple regression is a method of predicting the current value through several independent variables, and the amount of solar power can be predicted more accurately by including various environmental variables such as insolation, temperature, and humidity. It is more advantageous regarding prediction accuracy because it can reflect various variables than simple regression.

The ARIMA model predicts by reflecting time series data's irregularity, trend, and seasonality. This model combines the concepts of autoregressive (AR), difference (I), removes data abnormality, and moving average (MA), which reflects past errors. When the seasonality pattern is clear, the SARIMA (Seasonal ARIMA) model that expands it is mainly used [11], [12], [13], [14].

The main goal of time series data prediction is to predict future time point values based on past data distribution. Deep learning is widely used for time series data prediction as it performs strongly for complex pattern recognition and prediction problems. Long Short-Term Memory (LSTM) improves the long-term dependence problem of recurrent neural networks (RNNs), effectively modeling the time dependence of time series data. RNNs have strength in processing time order information but have limitations in learning long-term dependence, and to solve this problem, LSTM supplements the long-term dependence problem by introducing cell state and several gate mechanisms [15], [16], [17].

Autoencoders aim to compress and restore data and are mainly used for dimensionality reduction and feature extraction. General autoencoders are a type of neural network with the same input and output, and unlike supervised learning, they do not require labeling. They can be classified as part of unsupervised learning or self-supervised learning. Autoencoders are designed with the intention of bottlenecks to learn how to compress data in low dimensions, and depending on the purpose and data characteristics, various variations such as Basic Autoencoders, sparse autoencoders, variational autoencoders, and denoising autoencoders have been studied [18], [19].

In this paper, the LSTM-AE model was used to predict the variability more accurately and the output change in advance for the wide use of renewable energy to respond appropriately. The photovoltaic power generation data has multivariate variables. LSTM can be calculated by accepting these multivariate variables at once. LSTM also predicts power generation by reflecting the meteorological pattern and learning the temporal pattern and long-term dependence of time series data. These LSTM cells structurally form an autoencoder (AE) structure and predict power generation through dimensional reduction and restoration.

II. Materials and Method

A. Data Collection of Solar Power Generation

Photovoltaic power generation is performed through photovoltaic panels, which convert light energy into electrical energy and use it. Photovoltaic panels generate power by photoelectric effects, which absorb photons and generate electric fields between semiconductors to form electric currents. Variables that affect photovoltaic power generation can be affected by various external factors such as weather, geography, time, and the state of the panel, so it is necessary to respond to variables over time [20], [21], [22]. Photovoltaic power generation data should also be measured by considering a time series, and predictions that feel the effects of measured photovoltaic power generation data and external environmental variables should be performed simultaneously. Environmental variables can change the amount of insolation reaching the photovoltaic panel and the efficiency of the panel, and classification and integration of variables are required for accurate photovoltaic power generation prediction. The data measurable on the panel is extracted based on the date and time, as well as power-based data such as DCV, DCA, DCP, and ACP, as well as data due to environmental factors such as Horizontal Solar Radiation, Vertical Solar Radiation, Module Temperature, and Outside Temperature. In particular, the amount of horizontal and vertical solar radiation in solar power generation is a significant environmental variable that directly affects solar power generation depending on the location, direction, inclination angle, and intensity of insolation.

Horizontal Solar Radiation and Vertical Solar Radiation are factors that have an essential influence on solar power generation efficiency. These two values directly affect the amount of power generated depending on the location, direction, inclination angle, and intensity of the insolation of the solar panel. The amount of horizontal insulation significantly affects the main output of a horizontal solar panel system fixed to the ground in general. It is an essential variable in predicting the amount of power in an area where a horizontal panel system is mainly installed. On the other hand, the amount of vertical insulation primarily affects the output of inclined panels or panels installed on the wall of a building. These vertical panel systems are often used in building integrated photovoltaic (BIPV) and are designed to increase solar utilization in urban environments. As such, the amount of insolation on the horizontal and vertical surfaces must be individually considered to optimize solar power generation efficiency according to the panel arrangement conditions and regional sunlight conditions, enabling more precise prediction of power generation.

Since the amount of insolation on the horizontal and vertical surfaces also fluctuates according to the changes in the solar altitude and angle over time, it is essential to reflect the seasonal pattern and the change in the amount of insolation over time when modeling the amount of power. Since photovoltaic power generation data is fundamentally affected by various external environmental variables, noise, missing values, and outliers likely exist in the collected data.

When these factors are included in the model, not only the prediction performance may deteriorate, but a big problem may occur in the reliability of the result, so the preprocessing process plays a key role in refining the data and increasing the reliability, which can be said to be an essential step in improving the accuracy of the power generation prediction model. Missing values and outliers can cause significant problems when dealing with time series data. Missing values mean the case of missing values in a specific time zone, and such data may make information missing in the model training process be recognized as an incorrect pattern. Outliers are abnormal values in which a specific data point deviates from a regular pattern and may occur due to malfunctions of data collection equipment, environmental factors, or other unavoidable reasons. When missing and outliers are included in the model training, the model learns patterns different from the actual data, which can significantly reduce prediction accuracy.

In this paper, ARIMA (Autoregressive Integrated Moving) was used as a model to process missing and outliers in data. Pre-processing using the ARIMA model predicts patterns by analyzing data trends, seasons, residuals, and missing and outliers. It is a statistical technique used primarily in unique time series data to maintain data continuity and preserve information necessary for model learning by correcting for missing and outliers. The ARIMA model is relatively simple, and even if a relatively small amount of data is built into a model and reflected in the preprocessing, the model's operation method can be intuitively reflected, data can be refined, and stable data can be produced.

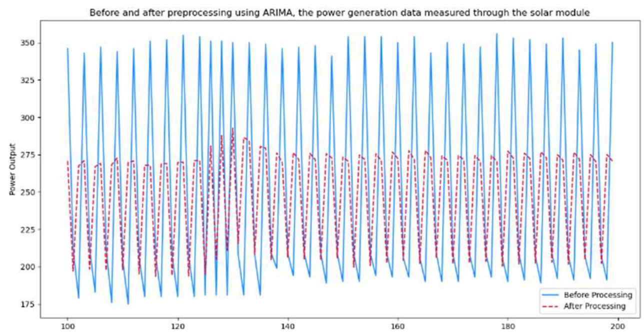

Fig. 1 Before and after preprocessing using ARIMA

Fig. 1 shows some of the multivariate photovoltaic data before and after preprocessing. ARIMA automatically processes data abnormality to ensure normality. Abnormalities in time series data mean that the statistical characteristics of data are not constant over time. Abnormal time series have characteristics that statistical properties such as mean, variance, and self-covariance change over time, which may lead to data analysis and prediction modeling difficulties. Abnormal data in time series data require preprocessing to secure normality. The ARIMA model performs differences to be used for prediction, and through this, a stable time series pattern is formed.

B. Dimension reduction with LSTM

Long-short-term memory (LSTM) is a recurrent neural network that performs well in learning patterns from continuous data, such as time series data. LSTM is designed to solve the long-term dependence problem handled by RNNs and learns while maintaining important information even on long sequences. Unlike RNNs, LSTM includes memory cells and gates that control information flow, allowing the preservation of long sequences without loss of past data information [23], [24], [25], [26], [27].

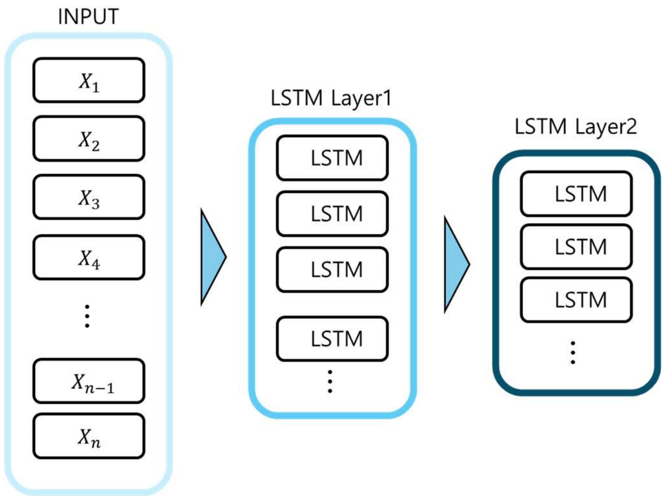

Fig. 2 LSTM in Autoencoder Structure

As shown in Fig 2, the LSTM layer accepts multivariate time series data preprocessed as input values, utilizes them as input values, and reduces the dimension so that only primary time series information can be extracted as features while preserving long sequences without losing information on past data. LSTM has an excellent ability to predict current and future patterns using past data. It processes weather and power generation data through this and performs long-term time-dependence learning. Solar power generation prediction is reflected in the prediction depending on environmental data (weather conditions, etc.). LSTM shows excellent performance in learning nonlinear relationships, capturing complex interactions between photovoltaic power generation and weather variables [28], [29], [30].

This paper uses LSTM to process multivariate time series data simultaneously. Various module data (direct current voltage, direct current, alternating current voltage, alternating current, direct current power generation, etc.) and meteorological data (horizontal, vertical solar radiation, module temperature, external temperature, etc.) of the photovoltaic power generation system are included. LSTM learns complex interactions while processing these data simultaneously. The LSTM used in the model has a structure consisting of an input gate, an oblivion gate, and an output gate and controls the process of selectively remembering or deleting information. This gate mechanism is related to how data from the previous state is transmitted to the current state, which determines how to process the information from the past. The LSTM of this structure can maintain important information within the time series data and efficiently delete information not required in the learning process.

C. Data Prediction Using Autoencoder Structure

The autoencoder derives prediction results by predicting the context of the data using a model that reconstructs and restores the data. In the learning process, the autoencoder optimizes to minimize the difference between the input data and the restored data and learns essential features of the input data.

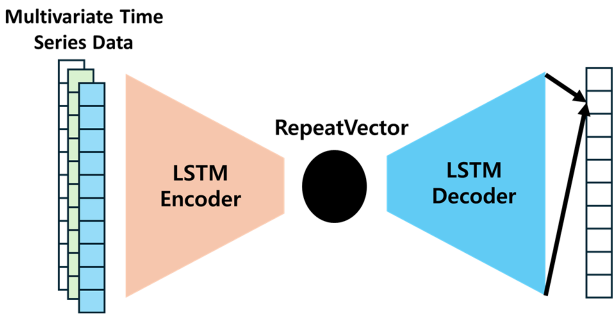

Fig. 3 LSTM-AE Structure

Fig 3 shows a data processing process using an LSTM model and an autoencoder. Using the LSTM model, the encoder part of the autoencoder is reproduced to compress the input data to express implicit characteristics. The LSTM model predicts and restores data while using the compressed feature data again by forming an autoencoder structure, which means the autoencoder's decoder. In this paper, dimension reduction and restoration were performed using the LSTM of two encoders and the LSTM of two decoders.

III. Results and Discussion

A. Experimental Environment

In this paper, data collection, including DC Power, AC Power, and various environmental variables, was performed to predict the amount of solar power, and the correlation between these data was analyzed. Data on highly correlated power generation were selected and used to learn the prediction model. For a more accurate prediction, weather information such as temperature, humidity, precipitation, and cloudiness provided by the Meteorological Administration was collected as time series data and included in the learning data. The power generation data collected by the photovoltaic power plant is based on the power generation data of the inverter measured in the Jeollanam-do area, and through this, the prediction algorithm was learned and evaluated.

A hybrid LSTM (Long Short-Term Memory) and Autoencoder model was used as the prediction model. LSTM has a structure for processing long-term dependencies that may effectively learn complex patterns of time series data. Autoencoder has an advantage in extracting and restoring features by compressing high-dimensional data characteristics. The hybrid model used in this study processes input time series data in an encoder-decoder structure.
It consists of two LSTM layers. The first LSTM layer generates the output of all time steps and transfers it to the next LSTM layer. The second LSTM layer outputs the encoded compression vector to express data in a reduced form, which is used as decoder input. The first decoder LSTM layer receives the encoder's output as a repetition vector and inputs the data, and the second decoder LSTM layer outputs a reconstructed sequence. In this process, the input and output data sizes are maintained the same.

B. Dataset Type and Environment Variables

The amount of photovoltaic power generation may vary due to various constraints. In the data in this paper, environmental variables, excluding the decrease in the amount of power generated due to the degradation of the photovoltaic module, were calculated. Environmental data that directly affects the amount of photovoltaic power generation use direct current and alternating current power generation data measured by the photovoltaic module and the module's temperature, external temperature, and temperature.

TABLE I
Power Plant System Power Data Schema

Column	Summary
PW_DATE	Collected DATE
PW_TIME	Collected TIME
PW_DCV	DC Voltage
PW_DCA	DC Ampere
PW_DCP	DC Power
PW_ACP	AC Power
PW_TOTPOWER	Total Power

Table 1 shows the schema for data collected by photovoltaic power plants. PW_DATE and PW_TIME store temporal information for time series data. PW_DCV, PW_DCA, PW_DCP, PW_ACP, and PW_TOTPOWER contain power generation data. To secure the reliability of the data measurable by the module and the prediction results, additional data were obtained through climate statistics that the Meteorological Administration can collect.

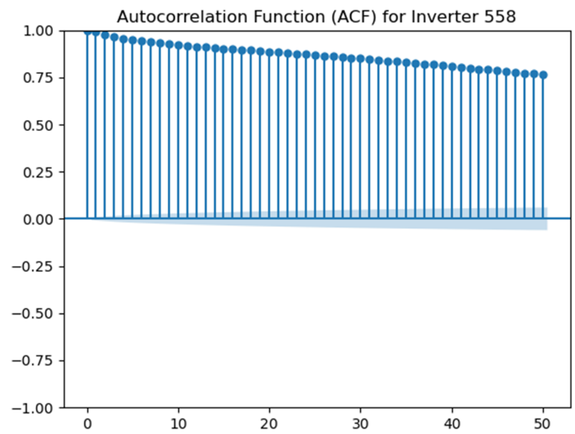

Fig. 4 Autocorrelation Graph

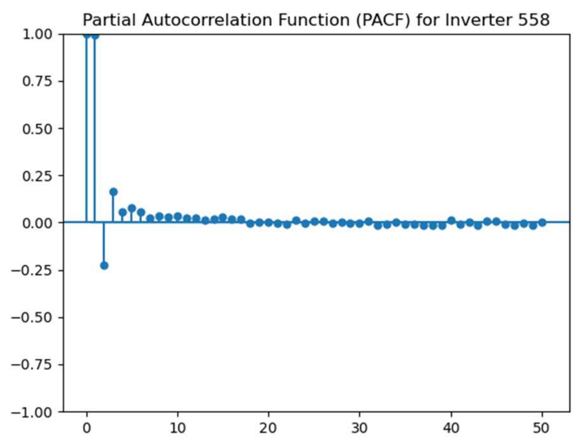

Fig. 5 Partial Autocorrelation Graph

ARIMA (Autoregressive Integrated Moving Average Model) is applied to data used for prediction by performing parallax correlation analysis. ARIMA maintains a continuous pattern because missing values of time series data can cause the time interval of the data to become uneven. While precipitation and temperature data among weather data can analyze a direct correlation with the amount of solar power generated, air volume and cloud volume data are difficult to apply directly due to the lack of specific values measured daily. Accordingly, the air volume and cloud volume data are used to adjust the model's hyperparameters, and more sophisticated data collection and learning will be performed through future research.

Two scenarios were observed due to analyzing some of the DC power data closely related to the amount of photovoltaic power generation. In the period (a) graph, the amount of power generation showed a relatively constant range of fluctuations, and the tendency of stable power generation appeared. On the other hand, in the graph of the period (b), the volatility of power generation was more significant, and a sharp drop and irregular fluctuation were observed. This can be interpreted as a result of being affected by weather conditions such as clouds and weather changes.

ARIMA (Autoregressive Integrated Moving Average Model) was applied as the prediction model. To this end, the pattern of time series data was identified by performing correlation analysis by lag. In addition, since missing values in time series data can lead to an imbalance in time intervals, the reliability of the prediction was increased by maintaining continuous patterns and supplementing missing values through the ARIMA model.

C. Configure Environment Variable Correlation Analysis Results and Data Input

Based on the prepared data, a correlation analysis was conducted using the Pearson correlation coefficient to analyze the relationship between data collected from the power plant to improve the prediction accuracy. A value with a high correlation coefficient with the AC power to be predicted was selected as the input value. As a result of the correlation analysis, the highest value of AC power and correlation coefficient among inverter data is DC power, and the converted value of DC power is AC power, so the relationship between the two is directly related to the efficiency of the inverter. Additionally, the amount of insolation is the highest correlation coefficient in the environmental sensor data to control the difference in the installation location of the inverter and weather variables.

Pearson correlation coefficient is a method of measuring the linear correlation between two variables. The two variables do not consider time and calculate the correlation at the same time point. The value is expressed as between -1 and 1; a positive correlation closer to 1 means a negative correlation and a closer to 0 means no correlation.

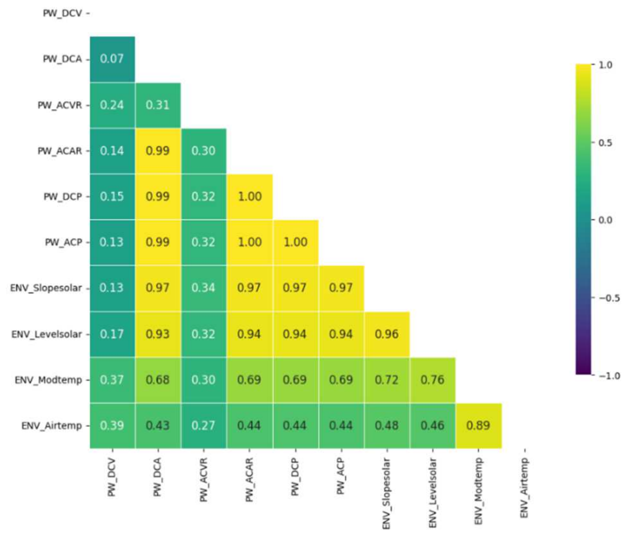

Fig. 6 Heatmap for correlation analysis using Pearson correlation coefficients of multivariate variables

Fig 6 shows the correlation of each environmental variable in a matrix form using the Pearson correlation coefficient. DC electricity is converted into AC electricity through an inverter. Therefore, the direct power generation (DCP) and AC power generation (ACP) values, which are the direct power generation values, have almost the exact correlation. The Pearson correlation coefficient was used in this paper to identify variables closely related to power generation.

As a result of the correlation analysis, the highest value of AC power and correlation coefficient among inverter data is DC power, and the converted value of DC power is AC power, so the relationship between the two is directly related to the efficiency of the inverter. Additionally, the amount of insolation is the highest correlation coefficient in the environmental sensor data to control the difference in the installation location of the inverter and weather variables.

D. Experimental Results

An LSTM-AE hybrid model was constructed, and MAE, RMSE, MAPE, and R^2 were used to evaluate the time series data prediction performance. After analyzing the correlation of the pre-processed data, prediction was performed using highly relevant data.

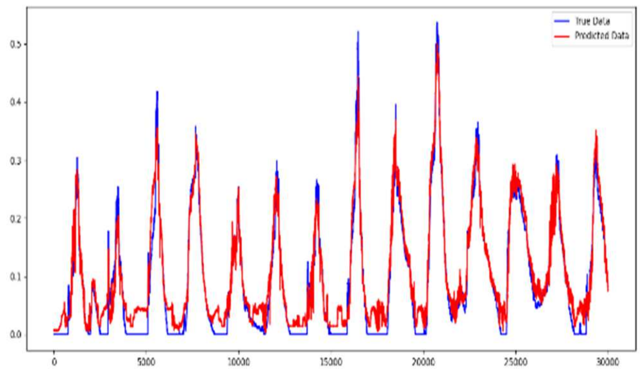

Fig. 7 Prediction Results Graph 1/3 Points

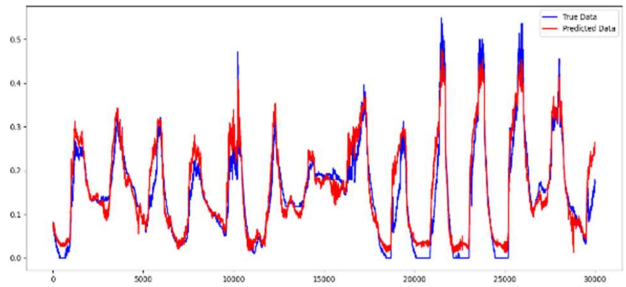

Fig. 8 Prediction Results Graph 2/3 Points

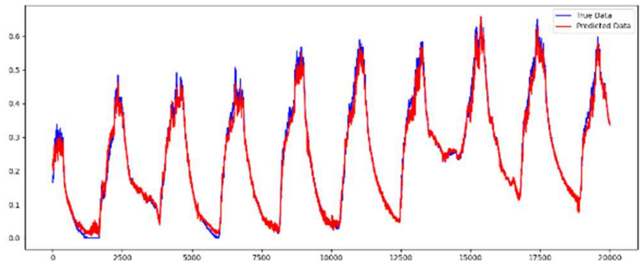

Fig. 9 Prediction Results Graph 3/3 Points

Figs 7, 8, and 9 show the prediction result. MAPE was 6.28, RMSE was 20.65, MAE was 16.45, and silver 0.88. Using the learned model, the power generation data of the photovoltaic module was analyzed by predicting the power generation data and confirming the decrease in the AC power generation conversion rate to DC by year. After that, to check the change in the monthly prediction result based on the prediction result and determine whether the model is overfitting, the period with the optimal power generation efficiency was analyzed based on the prediction result of the start point and end point.

Looking at the graph's trend through qualitative evaluation shows that the predicted value follows the trend of the actual value well. The predicted value is lower than the measured value at the maximum point, and the estimated value at the minimum point is lower than the predicted value. Additional error analysis was performed to accurately grasp the relationship between the actual value and the predicted value, and the error between the predicted value derived from the model and the actual power generation was statistically analyzed.

As the number of operating days increases, the difference between the predicted value and the measured value gradually increases, and an abnormal phenomenon that increases values outside the standard deviation range could be observed. This means that the inverter efficiency is slowly decreasing over time. To analyze this, it is necessary to diagnose the decrease in inverter efficiency using the power generation data collected from the inverter of different years and the environmental sensor data. In particular, it is also required to understand the trend of decreasing the inverter's efficiency through regression analysis between DC and AC power generation.

IV. Conclusions

This paper predicted the amount of solar power using a hybrid model of LSTM (Long Short-Term Memory) and Autoencoder to indicate the amount of solar power using deep learning. The LSTM used in the hybrid model performs excellently in processing sequence data. The entire architecture utilizes the structure of the Autoencoder and compresses it so that only the data features can be used in the encoder part. After that, the data is restored by using two LSTMs in the same way in the decoder part, and the seasonal performance trend is restored, and the data is predicted.

For accurate prediction, the ARIMA (Autoregressive Integrated Moving Average Model) algorithm was used to perform preprocessing to remove outliers and missing values in time series data. Through PACF (Partial Autocorrelation Function) correlation analysis, data with a high correlation with the actual power generation were used as learning data for prediction.

In this paper, we compared the data at each point in time by calculating the error between the predicted power generation and the actual value derived by the prediction model. The exact value, trend, and seasonality are predicted similarly, but the power generation is generally lower than the actual value. The accuracy was measured at about 94%. A study will be conducted through future research to determine the diagnosis and deterioration timing of the failure due to the occurrence of outliers and missing values in power generation.

References

[1] M. Lee, P. Lee, D. Jeong, M. Han, and S. P. Jung, "RE100 for 100% renewable electricity: Status and prospects," *J. Korean Soc. Environ. Eng.*, vol. 45, no. 3, pp. 161–169, Mar. 2023, doi:10.4491/ksee.2023.45.3.161.

[2] D.-H. T. Kim, C. Park, J.-B. Park, and J.-H. Roh, "An analytical approach for transmission use-of-system charges for a renewable energy power purchase agreement," *Trans. Korean Inst. Electr. Eng.*, vol. 72, no. 7, pp. 801–808, Jul. 2023, doi: 10.5370/kiee.2023.72.7.801.

[3] S. Sengupta, A. Sahay, and R. D. Hisrich, "The social–market convergence in a renewable energy social enterprise," *J. Clean. Prod.*, vol. 270, p. 122516, Oct. 2020, doi: 10.1016/j.jclepro.2020.122516.

[4] A. Kuś and D. Grego-Planer, "A model of innovation activity in small enterprises in the context of selected financial factors: The example of the renewable energy sector," *Energies*, vol. 14, no. 10, p. 2926, May 2021, doi: 10.3390/en14102926.

[5] C. D. Diale, M. G. Kanakana-Katumba, and R. W. Maladzhi, "Ecosystem of renewable energy enterprises for sustainable development: A systematic review," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 6, no. 1, pp. 401–408, Jan. 2021, doi: 10.25046/aj060146.

[6] H. Choi, "Stock prediction analysis through artificial intelligence using big data," *J. Korea Inst. Inf. Commun. Eng.*, vol. 25, no. 10, pp. 1435–1440, 2021.

[7] Y. Lim, W. Shin, and H. Kim, "Prediction model for gastric cancer in patients with chronic gastritis using RNN based on medical big data," in *Proc. Korean Inst. Inf. Sci. Eng. Conf.*, vol. 2020, no. 7, pp. 1289–1291, 2020.

[8] Y. Kim, S. Kim, and H. Yeo, "A study on traffic prediction using hybrid approach of machine learning and simulation techniques," *J. Korea Inst. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 100–112, Oct. 2021, doi: 10.12815/kits.2021.20.5.100.

[9] H. Kim and M. Lee, "Research trends in machine learning applications in marketing," *Korean J. Mark.*, vol. 36, no. 1, pp. 1–25, Feb. 2021, doi: 10.15830/kjm.2021.36.1.1.

[10] S. Lee, W. Choi, and M. Kim, "Development of weather forecasting algorithm for improvement of power generation prediction accuracy," in *Proc. Korean Inst. Electr. Eng. (KIEE) Summer Conf.*, pp. 843–844, Jul. 2021.

[11] H. Jeong et al., "A research of prediction of photovoltaic power using SARIMA model," *J. Korea Multimed. Soc.*, vol. 25, no. 1, pp. 82–91, 2022, doi: 10.9717/kmms.2022.25.1.082.

[12] R. W. Divisekara, G. J. M. S. R. Jayasinghe, and K. W. S. N. Kumari, "Forecasting the red lentils commodity market price using SARIMA models," *SN Bus. Econ.*, vol. 1, no. 1, Nov. 2020, doi:10.1007/s43546-020-00020-x.

[13] S. Ma, Q. Liu, and Y. Zhang, "A prediction method of fire frequency: Based on the optimization of SARIMA model," *PLoS ONE*, vol. 16, no. 8, p. e0255857, Aug. 2021, doi: 10.1371/journal.pone.0255857.

[14] N. Deretić et al., "SARIMA modelling approach for forecasting of traffic accidents," *Sustainability*, vol. 14, no. 8, p. 4403, Apr. 2022, doi:10.3390/su14084403.

[15] M. Abumohsen et al., "Hybrid machine learning model combining of CNN-LSTM-RF for time series forecasting of solar power generation," *e-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 9, p. 100636, Sep. 2024, doi: 10.1016/j.prime.2024.100636.

[16] H. Munawer Al-Otum, "Classification of anomalies in electroluminescence images of solar PV modules using CNN-based deep learning," *Sol. Energy*, vol. 278, p. 112803, Aug. 2024, doi:10.1016/j.solener.2024.112803.

[17] I. Jebli, F.-Z. Belouadha, M. I. Kabbaj, and A. Tilioua, "Deep learning based models for solar energy prediction," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 6, no. 1, pp. 349–355, Jan. 2021, doi: 10.25046/aj060140.

[18] L. P. Cinelli, M. A. Marins, E. A. Barros da Silva, and S. L. Netto, *Variational Methods for Machine Learning with Applications to Deep Networks*. Springer International Publishing, 2021. doi:10.1007/978-3-030-70679-1.

[19] V. Suresh et al., "Probabilistic LSTM-autoencoder based hour-ahead solar power forecasting model for intra-day electricity market participation: A Polish case study," *IEEE Access*, vol. 10, pp. 110628–110638, 2022, doi: 10.1109/access.2022.3215080.

[20] D. Kim, S.-M. Kim, J. Suh, and Y. Choi, "Anomaly detection of photovoltaic systems installed in renewable energy housing support project sites by analyzing power generation data," *J. Korean Sol. Energy Soc.*, vol. 42, no. 1, pp. 33–46, Feb. 2022, doi:10.7836/kses.2022.42.1.033.

[21] J. Seo, T. Lee, W. Lee, and J. Park, "A study on the outlier data estimation method for anomaly detection of photovoltaic system," *J. IKEEE (Inst. Korean Electr. Electron. Eng.)*, vol. 24, no. 2, pp. 403–408, 2020.

[22] Y. Yoon and S. Jeong, "Deterioration and abnormality condition diagnosis through measuring the DC capacitor capacity of PV inverter," *J. Inst. Internet Broadcast. Commun.*, vol. 24, no. 5, pp. 135–140, 2024.

[23] L. Wang et al., "Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model," *Energy*, vol. 262, p. 125592, Jan. 2023, doi:10.1016/j.energy.2022.125592.

[24] A. Djaafari et al., "Hourly predictions of direct normal irradiation using an innovative hybrid LSTM model for concentrating solar power projects in hyper-arid regions," *Energy Rep.*, vol. 8, pp. 15548–15562, Nov. 2022, doi: 10.1016/j.egyr.2022.10.402.

[25] L. Ren et al., "A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life," *IEEE Trans. Ind. Inform.*, vol. 17, no. 5, pp. 3478–3487, May 2021, doi:10.1109/tii.2020.3008223.

[26] N. L. M. Jailani et al., "Investigating the power of LSTM-based models in solar energy forecasting," *Processes*, vol. 11, no. 5, p. 1382, May 2023, doi: 10.3390/pr11051382.

[27] S.-C. Lim et al., "Solar power forecasting using CNN-LSTM hybrid model," *Energies*, vol. 15, no. 21, p. 8233, Nov. 2022, doi:10.3390/en15218233.

[28] P. Li, Y. Pei, and J. Li, "A comprehensive survey on design and application of autoencoder in deep learning," *Appl. Soft Comput.*, vol. 138, p. 110176, May 2023, doi: 10.1016/j.asoc.2023.110176.

[29] C. Zhang et al., "Autoencoder in autoencoder networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 2, pp. 2263–2275, Feb. 2024, doi: 10.1109/tnnls.2022.3189239.

[30] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenom.*, vol. 404, p. 132306, Mar. 2020, doi:10.1016/j.physd.2019.132306.