Short-Term Electricity Demand Forecasting Based on Cloudy and Clear Sky Solar Irradiance Data

Karma Dorji
School of Renewable Energy and Smart Grid Technology (SGtech), Naresuan University, Phitsanulok 65000, Thailand
karmad65@nu.ac.th

Sorawut Jittanon
School of Renewable Energy and Smart Grid Technology (SGtech), Naresuan University, Phitsanulok 65000, Thailand
sorawutj66@nu.ac.th

Temsiri Prompook
School of Renewable Energy and Smart Grid Technology (SGtech), Naresuan University, Phitsanulok 65000, Thailand
temsirip65@nu.ac.th

Yirga Belay Muna
School of Renewable Energy and Smart Grid Technology (SGtech), Naresuan University, Phitsanulok 65000, Thailand
yirgabelay.m@nu.ac.th

Chakkrit Termritthikun
School of Renewable Energy and Smart Grid Technology (SGtech), Naresuan University, Phitsanulok 65000, Thailand
chakkritt@nu.ac.th (corresponding author)

Received: 2 May 2025 | Revised: 17 June 2025 | Accepted: 28 June 2025

 DOI: https://doi.org/10.48084/etasr.11889

ABSTRACT

Short-term electricity demand forecasting is a critical task for real-time decision-making. This research investigates the impact of solar irradiance components on electricity demand forecasting. The study utilized two years of hourly electricity data in addition to three key solar irradiance components: Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), and Diffuse Horizontal Irradiance (DHI), as clear sky indices and exogenous features. Extensive feature selection was performed to discover the best feature that enhances the forecasting accuracy of the model. Temporal Fusion Transformer (TFT) was employed as the primary model to study the influence of these exogenous variables on the electricity demand forecasting. The results demonstrated that use of solar irradiance data in the electricity demand forecasting improves the forecasting accuracy of the model. Particularly the combination of DHI and DNI clear sky indices produced the most accurate predictions. However, the addition of GHI introduced redundancy and reduced performance. The performance of the TFT model was also compared with various benchmark models. TFT outperformed the other compared models in various forecast horizons.

Keywords-electricity demand forecasting; solar irradiance; clear sky index; deep learning; prosumer building

I. INTRODUCTION

Due to phenomena, such as urbanization, growing population numbers, and industrialization, the energy consumption worldwide has increased [1]. Forecasting the energy demand is becoming increasingly complex because of the rising electricity demand and renewable energy transition, requiring strategies to accommodate green energy, manage dispatchable demand, and store heating and electricity [2]. Forecasting can be categorized by time horizons in long-term, mid-term, and short-term or by forecasting techniques, including point, probabilistic, recursive, and multi-horizon forecasting [3]. Short-term demand forecasting has gained popularity owing to the electricity providers' to bid for energy supply and can additionally aid on the decision-making process of power plants [4]. Traditionally, statistical models, such as Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and exponential smoothing have been employed for electricity demand forecasting due to their effectiveness in modeling linear trends and seasonality in historical load data. However, these models are ineffective in handling complex, nonlinear relationships and exogenous factors, such as weather conditions [5]. To address this limitation Machine Learning (ML) and Deep Learning (DL) methods, including SVR, Random Forests, XGBoost, LSTM networks, TFT, and Time-series Dense Encoder (TiDE) are introduced [6, 7]. These models outperform traditional statistical models due to their flexibility, they can capture long-term trends, feature interactions, and can handle exogenous variables, significantly improving the models' forecasting performance [8].

Feature engineering based on environment, weather, and temporal features is essential in enhancing the accuracy of the models [9]. Weather conditions are recognized as key factors of the electricity consumption, as these conditions directly influence the heating, cooling, and lighting needs. It has been shown that incorporating meteorological features, such as temperature, humidity, dew point, wind speed, wind direction, and solar irradiance, can significantly improve the accuracy of the load forecasting models [10]. Authors in [11] propose a hybrid model, iSSA-LSSVM, to enhance the prediction accuracy for short-term load forecasting. This model utilizes ambient temperature, atmospheric pressure, relative humidity, and exhaust steam pressure as input variables and electrical power output as the target, outperforming three benchmark models by achieving a lower MSE and faster convergence. Studies included the effect of weather data, splitting the time horizon and climate and holiday features in energy forecasting [9, 12].

Solar irradiance, typically measured through indicators, like GHI, DNI, and DHI, has become important in regions with high solar Photovoltaic (PV) activity [13]. Despite these advancements, most studies still consider weather variables as continuous input without specifically distinguishing the effects under different sky conditions, such as cloudy and clear skies, which can introduce unique variations in irradiance and subsequently alter the load patterns. Solar irradiance at the top of the atmosphere is constant, averaging around 1366.1 ± 3.4% W/m², with variation due to the Earth-Sun distance. As solar radiation propagates through the atmosphere, it undergoes absorption and scattering owing to interactions with atmospheric particles. The amount of irradiance reaching a horizontal surface after being influenced by the presence of air molecules, pollutants, and cloud cover is called DHI. As cloudiness increases, DHI becomes a more significant component of the total irradiance. The radiation received on a horizontal surface is GHI, whereas DNI denotes the direct solar radiation travelling straight through the atmosphere, and θ is the solar zenith angle [14]:

$$GHI = DNI \times \cos(\theta) + DHI \quad (1)$$

To forecast the solar irradiance, data-driven, image-based, numerical weather prediction models and hybrid approaches are utilized. Historical data, satellite images, computer-based programs, and their combinations can be used as input. Clear-Sky Index (CSI) and Clearness Index (CI) are crucial in solar irradiance forecasting, with Clear-Sky Models (CSMs) employing radiative transfer models and regional meteorological factors, including the ozone content and water vapor content [15]. The CSI for DHI, DNI, and GHI is calculated using:

$$CSI = \frac{c_x}{I_y} \quad (2)$$

where c_x is the measured irradiance (DHI, DNI, and GHI) under cloudy or real conditions, and I_y is the modeled irradiance under clear sky conditions.

Solar irradiance is primarily used for solar power applications [16]. No prior studies have been conducted on the effect of solar irradiance on the electricity demand forecasting. CSIs are used as exogenous features of the forecasting model to determine if the fluctuations in solar irradiance, driven by weather and sky conditions, influence the electricity consumption patterns of the building. The study's main objective is to:

- Investigate the effect of solar irradiance on short-term electricity demand forecasting in a prosumer building.
- Find the relationship between the solar irradiance and electricity demand to improve the forecasting accuracy, allowing for a more efficient energy usage and prediction.

II. METHODOLOGY

Figure 1 provides the complete workflow. It details data collection, pre-processing, feature extraction, feature selection, model selection, and comparison.

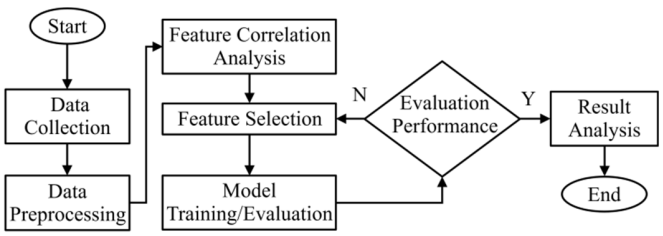

Fig. 1. The workflow of the study.

A. Data Collection

The study utilizes hourly electricity consumption measurements from the NDIR prosumer building at the SGtech, Naresuan University, and solar irradiance data from the OpenWeather web platform for the location of Naresuan University, Phitsanulok, Thailand, with a latitude of 16.74° N and longitude of 100.19° E. The dataset covers a two-year period from 01/01/2022 to 31/12/2023. The hourly electricity consumption is recorded using a smart metering system that logged the building's total demand in kW, along with timestamps. From the datetime feature, the temporal variables, namely workday (0-6) and hour-group (0-3), are extracted. The solar irradiance data are obtained at the same hourly resolution and matched temporally with the demand data to enable weather-aware forecasting. Table I summarizes the features used in the forecasting models.

TABLE I. FEATURE SPECIFICATION OF THE DATASET.

Features	Data type	Description	Range/length
Date	Datetime	2 years of data	2022 - 2023
Demand (kWh)	Float	Target variable	[0,∞)
Workday	Integer	Day of the week	{0-6}
Hour-group	Integer	Divide the 24 h into four groups of 6 h each	{0,1,2,3}
GHI CSI	Float	The ratio of GHI to clear-sky GHI	0 to 1
DNI CSI	Float	The ratio of DNI to clear-sky DNI	0 to 1.1
DHI CSI	Float	The ratio of DHI to clear-sky DHI	0 to 3.7

B. Model Implementation

TFT is a transformer-based DL multi-horizon forecasting model, which efficiently models multivariate time series data. It is interpretable, captures long-term dependencies, handles various exogenous features, and extracts information on how these features interact. TFT captures short-term and long-term dependencies using gated residual networks and temporal self-attention mechanisms [17]. Furthermore, two statistical models, ARIMA and Prophet, two ML models, XGBoost and CatBoost, and an additional DL model, TiDE, are used to analyze the results.

C. Evaluation Metrics

Three evaluation metrics are utilized to evaluate the performance of the models for a sample of n observations y corresponding predictions $\hat{y}$. The Mean Absolute Error (MAE) represents the average absolute difference between predicted and actual values, treating all errors equally, as provided by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

The Mean Squared Error (MSE) calculates the average of the squared differences between the predicted and actual values, giving more weight to larger errors due to squaring, as shown in:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

The Root Mean Squared Error (RMSE) is the square root of MSE. It provides error metrics in the same units as the target variable and emphasizes larger errors, as represented in:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

D. Feature Correlation Analysis

A feature correlation analysis is conducted to determine which features strongly influence the target variable's prediction. Figure 2 provides the visualization of the correlation heatmap.

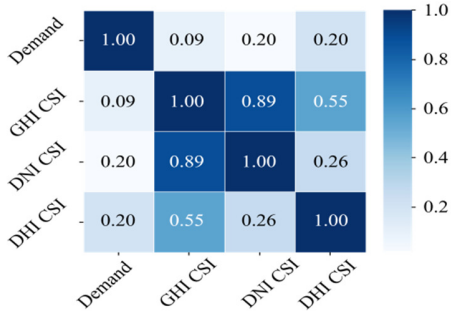

Fig. 2. Correlation heatmap of the demand and CSI.

A correlation value of 1 indicates a strong positive correlation, and a value closer to 0 indicates a weaker or no correlation. The DNI index presents a strong correlation to the GHI index, with a correlation value of 0.89. The DNI index has a weaker correlation to the DHI index, with a correlation value of 0.26. The DHI index shows moderate positive correlations with GHI, with a correlation value of 0.55. The strongest relationship is between the GHI and DNI indices when comparing the correlation between each index, as DNI is a major component of GHI. DHI has weaker correlations with DNI since diffuse radiation behaves independently of the direct radiation. All correlations are positive, meaning that when one index increases, the others also increase, indicating that these indices generally move in the same direction under changing atmospheric conditions. DHI has the strongest correlation to demand, with a correlation value of 0.20, followed by GHI, which has a value of 0.09, and DNI, which has a correlation value of 0.02.

III. RESULTS AND DISCUSSION

A. Diurnal Patterns of Solar Radiation Components

The diurnal curve presents the average behavior of solar radiation components over the observed two-year period. The analysis reveals distinct patterns in the three components of solar radiation throughout the day, as illustrated in Figure 3. DHI has the most pronounced variations during daytime, particularly between 05:00 and 18:00, with a peak around noon, reaching values above 2. A DHI CSI value exceeding 2 indicates that clouds scatter sunlight more effectively than clear skies due to their optical properties, thickness, and coverage. In contrast, DNI remains relatively low (<0.5), highlighting that

the frequent cloud cover has a negative impact. GHI exhibits moderate values (<0.75), with a slight increase during sunrise (04:00-05:00) and sunset (18:00-19:00), as during these hours, the solar angles are low, and all components are weak.

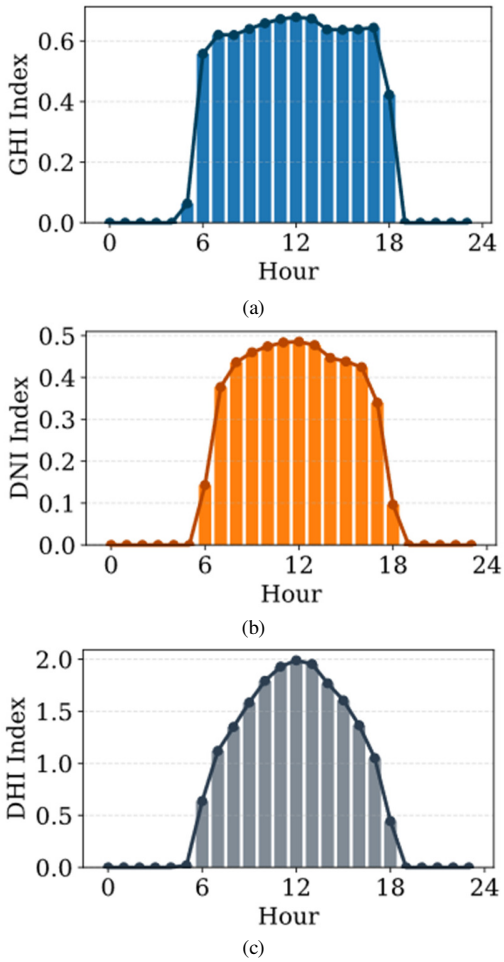

Fig. 3. Hourly averaged diurnal variation of CSI during 2022-2023.

B. Feature Performance

The CSI features are configured as presented in Table II. Seven setups are designed; each is trained using the TFT model, and their accuracy is measured based on the MAE, MSE, and RMSE scores, as presented in Table III. Table III displays the comparative test results of various benchmark models to assess the effectiveness of these configurations. Forecasts are performed for multi-horizons of 24, 48, 72, 120, and 168 h. The DHI CSI performs better than DNI, which in turn performs better than GHI. Additionally, paired features perform better than single features. The combination of the DNI and DHI produces the best forecasting performance with the lowest MAE of 2.45, lowest RMSE of 4.23, and MSE of 18.17. Combining all three features performs worse than pairing two features, suggesting possible overfitting or redundancy in the information.

TABLE II. COMBINATION OF INPUT FEATURES USED FOR THE EXPERIMENT

Scenario	Demand	Future covariate		Past covariate		
		Workday	Hour-group	GHI	DNI	DHI
1	✓	✓	✓	✓		
2	✓	✓	✓		✓	
3	✓	✓	✓			✓
4	✓	✓	✓	✓	✓	
5	✓	✓	✓	✓		✓
6	✓	✓	✓		✓	✓
7	✓	✓	✓	✓	✓	✓

TABLE III. SCENARIO COMPARISON

Input features	MAE		MSE		RMSE	
	Avg	std	Avg	std	Avg	std
GHI	4.39	1.77	55.13	35.49	7.02	2.41
DNI	3.22	0.51	33.64	11.59	5.71	1.00
DHI	2.61	0.26	21.82	5.12	4.64	0.54
(GHI + DNI)	2.62	0.35	20.88	5.28	4.53	0.54
(GHI + DHI)	2.87	0.53	26.47	10.03	5.05	0.97
(DNI + DHI)	2.45	0.19	18.17	3.89	4.23	0.46
(GHI + DNI + DHI)	3.34	0.79	35.76	15.69	5.85	1.23

All three CSIs result in a MAE of 3.34, MSE of 35.76, and RMSE of 5.85, significantly higher than the paired feature combinations. This indicates that the best forecasting performance is achieved when the DNI and DHI CSI are combined, producing the most accurate forecasts. When the CSIs are used as a single feature, the DHI produces the best forecasting accuracy compared to the other two indices, with a MAE of 2.61, MSE of 21.82, and RMSE of 4.64. DHI provides a stronger learning signal for the model. From Figure 3, it is clear that the DHI has a much smoother and symmetric curve, making it easier for the TFT attention mechanism to capture the trend and the temporal patterns, leading to better forecasting accuracy.

C. Performance Evaluation of TFT

The performance of TFT is compared to other benchmark models, as illustrated in Table IV. TFT is compared to two statistical models (ARIMA and Prophet), two ML models (XGBoost and CatBoost), and a DL model (TiDE).

TFT outperformed all models in the MAE and MSE scores for all forecast horizons except for the 120 and 72-h categories regarding MSE, where CatBoost outperformed TFT. ARIMA and Prophet performed the least, as these models underuse or ignore added solar irradiance features and fail to recognize the long-term patterns due to their limited memory. Additionally, statistical models cannot handle exogenous nonlinear time series data. The CatBoost and XGBoost models have stronger performance in the short term horizon as they benefit from the added DNI and DHI features. However, for the longer time horizon their performance diminishes as these models are not designed to handle long-term time dependency. Finally, the TiDE model outperformed the rest in MAE, except the TFT. TiDE underperformed due to the need for additional tuning regarding long-term dependencies.

TABLE IV. TFT COMPARISON TO BENCHMARK MODELS

Metric	Statistical		Regression		DL	
	ARIMA	Prophet	XGBoost	CatBoost	TiDE	TFT
24 h						
MAE	5.48	4.04	1.40	1.55	1.26	1.06
MSE	50.80	22.23	3.71	3.76	2.65	2.04
48 h						
MAE	4.90	4.01	1.84	1.77	2.81	1.35
MSE	45.69	21.75	6.97	5.50	21.45	4.67
72 h						
MAE	5.15	3.82	1.95	1.89	3.16	1.86
MSE	42.67	22.67	7.90	6.53	24.94	8.20
120 h						
MAE	4.90	3.47	2.11	2.14	2.91	2.06
MSE	39.53	20.60	8.96	8.20	19.94	9.71
168 h						
MAE	4.62	3.83	1.92	1.99	2.31	1.56
MSE	37.50	22.91	7.70	7.32	13.89	5.86

The experiment indicates that DHI and DNI strongly correlate with the prediction of energy demand. This can be attributed to the relationship of DNI to the actual sunlight. DHI captures the cloud cover and directly influences the energy demand. GHI has a lower correlation since it constitutes the combination of DNI and DHI, which makes the data redundant for the forecasting models.

D. Forecasting Model Performance on all Forecast Horizons

According to Figure 4a, regarding the 24-h forecast horizon, the TFT model outperforms Prophet and XGBoost. TFT's line is the closest to the actual target value, followed by XGBoost and Prophet. The TFT model consistently outperforms the rest of the models for all forecast horizons.

E. Impact of Solar Irradiance Components on Electricity Consumption

Alongside forecasting performance, this study investigates the relationship between the solar irradiance components and electricity consumption behavior. The results suggest that the CSI of DNI and DHI are closely associated with the electricity demand in contrast to GHI. Whereas previous studies used general weather features, such as temperature or total solar radiation, this study leverages CSI to separate the effects of individual solar components [10, 12]. This approach offers an insight into the effects of the solar irradiance variation on the short-term electricity demand and supports more effective feature selection in forecasting.

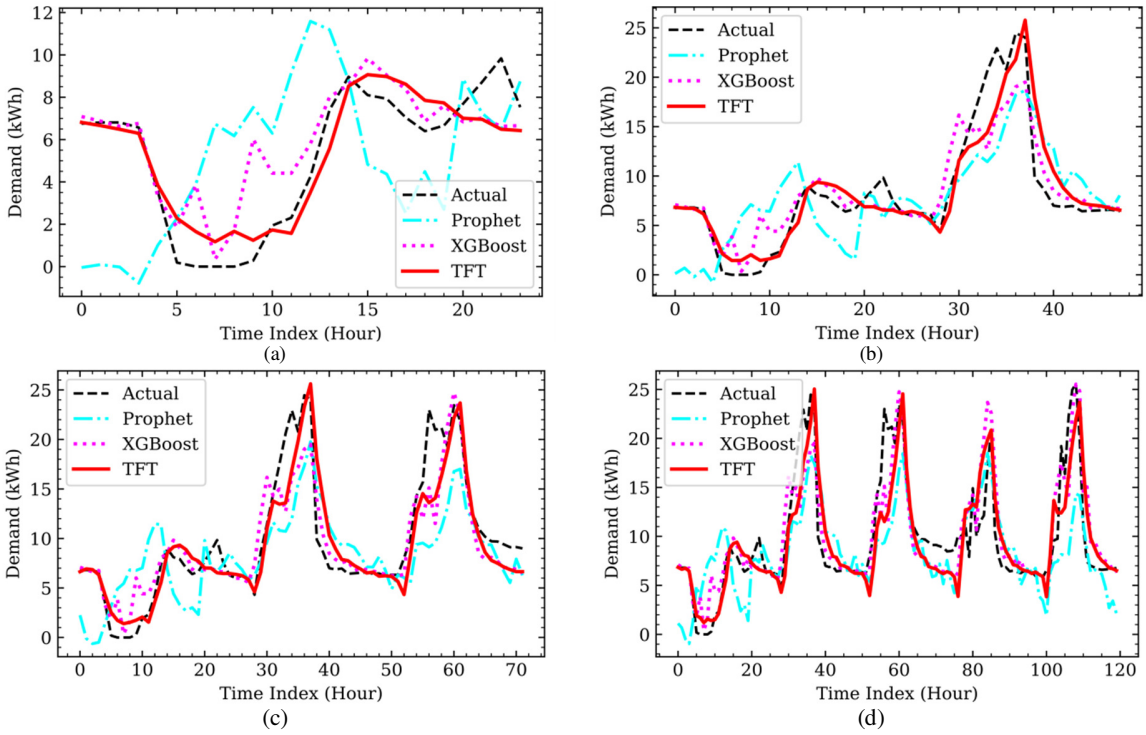

Fig. 4. Comparison between actual electricity demand and predictions from Prophet, XGBoost, and TFT across forecast horizons of (a) 24, (b) 48, (c) 72, and (d) 120 h.

IV. CONCLUSIONS

This study investigated the role of the solar irradiance components Direct Normal Irradiance (DNI), Global Horizontal Irradiance (GHI), and Diffuse Horizontal Irradiance (DHI) in clear sky conditions for short-term demand forecasting regarding the NDIR prosumer building. The implementation of seven separate scenarios of the collected features to the TFT model demonstrated that the combination of DNI and DHI Clear Sky Indices (CSI) significantly improves the forecasting accuracy of the electricity demand. In contrast, combining GHI, DNI, and DHI resulted in diminishing returns due to the information redundancy caused by the relationship between GHI with DNI and DHI.

The performance of TFT was compared against other benchmark models. TFT outperformed all benchmark models in MAE and MSE scores for all forecast horizons, as TFT effectively leverages the solar irradiance features and captures the long-term dependencies. Statistical models, performed poorly due to their inability to handle nonlinear exogenous variables. The ML models show competitive short-term results but struggle with long-term forecasting due to limited long-term time dependency modeling. Finally, this experiment concludes that the careful selection of solar irradiance features can enhance the accuracy of forecasting models, particularly for electricity demand forecasting tasks.

ACKNOWLEDGEMENTS

This research is part of a PhD dissertation conducted at Naresuan University, Thailand, under a scholarship provided by His Majesty the King of Bhutan in collaboration with Naresuan University. The Global and Frontier Research University Fund, Naresuan University, under Grant R2567C002, partly supported this work. The authors thank Mr. Roy I. Morien of the Naresuan University Graduate School for editing this manuscript's grammar, syntax, and general English expression.

REFERENCES

[1] N. Mounir, H. Ouadi, and I. Jrhilifa, "Short-term electric load forecasting using an EMD-BI-LSTM approach for smart grid energy management system," *Energy and Buildings*, vol. 288, Jun. 2023, Art. no. 113022, https://doi.org/10.1016/j.enbuild.2023.113022.

[2] M. G. Pinheiro, S. C. Madeira, and A. P. Francisco, "Short-term electricity load forecasting—A systematic approach from system level to secondary substations," *Applied Energy*, vol. 332, Feb. 2023, https://doi.org/10.1016/j.apenergy.2022.120493.

[3] Z. Ni, C. Zhang, M. Karlsson, and S. Gong, "A study of deep learning-based multi-horizon building energy forecasting," *Energy and Buildings*, vol. 303, Jan. 2024, Art. no. 113810, https://doi.org/10.1016/j.enbuild.2023.113810.

[4] A.-R. Kim, D. Park, C.-S. Lee, and H. Chang, "A Comparative Analysis of Time Series Forecasting Methods for Short-Term Electricity Demand Prediction," in *Proccedings of 14th International Conference on Information and Communication Technology Convergence*, Jeju Island, South Korea, Jul. 2023, pp. 1235–1238, https://doi.org/10.1109/ICTC58733.2023.10392314.

[5] L. Yue, R. Zhang, J. Ding, and Q. Liu, "Real-Time Statistical Weather Estimation and Prediction for Tropical Cyclone Intensity in an Interpretable Manner via Causal Inference," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–11, Aug. 2024, https://doi.org/10.1109/TGRS.2024.3451725.

[6] N. Klungsida, P. Maneechot, N. Butploy, and K. Khiewwan, "Forecasting Energy Consumption from EV Station Charging Using RNN, LSTM and GRU Neural Network," *Journal of Renewable Energy and Smart Grid Technology*, vol. 19, no. 1, pp. 1–6, Apr. 2024, https://doi.org/10.69650/rast.2024.254636.

[7] B. Lim and S. Zohren, "Time-series forecasting with deep learning: a survey," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, Feb. 2021, Art. no. 20200209, https://doi.org/10.1098/rsta.2020.0209.

[8] N. E. Benti, M. D. Chaka, and A. G. Semie, "Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects," *Sustainability*, vol. 15, no. 9, Apr. 2023, Art. no. 7087, https://doi.org/10.3390/su15097087.

[9] S. Chen, Y. Ren, D. Friedrich, Z. Yu, and J. Yu, "Prediction of office building electricity demand using artificial neural network by splitting the time horizon for different occupancy rates," *Energy and AI*, vol. 5, Sep. 2021, Art. no. 100093, https://doi.org/10.1016/j.egyai.2021.100093.

[10] S. Jittanon, Y. Mensin, N. Ketjoy, and C. Termritthikun, "Net Metering Prediction in Prosumer Building With Temporal Fusion Transformer Models," *IEEE*, vol. 12, pp. 109457–109469, Aug. 2024, https://doi.org/10.1109/ACCESS.2024.3438945.

[11] Z. Mustaffa and M. H. Sulaiman, "A Hybrid Prediction Model for Short-Term Load Forecasting in Power Systems," *ECTI Transactions on Computer and Information Technology*, vol. 18, no. 4, pp. 568–578, Oct. 2024, https://doi.org/10.37936/ecti-cit.2024184.257667.

[12] N. Tata, S. S. Machiraju, V. Akshay, D. M. Menon, N. B. Sai Shibu, and D. Arjun, "Prediction of Energy Consumption Using Statistical and Machine Learning Methods and Analyzing the Significance of Climate and Holidays in the Demand Prediction," in *Advances in Computing and Network Communications*, Singapore, Singapore, Jun. 2021, pp. 117–126, https://doi.org/10.1007/978-981-33-6987-0_10.

[13] S. A. Haider, M. Sajid, H. Sajid, E. Uddin, and Y. Ayaz, "Deep learning and statistical methods for short- and long-term solar irradiance forecasting for Islamabad," *Renewable Energy*, vol. 198, pp. 51–60, Oct. 2022, https://doi.org/10.1016/j.renene.2022.07.136.

[14] N. El-Amarty, M. Marzouq, H. El Fadili, S. D. Bennani, and A. Ruano, "A comprehensive review of solar irradiation estimation and forecasting using artificial neural networks: data, models and trends," *Environmental Science and Pollution Research*, vol. 30, no. 3, pp. 5407–5439, Nov. 2022, https://doi.org/10.1007/s11356-022-24240-w.

[15] D. S. Kumar, G. M. Yagli, M. Kashyap, and D. Srinivasan, "Solar irradiance resource and forecasting: a comprehensive review," *IET Renewable Power Generation*, vol. 14, no. 10, pp. 1641–1656, Jul. 2020, https://doi.org/10.1049/iet-rpg.2019.1227.

[16] R. Wattan, I. Masiri, S. Buntoung, C. Phoemwong, and S. Janjai, "An empirical model for estimating the monthly average daily global solar radiation from ground- and satellite-based meteorological data," *Journal of Renewable Energy and Smart Grid Technology*, vol. 18, no. 1, pp. 29–35, Jun. 2023.

[17] K. Dorji, S. Jittanon, P. Thanarak, P. Mensin, and C. Termritthikun, "Electricity Load Forecasting using Hybrid Datasets with Linear Interpolation and Synthetic Data," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 17931–17938, Dec. 2024, https://doi.org/10.48084/etasr.8577.