Accepted in 56th North American Power Symposium (NAPS 2024), Texas, USA

Multivariate Optimal Hybrid Deep Learning Model for Forecasting of Day-Ahead Solar Irradiance with Meteorological Constraints

Sohag Kumar Saha† (Graduate Student Member, IEEE) and Satish M. Mahajan‡ (Life Senior Member, IEEE)
†*Dept. of Electrical & Computer Engineering (ECE), Tennessee Technological University*
‡*Center for Energy Systems Research (CESR), Tennessee Technological University*
Cookeville, TN, USA
Email: {ssaha42, smahajan}@tntech.edu

***Abstract*—With the growing integration of solar power generation into smart grids, accurate solar irradiance forecasting is of paramount importance for efficient grid operation and renewable energy management. In this paper, a data decomposition approach with two different methods of deep learning and grid search optimization was combined to develop a better hybrid model to forecast solar irradiance. The proposed model combines spatial and temporal information to improve forecasting accuracy by considering the impact of meteorological constraints such as global horizontal irradiance, temperature, relative humidity, wind speed, cloud type, direct normal irradiance, diffuse horizontal irradiance, and solar zenith angle. In addition to the model architecture, this work incorporates hyper-parameter optimization to fine-tune the parameters of the model for optimal performance. The design of the model ensures that the system adapts to the specific characteristics of the solar irradiance data and meteorological conditions under consideration. The proposed hybrid model was evaluated using real-world data to outperform traditional forecasting methods in terms of accuracy. The results of the proposed hybrid model indicate better prediction ability as measured by four parameters (lower RMSE and MAE, fewer epochs, and a higher R^2 co-efficient).**

***Index Terms*—Forecasting Model, Solar Irradiance, Hyper-parameter Optimization, CEEMDAN-CNN-LSTM-Grid Search, and Deep Learning**

I. Introduction

THE rapid advancement of renewable energy technologies, particularly solar power, has led to an increasing integration into the smart grids [1]. Accurate prediction of solar irradiance is pivotal for optimizing the performance of solar energy systems and aiding in efficient energy production and grid management. This transition towards renewable energy sources is driven by the urgent requirements to mitigate climate change and reduce dependence on fossil fuels. Accurate forecasting requires a comprehensive model that can capture both the spatial and temporal dependencies of these factors. Traditional forecasting methods often fall short in this regard, leading to inaccuracies and inefficiencies. A number of different approaches, such as numerical weather prediction models, statistical models, machine learning models, image-based models, and hybrid models, have been extensively addressed in [2]. There are five different ways to predict solar irradiance mentioned in [3], such as persistence, physical, classical statistics, machine learning, and hybrid approaches. The persistence technique is typically used to assess the performance of a model by comparing the predicted outcomes with it. Physical models utilize meteorological variables to formulate conservation equations, facilitating the prediction of future atmospheric states and weather phenomena. However, the significant computational cost makes it less suitable for short-term forecasting [4]. Each method has its limitations and offers unique benefits, accuracy, and computational efficiencies compared to the others.

Several time series prediction techniques based on classical statistical methods, including Auto-regressive Moving Average (ARMA), Auto-regressive Integrated Moving Average (ARIMA), Lasso and Markov models [5] have been employed to develop prediction models based on the stochastic nature of solar irradiance. Nevertheless, the forecast accuracy of these time series schemes was low due to the non-stationary nature of the solar irradiance time series data. In addition, the accuracy of the prediction of Global Horizontal Irradiance (GHI) is influenced by a range of meteorological parameters, including temperature, relative humidity, wind speed, and cloud cover [6]. To mitigate the shortcomings of those nonlinear prediction approaches, machine learning-based systems could be implemented for better prediction accuracy of solar irradiance. Deep learning, a subset of machine learning, utilizes algorithms inspired by the structure and function of the brain's neural networks. Various deep learning models, including Convolutional Neural Networks (CNNs), Feedforward Neural Networks (FFNNs), Recurrent Neural Networks (RNNs) like Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU), and Attention-based Transformers, have been successfully employed for predicting irradiance [7].

In [8], a hybrid model consisting of CNN and LSTM is proposed to forecast PV power, compared to individual models of CNN and LSTM, and performed better in terms of precision. The work in [9] showed that the CNN-LSTM model with a sine-cosine algorithm (SCA) outperforms CNN-LSTM models without SCA. A hybrid model including Variational Mode Decomposition (VMD)-CNN-LSTM-Multilayer Perceptron (MLP) in [10] was proposed for the hourly and step-by-step prediction of solar irradiance and showed better interference resistance and accuracy. The work in [3] introduced the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)–CNN–LSTM based solar irradiance forecasting model, which outperforms several other models and approaches. In [11] authors proposed a system using an Improved-CEEMDAN and a Deep Residual Network (DRESNET) with Bidirectional LSTM. An attention-based framework in [12], was developed for multivariate time series forecasting using the Temporal Fusion Transformer (TFT) to capture long-distance dependencies to enhance performance. Furthermore, optimization of hyperparameters in the deep neural network is essential, as the performance of the model is strongly dependent on the architecture of the system [13]. Based on prior research in solar irradiance prediction, it can be inferred that the majority of studies have manually developed deep learning frameworks through a process of trial and error. This procedure is computationally intensive due to its high computational complexity, resulting in extended execution times for a manually tailored architecture with non-optimal results [14].

Therefore, in this study, we present a hybrid model along with a data decomposition approach and a deep learning architecture that can produce accurate forecasting results of solar irradiance without requiring manual tuning of the deep learning network. The main contributions of the proposed hybrid approach for forecasting of global horizontal solar irradiance are outlined as follows:

- Introduction of a new hybrid learning model that seamlessly integrates the strengths of CEEMDAN, CNN, LSTM, and grid search optimization to forecast GHI.
- The proposed hybrid model enhances forecasting accuracy with the integration of spatial & temporal information. By considering weather constraints, the model accurately represents the complex interaction of these factors, resulting in more reliable forecasts.
- The proposed hybrid model was empirically evaluated using real-world data from Crossville, TN, USA, which was fetched from the solar radiation database of the National Renewable Energy Laboratory (NREL).
- The result shows better performance compared to traditional forecasting methods in terms of statistical parameters (RMSE, MAE, R^2), least prediction errors, and fewer epochs.

The remainder of the paper is organized as follows. Section II describes the problem statement and motivation for this research. Section III provides the dataset preparation, data information, location, and data pre-processing. Section IV provides the proposed methodology and a detailed description of the proposed CEEMDAN–CNN–LSTM-Grid Search optimization model. Section V presents the evaluation of the model using real-world data based on various case studies. Finally, Section VI concludes the paper and addresses potential future work opportunities.

II. Problem Statement and Motivation

Leveraging historical solar irradiance data, weather parameters, and relevant geographical features, our proposed model aims to provide precise short-term and long-term forecasts. The problem that is addressed in this paper is a multi-variate time series-based solar irradiance prediction issue which is a challenging task. The term ”multivariate time series” implies that the prediction model takes into account multiple variables and meteorological factors that influence solar irradiance. The forecasting performance of irradiance is mainly dependent on GHI, along with several important weather parameters [15]. The input of multiple weather parameters from the past is used for solving this multivariate time series problem within a window function of finite look-back of temporal information. The study involves rigorous feature selection, model training, and validation to identify the most effective algorithms for capturing the complex non-linear relationships inherent in solar irradiance dynamics. The envisioned outcome is a robust forecasting framework that significantly improves the reliability and efficiency of solar energy systems, contributing to the sustainable integration of solar photovoltaic (PV) power into the broader energy landscape.

III. Dataset Description

The performance of the deep learning-based model heavily depends on a reliable dataset. In this work, a dataset consisting of three years of solar GHI with meteorological parameters-including temperature, wind speed, relative humidity, Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI), clearsky DNI, and cloud type from Crossville, Tennessee, USA, was used. This dataset was originally captured by NREL, which has developed a solar resource maps in conjunction with different climate data from the United States and other locations around the world [16]. There are different intervals of data available in the database in [17], and to get more precise actual scenarios, the 5-minute interval data with data points of 105,120 per year of 2020–2022, have been selected. The prepared data sets have known temporal features such as minutes, hours, days, months, and years. For seasonal analysis, the data set for each is divided into four seasons; winter (Dec.-Feb.), Spring (Mar.-May), Summer (June-Aug.), and Fall (Sept.-Nov.). After fetching the data from the NREL database, it goes through preprocessing stages, such as checking for missing data, removing any data outliers, normalizing the data, identifying the correlation of the target data with other meteorological information, and splitting the datasets for training, validation, and testing.

The Pearson correlation matrix helps to identify features strongly correlated with the target variable (GHI), helping to reduce the dimensionality by removing features with low correlation to the GHI or high correlation to each other. This matrix provides insights into the linear relationships between input features and GHI, helping to interpret how changes in certain features might impact GHI. By focusing on the most influential features, the matrix simplifies the model, leading to better generalization and faster computation times. The correlation matrix of the dataset is shown in Fig. 1 and each cell in the matrix shows the correlation coefficient between two variables. From this plot, it can be seen that DNI, clearsky DNI, DHI, and temperature have strong positive correlations with GHI forecasting, whereas relative humidity and solar zenith angle have strong negative correlations with GHI, and wind speed and cloud type have very weak correlations with the target variable.

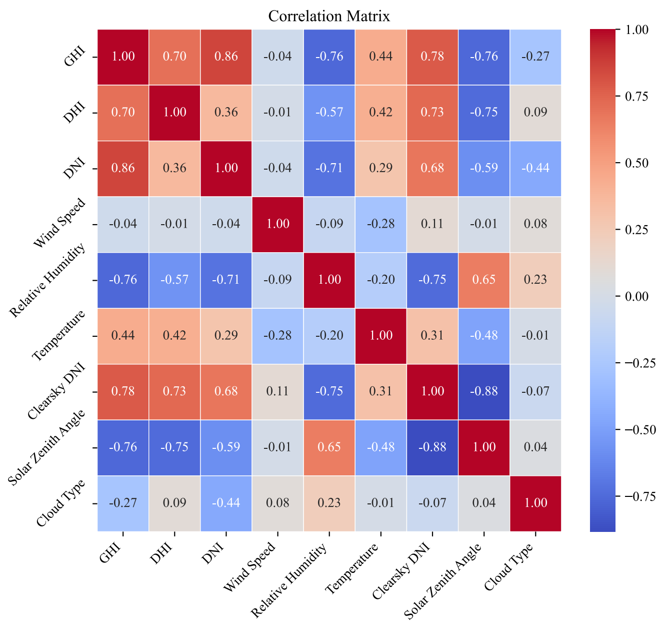

Fig. 1. The Pearson correlation matrix i.e. heatmap to show the forecasting relationship of target data (GHI) with associated meteorological parameters

IV. Research Methodology

The proposed hybrid methodology used for the prediction of solar irradiance in this paper combines the time-series data decomposition or signal processing and deep learning techniques. The proposed methodology for forecasting of GHI combines CEEMDAN to decompose the time series data into Intrinsic Mode Functions (IMFs), capturing both trends and oscillations. The CNN network extracts spatial features, while the LSTM network handles temporal dependencies for sequential data. In addition, the hyperparameter tuning using grid search optimization improves the model performance by selecting the optimal configuration for more accurate predictions. Each of the techniques employed in the proposed model including CEEMDAN, CNN, LSTM, and Grid Search Optimization for hyperparameter tuning is discussed briefly in this section.

A. CEEMDAN

The CEEMDAN is an enhanced version of the original Empirical Mode Decomposition (EMD) and its extension Ensemble Empirical Mode Decomposition (EEMD) and performs well with non-stationary and non-linear time series data [18]. It exhibits superior spectral separation of the modes, reduces the number of iterations, minimizes reconstruction errors, and enhances the decomposition process with reduced computational costs. As in [18] the equations for calculating IMFs and residual can be written as in Eq. (1) to Eq. (3). Consider a raw data $f(t)$ and k-th IMF be $(\overline{IMF}_k)(t)$. $\mathrm{EMD}_j(t)$ denotes the j-th IMF of EMD decomposition. The SNR of each step is denoted as ε_k, which is the standard deviation of Gaussian white noise with standard normal distribution is $\omega^i(t)$. The original time series data, denoted as $f^i(t)$ is augmented by the addition of white noise $\omega^i(t)$ at $i = 1, 2, 3, \ldots, n$.

$$\overline{\mathrm{IMF}}_k(t) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{EMD}_1\left(r_{k-1}(t) + \varepsilon_{k-1}\mathrm{EMD}_{k-1}(\omega^i(t))\right), \quad \text{where } k = 2, 3, \ldots, K \tag{1}$$

$$r_k(t) = r_{k-1}(t) - \overline{\mathrm{IMF}}_k(t) \tag{2}$$

$$f(t) = \sum_{k=1}^{K}(\overline{\mathrm{IMF}}_k(t) + r(t)) \tag{3}$$

The procedure of this algorithm ends when the residue has two extreme points and cannot be decomposed further. The final residue is $r(t)$, and the actual information $f(t)$ correlates as shown in Eq. (3), where K is the number of total modes.

B. CNN

Convolutional Neural Networks (CNNs) are suitable for forecasting of time series irradiance because of their inherent ability to detect spatial features in data [19]. Solar irradiance data often exhibits spatial dependencies, as the distribution of sunlight across geographical locations can significantly impact energy generation. CNNs excel at learning hierarchical representations of spatial features, making them effective in recognizing complex patterns in solar irradiance maps or images. Using convolutional layers, CNNs automatically learn relevant features, which are important for accurate irradiance predictions. Multiple-level data processing and dimensionality reduction are the foundation of CNNs, which turn input data into meaningful features. Furthermore, the shared weight characteristics of convolutional and pooling layers reduce the number of parameters in the CNN model, thereby mitigating training challenges and the risk of overfitting.

C. LSTM

The LSTM is a deep neural network, well-suited for forecasting of solar irradiance because of its ability to detect model temporal features in time series data [8]. Solar irradiance exhibits dynamic patterns that vary over time, influenced by factors such as the time of day, season, and weather conditions. LSTM learns patterns in the historical time series of irradiance, allowing them to make predictions that account for the inherent temporal dynamics. This makes LSTMs particularly valuable for predicting changes in solar irradiance throughout the day and across different seasons, enhancing the accuracy of forecasts for renewable energy applications. The ability of LSTMs to model complex temporal dependencies makes them a powerful tool in the context of solar energy prediction.

D. Grid Search for Hyperparameter Optimization

The grid search is a hyperparameter optimization algorithm used to systematically explore a predefined hyperparameter space for deep learning models. The procedure involves defining a grid of hyperparameter values, creating all possible combinations, and training and evaluating the model for each combination. The algorithm exhaustively searches the hyperparameter space, assessing model performance using a chosen metric. The optimal set of hyperparameters is determined based on the best performing configuration. The grid search is straightforward to implement and provides a systematic approach to finding the best hyperparameters for a given model [13]. The proposed model is trained using the identified optimal hyperparameters.

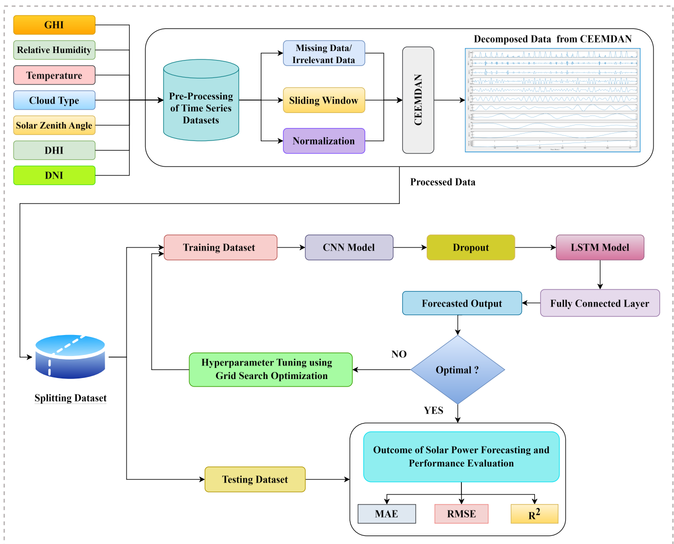

Fig. 2. The Block Diagram of the Proposed Hybrid Deep Learning Model

E. Proposed Hybrid Model

The objective of this work is to predict solar irradiance with higher accuracy with the least forecast errors. The complete block diagram of the proposed model is shown in Fig. 2. The proposed system presents a comprehensive approach that integrates diverse input parameters, including solar irradiance, temperature, relative humidity, cloud type, DHI, DNI and solar zenith angle. The system leverages a sophisticated data preprocessor to enhance the quality of the input features before subjecting them to the CEEMDAN algorithm. This multi-stage decomposition process facilitates the extraction of temporal and spectral characteristics, effectively capturing the nuanced patterns inherent in solar irradiance data. The subsequent integration of these features into CNN and LSTM models allows the exploitation of spatial and temporal dependencies, respectively. The adoption of a grid search algorithm for hyperparameter optimization is such that the models are fine-tuned to their optimal configurations. The iterative nature of the system involves continuous testing against validation data and dynamic adjustment of hyperparameters.

V. Results and Discussion

The proposed model was designed using Python programming and the used libraries are PyEMD, Tensorflow, Keras, Scikit-learn, Numpy, Pandas, and Matplotlib. The simulation was performed on a Lenovo Legion 7i machine with the Intel Core i7, 2.3 GHz, 11800H processor, 32 GB DDR4 RAM & Nvidia GeForce RTX 3060 graphics. The dataset was decomposed using the CEEMDAN algorithm by M. Torres [18], in which the decomposed data is shown in Fig. 3.

After data decomposition, the model was trained to learn the time series data sequence with the window function and the CNN-LSTM-based deep learning network. The hyperparameter of the learning model was optimized using the grid search method to achieve better performance and the least errors in the prediction values. The model was trained with 2 years of multivariate data and then validation and testing are performed on a seasonal basis. There are four different seasons in the selected location of Crossville, TN, and each season shows the forecasting performance of the model. Although the forecasting model computed the predicted irradiance for whole years due to space limitations, the selected three days of predicted versus actual data are shown and evaluated the performance of the designed system. In addition, only two seasons of forecasted data are presented in this paper, although all four seasons of predicted data were evaluated.

A. Forecasting Performance in Summer

The selected three-day summer horizon was June 16th to June 18th to compare the actual and predicted results, which is shown in Fig. 4. In the summer, the solar irradiance is generally highest as compared to other seasons and it is important to predict the solar irradiance with greater accuracy. It is found that, as there are a lot of peaks in the actual solar irradiance in summer, the forecasting model is trying to reach the peak and there is a little deviation but still performed better as compared to other standard algorithms such as CNN, GRU, and LSTM models.

B. Forecasting Performance in Winter

The model predicted the irradiance for three days (December 23rd to December 25th) in the winter season with good accuracy. The forecasted output of the proposed method is compared with popular forecasting approaches such as: CNN, LSTM, GRU and outperformed those approaches, as shown in Fig. 5.

C. Performance Parameters

The performance of the forecasting problem is determined by errors in prediction and measured by the calculation of the prediction errors of the model as compared with the ground truth time series data. When forecasting models are evaluated, particularly those using hybrid algorithms, several performance parameters are commonly used to assess their effectiveness. Some key performance metrics are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and Mean Squared Error (MSE) [20].

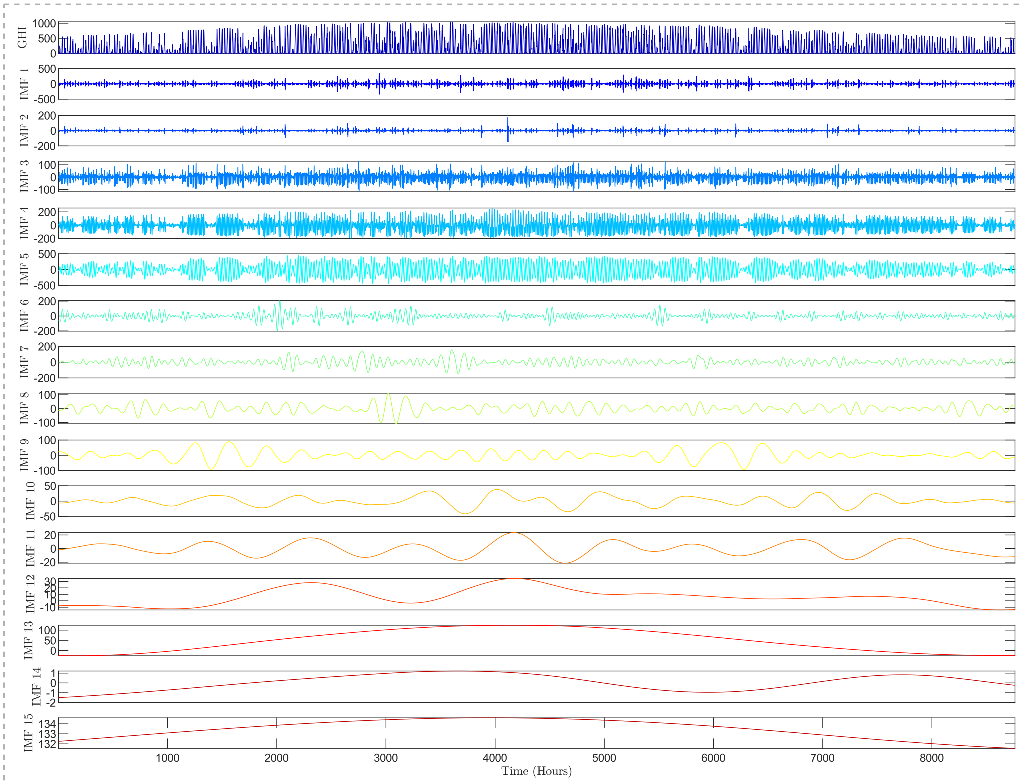

Fig. 3. Time Series Data Decomposition into IMFs using the CEEMDAN algorithm, whereas IMF1 has the highest, and IMF15 has the lowest frequency

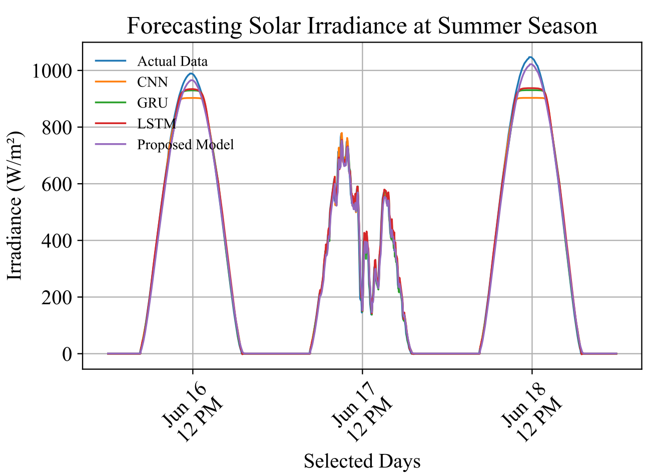

Fig. 4. Forecasting analysis of solar irradiance in the Summer season

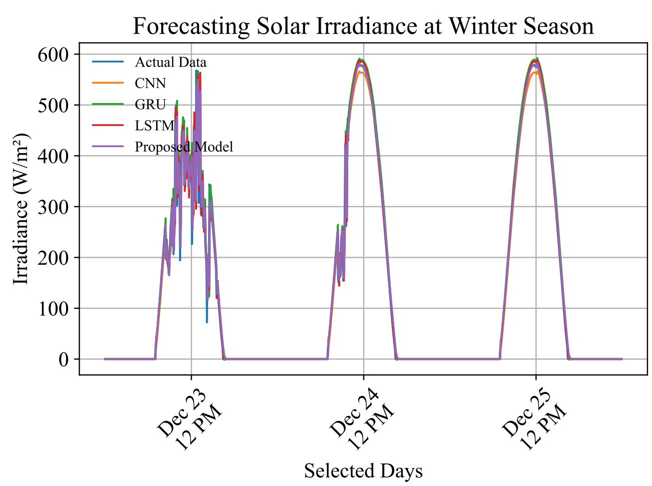

Fig. 5. Forecasting analysis of solar irradiance in the Winter season

MAE is the average absolute difference between the actual and predicted values. It provides a measure of the average magnitude of errors without considering their direction. MSE calculates the average of the squared differences between actual and predicted values. RMSE is the square root of the MSE. R^2 measures the proportion of variance in the dependent variable that is predictable from the independent variable. A higher R^2 indicates a better fit. The performance of the model was measured using RMSE, MAE, and R^2. The lowest value of RMSE and MAE is preferred for achieving better forecasting results along with high R^2 value. The unit of RMSE and MAE is W/m^2 and R^2 is dimensionless. The residual error can be written as: $\varepsilon_i = \text{GHI}_{\text{ground truth},i} - \text{GHI}_{\text{predicted},i}$, which implies that, $\varepsilon_i = y_i - \hat{y}_i$, where y_i, $\hat{y}_i$ and $\bar{y}_i$ are denoted as actual value, predicted value, and the mean value of the data points, respectively. The mathematical formulas as in [20] for the calculation of RMSE, MAE, and R^2 are given in Eq. (4) to Eq. (6) respectively, where N is the number of data points.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i^2} \tag{4}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\varepsilon_i| \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{6}$$

The relative performance of the proposed model and that of other popular algorithms (CNN, LSTM, and GRU), compared to the predicted output with the actual data from NREL is presented in Table I. It is evident that the proposed system performed better, exhibited the least errors during prediction, and a comparatively higher R-square value for better fit. Despite being computationally complex, the proposed model can handle time-series data complexity, resulting in better feature extraction, faster convergence with fewer epochs, and better predictive accuracy than the other model compared with.

TABLE I
COMPARISON OF PERFORMANCE OF THE PROPOSED MODEL

Parameters	CNN	LSTM	GRU	Proposed
RMSE	43.1879	38.2384	43.2561	37.1773
MAE	14.3241	14.2120	14.4772	14.0553
R^2	0.9741	0.9748	0.9742	0.9837
Epochs	20	30	25	15

VI. CONCLUSIONS

This work proposed a multivariate hybrid system to solve the solar irradiance forecasting problem and developed the model utilizing the CEEMDAN, CNN, LSTM, and Grid-Search-based hyperparameter optimization algorithms. The use of multiple weather constraints in the dataset helped predict accurately, which is essential for load balancing and stability in smart grid systems. Through rigorous case studies in different seasons throughout the year, the proposed model exhibits better performance based on forecasting accuracy and the least errors in prediction. The performance of the proposed model was evaluated via statistical parameters (RMSE, MAE, and R^2) and the model exhibited better prediction, fewer errors, and lower epochs, resulting in faster convergence as compared to the other popular forecasting techniques. In the forecasting problem, there is always room for improvement in the future by considering much larger training datasets and a more robust learning model.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support provided by the Center for Energy Systems Research (CESR), Tennessee Technological University, Cookeville, TN, USA. This work utilized the National Solar Radiation Database (NSRDB) which was developed by the National Renewable Energy Laboratory (NREL).

REFERENCES

[1] T. Ding, W. Jia, M. Shahidehpour, O. Han, Y. Sun, and Z. Zhang, "Review of optimization methods for energy hub planning, operation, trading, and control," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 3, pp. 1802–1818, 2022.

[2] E. B. Ssekulima, M. B. Anwar, A. Al Hinai, and M. S. El Moursi, "Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review," *IET Renewable Power Generation*, vol. 10, no. 7, pp. 885–989, 2016.

[3] B. Gao, X. Huang, J. Shi, Y. Tai, and J. Zhang, "Hourly forecasting of solar irradiance based on ceemdan and multi-strategy cnn-lstm neural networks," *Renewable Energy*, vol. 162, pp. 1665–1683, 2020.

[4] X. Mi, H. Liu, and Y. Li, "Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine," *Energy conversion and management*, vol. 180, pp. 196–205, 2019.

[5] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod, and A. O. Salau, "Machine learning based solar photovoltaic power forecasting: a review and comparison," *IEEE Access*, vol. 11, pp. 40 820–40 845, 2023.

[6] K. Benmouiza and A. Cheknane, "Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models," *Energy Conversion and Management*, vol. 75, pp. 561–569, 2013.

[7] M. J. Sammar, M. A. Saeed, S. M. Mohsin, S. M. A. Akber, R. Bukhsh, M. Abazeed, and M. Ali, "Illuminating the future: A comprehensive review of ai-based solar irradiance prediction models," *IEEE Access*, 2024.

[8] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Applied Energy*, vol. 251, p. 113315, 2019.

[9] S. M. J. Jalali, S. Ahmadian, A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "Automated deep cnn-lstm architecture design for solar irradiance forecasting," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 54–65, 2021.

[10] J. Liu, X. Huang, Q. Li, Z. Chen, G. Liu, and Y. Tai, "Hourly stepwise forecasting for solar irradiance using integrated hybrid models cnn-lstm-mlp combined with error correction and vmd," *Energy Conversion and Management*, vol. 280, p. 116804, 2023.

[11] S. Ghimire, R. C. Deo, D. Casillas-Perez, and S. Salcedo-Sanz, "Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction," *Renewable Energy*, vol. 190, pp. 408–424, 2022.

[12] S. Sakib, M. K. Mahadi, S. R. Abir, A.-M. Moon, A. Shafiullah, S. Ali, F. Faisal, and M. M. Nishat, "Attention-based models for multivariate time series forecasting: Multi-step solar irradiation prediction," *Heliyon*, vol. 10, no. 6, 2024.

[13] H. Alibrahim and S. A. Ludwig, "Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization," in *2021 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2021, pp. 1551–1559.

[14] S. M. J. Jalali, S. Ahmadian, A. Khosravi, M. Shafie-khah, S. Nahavandi, and J. P. Catalão, "A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8243–8253, 2021.

[15] K. Y. Bae, H. S. Jang, and D. K. Sung, "Hourly solar irradiance prediction based on support vector machine and its error analysis," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 935–945, 2016.

[16] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The national solar radiation data base (nsrdb)," *Renewable and sustainable energy reviews*, vol. 89, pp. 51–60, 2018.

[17] "NSRDB: National Solar Radiation Database by NREL," accessed on May 15, 2024. [Online]. Available: https://nsrdb.nrel.gov/data-viewer

[18] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2011, pp. 4144–4147.

[19] B. Gao, X. Huang, J. Shi, Y. Tai, and J. Zhang, "Hourly forecasting of solar irradiance based on ceemdan and multi-strategy cnn-lstm neural networks," *Renewable Energy*, vol. 162, pp. 1665–1683, 2020.

[20] Y. Huang, A. Wang, J. Jiao, J. Xie, and H. Chen, "Short-term pv power forecasting based on ceemdan and ensemble deeptcn," *IEEE Transactions on Instrumentation and Measurement*, 2023.