A REPORT ON

AI in renewable energy forecasting

BY

Yashwanth Karuparthi

2021A7PS0136U

Tejas Kambli

2021A7PS0157U

BITS Pilani, Dubai Campus

Dubai International Academic City (DIAC)

Dubai, UAE

BITS Pilani, Dubai Campus Dubai

International Academic City,

Dubai - UAE

Date of Submission: 22.12.2024

Title of the Project: AI in renewable energy forecasting

Name of the student: Yashwanth Karuparthi **IDNo.:**2021A7PS0136U

Discipline of Student: Computer Science(CS)

Name of the Project Faculty: Dr. Shazia Hasan

Key Words: Solar Energy, Renewable Energy, XGBoost, Prophet, Digital Twin

Project Area(s): Solar Irradiance forecasting, Time series forecasting, Hybrid Deep Learning

AI in renewable energy forecasting

Yashwanth Karuparthi
Computer Science and Engineering, Birla Institute of Technology and Sciences Pilani, Dubai Campus, UAE

Tejas Kambli
Electrical Electronics and Engineering, Birla Institute of Technology and Sciences Pilani, Dubai Campus, UAE

Corresponding author: shazia.hasan@dubai.bits-pilani.ac.in

Abstract

Accurate prediction of solar irradiance is important as it is directly correlated to Photovoltaic (PV) generation. Having future prediction of energy generation also helps optimize resource allocation and Reduce energy waste as it helps to support a stable and sustainable energy grid. Artificial Intelligence (AI) has been used to tackle the challenges of intrinsic variability of the data that is used for this task. With the increasing demand for Renewable Energy Sources like solar energy, precise forecasting is essential for efficient planning of energy production and consumption. Traditional Deep Learning (DL) forecasting methods often fail to account the complex and nonlinear relationship between multiple influencing factors. While the models that adopt multivariate approach are limited by their complexity and lack of robustness. This study compares and proposes a hybrid model with gradient boosting algorithm and Bidirectional Long-Short Term Memory (BiLSTM) to utilize temporal dependencies in the data along with the information on the complex feature interactions within different weather factors that influence Global Horizontal Irradiance (GHI) to enhance the forecasting accuracy. United Arab Emirates (UAE) based dataset was used in this study that was sourced from the National Renewable Energy Laboratory (NREL). The models were compared with standalone BiLSTM and Gradient Boosting Decision Trees on multiple evaluation metrics. The results when evaluated with different evaluation metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared values show that the hybrid model between CatBoost and BiLSTM achieved highest accuracy among other variants of gradient boosting algorithms. This study highlights the potential of developing hybrid models that are robust as well as efficient in complexity that could be used for improved energy management and sustainability.

I. Introduction

The transition to renewable energy sources is essential for addressing climate change and achieving a sustainable energy future. However, maintaining a steady energy supply is challenging due to the highly variable nature of renewable energy sources like wind and sun. These issues can be effectively addressed by data-driven AI models, which use both historical and current data to produce precise, flexible predictions.Solar irradiance refers to the power received from the Sun in the form of electromagnetic radiation per unit area that it has fallen. GHI, which is the total amount of solar radiation received per unit area on a horizontal surface, is one of the critical factors to predict PV energy generated by the solar cell.

Machine Learning (ML) techniques such as ensemble learning and support vector machines predict the results by utilizing the patterns in the datasets. DL models such as Recurrent Neural Networks (RNNs) and convolutional neural networks (CNNs) are based on these ideas and enable automatic feature extraction and generalizability to non-linear data. These models have been widely used in the energy sector due to their robustness and ability to handle large and heterogeneous datasets to generate accurate forecasts.

However, data-driven models present certain limitations as they can be biased or lack explainability, and need large datasets. Traditional forecasting methods often rely on univariate time series approaches, which fail to account for multifaceted influences of weather conditions and environmental factors on solar irradiance. While advanced models attempt to adapt on multiple variables, they often face the challenge of being computationally complex and lack of robustness.

This study explores the performance of various gradient boosting algorithms when used as a hybrid model with BiLSTM for solar irradiance forecasting by incorporating weather data along with GHI. The integration of these techniques aims to combine the individual strengths of each models to produce computationally efficient solution for solar irradiance prediction.

This paper is structured into 7 sections: Section 2 provides an overview of the related works that utilize hybrid BiLSTM models. Section 3 states the problem statement that this paper is sought to address, while section 4 describes the dataset exploring the features used for this study. We discuss the implementation of the proposed model in Section 5. The results obtained from the implementation are presented and discussed in Section 6 which leads us to Section 7 where we conclude the hypothesis of this research. Finally, we define the future scope for this study in section 8.

II. Literature review

Saxena et al. [1] proposed a hybrid model by integrating K-Nearest Neighbors (KNN) with Support Vector Machine (SVM) to leverage the structural diversity of KNN and data diversity through SVM and their work showed high performance compared to conventional Long-Short Term Memory (LSTM) models, with accuracy of 98% for forecasting PV power generation. The study used one-year dataset from Jodhpur and has hourly average temperature, sunlight duration, Global Solar Irradiance (GSI) and PV energy generation. The model excels in real-time dataset due to non-linear mapping of data to higher-dimensional spaces. The hybrid model showed lower RMSE values of 0.001, which is lower than 0.14 for LSTM model during the training phase. When comparing accuracy, sensitivity, and specificity for specific parameters, the LSTM model nearly surpassed KNN-SVM in overall accuracy and specificity by 2.99% and 3.91%, respectively. However, KNN-SVM shows a higher sensitivity by 2.46% when used with Hourly Total Global Solar Radiation (HTGSR), which could indicate it is more effective in correctly identifying positive cases. However the model remains sensitive to the data preprocessing methods used as it gave higher sensitivity when used with HTGSR compared to using only Global Solar Irradiance.

Optimizing hyper-parameters is essential for increasing machine learning models' efficiency and forecast accuracy, particularly in RNN, meanwhile model parameters are often adjusted using extensive simulations and empirical information, this process may be resource- and time-intensive. Bayesian optimization (BO) uses probabilistic models to direct the search and enable methodical exploration of the hyper-parameter space. The majority of BO approaches like BO with truncated-Newton (BO-TNC) and BO with limit-BFGS-bound (BO-L-BFGS-B) depend on gradient-based methods, which can increase the computational cost. This is addressed by Yaru Li et al. [2] where the BO-PSO (particle swarm optimization within the BO framework) approach was used to parallelizable particle motions instead of gradient computations. The research analyzed the models such as RNN-BO-PSO, the LSTM-BO-PSO, and the previously mentioned models on the normalized power load data with a sampling of 15 minutes where three main parameters were tuned using the algorithms like, Feature length, Number of network units, and Batch size of training data. LSTM-BO-PSO showed the highest R-squared value of 0.9951. Though the research has obtained higher accuracies with the optimization techniques, it did not utilize the models on multiple variables, thus not providing results to prove the model to be robust for real-world applications. Another limitation can be pointed at the utilization of BO as the model can over-exploit the feature space, leading it to get stuck in local optima without fully exploring the other promising regions in the hyperparameter space.

M. Tortora et al. [3] introduced a Multi-Level Fusion Transformer-Based Model that combines the strength of both AI and physics-based methods. Forecasting methods typically fall into physics-based and data-driven approaches, with AI-based models capable to capture complex data patterns. However, these models tend to not capture the underlying principles of

PV generation, which they aim to bridge the gap by leveraging self-attention transformer architecture with a multi-level fusion of historical PV and weather data to produce multivariate and multistep day-ahead forecasts. The authors used the Ausgrid benchmark dataset which includes solar capacity, total energy consumption, postal codes, solar conditions like Direct Normal Irradiance (DNI), GHI, and Diffuse Horizontal Irradiance, along with the weather data like temperature, pressure, humidity, and wind speed. MATNet produces the best accurate results with the lowest RMSE score of 0.0460 and 0.0245 MAE score, compared to LSTM-based MATNet giving 0.0495 and MAE score of 0.0267. However, the study fails to account for outliers in the weather data, since the utilized weather dataset didn't have such conditions to train the model in such scenarios.

Yahia Said et al. [4] proposed a hybrid forecasting model by combining LSTM with an autoencoder to enhance solar energy predictions. The LSTM captures temporal patterns in historical solar energy data, while the autoencoder extracts spatial features. The findings from this study demonstrate that extracting temporal features before spatial features would yield better performance. The data used include the historical time series of active power with weather conditions like temperature, humidity, wind speed, GHI, and DHI. The results show that the sequence-to-sequence LSTM-autoencoder model produced predictions with low MAE values of 0.217 and RMSE values of 0.602, compared with 0.312 MAE score and 0.645 RMSE score shown by sequence-to-sequence LSTM and 0.219 MAE score and 0.613 RMSE score shown by LSTM-autoencoder model. However, the paper's methodology lacks the generalization ability to extract temporal features before spatial features when geographical context like weather, latitude, and longitude are used in the data.

Shahram Hanifi et al [5] introduced a hybrid model between LSTM and CNN by combining with wavelet packet decomposition (WPD) to use in wind power forecasting. The data used include wind speed, and power generation along with the configuration of the wind turbine blades. WPD is utilized to decompose wind power data into various frequency sublayers. It works by applying both low-pass and high-pass filters to decompose a signal into low and high-frequency components to allow for a multi-level analysis of the frequency of the signal's components. The authors employed Sequential Model-Based Optimization (SMBO) with the Tree Parzen Estimator (TPE) for hyperparameter tuning, enhancing the efficiency of the model. The optimized LSTM captures both long-term and short-term dependencies in low-frequency sub-layers, while the CNN focuses on high-frequency sub-layers characterized by short-term dependencies. However, this research is limited to the use of non-geographical features. The proposed model has resulted in high accurate results considering the mentioned data with the respective MSE, RMSE, and R-square values being 0.633, 0.79, and 0.99.

Politecnico di Milano et al [6] proposed a Bayesian Hyperparameter Optimization (BHO) approach with a Gaussian process on the stacked BiLSTM model. BiLSTM captures dependencies from both past and future contexts in the sequence by processing sequence data in both forward and backward directions. By combining two LSTM layers, where one reads the sequence from start to end and another from end to start, BiLSTM can better understand complex temporal relationships. In the study, the method was validated using two public datasets with different time granularities, highlighting both accuracy and computational efficiency improvements over other state-of-the-art methods. Their analysis revealed that decreasing time granularity reduced model creation time without sacrificing accuracy, which was quantified using Floating-point Operations (FLOPs) as a metric for computational volume. The model gave the respective MAE% and RMSE% values as 1.306 and 1.696. However, the model showed a tradeoff between accuracy and computational load, emphasizing the importance of careful parameter selection for optimal performance. BiLSTM model also exhibits limitations when used with multivariate data since it is developed to capture the patterns based on historical sequences, but they struggle to model the relationships among the different variables effectively.

Dahmani et al. [7] investigated Single Neural Networks (SNN) and Bootstrap Aggregated Neural Networks (BANN) to predict hourly global radiation. The study used a dataset with 3,606 data points collected from Bouzareah, Algeria, which has weather conditions similar to that of Dubai. The features used include temporal features, mean temperature, daylight duration, wind speed, wind direction, relative humidity, rainfall, and daily worldwide solar radiation. To enhance model precision and robustness, they applied bootstrap aggregation, which generated multiple training datasets through bootstrap resampling with replacement. This approach helped mitigate limitations caused by an incomplete training dataset, allowing each SNN model to learn from diverse samples. The BANN was constructed by averaging the outputs of individual neural networks (INNs), creating a more resilient ensemble. The ANN is architectured using varying number of neurons ranging from three to twenty-five. Four activation functions were tested in the hidden layers like logistic sigmoid, hyperbolic tangent, sine, and exponential, with a pure-linear function for the output layer. Results showed that BANN achieved a lower RMSE of 62.49 Wh/m^2 compared to the SNN's 68.40 Wh/m^2, demonstrating greater accuracy and robustness in solar radiation prediction. SNN, INN and BNN showed respective R-squared values of 0.9620, 0.9618, and 0.9680, showing that BANN can accurately predict the diffused solar radiation using weather conditions that are similar to that of Dubai. However, the observations from the fit curves show that the model may suffer from systematic bias and it could consistently overestimate or underestimate the target values.

A. Verdone et al. [8] explored the application of Spatio-Temporal Graph Neural Networks (STGNNs) with 1D CNN for energy forecasting, leveraging their ability to integrate topological data related to the spatial distribution of energy plants and temporal data from time series. The study utilized synthetically produced time series data having features such as energy production data along with wind speed and temperature data, where the data was sampled for 1 hour. Each time step was normalized to stay in the interval between 0 and 1 given the minimum and maximum values of each time series. The proposed Graph Convolutional Network + 1D CNN (GCN1D) model avoid the vanishing gradient problem which is associated with RNN by utilizing a Graph Convolutional Network (GCN), which allows for the integration of both topological data, defined by the distribution of the plants in the territory, and temporal data of the time series. The model outperformed LSTM and GCN-LSTM model with an r-squared, MSE, MAE values of 0.984, 105, and 1245 respectively. The LSTM model gave the r-squared, MSE, MAE values of 0.952, 311, and 2430 respectively. While GCN-LSTM gave 0.975, 162, and 1965 for r-squared, MSE, and MAE values respectively. The potential drawback in this model can be pointed at the absence of dropout layer, due to which the model may become overly reliant on specifi neurons or layers, leading to overfitting.

Vartholomaios, A. et al. [9] employed a hybrid model combining Extra Trees (ExT) and Prophet model to predict one step ahead solar and wind energy forecasting. The study used biseasonal features, temporal features such as sunrise, morning, noon, sunset and night times along with solar + wind energy source generation data. The study also experimented with features obtained from feature engineering where rolling window features were extracted from the energy generation data, based on key statistics like mean, minimum, maximum, standard deviation, and variance. 48 and 72 hours ahead energy values were also considered to train the model. The modeling process begins by first adding custom seasonalities and regressor to the Prophet model and then fitting the model to the training data, where the seasonal component is calculated. The seasonal component is subtracted and the deseasoned series is used as input to the ExT model, where the final forecast is produced by the residual forecasts along with the extrapolated seasonal patterns from the Prophet model. The proposed hybrid model with feature engineering performed showed the best MAE and RMSE values of 0.041 and 0.067 respectively when used with solar energy data, and 0.069 and 0.088 MAE and RMSE values when used with wind energy data, compared to only using the Prophet model which gave 0.055 and 0.08 MAE and RMSE values with solar generation data and 0.083 and 0.104 MAE and RMSE values with wind energy data. By combining the model to get the hybrid model, the resultant model can capture non-linear patterns in residual time series as well as estimate the seasonal and trend components of time series. These capabilities including the ease of use of the model makes the proposed method outperform the competition for time-dependent data having multiple seasonal effects. However, the random splitting threshold used in ExT can make the model depend on greater number of estimators, which can cause less efficient feature extraction.

Wang et al. [10] proposed a novel ensemble prediction model known as the Gradient Boosting Dendritic Network (GBDD) that leverages a gradient-boosting strategy combined with a dendritic network architecture to predict ultra-short-term PV power. The study utilizes the ability of Dendritic Network (DD), which employs Hadamard products to perform multiple logical operations, which proves to be efficient compared to mapping operation functions. The study also claims that DD operates as a white-box model, adding explainable capability to the hybrid model. The GBDD model has 5 inputs with a maximum epoch range set to 50 and the learning rate used is 0.006. The study used 9 days of historical weather data and PV power output data such as DHI, GHI, radiation diffuse tilted (RDT), radiation global tilted (RGT), daily rainfall, relative humidity, temperature, wind direction, and wind speed. The model was evaluated using MAE, MAPE, and RMSE scores, and was compared with LSTM, and GBLSTM. The proposed model gave 0.0771, 6.6027, and 0.1246 scores of MAE, MAPE, and RMSE, while LSTM gave 0.1207, 9.8891, and 0.1700 values of previously mentioned evaluation metrics in order. GBLSTM on the other hand gave improved scores with the scores as 0.0998, 9.2795, and 0.1544. These results showcase that the proposed model not only can outperform the models used for comparison, but also be used as a hybrid model for other prediction methods to improve the model predictions. The high performance of the model can be accounted to the use of the Maximum Information Coefficient (MIC) for variable selection, that allows the model to capture both linear and non-linear relationships between input features and PV output. The introduction of the Hampel filter to smooth the residuals allow the DD base learner to focus on broader trends rather than getting distracted by the noise in the the dataset and hence helps in preventing overfitting in the initial model, which is DD in the proposed model. While the Hampel filter helps reduce noise in the dataset, it could potentially over-smoothen the data and diminish the model's ability to capture rapid fluctuations, which are commonly present in the data for PV generation. The model is proved to perform well with larger datasets, but the study failed to test the model with real-time environment.

III. Problem Statement

Traditional forecasting approaches often fail to capture the complexities and variabilities that are in renewable energy generation hence leading to unreliable predictions and suboptimal decision-making while existing models in the literature rely on simple, uni variate approach for time series analysis tasks, which fail to account for the multiple factors that effect the target variable and neglect the nonlinear relationships that are inherent in solar irradiance data. While advanced models are also being used to be able to use multiple variables they often have high computational complexity or lack robustness. This study aims to address these challenges by proposing a hybrid gradient boosting and BiLSTM model to adapt to adapt to multiple variables and create comprehensive and robust models that are computationally efficient for solar irradiance forecasting.

IV. Dataset

The dataset used in this study is derived from NREL and includes Global Horizontal Irradiance (W/m^2) and weather data based on the United Arab Emirates (UAE). GHI is the The weather data consists of temperature (◦C), pressure (mbar) and wind speed (m/s), which are critical factors that influence the solar irradiance. The study utilized one month of data to train and test the models proposed. The choice of the UAE as the study region reflects its high solar potential and the growing interest in optimizing renewable energy solutions in arid and semi-arid climates.

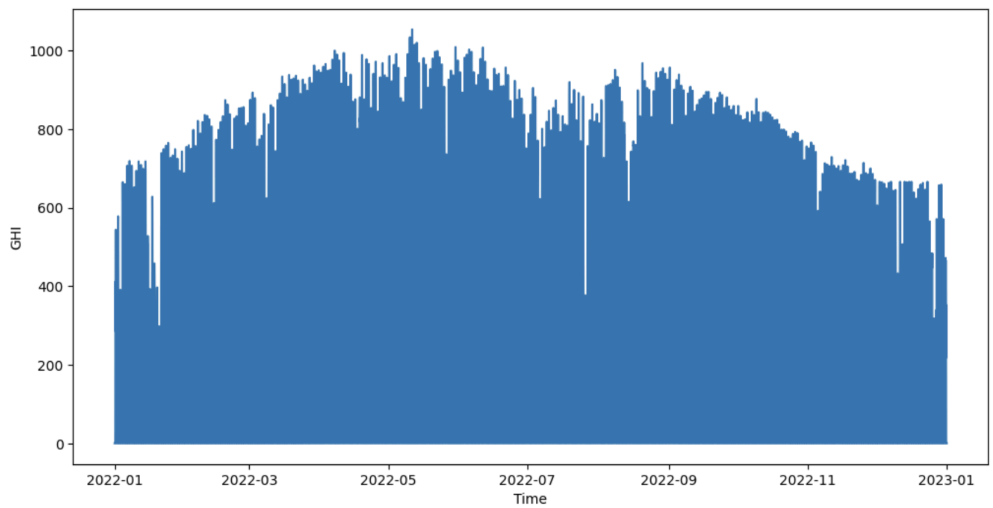

Figure 1. *Graph showing GHI trend in 2022*

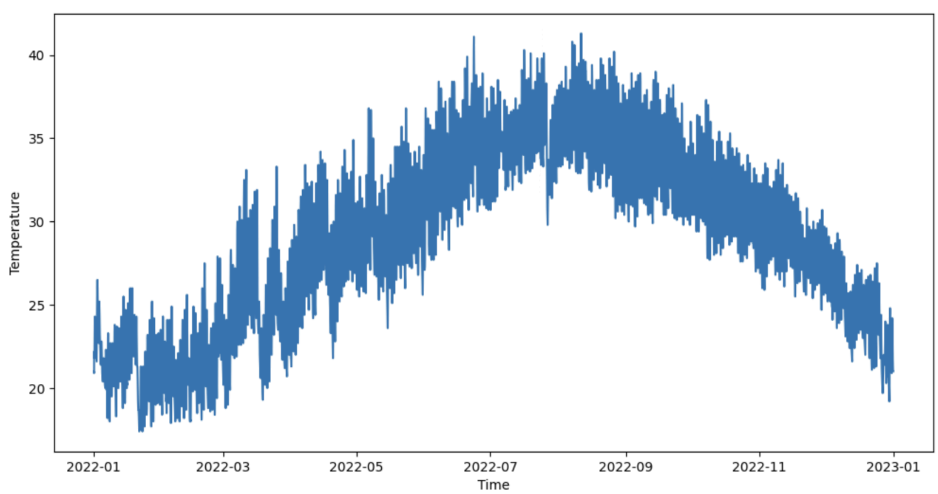

Figure 2. *Graph showing Temperature trend in 2022*

The initial datasets were obtained in multiple files, which were merged and cleaned to create a unified dataset. The resultant dataset is split into training and testing subsets, with 80% allocated for training and 20% for testing. One month data was considered for training and testing the model. This paper considered the data from April of 2022 to be utilized in this work. Data from April 2022 was specifically selected for this study, as it marks the onset of summer as can be seen in the, which is characterized by a gradual increase in temperature during this period, as can be seen in figures 1 and 2, and it is the most important time when predicting solar irradiance is critical which as it is used for also predicting the PV generation.

V. Methodology

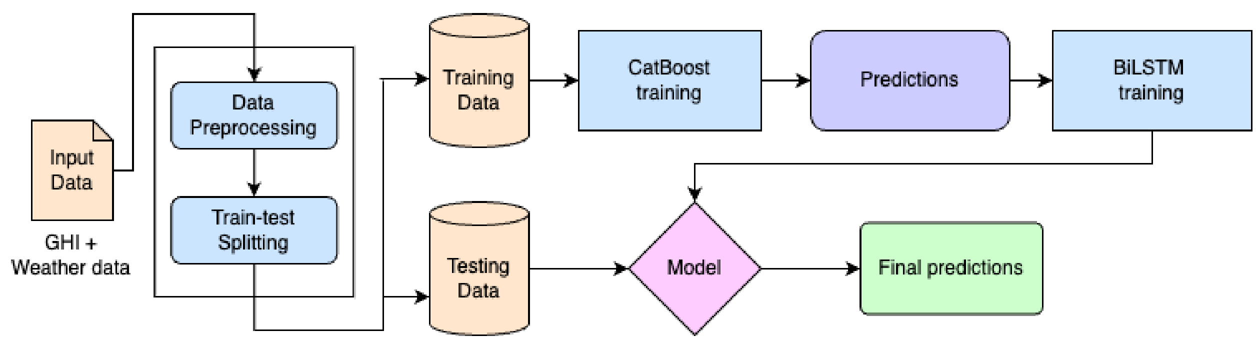

Figure 3. *Model architecture of the Catboost-BiLSTM model*

Figure 3 illustrates the architecture of the model that was used in this study. In this study, hybrid models between the various implementations of gradient boosting decision tree algorithms, like XGBoost, CatBoost, LightGBM and BiLSTM were utilized. A stacked BiLSTM architecture was used which was combined with the Gradient Boosting Decision Trees (GBDT) to synergize the strengths of both the individual models, as BiLSTM can process temporal dependencies while GBDT excels in capturing non-linearity in the data.

5.1 LSTM

Traditional RNNs that work on backpropagation calculate the gradients as a product of derivatives across multiple layers. This can be an issue when the derivatives become small and their repeated multiplication lead to the gradients shrinking exponentially as they move backward through the network. LSTM networks overcome this limitation of vanishing gradients through the use of gating mechanism, which control the information flow through the network. These gates include forget gate, input gate, and output gate. These gates allow the model to selectively remember the information that are important.

5.2 BiLSTM

BiLSTM is an extension of the standard LSTM architecture that processes the data in both forward and backward directions. With the presence of two LSTM layers, one layer processes the input sequence from start to end, and the other processes it from end to start. The output from both the layers are concatenated to get the predicted values. This mechanism enables BiLSTM to efficiently capture the temporal patterns and contextual information from both past and future sequences, rather than relying only on the historical data.

5.3 GBDT and its variants

While BiLSTM can learn temporal dependencies from the time series data, they lack at capturing highly nonlinear relationships when multiple variables are provided which are critical for predictions, which is addressed by GBDT algorithms. It consists of decision trees that are used as weak learners, where each learned is sequentially built to reduce the error of the previous tree. The process of boosting helps in optimizing gradient descent, where the loss function is minimized iteratively. The structure of the decision tree model can be fine tuned to match the specific characteristics of the dataset which makes them effective in handling various types of data such as continuous and discrete data. Least squares regression tree was used to process the data as it adapts to the impurity function of the regression tree.

5.4 XGBoost

XGBoost is a high efficient model that uses both gradients and second-order derivatives for better optimization of the loss function. It uses L1 (Lasso) and L2 (Ridge) regularization techniques to reduce the overfitting and enhance model convergence. XGBoost is also efficient in handling large dataset and uses max-depth pruning which speeds up the training process.

5.5 LightGBM

LightGBM, unlike the previously discussed gradient boosting techniques that build trees level by level, it uses a leaf-wise approach where it splits the leaf that has maximum loss reduction. It utilizes histogram-based algorithm to convert continuous features into discrete bins.

5.6 CatBoost

CatBoost, which is particularly used with categorical datasets, can be used for numerical data as well. It builds balanced or symmetric trees which leads to faster convergence and more stable results. It also optimizes the convergence by using ordered boosting technique which enhance the performance when used with numerical dataset where interaction between features can be highly complex. Ordered boosting is a technique where overfitting is reduced by creating several permutations of data using only past observations when calculating the leaf values.

5.7 CatBoost-BiLSTM model with teacher forcing mechanism

The combination of CatBoost and BiLSTM leverages the unique strengths of both the models to improve the forecasting accuracy. To first capture the relationship between the features, the CatBoost model is trained on the training data. The output from the model is then normalized using minmax scaler before using it as input for the BiLSTM model which contains valuable information about the underlying complex patterns of the data. This mechanism of using outputs from the previous model for training is called Teacher Forcing mechanism. Through this sequential mechanism, it prevents error accumulation over long sequences and helps captures both time dependent and nonlinear information from the data.

The BiLSTM model architecture uses stacked architecture with dropout layers which helps the model to learn more intricate and deeper temporal patterns. Dropout layer is essential to reduce overfitting, as it works by randomly dropping out or setting a fraction of the input units to zero during each forward pass while training. This layer also helps reduce co-adaptation of neurons where they may begin to depend on other neurons for making predictions and lead to poor generalization.

The architecture includes two Bidirectional LSTM layers, each consisting of 500 neurons and utilizing the ReLU activation function. Two dropout layers with dropout rate of 0.3 are added to reduce overfitting, followed by the inclusion of a dense layer for final predictions.

The model is trained for 20 epochs with a batch size of 64, optimizing the learning process using the MSE loss function. The Adam optimizer is employed for hyperparameter tuning, ensuring efficient convergence during training.

Table 1 Hyperparameters of the proposed CatBoost-BiLSTM model for univariate and multivariate data

Hyperparamaters	**Specifications**
No. of BiLSTM Layers	2
No. of dropout layers in BiLSTM model	2
No. of neurons	500
Dropout rate	0.3
No. of epochs	20
Batch size	64
N_estimators for gradient boosting algorithms	100
Learning rate for gradient boosting algorithms	0.001
Max depth for gradient boosting algorithms	5

VI. Results and discussion

Inorder to understand the effect of weather variables on prediction performance, this study initially trained the model using only GHI data and evaluated on metrics like MAE, MSE, RMSE, and R-squared values.

6.1 MAE

MAE represents the average absolute differences between the predicted and the actual values and it assesses how far the prediction values deviate from the true values on an average. It is calculated as the mean of the absolute values of the residuals (errors). MAE treats all the error equally without giving additional weights to larger deviations which makes it simple and interpretable measure of accuracy useful in situations where all errors are of similar importance.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

6.2 MSE

MSE shows the average of the squared differences between predicted and actual values. It penalizes the larger deviations more heavily by squaring the errors, which makes it suitable for detecting significant prediction errors. However, due to squaring of the errors, the value of MSE is not in the same unit as the target variable.

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

6.3 RMSE

RMSE is the square root of MSE which scales the error back into the same unit as the target variable, hence making it easier to interpret than MSE. It captures both the magnitude and variability of the errors as it penalizes the large prediction errors heavily and making it useful for understanding size of significant errors.

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - p_i)^2}$$

6.4 R squared

R-squared which is also known as coefficient of determination measures the proportion of variance in the predictions as compared to the target variables and it ranges from 0 to 1, where values closer to 1 indicates the model adapts to most of the variability in the data while value closer to 0 suggests the model fails to capture the variability. It tells about the goodness of fit of the model and shows how well the predictions align with the actual data.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

Table 2 presents the evaluation metrics of the models using only GHI data. BiLSTM model gave reasonable performance with an MAE of 46.7822, MSE of 5042.845, RMSE of 65.384, and an R-square value of 0.8934. While the XGBoost gave the best performance, with lowest MSE value of 6908.263 and highest R-square score of 0.9583. In comparison, GBDT-BiLSTM model exhibited 0.9329 R-square value and a higher RMSE of 82.7274, indicating relatively less precision in predictions. The LightGBM-BiLSTM model also performs well, achieving the lowest RMSE of 63.8594 and 0.9523 R-square score. On the other hand, the CatBoost-BiLSTM model gave an MAE of 58.7375 and RMSE of 103.6512. CatBoost with only GHI shows robust performance but lags slightly behind the other hybrid models in terms of error metrics.

Table 2 *Evaluation scores of the models using only GHI data*

Model	MAE	MSE	RMSE	R-square
BiLSTM	46.7822	5042.845	65.384	0.8934
GBDT-BiLSTM	56.2412	8028.9234	82.7274	0.9329
XGBoost-BiLSTM	**49.2183**	**6908.263**	**77.828**	**0.9583**
CatBoost-BiLSTM	58.7375	7048.3828	103.6512	0.9253
LightGBM-BiLSTM	50.9374	4238.2343	63.8594	0.9523

Table 3 displays the evaluation metrics of the models using both GHI and weather data as inputs. The standalone BiLSTM model demonstrates moderate performance, with an MAE of 62.9802, MSE of 7647.0178, RMSE of 87.4472, and an R-square value of 0.9344. These results suggest that the model is able to capture temporal dependencies effectively but does not fully leverage the nonlinear relationships between features. Among the hybrid models, the CatBoost-BiLSTM achieves the best overall performance, with the lowest MAE of 50.5336, MSE of 5912.151, RMSE of 76.8905, and R-square value 0.9493 which is the highest values that was observed among the models that are tested. So the combination of CatBoost for feature extraction and BiLSTM for temporal pattern learning has proved to significantly enhance the predictions. LightGBM-BiLSTM model shows comparatively better performance, as it gave an

MAE of 62.7963, MSE of 9245.9335, and RMSE of 96.1557. The R-square value of 0.9207 suggests that it is strongly able to generalize on the training data, but has slightly lower predictive accuracy compared to CatBoost-BiLSTM, while GBDT-BiLSTM model gave MAE of 70.5263 and RMSE of 101.1741. The higher error values and lower accuracy values when using GHI and weather data compared to only GHI can be observed, accounting weather data simulates real-world scenarios.

Table 3 *Evaluation scores of the models using GHI and weather data*

Model	MAE	MSE	RMSE	R-square
BiLSTM	62.9802	7647.0178	87.4472	0.9344
GBDT-BiLSTM	70.5263	10236.2026	101.1741	0.9122
CatBoost-BiLSTM	**50.5336**	**5912.151**	**76.8905**	**0.9493**
XGBoost-BiLSTM	67.9733	10743.5852	103.6512	0.9078
LightGBM-BiLSTM	62.7963	9245.9335	96.1557	0.9207

Looking at the loss curves produced by the 3 best performing models during training in Figure 4, it can be observed that CatBoost-BiLSTM has generalized very well on unseen data, proving its robustness. Both BiLSTM and LightGBM-BiLSTM started with low loss values, but have struggled to generalize on training data. However, They have shown low loss values at the end of 20th epoch, but not as par as CatBoost-BiLSTM model.

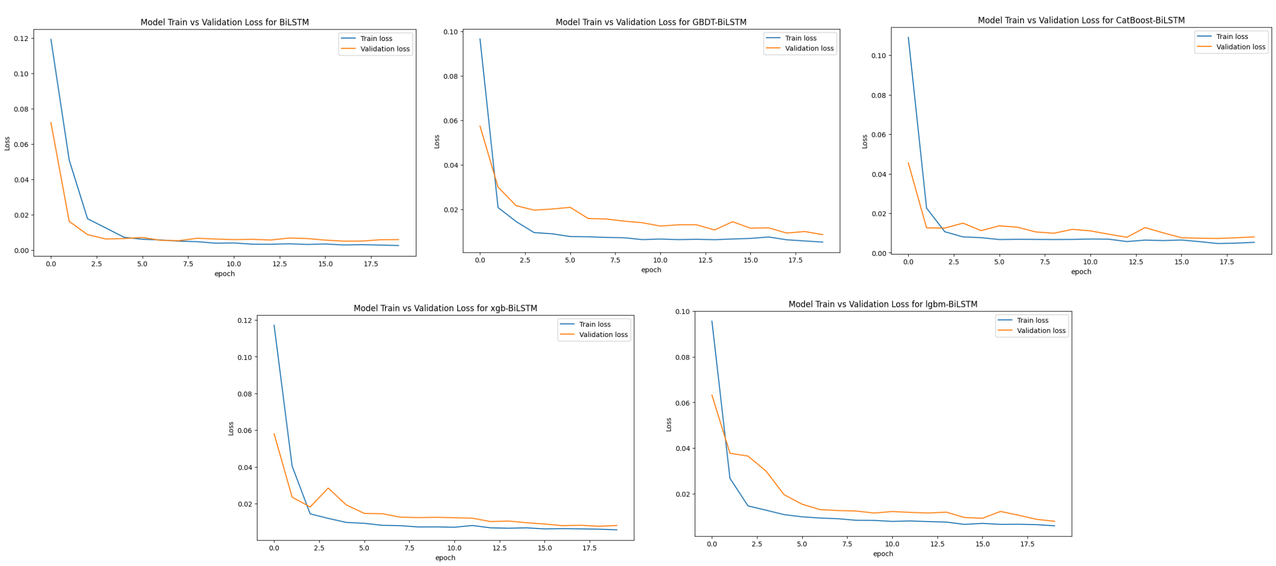

Figure 4. *Learning Curves for Implemented Models*

Comparing the actual vs predicted graph from figure 5 and 6, which demonstrates the model's performance in accurately capturing trends over the testing period, the predicted values from CatBoost-BiLSTM closely align with the actual values. As can be seen in the graph, CatBoost-BiLSTM is able to accurately match the peaks while predicting on the unseen data. It is also able to estimate the growing values using the one month data. This indicates the high degree of accuracy of the model's forecasting capabilities.

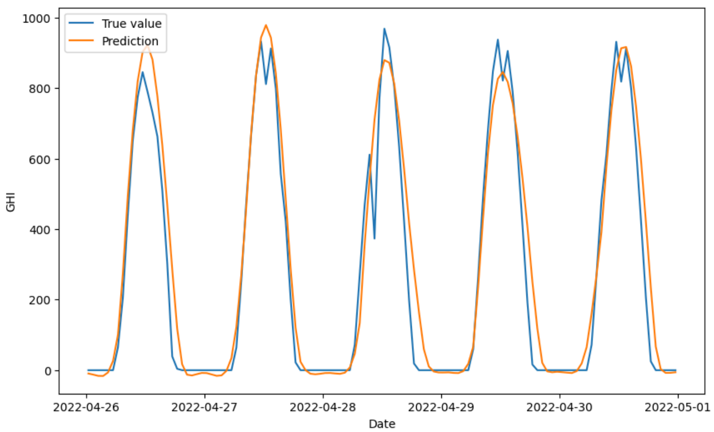

Figure 5. *Actual vs. Predicted GHI graph using CatBoost-BiLSTM model*

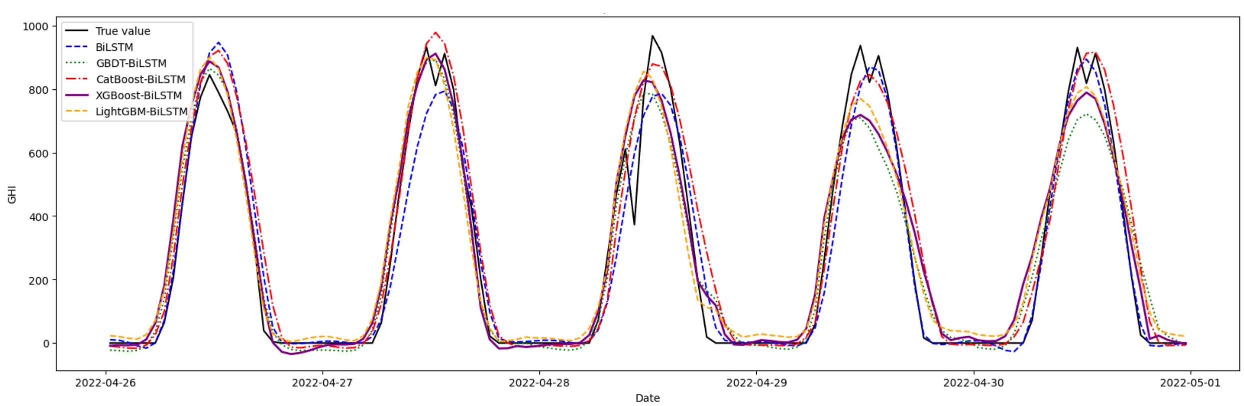

Figure 6. *Actual vs. Predicted GHI graph with different models*

VII. Conclusion

This study addresses the challenge of adopting and accurately predicting solar irradiance which are critical for integrating renewable energy into the modern energy grids. Traditional methods of simple univariate RNN models fall short of capturing complexities that are inherent in solar irradiance data, including its dependence on multiple interacting variables and nonlinear relationships.

This paper proposes and compares different variants of gradient boosting models that was hybridized with BiLSTM. The methodology leverages the ability of GBDT variants like CatBoost, LightGBM, and XGBoost with Teacher Forcing mechanism to capture nonlinear relationships and complex feature interactions, combined with the temporal dependency learning strength of BiLSTM. Teacher Forcing mechanism in training BiLSTM further enhances model performance by reducing error accumulation over long sequences.

The results demonstrate that the CatBoost-BiLSTM model achieves the best overall performance in terms of accuracy and robustness when using both GHI and weather data, like temperature, pressure and wind speed, as inputs, with the lowest error metrics and the highest R-square value. This shows the importance of hybridizing feature extraction and temporal modelling techniques to improve prediction capabilities. Although other hybrid models such as LightGBM-BiLSTM and GBDT-BiLSTM also performed well enough, they gave slightly lower precision and generalization capabilities compared to CatBoost-BiLSTM.

In conclusion, the proposed hybrid model not only addresses the shortcomings of existing methods but also provides a computationally efficient and robust solution for solar irradiance forecasting.

VIII. Future Work

Future work can focus on expanding the time horizon of the data to further improve the model accuracy for larger datasets. Whale Optimization Algorithm [11] can be utilized for efficient hyperparameter optimization for BiLSTM, as it effectively balances the exploration and exploitation while choosing the next hyperparamters. Additionally, Digital Twin can be used to simulate the power generation from solar cell by utilizing the prediction data of GHI as inputs to estimate future energy production.

References

1. Saxena, Nishant, et al. “Hybrid KNN-SVM Machine Learning Approach for Solar Power Forecasting.” *Environmental Challenges*, vol. 14, 1 Jan. 2024, pp. 100838–100838, https://doi.org/10.1016/j.envc.2024.100838.
2. Li, Yaru, et al. “A New Hyper-Parameter Optimization Method for Power Load Forecast Based on Recurrent Neural Networks.” *Algorithms*, vol. 14, no. 6, 1 June 2021, p. 163, www.mdpi.com/1999-4893/14/6/163, https://doi.org/10.3390/a14060163.
3. Tortora, Matteo, et al. “MATNet: Multi-Level Fusion and Self-Attention Transformer-Based Model for Multivariate Multi-Step Day-Ahead PV Generation Forecasting.” *ResearchGate*, 17 June 2023.
4. Said, Yahia, and Abdulaziz Alanazi. “AI-Based Solar Energy Forecasting for Smart Grid Integration.” *Neural Computing and Applications*, 13 Dec. 2022, https://doi.org/10.1007/s00521-022-08160-x.
5. Shahram Hanifi, et al. “Offshore Wind Power Forecasting Based on WPD and Optimised Deep Learning Methods.” *Renewable Energy*, vol. 218, 1 Dec. 2023, pp. 119241–119241, https://doi.org/10.1016/j.renene.2023.119241.
6. Panagiotis Eleftheriadis, et al. “Bayesian Hyperparameter Optimization of Stacked Bidirectional Long Short-Term Memory Neural Network for the State of Charge Estimation.” *Sustainable Energy, Grids and Networks*, vol. 36, 1 Dec. 2023, pp. 101160–101160, https://doi.org/10.1016/j.segan.2023.101160.
7. Dahmani, A. “Prediction of Hourly Global Solar Radiation: Comparison of Neural Networks / Bootstrap Aggregating.” *Fkit.hr*, 2022, silverstripe.fkit.hr/kui/issue-archive/article/992.
8. Verdone, Alessio, et al. “Explainable Spatio-Temporal Graph Neural Networks for Multi-Site Photovoltaic Energy Production.” *Applied Energy*, vol. 353, 1 Jan. 2024, pp. 122151–122151, https://doi.org/10.1016/j.apenergy.2023.122151.
9. Argyrios Vartholomaios, et al. “Short-Term Renewable Energy Forecasting in Greece Using Prophet Decomposition and Tree-Based Ensembles.” *Communications in Computer and Information Science*, 1 Jan. 2021, pp. 227–238, https://doi.org/10.1007/978-3-030-87101-7_22.
10. Wang, Chunsheng, et al. “Gradient Boosting Dendritic Network for Ultra-Short-Term PV Power Prediction.” *Frontiers in Energy*, 10 Jan. 2024, https://doi.org/10.1007/s11708-024-0915-y.
11. Y. Song, H. Xie, Z. Zhu, and R. Ji, “Predicting energy consumption of chiller plant using WOA-BiLSTM hybrid prediction model: A case study for a hospital

building," *Energy and buildings*, vol. 300, pp. 113642–113642, Dec. 2023, doi: https://doi.org/10.1016/j.enbuild.2023.113642.