 doi:10.20944/preprints202208.0345.v1

Type of the Paper (Article)

SENERGY: A Novel Deep Learning-Based Auto-Selective Approach and Tool for Solar Energy Forecasting

Ghadah Alkhayat [1], Syed Hamid Hasan [2], Rashid Mehmood [3,*]

[1] Department of Computer Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; grasheedalkhyat@stu.kau.edu.sa
[2] Department of Computer Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; shhasan@kau.edu.sa
[3] High Performance Computing Centre, King Abdulaziz University, Jeddah 21589, Saudi Arabia;
* Correspondence: RMehmood@ksa.edu.sa

Abstract: The sustainability of the planet and its inhabitants is in dire danger and is among the highest priorities on global agendas such as the Sustainable Development Goals (SDGs) of the United Nations (UN). Solar energy -- among other clean, renewable, and sustainable energies -- is seen as essential for environmental, social, and economic sustainability. Predicting solar energy accurately is critical to increasing reliability and stability, and reducing the risks and costs of the energy systems and markets. Researchers have come a long way in developing cutting-edge solar energy forecasting methods. However, these methods are far from optimal in terms of their accuracies, generalizability, benchmarking, and other requirements. Particularly, no single method performs well across all climates and weathers due to the large variations in meteorological data. This paper proposes SENERGY (an acronym for Sustainable Energy), a novel deep learning-based auto-selective approach and tool that, instead of generalising a specific model for all climates, predicts the best performing deep learning model for GHI forecasting in terms of forecasting error. The approach is based on carefully devised deep learning methods and feature sets through an extensive analysis of deep learning forecasting and classification methods using ten meteorological datasets from three continents. We analyse the tool in great detail through a range of metrics and methods for performance analysis, visualization, and comparison of solar forecasting methods. SENERGY outperforms existing methods in all performance metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Forecast Skills (FS), Relative Forecasting Error, and the normalised versions of these metrics. The proposed auto-selective approach can be extended to other research problems such as wind energy forecasting and predict forecasting models based on different criteria (in addition to the minimum forecasting error used in this paper) such as the energy required or speed of model execution, different input features, different optimisations of the same models, or other user preferences.

Keywords: Deep learning; solar radiation forecasting; model prediction; solar energy; multi climates data; generalizability; sustainability; Long Short-Term Memory (LSTM); Gated Recurrent Unit (GRU); Convolutional Neural Network (CNN); Hybrid CNN-Bidirectional LSTM; LSTM Autoencoder.

1. Introduction

The last century has seen many technological advancements that have enabled us to make unimaginable progress, particularly during the last few decades. This progress however has come at a rapidly increasing price. The sustainability of the planet and its inhabitants is in dire danger and is among the highest priorities on global agendas such as the Sustainable Development Goals (SDGs) of the United Nations (UN). Solar energy, among other clean, renewable, and sustainable energies, such as wind energy, is essential for environmental, social, and economic sustainability.

 doi:10.20944/preprints202208.0345.v1

Solar energy could generate larger than 10,000 times the world's total energy consumption with its Earth strike rate of 173,000 terawatts [1]. Therefore, solar energy has enormous potential for reducing global carbon emissions. For example, the installation of 113,533 domestic solar systems in California, USA, has lowered or prevented 696,544 metric tons of CO2 emissions [2]. Developing capacity for solar energy production is also critical for Saudi Arabia, which is among the top few oil producers and consumers in the world and is ranked sixth in the world in terms of its potential for producing solar energy [3]. The Sakaka 300-megawatt (MW) solar power station, Saudi Arabia's first utility-scale solar PV project, was connected to the national grid in November 2019. With a $302 million investment, the plant will cover a six-square-kilometer area in AlJouf. This is the first in a series of projects under Saudi Arabia's national renewable energy program to generate 9.5GW of renewable energy by 2023 [4].

The need for integrating solar energy into the electrical grid has motivated researchers globally to develop advanced methods for solar radiation forecasting. Accurate prediction of solar radiation is vital to ensure hybrid energy systems' reliability and permanency. Specifically, it reduces the risks and costs of managing the energy market and energy systems, which are attributed to the influence of climate changes and weather variability [5], [6]. The applications of solar radiation forecasting in solar energy systems vary according to the forecasting horizon, which ranges from very short to long term. They include real-time monitoring, demand and supply balancing, decision making, unit commitment, power plant maintenance scheduling, site selection, solar plant installation, grid operations planning, and others [7].

Solar energy and its generated electrical energy outputs will always be unsteady due to the variable and uncertain nature of weather. As a result, solar energy prediction is critical and difficult, necessitating the development of advanced methods. There are four types of methods used for this purpose: physical (such as numerical and simulation weather prediction models), statistical, those based on artificial intelligence (AI), and hybrid methods [8], [9]. Because of their ability to discover nonlinear relationships and provide superior performance, artificial intelligence methods such as machine and deep learning methods have grown in popularity. Machine learning including deep learning methods, in particular, have excelled in a wide range of scientific problems and applications domains, including computer vision and natural language processing [10]–[12], transportation [13], healthcare [14], education [15], and smart cities [16]. This is also true for solar energy forecasting, with many deep learning methods emerging in recent years that outperform the other three types of forecasting methods [10], [17].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

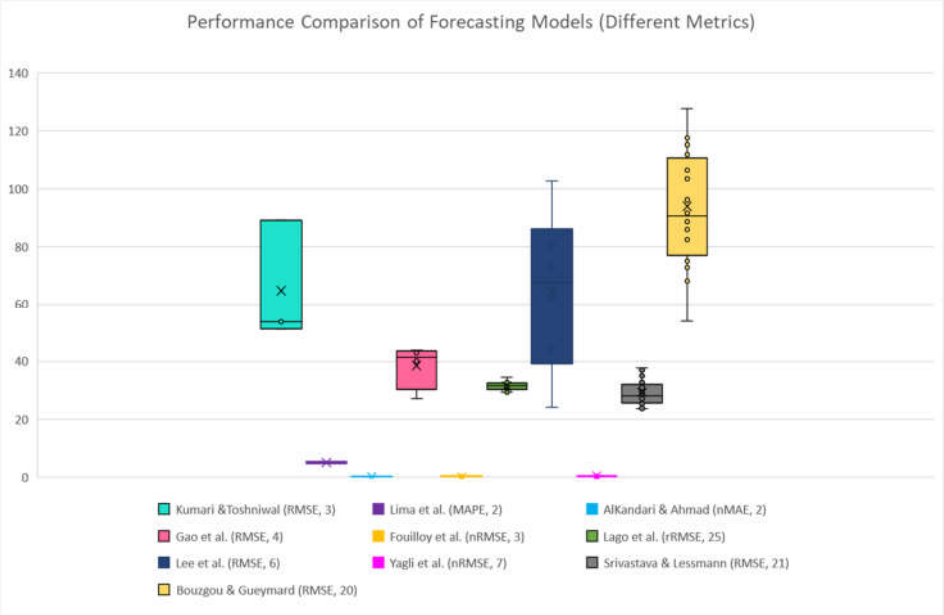


Figure 1. Performance Comparison of Solar Forecasting Models (Different Performance Metrics).

We have performed an extensive literature review (see Section 2 and [17]) on deep learning-based solar energy forecasting methods and have identified the key research gaps in this field. We explain the research gaps using **Figure 1**. The figure provides a performance comparison of different deep learning models. The compared works include Kumari and Toshniwal [18], Lima et al. [19], AlKandari and Ahmad [20], Gao et al. [21], Fouilloy et al. [22], Lago et al. [23], Lee et al. [24], Yagli et al. [25], Srivastava and Lessmann [26], and Bouzgou and Gueymard [27]. We will elaborate on the reasons for the selection of these methods in the later sections. Note in the figure that performance for different methods is plotted using different performance metrics as originally used by the authors in their published works. The metrics used in these works include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Normalized RMSE (nRMSE), Relative RMSE (rRMSE), and Normalised MAE (nMAE). Each work is plotted as a box plot, labelled with the authors' names, the performance metric used by the authors, and the total number of datasets used in their work. For example, Lima et al. [19] reported the performance of their proposed methods using MAPE with two datasets; it is labelled as Lima et al. (MAPE, 2). Ideally, the box plot should be closest to the x-axis to reflect a small value for the error metric. Also, the box plot should be vertically small to reflect small variations in the error metrics for different datasets.

The figure shows that different works have used different metrics and different numbers of datasets and that there is a large variation in their performances. The use of different metrics makes it difficult to compare the performance of different methods. A larger number of datasets may indicate better generalisability and validation of results; however, this is not necessarily true because it depends on the size of the datasets, variability in the climates and data characteristics, and the metrics used to measure the performance, and other factors. Even if the same performance metric was used by these works, comparing their performance in terms of RMSE or other existing metrics is difficult because these metrics do not always exhibit the variability in the input data such as variations in the types of climates, the proportion, and unpredictability of sunny and cloudy weather, variations in GHI (Global Horizontal Irradiance), etc. For example, although both Lima et al. [19], and AlKandari and Ahmad [20] used two datasets, it is hard to fairly compare them because the error metrics used by them are different (MAPE versus nMAE). Similarly, it

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

is hard to compare the results reported by Kumari and Toshniwal [18], and Srivastava and Lessmann [26], due to the large difference between the number of datasets (3 versus 21) despite the fact that both of them reported their results using the same metric (RMSE). Note that a large number of datasets do not necessarily show variations in the input data; one needs to look at the size of the datasets, the dataset climates, the variations in data, etc.

The challenges described above call for new approaches from the community for novel forecasting and evaluation methods. There is a need for independent and transparent evaluation and extensive testing of the published models [10], similar to what has been done in other fields such as computer vision. Some researchers have suggested the use of a single statistical index called the global performance indicator to overcome the difficulty of comparing different performance metrics [28], [29]. Moreover, some independent benchmarking exercises or conferences in the renewable energy fields have started to emerge. An example is the Global Energy Forecasting Competition in the USA, which to date has been organized three times in 2012, 2014, and 2017 [30]. While all of these works and proposals demonstrate that the community has made significant progress in developing high-performance solar energy forecasting methods. Much more sustained effort is required to improve forecasting model accuracies and generalizability, as well as extensive, transparent, and fair benchmarking of these models. Because of the large variations in meteorological data, no single forecasting method performs well across all climates and weathers. There is a need to close this gap so that forecasting methods can perform optimally across varying climates and data.

This paper proposes a novel deep learning-based auto-selective approach and tool that, instead of generalising a specific model for all climates, predicts the best performing deep learning model for GHI forecasting. We call this approach and tool SENERGY, an acronym for Sustainable Energy. The approach is based on carefully devised deep learning methods and feature sets through an extensive analysis of deep learning forecasting and classification methods using ten meteorological datasets from three continents. The models that we have used in this work include Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), Hybrid CNN-Bidirectional LSTM, and LSTM Autoencoder. We analyse the tool in great detail through a range of metrics and methods for performance analysis, visualization, and comparison of solar forecasting methods. SENERGY outperforms existing methods in all performance metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Forecast Skills (FS), Relative Forecasting Error, and the normalised versions of these metrics.

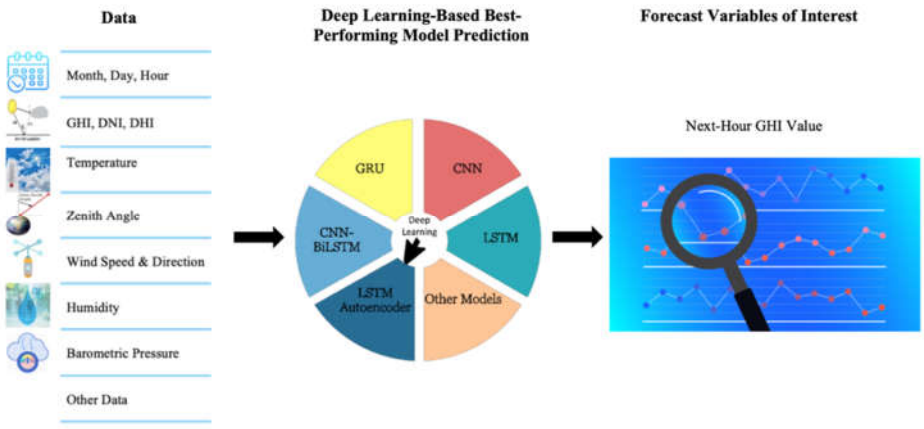


Figure 2. SENERGY: A High-Level Overview.

 doi:10.20944/preprints202208.0345.v1

Figure 2 shows a higher-level overview of the SENERGY approach. The figure shows that various forecasting temporal information (month, day, hour) along with the previous values of Global Horizontal Irradiation (GHI) and weather variables is supplied to the tool as inputs and the tool recommends the best forecasting model and uses this model to provide forecasted GHI. A detailed explanation of the design of the SENERGY approach and tool is given in Section 3.

The approach proposed in this paper to use machine or deep learning to automatically predict a best-performing model or configuration is not new and has been used in our earlier work for computations of Sparse Matrix-Vector (SpMV) products [31]–[33]. However, to the best of our knowledge, this is the first time that such an approach has been used in solar energy forecasting and is implemented in a tool for this purpose. The proposed auto-selective approach currently considers minimum forecasting error to predict the best performing deep learning model for GHI forecasting. It can be extended to predict forecasting models based on different additional criteria such as the energy required or speed of model execution, different input features, different optimisations of the same models, or other user preferences. Additional deep learning models for classification (to auto-select) or forecasting solar radiation can be incorporated into the tool to improve the performance and diversity of the tool. The approach is extensible also to other renewable energy sources and problems such as wind energy forecasting.

The contributions of this paper can be summarised as follows.

1. This paper proposes a novel approach and tool that uses deep learning to automatically predict the best-performing solar energy forecasting model. The approach is extensible to other performance metrics or user preferences and is applicable to other energy sources and problems.
2. We provide an in-depth analysis of five deep learning models for solar energy forecasting using ten datasets from three continents. This is the first time that such a combination of models, datasets, and analyses has been reported. Particularly, none of the earlier works have reported forecasting based on five deep learning-based models with such many locations in Saudi Arabia and provided a comparison with locations abroad (Toronto and Caracas).
3. We highlight the need for standardisation in performance evaluation of machine and deep learning modelling in solar forecasting by providing extensive analysis and visualisation of the tool and its comparison with other works using several performance metrics. We have not seen such an extensive evaluation of work earlier in solar energy forecasting. This paper is expected to open new avenues for higher depth and transparency in benchmarking of solar energy forecasting methods.

This paper is organized as follows: Section **Error! Reference source not found.** reviews the related works. Section **Error! Reference source not found.** presents the methodology used in the work including subsections describing the SENERGY development process, the data collection, data pre-processing, feature importance, and five forecasting model structures. Section **Error! Reference source not found.** also includes a description of the performance evaluation metrics and implementation of the models. In Section **Error! Reference source not found.**, the results are discussed and analysed in detail. Section **Error! Reference source not found.** concludes and provides future directions.

2. Literature Review

Deep learning models' promising achievements have attracted researchers to apply them in the field of solar radiation and solar energy forecasting. Their advantages include the ability to discover nonlinear relationships among inputs, generalization capability, and unsupervised feature learning in addition to superior performance. In our earlier work [17], we have done an extensive review of solar and wind energy forecasting

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

methods based on deep learning and proposed a taxonomy of this research field as shown in Figure 3. The most used deep learning-based architectures in the literature are the hybrid models followed by Recurrent Neural Network models including LSTM model and GRU and then, CNN in the third place. Based on numerous studies included in the review, we found that deep learning-based forecasting models always achieve relatively higher accuracies and generalization ability compared to other machine learning models and statistical methods, especially when they are combined with other algorithms in hybrid models. However, a definite conclusion cannot be drawn about the forecaster that has the best performance unless extensive testing is done using datasets from different climates and topographies that contain data about all seasons and weather conditions. Although deep learning models have proven their ability to provide competitive results in terms of forecasting accuracy, there is still room for improvement in terms of models' generalization and stability. More studies should focus on developing general forecasting models since developing a model for each location is infeasible. Few studies proposed forecasting models for a whole region as [34]–[38]. However, general forecasting models should be able to provide forecasting to locations from different climatic zones not only similar regions.

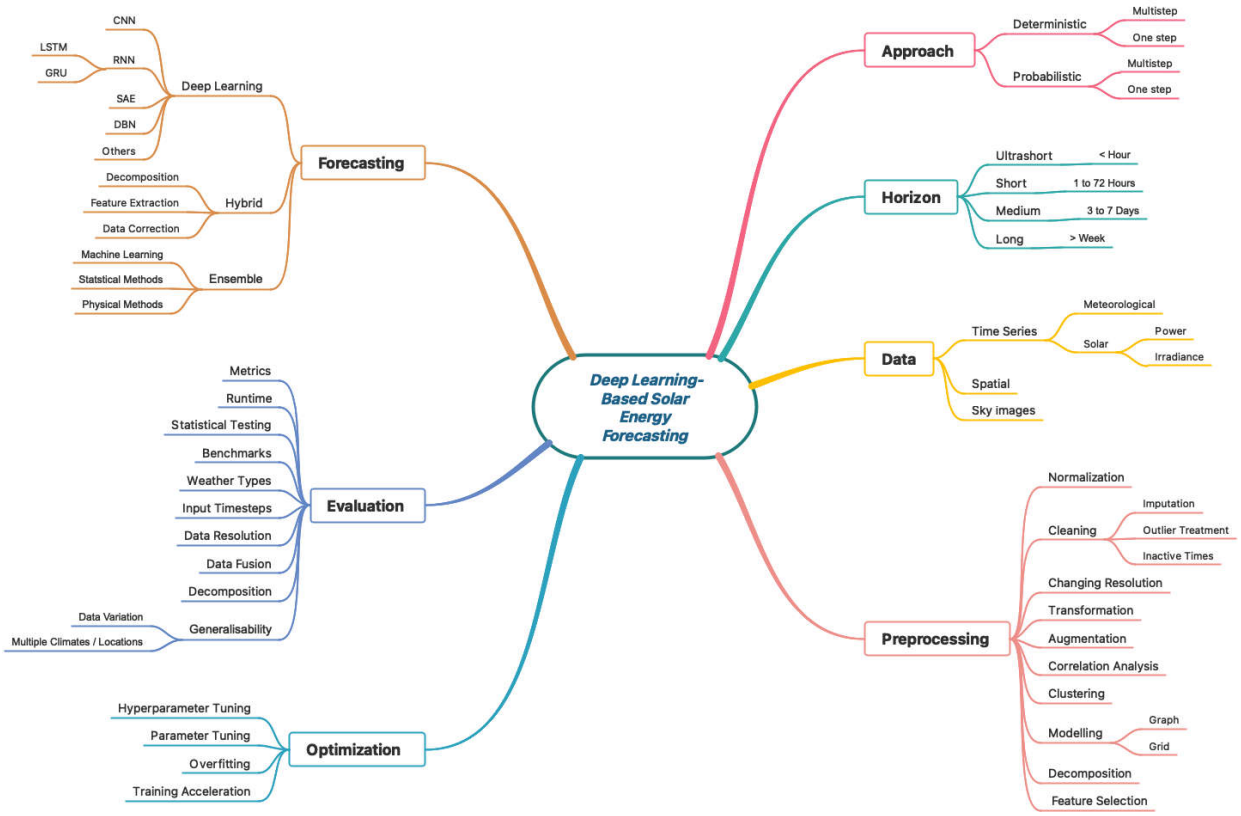


Figure 3. Deep learning-based solar energy forecasting taxonomy.

The current efforts focused on improving the generalizability of deep learning-based forecasters in the literature are still limited. Some researchers suggest ensemble learning to improve generalization. Ensemble learning takes the average prediction of several forecasting models instead of finding a single best-performing one. For example, Lima et al. [19] used ensemble learning along with a new integration technique based on Portfolio Theory. Their proposed solar irradiance ensemble forecasting model integrates the Multilayer Perceptron network (MLP) model, Support Vector Regression (SVR) model, Radial

 doi:10.20944/preprints202208.0345.v1

Basis Function model, and LSTM model. Weights are assigned adaptively to each model before calculating the final forecasting result. Such a self-adaptive model structure is a way to improve the forecasting performance in terms of accuracy and generalizability. The authors also compared their model performance using datasets collected from Brazil and Spain. Likewise, Khan et al. [39] combined Artificial Neural Network (ANN), LSTM, and eXtreme Gradient Boosting (XGBoost) in their ensemble model to improve the generalization of solar forecasting. Their method achieved more stable performance in several case studies than using ANN, LSTM, or bagging alone. AlKandari & Ahmad [20] proposed a solar power forecasting model employing an ensemble approach, which combines GRU, LSTM, and Theta models. They found that the ensemble technique of both machine learning and statistical models achieved better prediction accuracy than single models. They also compared their model performance using datasets from Kuwait and USA. Wang et al. [40] utilized classification along with ensemble learning in their proposed PV power ensemble forecasting framework. The classification is done to identify the daily pattern label of the forecasting day to improve the forecasting accuracy performed by multiple LSTM models. Singla et al. [41] utilized wavelet transform (WT) to decompose the input time series data into different sub-series, then, trained a bidirectional LSTM model for each one. The forecasted values of each sub-series from BiLSTM models are combined to deliver the final 24-h ahead GHI forecast. Pan and Tan [42] cluster analysis on data to get weather regimes before employing Random Forests to acquire prediction from different weather regimes. El-Kenawy et al. [43] developed an ensemble model for solar radiation forecasting, which consists of LSTM, NN, and SVM. This model's ensemble weights are optimized by Advanced Sine cosine algorithm that shows performance superiority over the average and K-Nearest Neighbors ensemble methods.

Comparing the performance of a proposed model using several datasets collected from locations with different climates is a practice in the literature that aims to improve forecasting models' performance generalization and stability. For example, Kumari and Toshniwal trained and tested their ensemble model, which consists of XGB Forest and Deep Neural Network (DNN) for hourly GHI forecasting using data collected from three locations in India that have humid subtropical, hot semi-arid, and subtropical climates [18]. Similarly, Gao et al. [21] tested their proposed CNN and LSTM hybrid model for hourly solar irradiance forecasting using four datasets from locations with Mediterranean, semi-arid, rainforest, and desert climates. Likewise, Kapa et al. [44] compared the performance of a DNN model for daily GHI forecasting on datasets collected from thirty-four cities in Turkey, which belong to very wet, humid, semi-humid, and semi-dry climates. Fouilloy et al. [22] also compared eleven machine learning and statistical models for solar irradiation forecasting using three datasets with different meteorological characteristics. The datasets sources are two locations in France Odeillo in the mountains and Ajaccio near the Mediterranean Sea in addition to Tilos island in Greece. Lago et al. [23] proposed a generalized model based on DNN for solar irradiance forecasting using data from twenty-five locations in the Netherlands while Lee et al. [24] compared several ensemble models using datasets from six distinct locations in the USA. Yagli et al. [25] evaluated sixty-eight machine learning algorithms using data from five climate zones for hourly solar forecasting. Srivastava and Lessmann [26] compared LSTM performance in twenty-one locations in Europe and USA from ten climate types to another three methods. Jeon and Kim [45] proposed a global LSTM model for the next-day solar irradiance prediction by training the model with data collected from Cape Town, Canberra, Colorado, and Paris, then evaluating it with data from Inchon. Bouzgou and Gueymard [27] trained an Extreme Learning Machine (ELM) model using data from twenty locations that belong to four different climates.

Table 1 highlights the approach used to improve forecasting performance generalization in each paper covered in this section along with the results and main findings. Since most of the researchers in this field are more concerned with improving forecasters' accuracies as their main goal, there is a need to explore new methods to improve generalization

 doi:10.20944/preprints202208.0345.v1

as a way toward achieving higher accuracy. In this paper, we combined two methods: the knowledge gained from comparing multiple forecasting models' performance on different climate data along with classification to recommend the best forecasting model for certain data inputs.

Table 1. Summary of the literature review.

Ref#	Ensemble model	Multiple climates	Results	Main findings
[18]	✓	✓	The ensemble model (XGBF-DNN) performed better than Smart Persistence, SVR, Random Forest (RF), XGBoost, and DNN models for all three locations Jaipur, New Delhi, and Gangtok. Hence, it can be generalized to predict hourly GHI for other geographical locations.	The ensemble model (XGBF-DNN) attained RMSE =53.79 for Jaipur, RMSE = 51.35 for New Delhi, and RMSE=89.13 for Gangtok
[19]	✓	✓	Errors resulting from the integration of forecast techniques (LSTM, MLP, RBF, SVR) had a better performance than the individual errors of each model for Brazil and Spain in hour-ahead PV power forecasting.	The ensemble model of LSTM, MLP, RBF, and SVR achieved MAPE =5.36% for Spain and 4.52% for Brazil
[39]	✓		The ensemble model of ANN, LSTM, and XGBoost performed better than ANN and LSTM models alone in PV energy generation forecast.	The ensemble model of ANN, LSTM, and XGBoost achieved RMSE= 0.74 and MAE=0.47 with 15-min data resolution and RMSE= 0.78 and MAE=0.59 with 1-hour data resolution.
[20]	✓	✓	The ensemble model of GRU, LSTM, and Theta achieved better performance with Shagaya dataset than with Cocoa because it contains relative weather data. The ensemble model achieved better accuracy than any single ML algorithm and theta model in day-ahead solar power generation forecast.	The ensemble model of GRU, LSTM, and Theta achieved nMAE= 0.0317 for Shagaya location in Kuwai while LSTM model alone achieved nMAE=0.0739 for Cocoa location in USA, which is slightly better than the ensemble model performance with nMAE=0.0877. Based on nMSE results, the ensemble model achieved better performance than individual models for both datasets.
[40]	✓		The ensemble model of LSTMs with time correlation under a partial daily pattern prediction framework attained better performance than BPNN model, SVM model, and persistent model in day-ahead PV power forecasting.	The ensemble model of LSTMs with time correlation under a partial daily pattern prediction framework attained RMSE= 5.68.
[41]	✓		The results show that the ensemble model of wavelet transform (WT) and bidirectional LSTM outperformed the naïve predictor, LSTM, GRU, BiLSTM and two different WT based BiLSTM in 24-h ahead solar irradiance forecast.	The results show that the ensemble model of wavelet transform (WT) and bidirectional LSTM outperforms the naïve predictor, LSTM, GRU, BiLSTM and two different WT based BiLSTM with annual average RMSE= 45.61and MAPE=6.48%.
[42]	✓		The ensemble model of RF with cluster analysis for day-ahead solar forecasting performed better than RF alone and gradient boosted regression trees. Classify the weather regimes with cluster analysis improved the model accuracy.	The ensemble model of RF with cluster analysis for day-ahead solar forecasting performed better than RF alone and gradient boosted regression trees with nRMSE=8.8
[43]	✓		The ensemble model of LSTM, NN, and SVM for solar radiation forecasting outperformed all the reference models. The best optimizing ensemble weights method is Advanced Sine and Cosine algorithm.	The ensemble model of LSTM, NN, and SVM for solar radiation forecasting outperformed all the reference models with RMSE=0.0018. The best optimizing ensemble weights method is Advanced Sine and Cosine algorithm.
[21]		✓	The proposed hybrid model of complete ensemble empirical mode decomposition adaptive noise (CEEMD), CNN, and LSTM to forecast hourly irradiance performed better	The proposed hybrid model of complete ensemble empirical mode decomposition adaptive noise (CEEMD), CNN, and LSTM achieved annual RMSE= 42.84 for Tamanrasset, 43.98 for Hawaii's

 doi:10.20944/preprints202208.0345.v1

			compared to the single LSTM, BPNN, SVM, the hybrid CEEMDAN-LSTM, CEEMDAN-BPNN, and CEEMDAN-SVM models.	Big Island, 40.60 for Denver, and 27.09 for Los Angeles.
[44]		✓	The proposed DNN for daily GHI prediction showed good performance with 34 cities, which represent all possible climatic conditions in Turkey. Using all inputs (extraterrestrial radiation, sunshine duration, cloud cover, maximum temperature, and minimum temperature) gave the best results.	The proposed DNN for daily GHI prediction achieved RMSE ranges from 0.52 to 1.29 for 34 cities, which represent all possible climatic conditions in Turkey.
[22]		✓	Statistical models performance of hourly solar irradiation forecasting with low to medium meteorological variabilities data is efficient while with high variability or longer forecasting horizons (4-hours ahead and more), bagged regression tree and RF approaches performed better than statistical models.	For a medium and low variability dataset (Tilos and Ajaccio), the best results for an hour-ahead forecasting come from SVR model with MAE= 71.27 and 54.58. For a high variability dataset (Odeillo), the best results for an hour-ahead forecasting come from RF.
[23]		✓	The proposed global DNN for hourly GHI forecasting, which was trained using data from 25 locations in the Netherlands (satellite-based measurements and weather-based forecasts) has a better average performance than other four local models.	The proposed global DNN for hourly GHI forecasting, which was trained using data from 25 locations in the Netherlands has a better average performance with relative RMSE= 31.31% than other four local models. The lowest relative RMSE=29.24 for Hoek v. H. site and the highest relative RMSE=34.55 for Deelen site
[24]	✓	✓	The ensemble models (Boosted Trees, Bagged Trees, RF, and Generalized RF) for short-term prediction of solar irradiance offered superior prediction performance compared to Gaussian process regression and SVR.	The ensemble model Generalized RF achieved the best MAPE results for 4 out of 6 datasets (MAPE equals to 19.76, 42.27, 31.79, and 58.58 for CA, TX, WA, and MN respectively)
[25]		✓	For the task of hourly solar forecasting, tree-based methods were found superior in average nRMSE under all-sky conditions, whereas variants of MLP and SVR were the best performers under clear-sky conditions. RF with Quantile Regression performed well under overcast skies at all 7 locations.	Tree-based methods are superior compared to other machine learning algorithms for all-sky conditions with nRMSE ranges from 15.46% to 33.36% based on location.
[26]		✓	The LSTM model outperformed Gradient Boosting Regression, FFNN, and Persistence methods in day-ahead GHI forecasting	LSTM model outperformed Gradient Boosting Regression, FFNN, and Persistence methods in day-ahead GHI forecasting with RMSE ranges from 23.6 to 37.78 for 21 locations
[45]		✓	The global LSTM model, which was trained with international data for next-day GHI prediction, achieved RMSE=30 with Inchon in Korea	The global LSTM model, which was trained with international data achieved RMSE=30 with Inchon in Korea
[27]		✓	ELM model, which was trained with data from 20 locations, has good performance for 15-min ahead, 1-h ahead, and 24-h ahead forecasting	ELM model achieved average RMSE= 93.82 for 20 locations for 1-h ahead forecasting

2.1. Research Gap

The literature review presented in this section (also see [17]) identified the major research gaps in deep learning-based solar energy forecasting methods research. Despite significant progress in developing high-performance solar energy forecasting methods, much more sustained effort is required to improve forecasting model accuracies and generalizability, as well as extensive, transparent, and fair benchmarking of these models. Because meteorological data varies so greatly, no single forecasting method performs well across all climates and weathers. There is a need to close this gap so that forecasting methods can perform optimally across varying climates and data. This paper proposes a novel approach and tool for automatically predicting the best-performing solar energy forecasting model using deep learning. The method is adaptable to other performance metrics or

 doi:10.20944/preprints202208.0345.v1

user preferences, as well as other energy sources and problems. To our knowledge, this is the first time such an approach has been used in solar energy forecasting and has been implemented in a tool for this purpose.

Using ten datasets from three continents, we conduct an in-depth analysis of five deep learning models for solar energy forecasting. None of the previous works reported such an integration of models, datasets, and analyses. None of the works reported forecasting based on five deep learning-based models with such a large number of locations in Saudi Arabia, nor did they provide a comparison with locations elsewhere (Toronto and Caracas). We highlight the need for standardisation in the performance evaluation of machine and deep learning modelling in solar forecasting by providing in-depth analysis and visualization of the tool, as well as comparisons with other works using various performance metrics. We have not seen a thorough evaluation of work in solar energy forecasting before. We anticipate that this work will pave the way for greater depth and transparency in benchmarking solar energy forecasting methods.

3. SENERGY: Methodology and Design

We first briefly describe the SENERGY development process on a high level in Section3.1, then move to the detailed steps in later sections. The datasets development process is described in Section 3.2, which includes data collection and data preprocessing for forecasting and model prediction. In Section 3.3, we discuss four feature importance methods: Pearson's correlation, Mutual Information, Forward Feature Selection and Backward Feature Elimination in Section, and LASSO feature selection. Then, five deep learning models, which are used in SENERGY, are explained in Section 3.4: Long Short-Term Memory, Gated Recurrent Unit, Convolutional Neural Network in Section, Hybrid CNN and Bidirectional Long Short-Term Memory, and Long Short-Term Memory Autoencoder. Finally, a description of the performance evaluation metrics is given in Section 3.5 and SENERGY implementation in Section 3.6.

3.1. Tool Development Process

The SENERGY development process is displayed in Figure 4, which starts with collecting datasets from multiple locations that have different climates, followed by data preprocessing, such as filling missing values, creating lagged features, and normalization. Then, the process continues with feature selection through Pearson's correlation, mutual information, forward feature selection, backward feature elimination, and LASSO methods. Afterward, preprocessed data is used for training and testing five deep learning-based forecasters. The forecasters' performance on the datasets is compared using several performance evaluation metrics. Based on performance comparison, the best model label is obtained, which is the forecaster that achieves the least forecasting error. Then, the best model label is added to become the target variable, and all the datasets are combined to train and test the best forecaster recommendation model. After completing the development process of SENERGY, the tool is able to receive new inputs, recommend the best forecaster based on inputs, and use the chosen forecaster to predict the next hour GHI.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

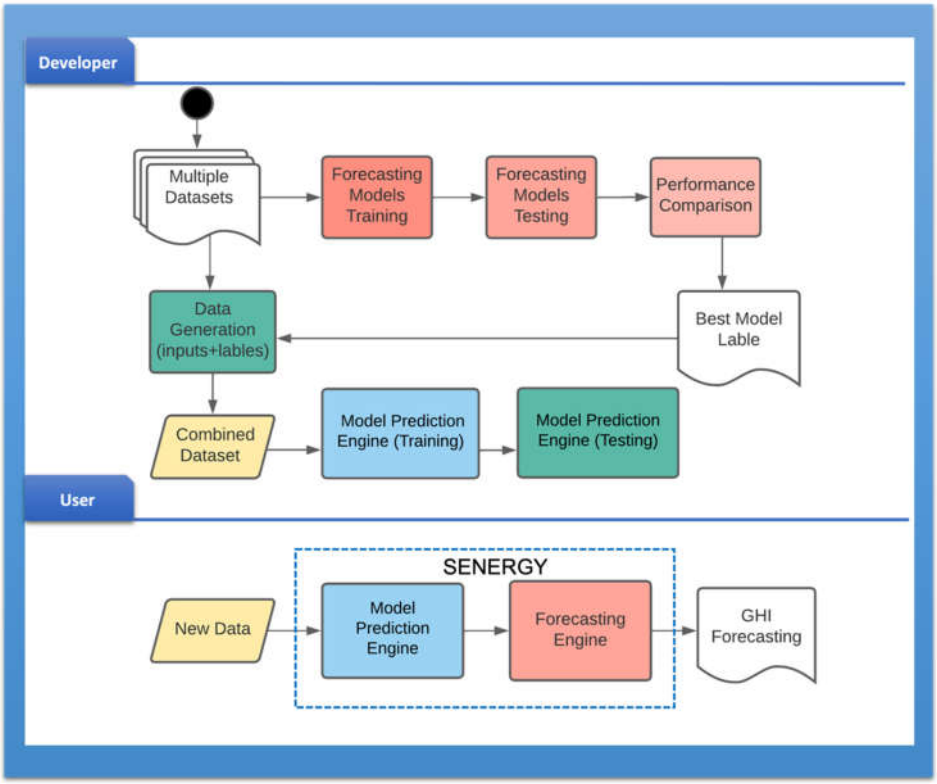


Figure 4. SENERGY development process.

3.2. *Datasets Development*

Here in this section, we describe first the data collection process (3.2.1), then, data preprocessing and feature engineering done for forecasting (3.2.2) and model prediction 3.2.3).

3.2.1. Data Collection

We used a total of ten datasets. Eight of them were collected from Solar monitoring stations in Saudi Arabia and the remaining two are Toronto dataset and Caracas dataset. The used datasets represent three different climates and contain records of five years, which ensure the inclusion of all various weather types, such as sunny, cloudy, rainy, etc. The datasets of Saudi Arabian locations were provided by King Abdullah City for Atomic and Renewable Energy (K.A.CARE)[46]. They contain the measurements of three components of solar radiation: Direct Normal Irradiance (DNI), Global Horizontal Irradiance (GHI), and Diffuse Horizontal Irradiance (DHI), in addition to related meteorological parameters. The datasets cover the period from 1 January 2016 to 31 December 2020. Ideally, each dataset should contain the observations of 1827 days (5 years) averaged into one-hour intervals. However, some days' observations are not available because of device malfunction or maintenance scheduling. The ground-based measurements were taken at eight Tier 1 solar monitoring stations with a resolution of 1 minute. Tier 1 stations provide the highest quality data, with the uncertainty of +/- 2% (sub-hourly). Table 2 presents information about these solar monitoring stations including the station name, latitude, longitude, and elevation. The climate classification of all locations is hot desert climate (BWh) according to Köppen classification obtained from ClimateCharts.net [47]. Figure 5 shows the solar stations' location on the Saudi Arabia map.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

Table 2. Saudi Solar monitoring stations information.

Station #	Station Name	Latitude (N)	Longitude (E)	Elevation (m)
1	Al-Baha University	20.1794	41.6357	1680
2	Al-Jouf College of Technology	29.77634	40.02318	680
3	Saline Water Conversion Corporation (Al-Khafji)	28.50676	48.45513	13
4	Arar Technical Institute	31.0274	40.90642	583
5	Hail College of Technology	27.65261	41.70826	928
6	Tabuk University	28.38287	36.48396	781
7	Taif University	21.43278	40.49173	1518
8	Wadi-Addawasir College of Technology	20.43008	44.89433	671

Figure 5. Solar monitoring stations' locations on Saudi Arabia map.

The datasets of Toronto, Canada; and Caracas, Venezuela were collected from National Solar Radiation Database accessed through the National Renewable Energy Laboratory (NREL) website [48]. These datasets were gathered by geostationary satellites unlike Saudi datasets, which were collected from ground stations. The climate classification of Toronto is humid continental (Dfb) and of Caracas is tropical (A) according to Köppen classification. Table 3 provides the source information of both datasets and Figure 6 shows Caracas and Toronto locations on the map.

Table 3. External datasets source information.

Location	Latitude (N)	Longitude (E)	Elevation (m)	Climate Class
Caracas, Venezuela	10.49	-66.9	942	A
Toronto, Canada	43.65	-79.38	93	Dfb

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

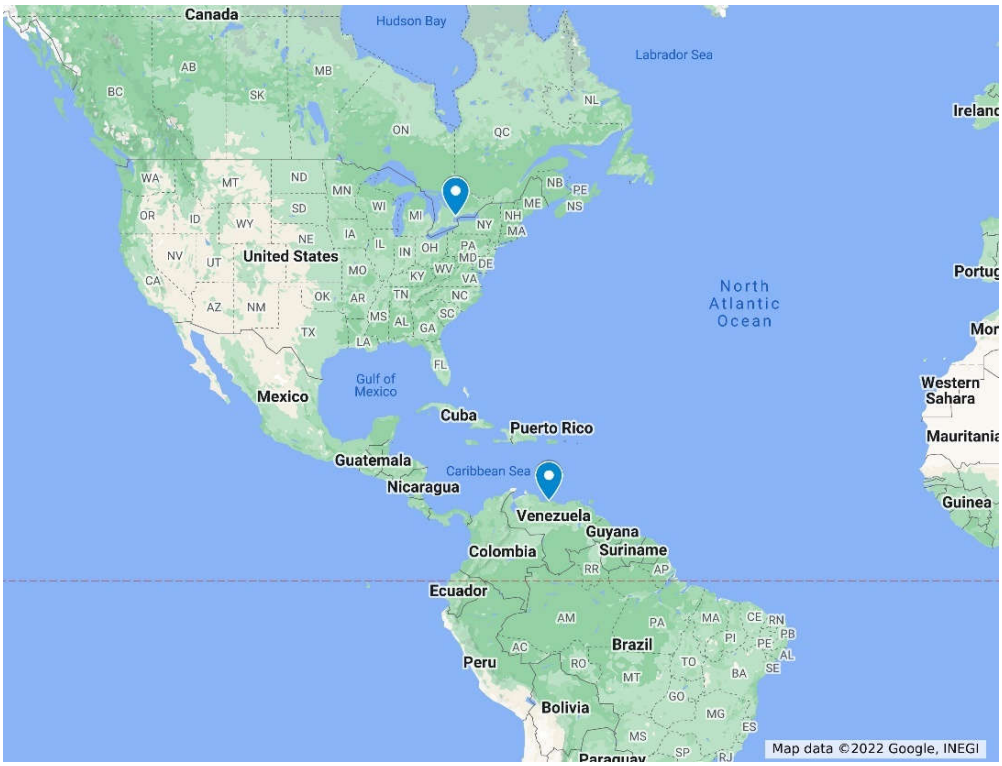


Figure 6. Toronto and Caracas locations on the map.

3.2.2. Datasets for Forecasting

In this section, a description of data preprocessing for forecasting is given. First, the data variables and the relationships among them are clarified, then, creating lagged features and Temporal features steps are explained. Next, filling missing values, deleting night hours records, and data normalization steps are described. Finally, detailed information about each dataset is given.

For GHI forecasting, researchers usually use historical values of GHI alone as inputs to make a prediction or include other meteorological variables, such as wind speed and air temperature. Sometimes forecasted values of the meteorological variables and GHI are also used as inputs, such as Numerical weather prediction (NWP) models' outputs [17]. In our work, the following nine measurements are chosen as inputs to GHI forecasting models. Figure 7 shows the relationship between GHI and the nine measurements in three datasets only as an example (Al-Baha, Al-Jouf, and Hail datasets).

- GHI: the total amount of shortwave radiation received from above by a surface horizontal to the ground. It is calculated using the following equation, which explains how GHI is related to DHI, DNI, and the Zenith Angle (ZA) [49].

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

- DHI: solar radiation that does not arrive on a direct path from the sun, but has been scattered by molecules and particles in the atmosphere and comes equally from all directions
- DNI: solar radiation that comes in a straight line from the direction of the sun at its current position in the sky. On a sunny day, GHI consists of 20% DHI and 80% DNI [49].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

- ZA: the angle between the sun's rays and the vertical.
- Air Temperature (AT). It has a positive correlation with solar radiation [50] as can be seen in Figure 7
- Wind Speed (WS) and Wind Direction (WD) at 3 meters
- Barometric Pressure (BP)
- Relative Humidity (RH). It has a negative correlation with solar radiation [50] as shown in Figure 7

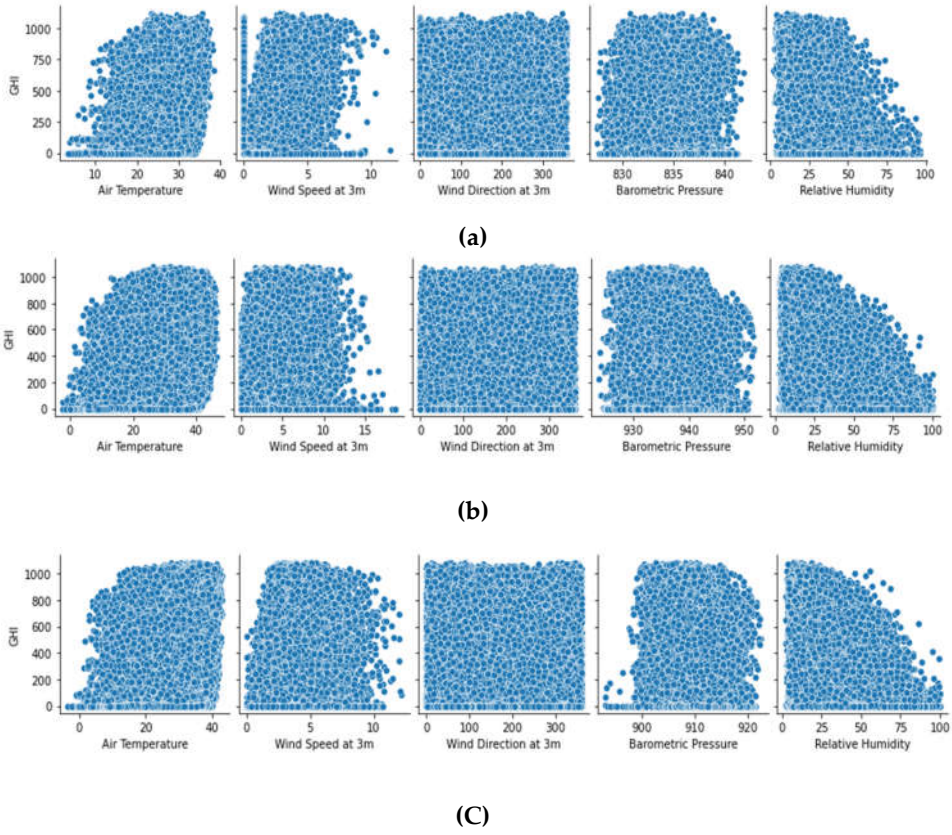


Figure 7. The relationship between GHI and the meteorological variables in: (a) Al-Baha; (b) Al-Jouf; and (c) Hail.

Using the previous three hours measurements (lag= 3 hours), we created a set of twenty-seven features. Table 4 shows the list of these features along with their unit. To create the lagged features, we used the shift method in Pandas library. Table 5 shows an example of using the shift method with GHI values to create lagged features. To guide the decision about the lag, we utilized the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for GHI as presented in Figure 8. The ACF shows a correlation of GHI with its 3 past values while the PACF shows a high correlation of GHI with its first lag only. However, such functions can measure only the linear relationship between an observation at time t and the observations at previous times.

Table 4. Forecasting datasets features.

Time t features	Time t-1 features	Time t-2 features	Time t-3 features	Unit
GHI (output)	GHI_lag1	GHI_lag2	GHI_lag3	Wh/m2
Hour_sin (HS)	DNI_lag1	DNI_lag2	DNI_lag3	Wh/m2

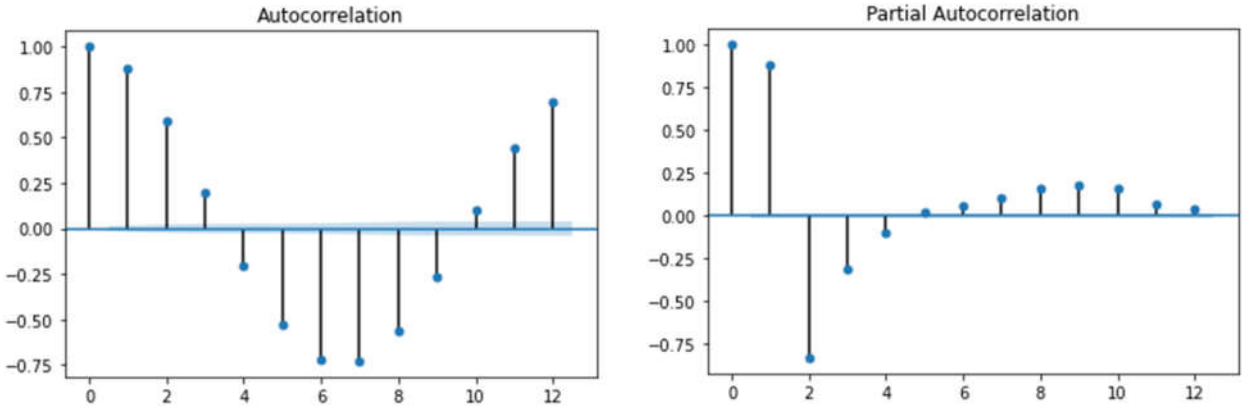
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

Hour_cos (HC)	DHI_lag1	DHI_lag2	DHI_lag3	Wh/m2
Day_sin (DS)	AT_lag1	AT_lag2	AT_lag3	° C
Day_cos (DC)	ZA_lag1	ZA_lag2	ZA_lag3	°
Month_sin (MS)	WS_lag1	WS_lag2	WS_lag3	m/s
Month_cos (MC)	WD_lag1	WD_lag2	WD_lag3	°
	RH_lag1	RH_lag2	RH_lag3	%
	BP_lag1	BP_lag2	BP_lag3	Pa (Saudi data)/ Millibar (others)

* Wh: watt-hour; m: meter; C: Celsius; s: second; Pa: pascal.

Table 5. Example of creating lagged features of GHI.

Tim stamp e	GHI at t	GHI at t-1	GHI at t-2	GHI at t-3
01/01/2016 7:00	0	0	0	0
01/01/2016 8:00	35.3	0	0	0
01/01/2016 9:00	236.2	35.3	0	0
01/01/2016 10:00	468.8	236.2	35.3	0
01/01/2016 11:00	609.6	468.8	236.2	35.3
01/01/2016 12:00	688.7	609.6	468.8	236.2
01/01/2016 13:00	686.8	688.7	609.6	468.8
01/01/2016 14:00	635.6	686.8	688.7	609.6
01/01/2016 15:00	522.7	635.6	686.8	688.7
01/01/2016 16:00	361.3	522.7	635.6	686.8
01/01/2016 17:00	166.2	361.3	522.7	635.6
01/01/2016 18:00	15.6	166.2	361.3	522.7



(**a**) (**b**)

Figure 8. (a) Autocorrelation Function; (b) Partial Autocorrelation Function of GHI and its lagged readings.

Temporal variables (month, day, hour) of the forecasting time (t) are also important inputs. Since they have cyclical nature, we decided to encode them into sine and cosine using the following equations [51]. The result of this transformation is additional six features (hour sine, hour cosine, day sine, day cosine, month sine, month cosine). The total features used for training the forecasting models are thirty-three as shown in Table 4.

$$\widetilde{X} = sin\left(\frac{2.\pi.X}{max(X)}\right) \quad (2)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

$$\widetilde{X} = cos\left(\frac{2.\pi.X}{max(X)}\right) \quad (3)$$

As mentioned earlier, there are missing records for many days in the Saudi datasets. To consider that during input-output construction, we eliminated any hour record that does not have the previous three consecutive hours' records [52]. Records of the years 2016, 2017, and 2018 were used for training while records of the years 2019 and 2020 were used for validation and testing respectively. However, for Arar, Al-Khafji, and Tabuk datasets, the number of missing days is large. Therefore, records of the year 2020 and the first four months of 2021 were used for testing sets of these locations. In Wadi-Addawasir, Arar, and Al-Baha dataset, a few DHI values are missing, and they were filled by Equation (1). Many values of wind direction and wind speed are missing in Wadi-Addawasir, Tabuk, and Taif datasets. However, the interpolation method cannot be used to fill these values because they are for consecutive hours. For such situation, usually, researchers in this field either use a regression model to predict the missing values or use another source of data, like nearby station [53]–[55]. Since regression model accuracy might affect the data quality, we decided to use a nearby station data to fill the missing wind speed/direction values in Wadi-Addawasir and Taif datasets. The source of such data is King Abdullah Petroleum Studies and Research Center (KAPSARC) [56]. The number of hourly records filled in Wadi-Addawasir dataset is 11978 hours while it is 7630 in Taif datasets. On the other hand, Tabuk dataset has only 529 missing hours' records. Therefore, we decided to eliminate these records since the year 2021 records are added to the dataset to compensate for the shortage. Comparing methods for filling missing values and studying their impact on the forecasting results, as done in [57], [58], would be an opportunity for future work.

Preprocessing steps also include deleting the records in which GHI equals zero, which represent nighttime hours. Moreover, all features were normalized to the range of [0,1] by min-max scaler, then denormalized to the normal range after the training process was completed. Table 6 presents information about each dataset including the total hourly records used for training, validation, and testing in addition to the number of missing days out of five years. It also indicates the mean, Standard Deviation (SD), and Variance (Var) of GHI of training, validation, and testing datasets.

(**a**) (**b**)

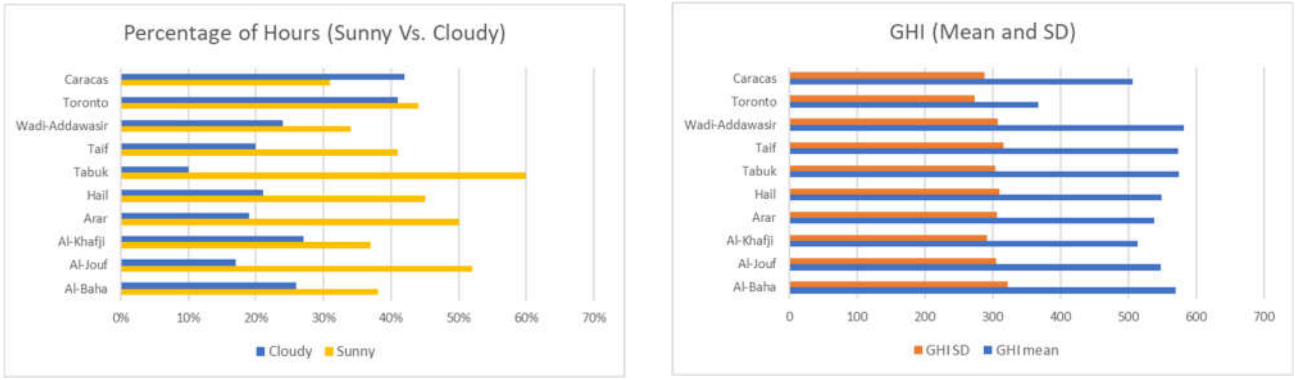
Figure 9 shows the percentage of cloudy and sunny hours of all datasets on the left chart while GHI mean and GHI SD are shown on the right chart.

Table 6. Forecasting datasets information.

Location	Total Hourly Records	Missing Days	GHI Mean	GHI SD	GHI Var
Al-Baha	Train: 6227	635 days	574.67	323.90	104896.29
	Val: 3056		552.10	325.90	106176.30
	Test: 2247		582.09	311.16	96780.11
Al-Jouf	Train: 8600	363 days	554.11	307.66	94643.25
	Val: 2991		547.92	306.49	93901.92
	Test: 2554		528.14	296.47	87858.12
Al-Khafji	Train: 4618	970 days (Year 2019)	504.81	288.56	83245.88
	Val: 2363		555.17	308.66	95231.29
	Test: 2110		486.59	275.73	75991.13
Arar	Train: 8339	575 days	546.71	310.06	96128.23

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

	Val: 3589		537.73	300.20	90097.23
	Test: 1357		485.46	295.04	86983.40
Hail	Train: 8723	271 days	552.26	311.69	97140.65
	Val: 3260		544.05	310.67	96486.20
	Test: 2561		543.77	303.82	92270.30
Tabuk	Train: 7576	542 days	593.27	310.35	96307.42
	Val: 3100		579.62	303.93	92342.88
	Test: 1937		498.03	261.73	68465.05
Taif	Train: 8618	272 days	580.83	321.62	103424.30
	Val: 3386		562.14	308.42	95094.37
	Test: 2543		567.62	308.47	95115.01
Wadi-Adda-wasir	Train: 9199	242 days	584.98	309.00	9547422
	Val: 3450		579.24	306.12	93684.80
	Test: 2551		578.02	301.69	90982.42
Caracas	Train: 10112	0 days	499.28	284.48	80922.07
	Val: 3428		505.95	288.71	83327.90
	Test: 3428		524.82	297.12	88255.24
Toronto	Train: 9892	0 days	381.15	273.39	74732.91
	Val: 3392		336.74	266.95	71242.70
	Test: 3388		366.77	278.11	77322.36
All	Train: 81904	3870 days	-	-	-
	Val: 32015				
	Test: 24676				



(**a**) (**b**)

Figure 9. (a) Percentage of hours (sunny Vs. cloudy) of 10 datasets; (b) GHI (mean and SD) of 10 datasets.

3.2.3. Datasets for Model Prediction

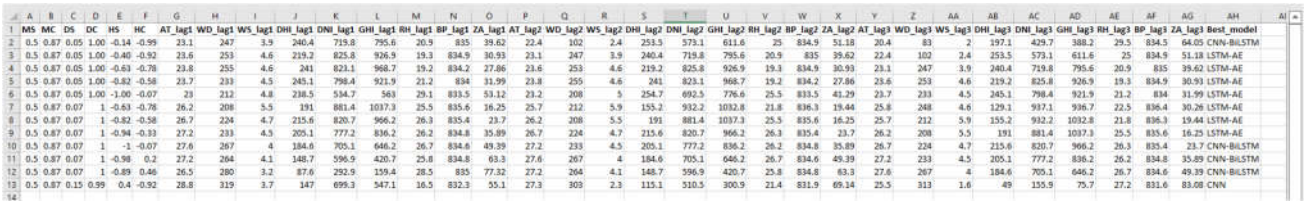
Preparing data for Auto-Selective Model Prediction Engine starts by combining all the ten datasets into one dataset, then adding a new column called "Best model" to the thirty-three features listed in Table 4. To determine the "Best model" for each record, first, we calculated the forecasting error of each model using Equation (4), which represents the absolute value of the difference between the actual GHI and the forecasted GHI. The best forecasting model for each record will be the model that achieves the least forecasting error.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

$$\text{Forecasting error} = |\text{actual GHI} - \text{forecast GHI}| \quad (4)$$

Figure 10 shows a snapshot of a few data records after adding "Best model" feature to the thirty-three features used for forecasting. Then, we used label encoding to convert this column to numeric values (0 for the CNN-BiLSTM model, and 1 for the LSTM-AE model). The total records used with the Auto-Selective Model Prediction Engine is 24576 (80% of them used for training and 20% for testing). The class distribution is 23% and 77% for the CNN-BiLSTM model, and LSTM-AE model respectively.

Figure 10. Snapshot of data inputs of Best Forecaster Recommendation model.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	MS	MC	DS	DC	HS	HC	AT_lag1	WD_lag1	WS_lag1	DHI_lag1	DNI_lag1	GHI_lag1	RH_lag1	BP_lag1	ZA_lag1	AT_lag2	WD_lag2	WS_lag2	DHI_lag2	DNI_lag2	GHI_lag2	RH_lag2	BP_lag2	ZA_lag2	AT_lag3	WD_lag3	WS_lag3	DHI_lag3	DNI_lag3	GHI_lag3	RH_lag3	BP_lag3	ZA_lag3	Best_model	
2	0.5	0.87	0.05	1.00	-0.14	-0.95	23.1	247	3.9	240.4	719.8	795.6	20.9	835	39.62	22.4	102	2.4	253.5	573.1	611.6	25	834.9	51.18	20.4	83	2	197.1	429.7	388.2	29.5	834.5	64.05	CNN-BiLSTM	
3	0.5	0.87	0.05	1.00	-0.40	-0.92	23.6	253	4.6	219.2	825.8	926.9	19.3	834.9	30.93	23.1	247	3.9	240.4	719.8	795.6	20.9	835	39.62	22.4	102	2.4	253.5	573.1	611.6	25	834.9	51.18	LSTM-AE	
4	0.5	0.87	0.05	1.00	-0.63	-0.78	23.8	255	4.6	241	823.1	968.7	19.2	834.2	27.86	23.6	253	4.6	219.2	825.8	926.9	19.3	834.9	30.93	23.1	247	3.9	240.4	719.8	795.6	20.9	835	39.62	LSTM-AE	
5	0.5	0.87	0.05	1.00	-0.82	-0.58	23.7	233	4.5	245.1	798.4	921.9	21.2	834	31.99	23.8	255	4.6	241	823.1	968.7	19.2	834.2	27.86	23.6	253	4.6	219.2	825.8	926.9	19.3	834.9	30.93	LSTM-AE	
6	0.5	0.87	0.05	1.00	-1.00	-0.07	23	212	4.8	238.5	534.7	563	29.1	833.5	53.12	23.2	208	5	254.7	692.5	776.6	25.5	833.5	41.29	23.7	233	4.5	245.1	798.4	921.9	21.2	834	31.99	LSTM-AE	
7	0.5	0.87	0.07	1	-0.63	-0.78	26.2	208	5.5	191	881.4	1037.3	25.5	835.6	16.25	25.7	212	5.9	155.2	932.2	1032.8	21.8	836.3	19.44	25.8	248	4.6	129.1	937.1	936.7	22.5	836.4	30.26	LSTM-AE	
8	0.5	0.87	0.07	1	-0.82	-0.58	26.7	224	4.7	215.6	820.7	966.2	26.3	835.4	23.7	26.2	208	5.5	191	881.4	1037.3	25.5	835.6	16.25	25.7	212	5.9	155.2	932.2	1032.8	21.8	836.3	19.44	LSTM-AE	
9	0.5	0.87	0.07	1	-0.94	-0.33	27.2	233	4.5	205.1	777.2	836.2	26.2	834.8	35.89	26.7	224	4.7	215.6	820.7	966.2	26.3	835.4	23.7	26.2	208	5.5	191	881.4	1037.3	25.5	835.6	16.25	LSTM-AE	
10	0.5	0.87	0.07	1	-1	-0.07	27.6	267	4	184.6	705.1	646.2	26.7	834.6	49.39	27.2	233	4.5	205.1	777.2	836.2	26.2	834.8	35.89	26.7	224	4.7	215.6	820.7	966.2	26.3	835.4	23.7	CNN-BiLSTM	
11	0.5	0.87	0.07	1	-0.98	0.2	27.2	264	4.1	148.7	596.9	420.7	25.8	834.8	63.3	27.6	267	4	184.6	705.1	646.2	26.7	834.6	49.39	27.2	233	4.5	205.1	777.2	836.2	26.2	834.8	35.89	CNN-BiLSTM	
12	0.5	0.87	0.07	1	-0.89	0.46	26.5	280	3.2	87.6	292.9	159.4	28.5	835	77.32	27.2	264	4.1	148.7	596.9	420.7	25.8	834.8	63.3	27.6	267	4	184.6	705.1	646.2	26.7	834.6	49.39	CNN-BiLSTM	
13	0.5	0.87	0.15	0.99	0.4	-0.92	28.8	319	3.7	147	699.3	547.1	16.5	832.3	55.1	27.3	303	2.3	115.1	510.5	300.9	21.4	831.9	69.14	25.5	313	1.6	49	155.9	75.7	27.2	831.6	83.08	CNN	
14																																			

3.3. *Feature Importance*

Nonlinearity and the "black-box" nature of deep learning models make it difficult to explain them and rank features based on importance. In this section, we use four conventional methods for feature selection: Pearson's correlation (3.3.1), Mutual Information (3.3.2), Forward Feature Selection and Backward Feature Elimination (3.3.3), and LASSO (3.3.4). However, we did not eliminate any feature listed in Table 4 based on the results of these four methods since there is no agreement among them. For example, a feature that is considered insignificant by a method would be selected as an important feature by another. Therefore, we used such methods to understand the relationship between variables and provide insight into the data. In Section 4.1.1, the effect of the lagged features on forecasting is studied by training the models using only the first lagged features, then repeating training after adding the second and third lagged features. To present the results of feature importance methods, four or five datasets out of ten were selected for the sake of brevity.

3.3.1. Pearson's Correlation

Pearson's correlation coefficient measures the linear relationship between two variables [59]. Figure 11 displays the correlation matrix for Al-Jouf while the same is displayed for Al-Khafji in Figure 12. The correlation matrices for Caracas and Toronto are shown in Figure 13 and Figure 14. Table 7 lists the most significant correlations between GHI and other features of the five datasets. The strongest positive correlation is between GHI and its last hour value while the strongest negative correlation is between GHI and hour cosine except for Toronto dataset, which is with Zenith Angel of lag 1. From Table 7, it is noticed that almost the same set of important features appear in the five datasets and thus, location or climate has a slight impact on feature correlation. For example, DNI of lag 2 is more important in Toronto dataset than in other datasets.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

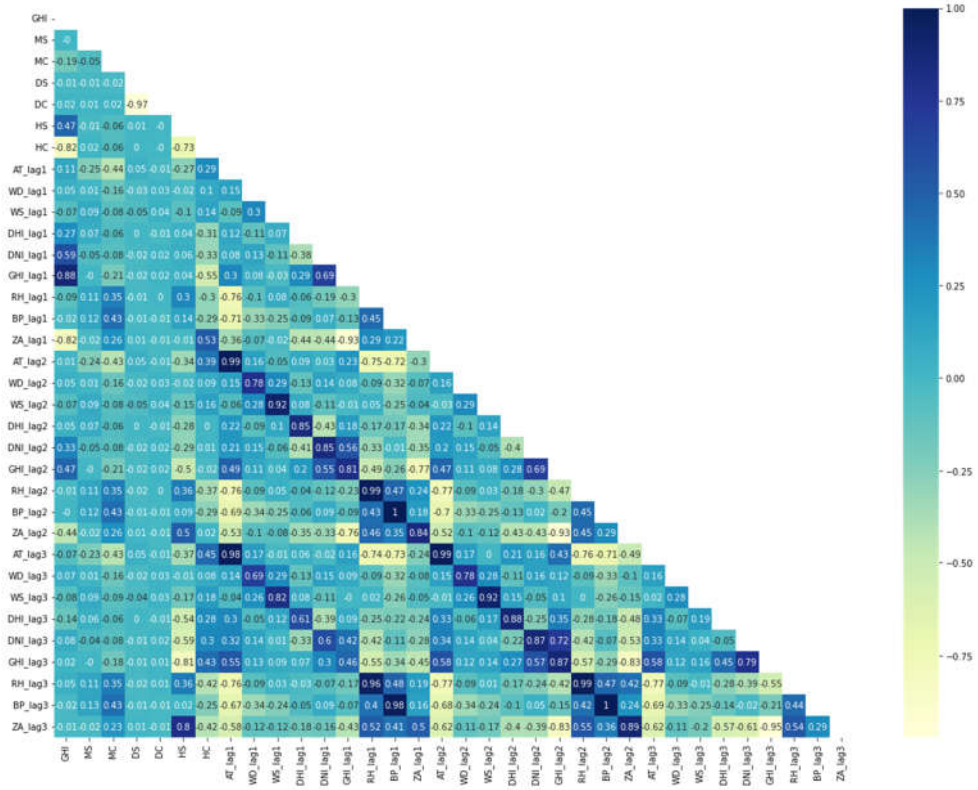


Figure 11. Al-Jouf dataset correlation matrix

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

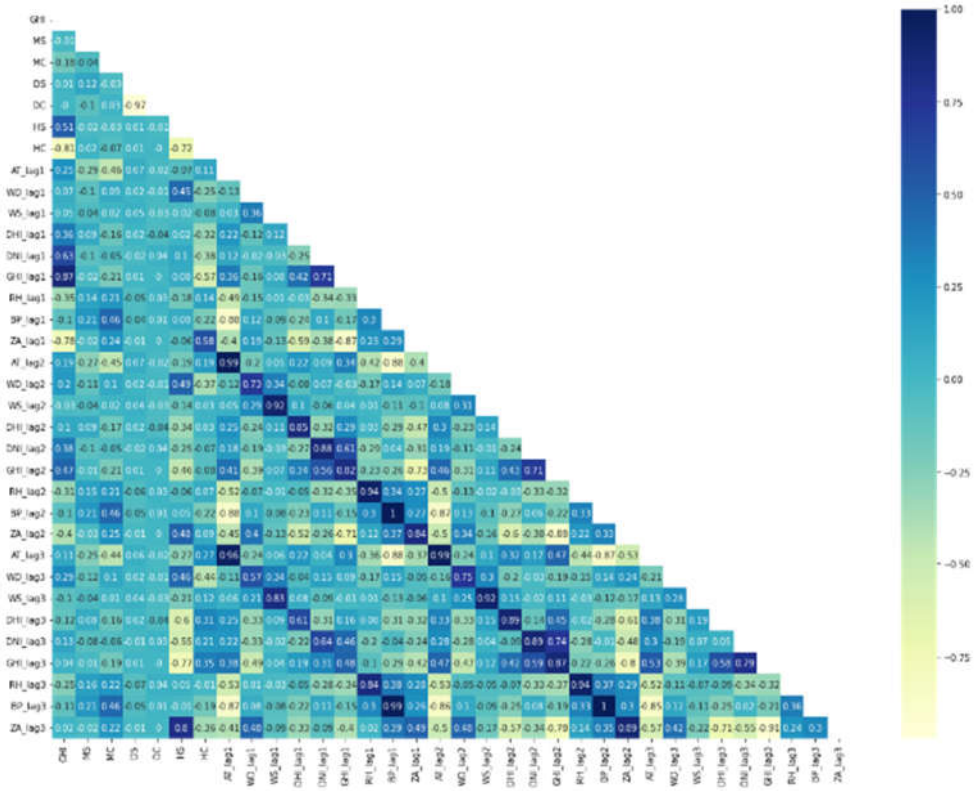


Figure 12. Al-Khafji dataset correlation matrix.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

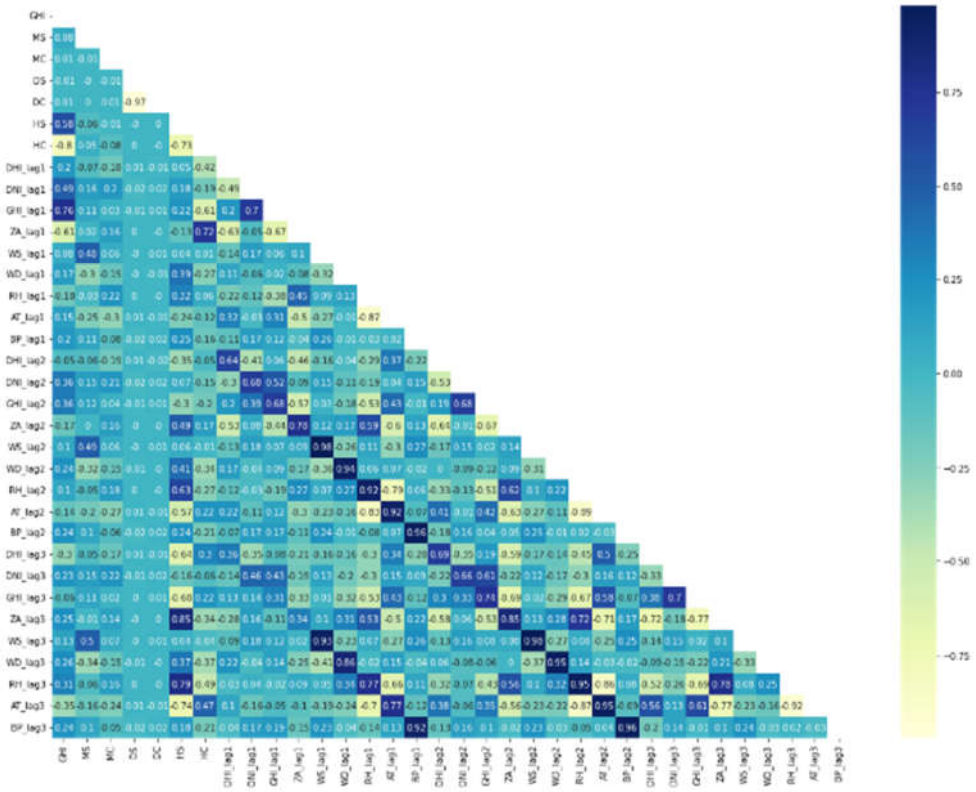


Figure 13. Caracas dataset correlation matrix.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

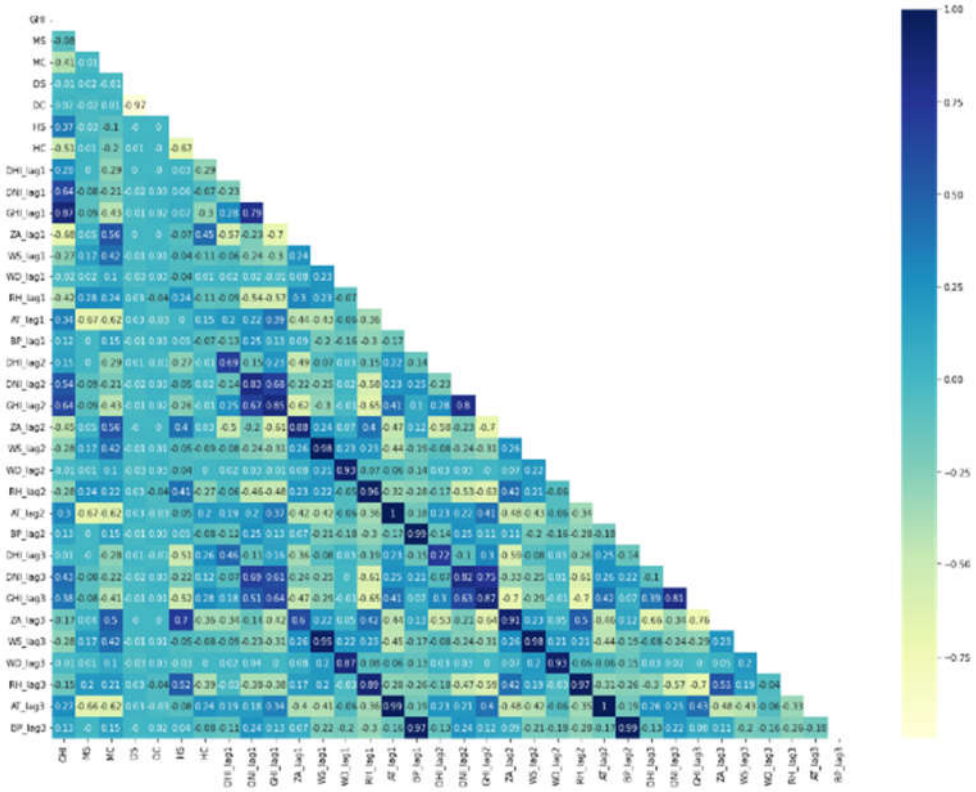


Figure 14. Toronto dataset correlation matrix.

Table 7. Significant Pearson's correlation (PC) between GHI and other features.

Al-Jouf		Al-Khafji		Wadi-Addawasir		Caracas		Toronto	
Feature	PC	Feature	PC	Feature	PC	Feature	PC	Feature	PC
GHI_lag1	0.88	GHI_lag1	0.87	HC	-0.91	HC	-0.80	GHI_lag1	0.87
HC	-0.82	HC	-0.81	GHI_lag1	0.86	GHI_lag1	0.76	ZA_lag1	-0.68
ZA_lag1	-0.82	ZA_lag1	-0.78	ZA_lag1	-0.80	ZA_lag1	-0.61	DNI_lag1	0.64
DNI_lag1	0.59	DNI_lag1	0.63	HS	0.53	HS	0.58	GHI_lag2	0.64
HS	0.47	HS	0.51	DNI_lag1	0.53	DNI_lag1	0.49	DNI_lag2	0.54
GHI_lag2	0.47	GHI_lag2	0.47					HC	-0.51

3.3.2. Mutual Information

Mutual Information (MI) measures the reduction in uncertainty for one variable given a known value of the other variable [60]. Figure 15 shows the MI values of all features for five datasets (Al-Jouf, Al-Khafji, Wadi-Addawasir, Caracas, Toronto). The most significant features for GHI prediction are GHI lagged observations and Zenith Angle lagged observations. Hour sine and cosine are also important in GHI prediction based on MI values. As in the case of Pearson's correlation, location or climate has a slight impact on MI values since the same set of features show significance in the five datasets with small variation. For example, Hour sine and cosine are less important in Toronto dataset

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

than others. GHI and Zenith Angle lagged observations are more important in Saudi locations than in Caracas or Toronto.

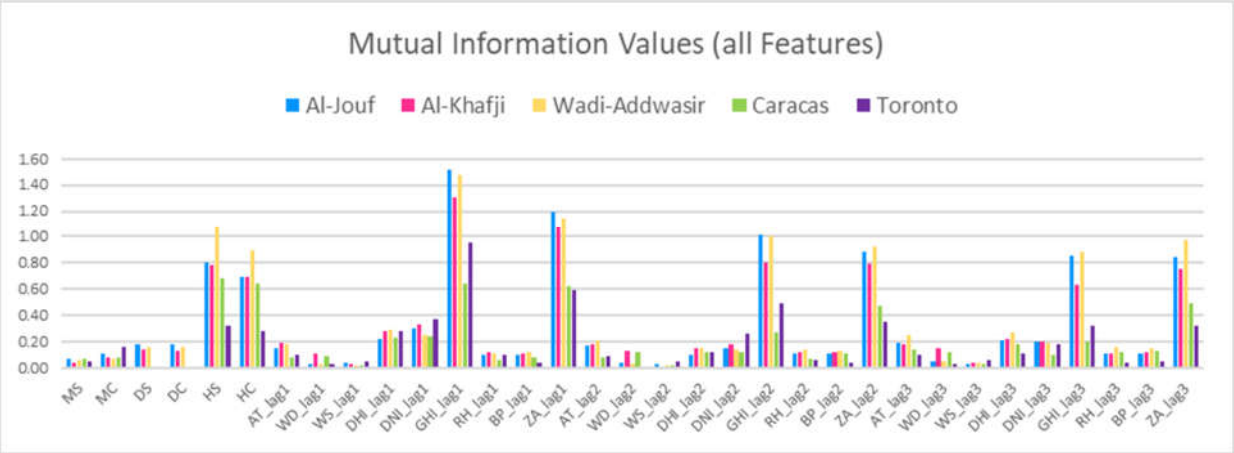


Figure 15. MI values of all features for Al-Jouf, Al-Khafji, Wadi-Addwasir, Caracas, and Toronto datasets.

3.3.3. Forward Feature Selection (FFS) and Backward Feature Elimination (BFE)

Forward feature selection is an iterative method, which starts with no feature in the model. In each iteration, the feature which best improves the model is added till an addition of a new variable does not improve the performance of the model. Backward feature elimination on the other hand starts with all the features and removes the least significant feature at each iteration. This process is repeated until no improvement is observed with feature removal [61]. Table 8 shows ten selected features by FFS & BFE for five datasets (Al-Jouf, Al-Khafji, Wadi-Addawasir, Caracas, Toronto). Features selected by both methods for the same dataset are italicized. From Table 8, we can see that some features are selected in all datasets, such as Hour sine, DHI, DNI, and GHI lagged observations while other features are rarely selected like Wind Speed, Relative Humidity, Barometric Pressure, and Air Temperature.

Table 8. Selected features by FFS & BFE.

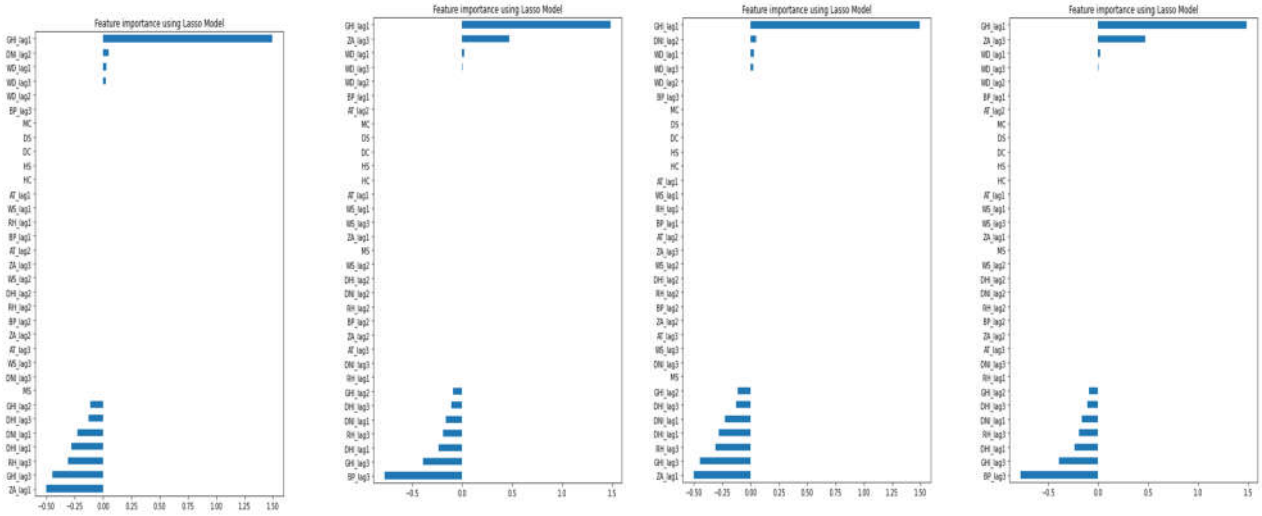
Al-Jouf		**Al-Khafji**		**Wadi-Addawasir**		**Caracas**		**Toronto**	
FFS	BFE	FFS	BFE	FFS	BFE	FFS	BFE	FFS	BFE
HS	HS	HS	HS	HS	HS	HS	HS	MS	HS
HC	HC	WS_lag1	DHI_lag1	HC	HC	HC	HC	HS	DHI_lag1
DHI_lag1	DHI_lag1	DHI_lag1	DNI _lag1	DHI_lag1	DHI_lag1	DHI_lag1	DHI_lag1	GHI_lag1	DNI_lag1
DNI_lag1	DNI_lag1	DNI_lag1	GHI_lag1	DNI_lag1	DNI _lag1	DNI_lag1	DNI_lag1	ZA_lag1	GHI_lag1
GHI_lag1	GHI_lag1	GHI_lag1	BP_lag1	GHI_lag1	GHI_lag1	GHI_lag1	GHI_lag1	WS_lag1	AT_lag1
ZA_lag1	ZA_lag1	ZA_lag1	DHI_lag2	DHI_lag2	DHI_lag2	RH_lag1	RH_lag1	WS_lag3	ZA_lag2
DHI_lag2	DHI_lag2	DHI_lag2	DNI_lag2	ZA_lag3	ZA_lag3	GHI_lag2	DNI_lag2	DNI_lag2	DHI_lag3
DNI_lag2	DNI_lag2	DHI_lag3	GHI_lag2	GHI_lag3	GHI_lag2	ZA_lag2	ZA_lag2	GHI_lag3	DNI_lag3
GHI_lag3	GHI_lag2	DNI_lag3	ZA_lag2	ZA_lag1	ZA_lag2	WS_lag3	WS_lag3	ZA_lag3	GHI_lag3
ZA_lag3	AT_lag1	GHI_lag3	GHI_lag3	RH_lag1	DNI_lag2	AT_lag3	AT_lag3	RH_lag3	AT_lag3

3.3.4. LASSO Feature Selection

The LASSO method regularizes model parameters by shrinking the regression coefficients, reducing some of them to zero. The feature selection phase occurs after the shrinkage, where every non-zero value is selected to be used in the model [62]. Figure 16 shows the selected features based on the LASSO method for five datasets (Al-Jouf, Al-Khafji,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

Wadi-Addawasir, Caracas, Toronto). The most significant features in all five datasets are GHI lagged observations, especially the last hour value (GHI_lag1), and last hour Zenith Angle value (ZA_lag1). DHI and DNI lagged observations are less important. Relative Humidity seems to have importance in Caracas dataset. Surprisingly, time-related features are insignificant in all five datasets.



(**a**) (**b**) (**c**) (**d**)

Figure 16. Selected features based on the LASSO method: (a) Al-Jouf; (b) Al-Khafji; (d) Caracas; (e) Toronto.

3.4. Models Development

In this section, five deep learning models have been explained: Long Short-Term Memory (3.4.1), Gated Recurrent Unit (3.4.2), Convolutional neural network (3.4.3), Hybrid CNN and Bidirectional LSTM (3.4.4), and LSTM Autoencoder (3.4.5). All five models are used for next-hour GHI forecasting while only LSTM model is used also for classification to serve as Auto-Selective Model Prediction Engine.

3.4.1. Long Short-Term Memory (LSTM)

LSTM is a special type of Recurrent Neural Network (RNN) capable of learning long-term dependencies and tends to work much better than traditional RNN on a variety of tasks. In addition to the hidden state, LSTMs have the cell state that carries the relevant information from the earlier steps to later steps. Along the way, the new information is added to or removed from the cell state via input and forget gates. The output gate determines if the current memory cell will be output. More details about LSTM can be found in [5], [63].

An LSTM model for the next hour GHI forecasting is implemented as shown in Figure 17. It consists of three LSTM layers for feature extraction and one dense layer to make GHI prediction. Each LSTM layer has 128 hidden states. Another LSTM model with a similar structure is implemented to work as Auto-Selective Model Prediction Engine with two differences. First, two dense layers are used for classification instead of regression with 8 and 2 neurons respectively. Second, the criterion function is Cross-entropy loss instead of Mean Squared Error loss (MSE).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

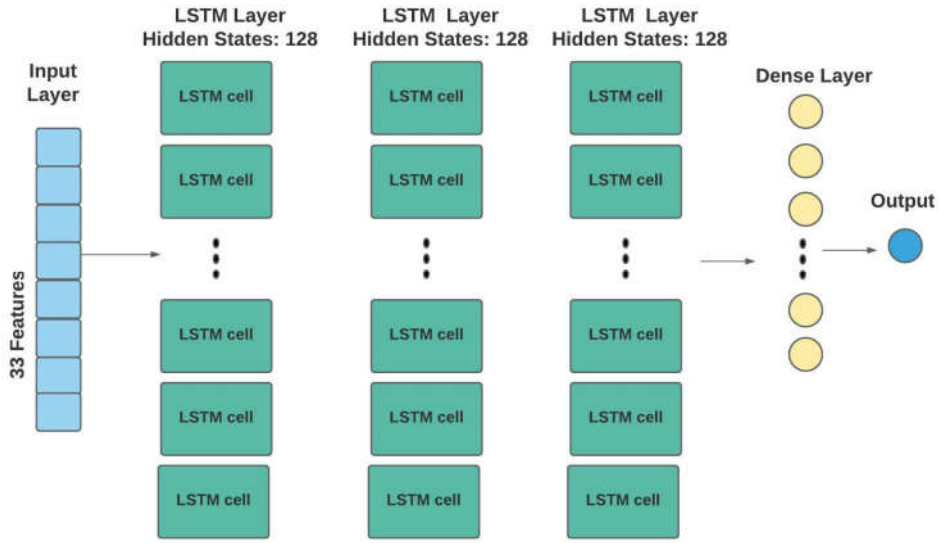


Figure 17. LSTM forecasting model.

3.4.2. Gated Recurrent Unit (GRU)

GRU is like LSTM, it also captures long-term dependencies, but it does it using reset and update gates without any cell state. While the update gate determines how much of the past information needs to be kept, the reset gate decides how much of the past information to forget. GRUs are often faster and require less memory than LSTMs because they require less computation [64].

A GRU model for the next hour GHI forecasting is implemented as shown in Figure 18. It consists of three GRU layers for feature extraction and one dense layer to make GHI prediction. Each GRU layer has 128 hidden states.

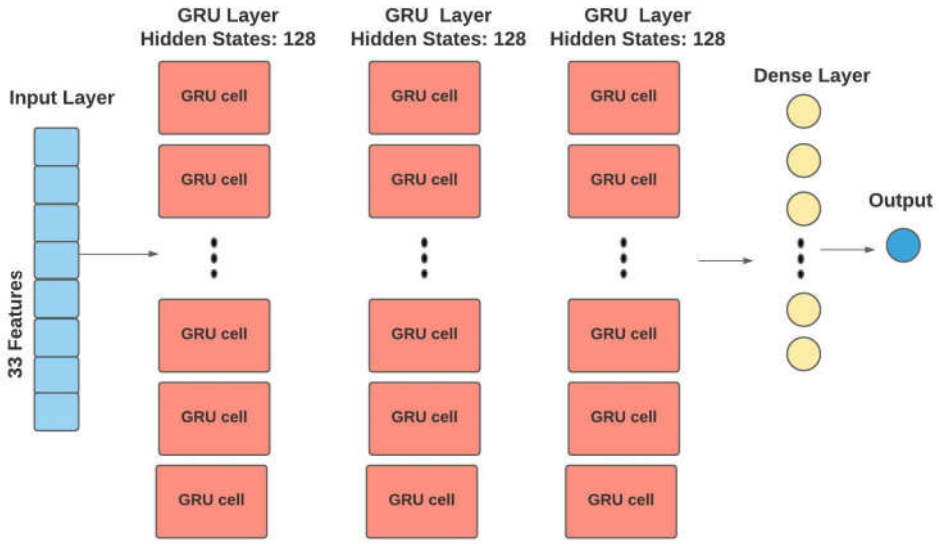


Figure 18. GRU forecasting model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

3.4.3. Convolutional neural network (CNN)

CNN is a type of neural network that is widely known in the computer vision field. It consists of several convolutional and pooling layers followed by fully connected layers. In convolutional layers, feature maps are created by applying convolution filters on inputs while these feature maps are down-sampled in pooling layers. After several convolution and down-sampling operations, features are flattened into 1D and passed to one or more fully connected layers to generate the output. More details about CNN can be found in [65], [66].

A CNN model for the next hour GHI forecasting is implemented as shown in Figure 19. It consists of two 1D-convolutional layers, one max-pooling layer, and two dense layers. In the first convolutional layer, 10 feature maps are created using a kernel of size 2 and stride of 2 while in the second convolutional layer, 5 feature maps are created. The max-pooling layer uses a kernel of size 2 and a stride of 1.

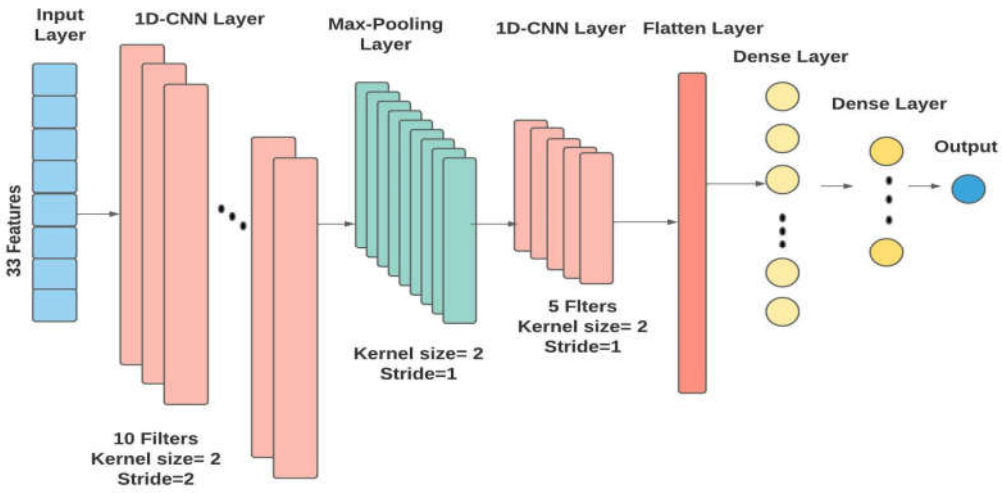


Figure 19. CNN forecasting model.

3.4.4. Hybrid CNN-Bidirectional LSTM (CNN-BiLSTM)

Bidirectional-LSTM (BiLSTM) is an adjusted version of LSTM that contains two layers: one to process inputs in a forward direction, and another to process inputs in a backward direction. This structure allows learning from past and future information. More details about BiLSTM can be found in [67], [68].

In CNN and BiLSTM structure, convolutional and pooling layers are followed by BiLSTM layers, then one or more dense layers to generate the output [69].

A CNN BiLSTM model for the next hour GHI forecasting is implemented as shown in Figure 20. It has the same design as the CNN model illustrated previously with an additional BiLSTM layer placed before the dense layers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

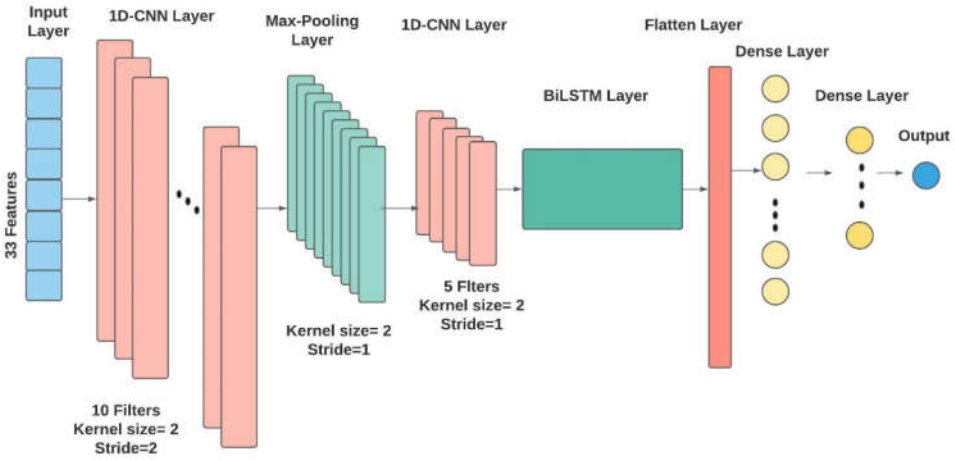


Figure 20. CNN-BiLSTM forecasting model.

3.4.5. LSTM Autoencoder (LSTM-AE)

Autoencoder is a neural network that consists of two parts encoder and decoder. The encoder receives inputs and compresses them into a feature vector called latent space while the decoder decompresses the feature vector into an output. This data reconstruction process helps the model extract the most important features. LSTM Autoencoder model is an Autoencoder in which both the encoder and decoder consist of LSTM layers to learn temporal dependencies in sequence data. More about LSTM-AE can be found in [70], [71].

An LSTM-AE model for the next hour GHI forecasting is implemented as shown in Figure 21. Both the encoder and decoder have two LSTM layers, followed by a dense layer to make GHI prediction.

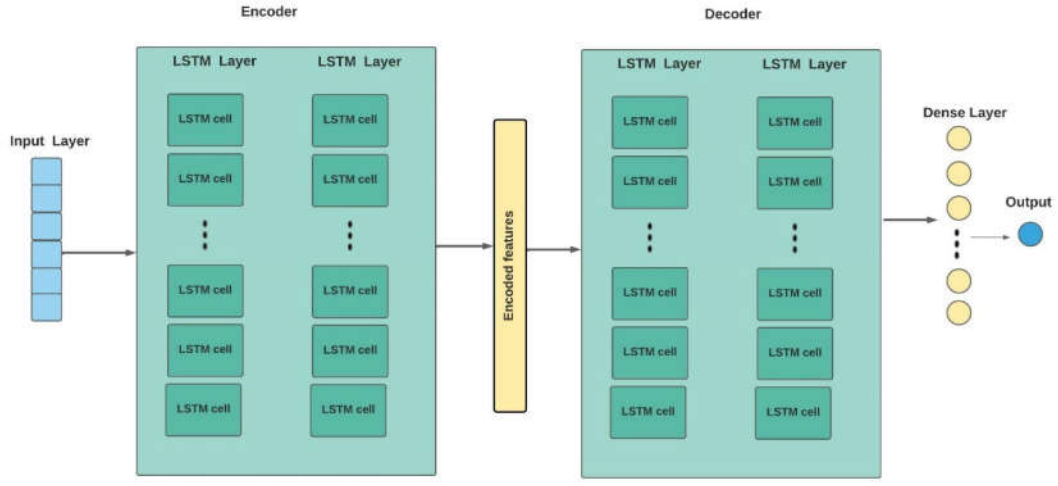


Figure 21. LSTM-AE forecasting model.

3.5. Performance Evaluation Metrics

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

In this paper, six performance evaluation metrics are used to evaluate the forecasting models.

Mean Absolute Error (MAE) is the mean of the absolute values of the individual forecast errors on overall examples (N) in the test set. Each forecasting error is the difference between the actual value (actual GHI) and the forecast value (forecast GHI). A lower value of MAE is better. It is calculated as follows [72].

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}|\text{actual GHI}_{\mathrm{i}} - \text{ forecast GHI}_{\mathrm{i}}| \tag{5}$$

Root Mean Square Error (RMSE) is the standard deviation of the residuals or the forecast errors. It measures how spread out the residuals are and how the data is concentrated around the line of regression. A lower value of RMSE is better. It is calculated as follows [72].

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\text{actual GHI}_{\mathrm{i}} - \text{ forecast GHI}_{\mathrm{i}})^2} \tag{6}$$

Coefficient of determination (R^2) is a statistical measure that determines the proportion of variance in the dependent variable that can be explained by the independent variable. It shows how well the data fit the regression model. R^2 value ranges from 0 to 1 and a higher coefficient indicates a better fit for the model. It is calculated as follows [72].

$$\mathrm{R}^2 = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\text{actual GHI}_{\mathrm{i}} - \text{ forecast GHI}_{\mathrm{i}})^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\text{actual GHI}_{\mathrm{i}} - \overline{\mathrm{GHI}})^2} \tag{7}$$

Mean Absolute Percentage Error (MAPE) is a measure of forecasting accuracy. This percentage indicates the average difference between the forecasted value and the actual value. The smaller the MAPE the better the forecast. It is calculated as follows [73].

$$\mathrm{MAPE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left|\frac{\text{actual GHI}_{\mathrm{i}} - \text{ forecast GHI}_{\mathrm{i}}}{\text{actual GHI}_{\mathrm{i}}}\right| \times 100\% \tag{8}$$

Normalized Metric (nMetric) is used to compare multiple forecasting methods applied to different datasets. The GHI range in a particular location affects the forecast results significantly. nMetric takes this fact into account by dividing the obtained Metric by the mean of GHI of the test dataset as shown in the equation below, which could allow a fairer comparison [6]. Normalization could be applied to any metric, such as MAE, RMSE, and MAPE.

$$\mathrm{nMetric} = \frac{\mathrm{Metric}}{\overline{\mathrm{GHI}}} \tag{9}$$

Forecast Skills (FS) is used to compare a proposed forecasting model performance metric with a reference model performance metric. A commonly used reference model in the literature is the persistence method. The evaluation metric could be RMSE, MAE, or others. FS is calculated as follows [6].

$$\mathrm{FS} = 1 - \frac{\mathrm{Metric}_{\,\mathrm{proposed}}}{\mathrm{Metric}_{\,\mathrm{persistence}}} * 100\ \% \tag{10}$$

Note that for performance analysis in Section 4, we have used both standard and normalized versions of MAE, RMSE, and MAPE.

3.6. Tool Implementation

In this paper, PyTorch, which is an open-source machine learning framework developed by Facebook's AI Research lab, was used as the platform to create deep learning models, where Python3 was employed as the programming language. The experiments

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

were performed on a laptop with Intel Core i7-11800 H CPU, NVIDIA GeForce RTX 3070 GPU, and 16 GB memory. However, all deep learning models were developed using GPU. The hyperparameters used in each model are listed in Table 9 in addition to the optimization methods.

Table 9. Models hyperparameter.

Model	Batch size	Layers	Learning rate	Number of epochs	Optimization
LSTM	256	3 hidden layers with 128 hidden states, 1 dense layer	0.001	100	Dropout= 0.2, ReLU function, Weight decay = 0.000001, Adam
GRU	256	3 hidden layers with 128 hidden states,1 dense layer	0.001	100	Dropout= 0.2, ReLU function, Weight decay = 0.000001, Adam
CNN	64	2 conv layers with 10 and 5 filters, 1 max-pooling layer, 2 dense layers	0.001	100	Dropout= 0.2, ReLU function, Adam, batch normalization
CNN-BiLSTM	64	2 conv layers with 10 and 5 filters, 1 max-pooling layer, 1 BiLSTM layer, 2 dense layers	0.001	100	Dropout= 0.2, ReLU function, Adam, batch normalization
LSTM-AE	256	4 LSTM layers with 128 hidden states, 1 dense layer	0.001	100	ReLU function, weight dcay=0.000001, Adam

4. SENERGY: Results and Evaluation

The performance of the two SENERGY components, Forecasting Engine and Auto-Selective Model Prediction Engine are evaluated in Section 4.1. and 4.2 respectively. The evaluation of both components is analyzed from several aspects, such as climate and location, sunny and cloudy weathers, and summer and winter seasons. Then, the achieved gain and loss in forecasting performance using SENERGY is discussed in Section 4.3. Finally, a comparison of SENERGY performance with other related works is provided in Section 4.4.

4.1. SENERGY: Forecasting Engine Performance

In this section, first, the effect of the lagged features on forecasting is analyzed (4.1.1). Then, the forecasting results of five deep learning models, which are described earlier in Section 3.4 are analyzed here. The analysis is provided using four aspects: climate and location (4.1.2), sunny and cloudy weather (4.1.3), summer and winter seasons (4.1.4), and forecasting error results (4.1.5). The results reported are the average of the evaluation metrics for fifty simulations, which were calculated for unseen data (testing datasets). The size of each testing dataset is given in Table 6 and the used performance evaluation metrics are described in Section 3.5.

4.1.1. Effect of Lagged Features on Forecasting

In Section 3.2.2, we explained how lagged features were created and why we decided to use lag equal to 3 (the last three hours of observations). In this Section, we use Toronto dataset to study the effect of using lag equals 1, 2, and 3 to examine the effect of such different lags on the forecasting results. Figure 22 shows the difference in MAE, RMSE, and MAPE for the five forecasting models when using lag equals to 1, 2, and 3 with Toronto dataset. With LSTM, GRU, and CNN-BiLSTM model, it is noticed that using lag 3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

made the results slightly worse, except with MAPE, which had improved. In contrast, the LSTM-AE model achieved better results in all error metrics with lag 2 over lag 1 and the best results with lag 3. Given the fact that GHI is only highly correlated with GHI for lag 1 (see Table 7), using lag equals 1 would give satisfactory results, especially if dimensionality might affect the model efficiency. Otherwise, it is worth trying different lagged features to see if that would result in better performance as in the case of the LSTM-AE model, especially because the climate in the data source might have an effect as well.

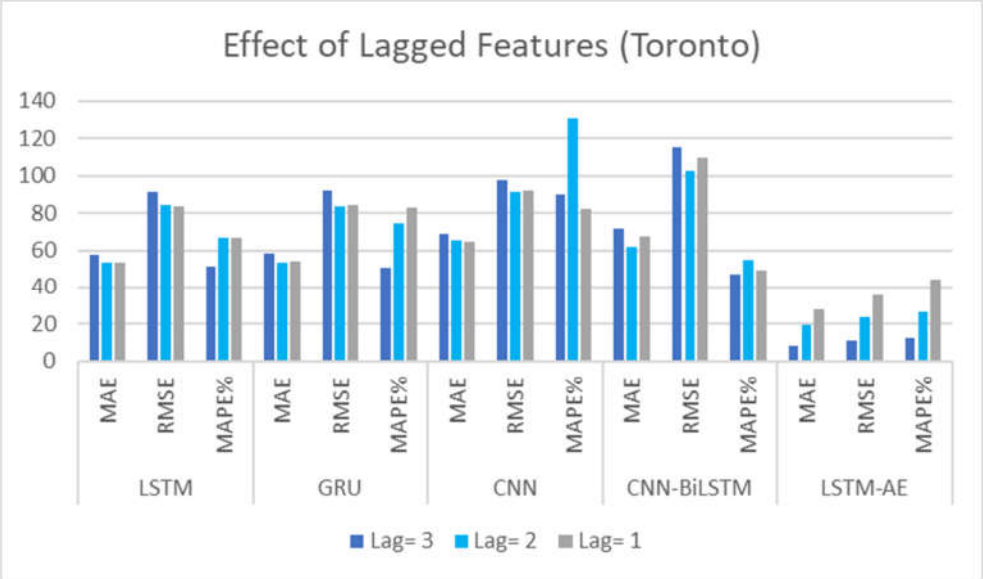
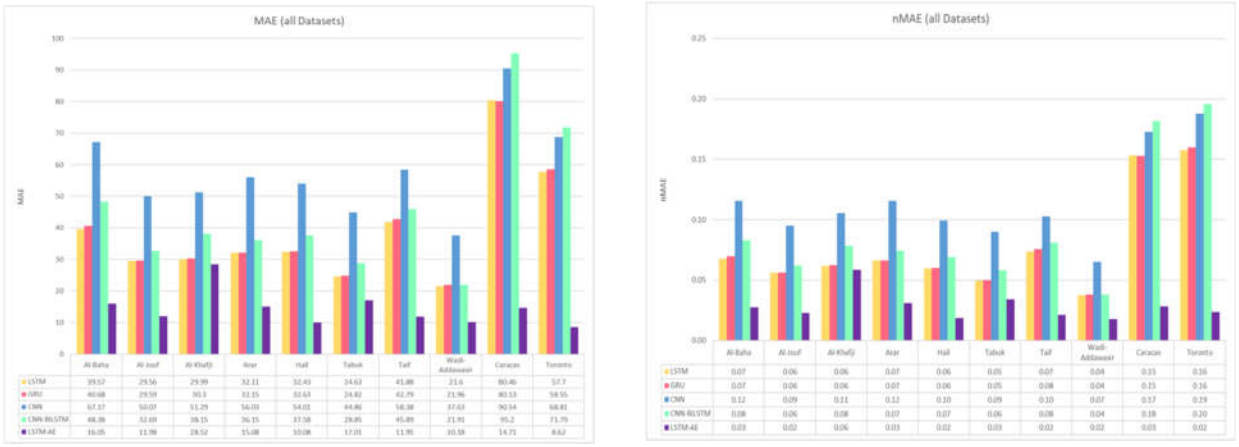


Figure 22. The effect of the lagged features on Toronto dataset.

4.1.2. Effect of Climate and Location on Forecasting

The performance of five deep learning-based forecasting models (LSTM, GRU, CNN, CNN-BiLSTM, and LSTM-AE) is compared in this Section for all the ten datasets for the task of next-hour GHI prediction. Forecasting results using MAE metric and its normalized value are plotted in Figure 23. From the figure, we can see that the best MAE and nMAE values are associated with Wadi-Addwasir while the worst values are associated with Caracas and Toronto, except for LSTM-AE model. The high performance related to Wadi-Addwasir dataset might be attributed to the completeness of this dataset compared to other Saudi datasets since it has the least number of missing days and the largest training set size. In contrast, the low performance associated with Caracas and Toronto datasets might be attributed to the high percentage of cloudy hours (or unclear sky condition) compared to other Saudi locations. The best model according to MAE and nMAE values is LSTM-AE model, which achieves nMAE equal to 0.02 with Wadi-Addwasir and Toronto datasets. This excellent performance is attributed to the ability of the model to reconstruct the inputs into a better representation in addition to extracting the temporal features. On the other hand, the worst performance is associated with CNN model with Saudi datasets while CNN-BiLSTM model is the worst for Caracas and Toronto. With time-series data, the temporal features are the most important features, which cannot be captured by the CNN model.

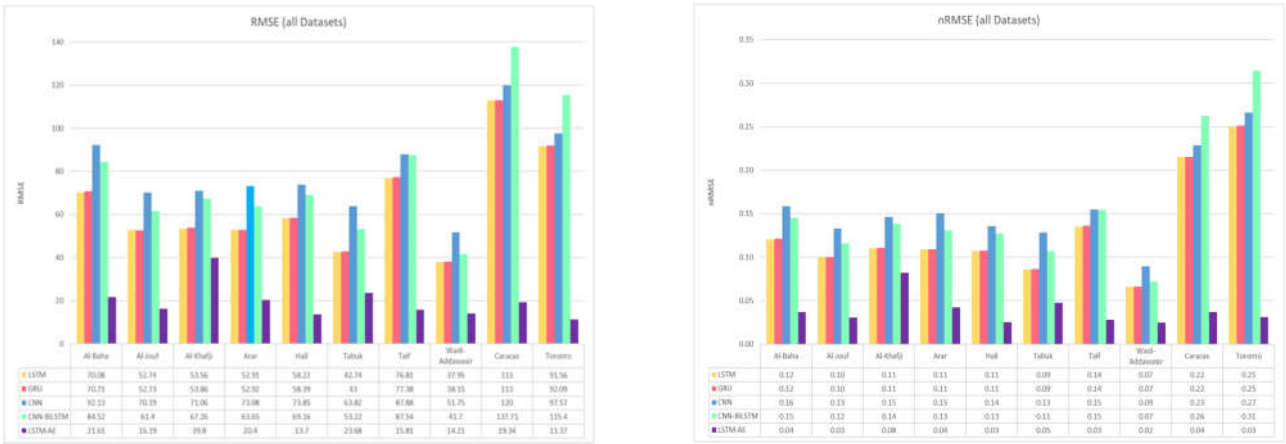
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1



(**a**) (**b**)

Figure 23. Forecasting results of 5 models for all datasets (a) MAE; (b) nMAE.

Forecasting results using RMSE metric and its normalized value are plotted in Figure 24. From the figure, we can see that the best RMSE and nRMSE values are associated with Wadi-Addwasir for all five models. This is also observed earlier with MAE and nMAE results. On the other hand, the worst values are associated with Caracas and Toronto for all models, except for LSTM-AE, which achieved the worst nRMSE value equal to 0.08 with Al-Khafji dataset. We mentioned earlier the advantage of Wadi-Addwasir dataset compared to other Saudi data and the disadvantage of Caracas and Toronto. Regarding Al-Khafji dataset, it has missing data equal to one year, which might explain the low performance of LSTM-AE model here. However, LSTM-AE model is the best model for all locations while CNN is the worst with Saudi datasets and CNN-BiLSTM model is the worst with Caracas and Toronto data. As mentioned earlier, the ability of LSTM-AE to reconstruct the inputs into a better representation in addition to extracting the temporal features might be the reason behind its superior performance.



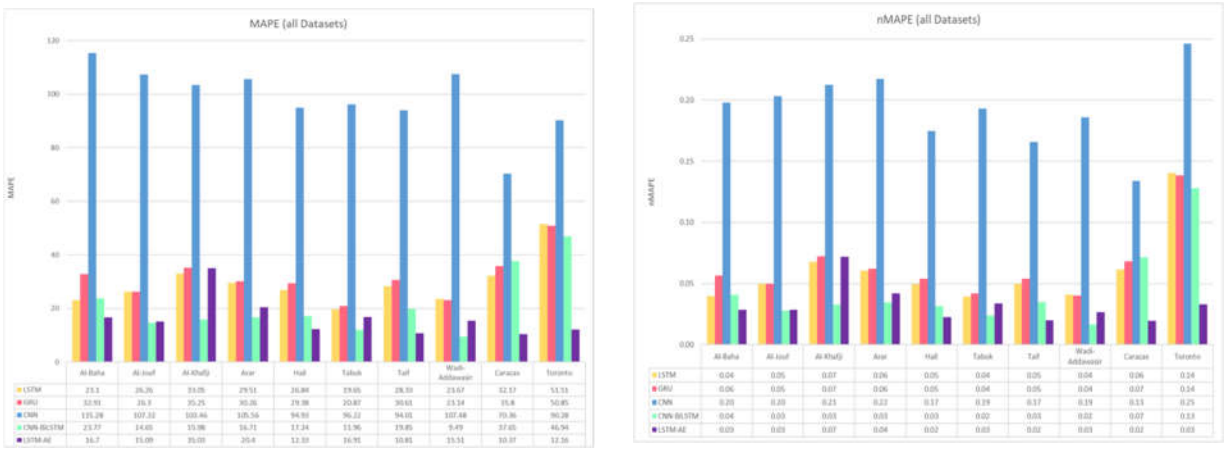
(**a**) (**b**)

Figure 24. Forecasting results of 5 models for all datasets (a) RMSE; (b) nRMSE.

Forecasting results using the MAPE metric and its normalized values are plotted in Figure 25. From the figure, we can see that the location effect on MAPE and nMAPE values are different from what was observed earlier with MAE and RMSE results. For example,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

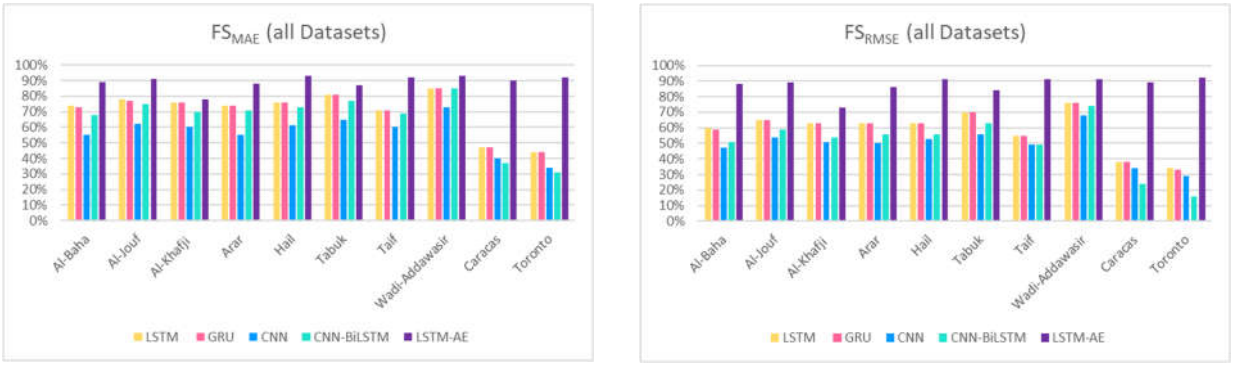
the best nMAPE value for LSTM is 0.04 achieved with Al-Baha, Tabuk, and Wadi-Addwasir while for GRU model it is also 0.04 achieved with Tabuk and Wadi-Addwasir. For CNN, the best nMAPE value is 0.13 achieved with Caracas. For CNN-BiLSTM, the best nMAPE value is 0.02 achieved with Tabuk and Wadi-Addwasir. For LSTM-AE, the best nMAPE value is 0.02 achieved with Hail, Taif, and Caracas. On the other hand, the worst values for all models are associated with Toronto, except for LSTM-AE model, which achieved the worst value, which is 0.07 with Al-Khafji. Comparing models performances, the best is LSTM-AE model for five datasets (Al-Baha, Hail, Taif, Caracas, Toronto) and CNN-BiLSTM model for four datasets (Al-Khafji, Arar, Tabuk, Wadi-Addwasir). Otherwise, the worst is CNN model for all locations. MAPE (refer to Equation (8)) is different from other metrics because it gives the forecasting error relative to the actual GHI, which might explain the different results observed with this metric.



(**a**) (**b**)

Figure 25. Forecasting results of 5 models for all datasets (a) MAPE; (b) nMAPE.

Figure 26 shows the FS results based on MAE and RMSE for all the forecasting models, which represent the performance improvement compared to the persistence method. The best FS results are achieved by the LSTM-AE model, which is 93% in MAE with Hail and Wadi-Addawasir datasets while it is 92% in RMSE with Toronto dataset.



(**a**) (**b**)

Figure 26. Forecasting results of 5 models for all datasets (a) FS_{MAE}; (b) FS_{RMSE}.

 doi:10.20944/preprints202208.0345.v1

In summary, looking at the performance from the models' perspective (refer to Figure 23, Figure 24, and Figure 25), it is evident that the LSTM-AE model achieved the lowest nMAE, nRMSE, and nMAPE, which is equal to 0.02. This excellent performance, as mentioned earlier, is attributed to the ability of the model to reconstruct the inputs into a better representation in addition to extracting the temporal features. The LSTM and GRU models come in second place while the CNN model achieved the worst results. With time-series data, the temporal features are the most important features, which cannot be captured by the CNN model. However, CNN-BiLSTM is the worst model for Caracas and Toronto according to MAE and RMSE results. In contrast, according to nMAPE metric, the CNN-BiLSTM model outperformed the LSTM-AE with four out of ten datasets (Al-Khafji, Arar, Tabuk, and Wadi-Addwasir) and both models achieved the same value with Al-Jouf.

Looking at the performance from the locations' perspective (refer to Figure 23, Figure 24, and Figure 25), we can notice that the best nMAE, nRMSE, and nMAPE results for all models are mostly associated with Wadi-Addwasir dataset. On the other hand, the worst results are linked with Toronto and Caracas datasets. As mentioned earlier, the high performance related to Wadi-Addwasir dataset might be attributed to the completeness of this dataset compared to other Saudi datasets since it has the least number of missing days and the largest training set size (see Table 6). The second-best performance is associated with Tabuk dataset. Despite the high number of missing records, it has the highest percentage of sunny hours and the lowest percentage of cloudy hours among other datasets (see Figure *9*). In contrast, the low performance associated with Toronto and Caracas datasets might be attributed to the high percentage of cloudy hours (or unclear sky condition) compared to other Saudi locations (see Figure *9*). This in turn means GHI varies from time to time and is hard to predict. We can infer that the most important factor that affects models' performance is the climate in the dataset source, followed by the completeness of the dataset to help the model learn the GHI variations accurately.

4.1.3. Effect of Sunny and Cloudy Weather on Forecasting

To examine the effect of weather type on models' performance, we plot in Figure 27 the actual vs. predicted GHI of one sunny and one cloudy day by all the five models for five locations: Al-Jouf, Al-Khafji, Wadi-Addawasir, Caracas, and Toronto. The first three-hour of GHI values after sunrise were used as inputs to the models. Therefore, the prediction starts from 11:00 am or 10:00 am depending on the sunrise time in the location of the data source. Similarly, the last time is 18:00 or 17:00 depending on the sunset time, which is the last time for GHI prediction of the day. From Figure 27, we can observe that predicting GHI on sunny days is more accurate than on cloudy days. It is also noticed that the LSTM-AE model is the most accurate model on sunny and cloudy days. Even if it is not very accurate as in the case of Toronto cloudy day, it is able to capture the trend line closely. In contrast, the CNN-BiLSTM model sometimes achieves a closer prediction than the LSTM-AE model, but it could not capture the trend line accurately like the LSTM-AE model as shown in the case of Toronto cloudy day. On the other hand, the CNN model achieves the worst prediction, especially on cloudy days.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

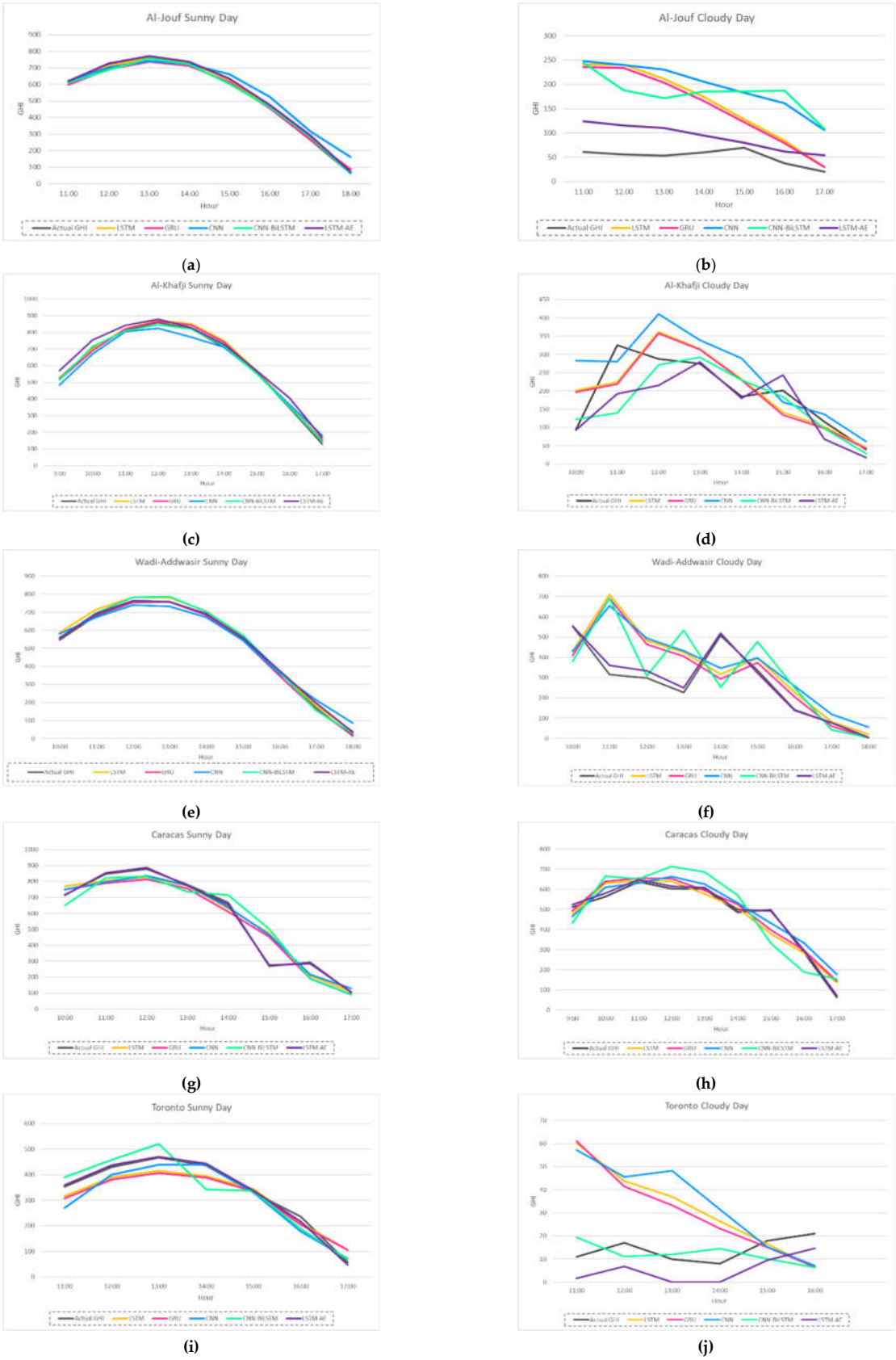


Figure 27. Sunny vs. Cloudy -- Actual Vs. predicted GHI of 5 models for: (a) Al-Jouf sunny; (b) Al-Jouf cloudy; (c) Al-Khafji sunny; (d) Al-Khafji cloudy; (e) Wadi-Addwasir sunny; (f) Wadi-Addwasir cloudy; (g) Caracas sunny; (h) Caracas cloudy; (i) Toronto sunny; (j) Toronto cloudy

 doi:10.20944/preprints202208.0345.v1

4.1.4. Effect of Summer and Winter Seasons on Forecasting

To examine the effect of seasons on models' performance, we first show, in Figure 28, the actual vs. predicted GHI of the coldest and hottest months (January and August) by all models for five locations: Al-Jouf, Al-Khafji, Wadi-Addawasir, Caracas, and Toronto. Like sunny and cloudy results, we found that the LSTM-AE model is the most accurate in January and August while CNN is the worst model. It is noticed also that the CNN-BiLSTM model performs poorly with specific datasets as in the case of Caracas and Toronto because the GHI readings are not stationary. Figure 29 shows the MAE for summer and winter for each dataset. The MAE metric is selected for no specific reason, we could have plotted RMSE and other metrics, or all the metrics considered in this paper. However, we plot one metric for the sake of brevity. We divided the year into two seasons for simplification and because Saudi Arabia does not experience four seasons. Summer includes May, June, July, August, September, and October while winter includes the remaining months. From Figure 29, we can see that winter MAE is higher than summer in all datasets, except for Taif, Caracas, and Toronto where MAE is higher in summer. Another observation is that the CNN model and CNN-BiLSTM model have the largest difference in MAE from summer to winter while the LSTM-AE model has a very slight difference.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

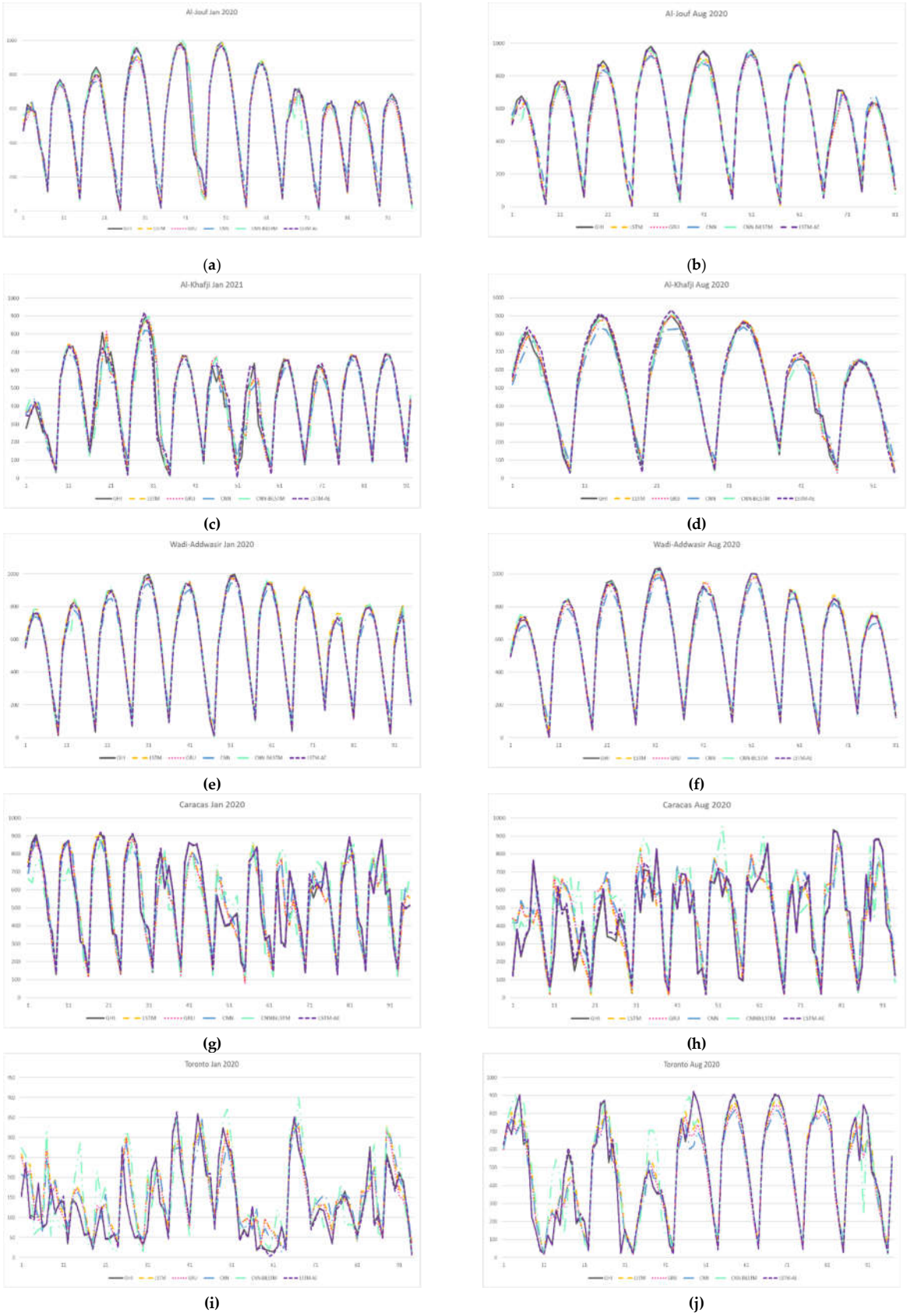


Figure 28. Summer vs. Winter -- Actual vs. predicted GHI of 5 models for: (a) Al-Jouf Jan; (b) Al-Jouf Aug; (c) Al-Khafji Jan; (d) Al-Khafji Aug; (e) Wadi-Addwasir Jan; (f) Wadi-Addwasir Aug; (g) Caracas Jan; (h) Caracas Aug; (i) Toronto Jan; (j) Toronto Aug.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

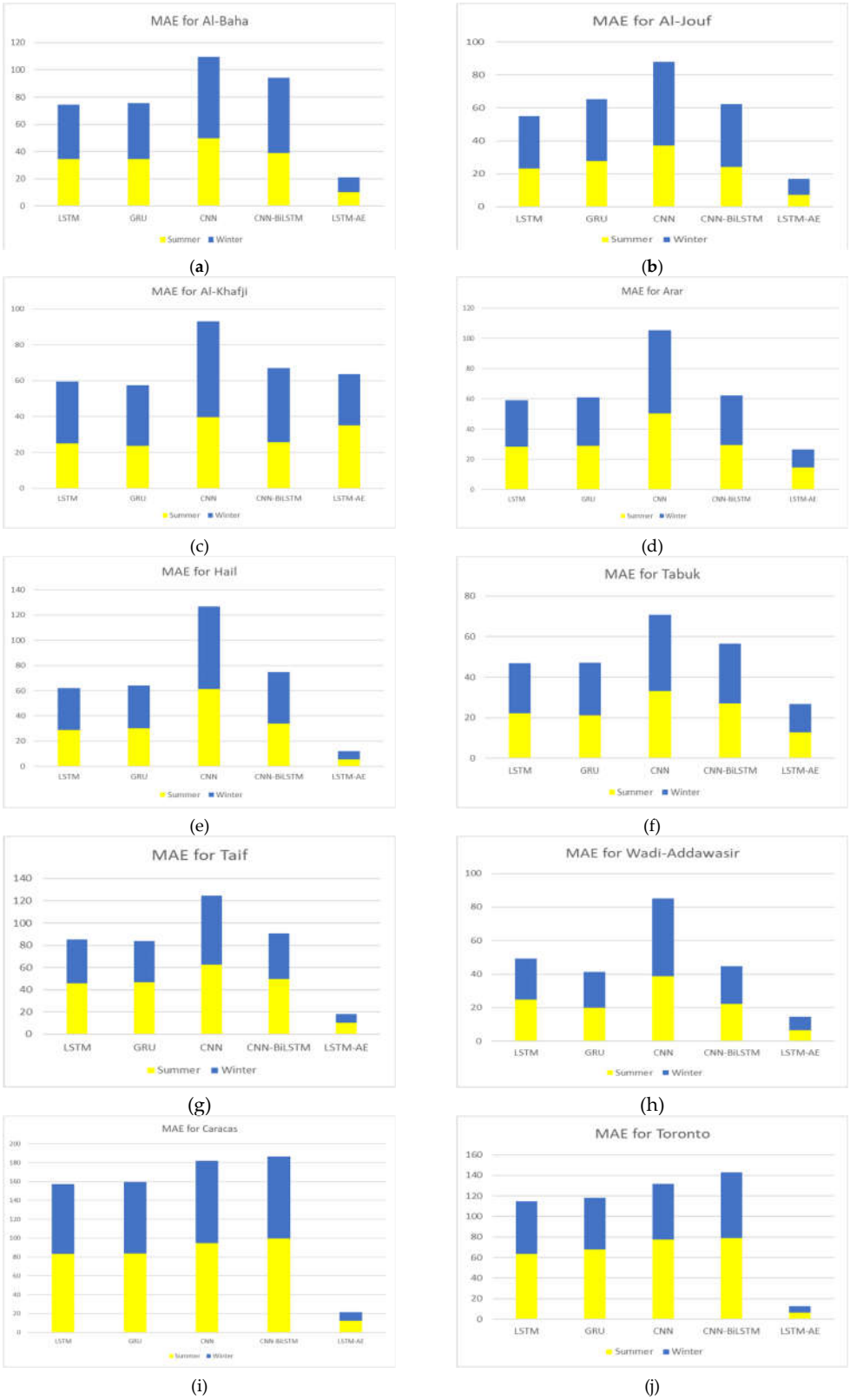
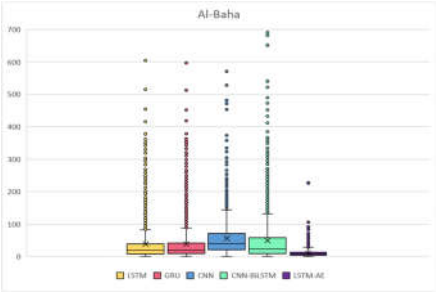


Figure 29. MAE of Summer Vs. Winter for 5 models for: (a) Al-Baha; (b) Al-Jouf; (c) Al-Khafji; (d) Arar; (e) Hail; (f) Tabuk; (g) Taif; (h) Wadi-Addwasir; (i) Caracas; (j) Toronto.

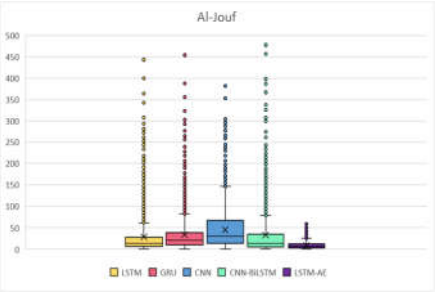
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

4.1.5. Digging Deeper into Forecasting Error for Each GHI Prediction

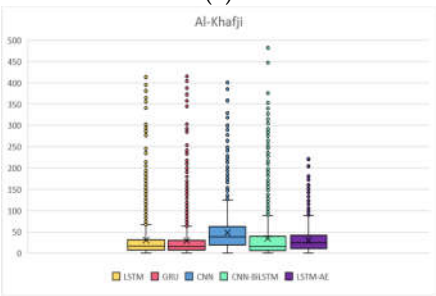
The forecasting error is defined earlier (see Equation (4)). To depict the forecasting error distribution and outliers of the five models for all datasets, a Boxplot of the forecasting error is displayed in Figure 30. Note that the plot for each location contains forecasting errors for each data item in the testing dataset that is used for prediction (see Table 6 for details about the testing sets sizes). All models' forecasting error interquartile range is below 100 except for Caracas dataset. It is clear from the figure that the forecasting error of the LSTM-AE model has the smallest interquartile range with the fewest outliers. Model-wise, the forecasting error of the CNN-BiLSTM model has the highest outliers while dataset-wise, Toronto and Taif datasets have the highest outliers.



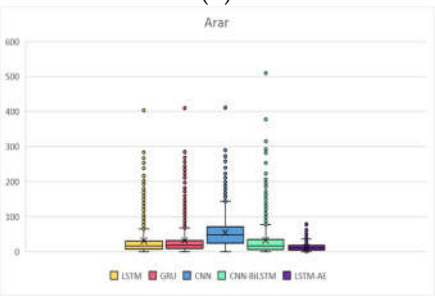
(a)



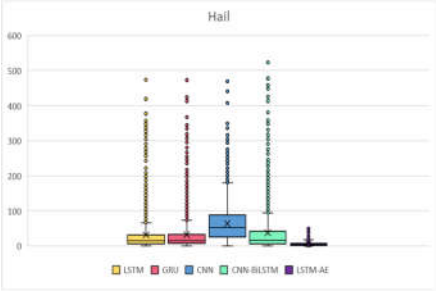
(b)



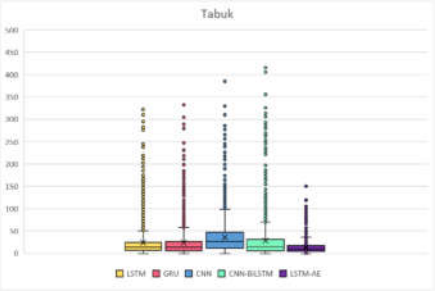
(c)



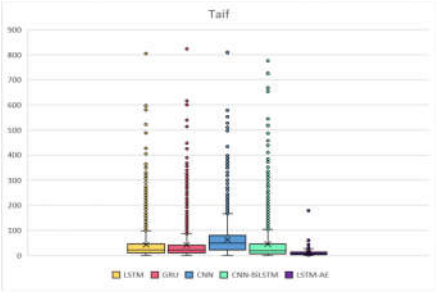
(d)



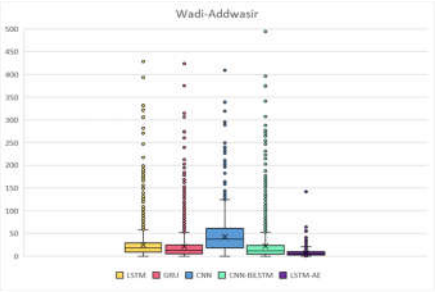
(e)



(f)



(g)



(h)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

(i) (j)

Figure 30. Boxplot of GHI forecasting error of 5 models for 5 models for: (a) Al-Baha; (b) Al-Jouf; (c) Al-Khafji; (d) Arar; (e) Hail; (f) Tabuk; (g) Taif; (h) Wadi-Addwasir; (i) Caracas; (j) Toronto.

The forecasting error is used to determine the "best model" label of each record in the testing datasets of all locations. The best model is the model that achieves the least forecasting error for each record. Figure 31 shows the achieved percentage of all the five models as the "best model" based on the forecasting error. The percentage is calculated by dividing the number of records in which a model is the best by the total number of records. It is clear from the pie chart that LSTM-AE model is the best model for 54% of the records while CNN-BiLSTM comes in second place with 17%. LSTM and GRU models achieved the least forecasting error for 11% of the records whereas CNN does so for 7% only.

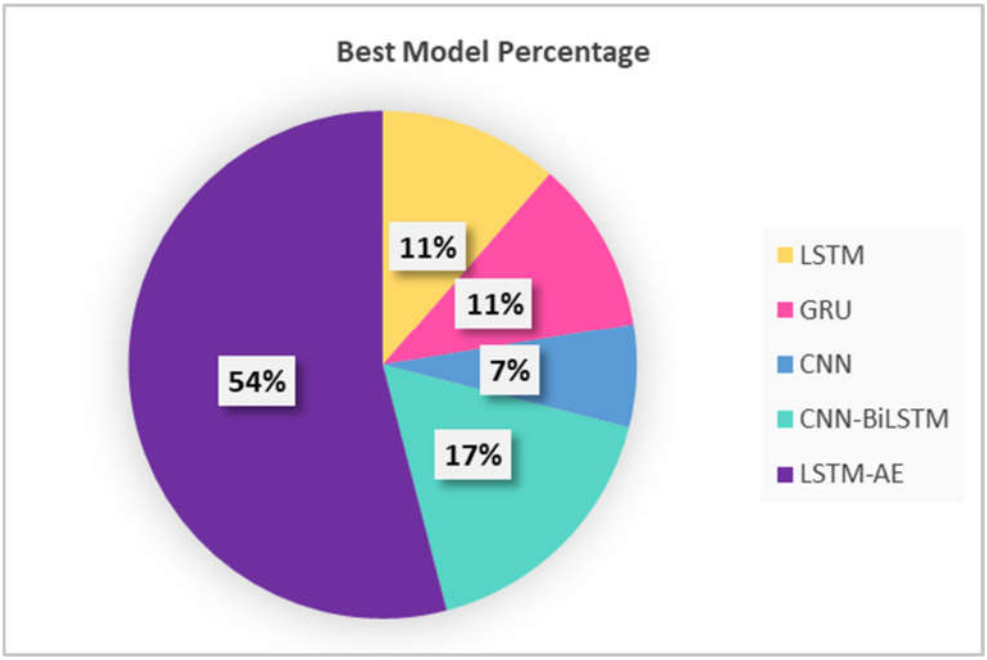


Figure 31. The achieved percentage of the models as "best model" based on the forecasting error.

4.2. SENERGY: Auto-Selective Model Prediction Engine Performance

In Section 4.1, we compared the five forecasters' performances on ten datasets and found that according to MAE and RMSE results, the LSTM-AE model is the best forecaster without competition. However, according to the MAPE metric, the LSTM-AE model is the best forecaster with half of the datasets while CNN-BiLSTM is the best with the other half. We also compared the five forecasters' performances using the forecasting error of each individual record (see Equation (4)) and we found that LSTM-AE model is the best model for 54% of the total records while CNN-BiLSTM is the best for 17%. The remaining models CNN, GRU, and LSTM together achieved 29% only (see Figure 31). This imbalance in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

data that comes from the forecasting models' performance variation would affect the classifier training negatively. Considering both the overall performance of the forecasters represented in MAPE metric and item-wise performance represented in the forecasting error, we decided to use only two models: LSTM-AE, and CNN-BiLSTM in SENERGY tool to mitigate the imbalanced data issue. Accordingly, we built an Auto-Selective Model Prediction Engine that chooses one out of the best two models based on the same inputs used for forecasting. We will incorporate in the tool additional models for GHI forecasting in the future. A description of the Auto-Selective Model Prediction Engine structure is given in Section 3.4.1. Figure 32 shows the confusion matrix of the Engine. As shown in the matrix, correctly classified CNN-BiLSTM records account for 8.4% while LSTM-AE records account for 72.57%.

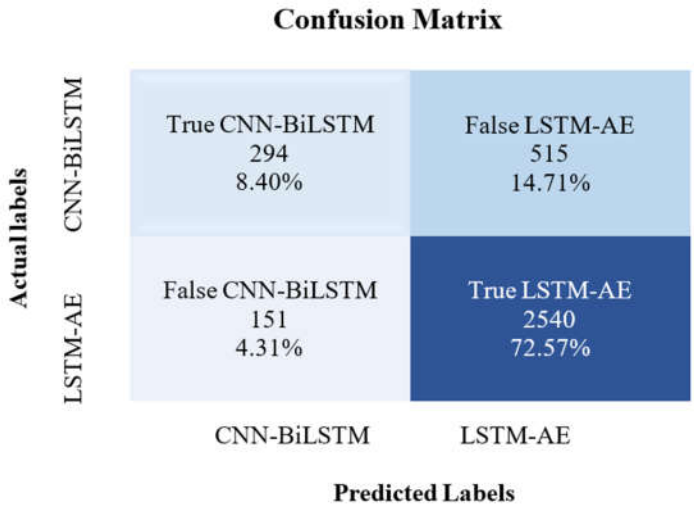


Figure 32. Auto-Selective Model Prediction Engine confusion matrix.

Table 10 presents the classification report of the Auto-Selective Model Prediction Engine. It shows the precision, recall, F1-score, and support of both models CNN-BiLSTM and LSTM-AE separately, then, the classification accuracy of the engine. The total number of records used for testing the engine is 3500 as shown in Support column. Out of which, CNN-BiLSTM model accounts for 23% (809/3500), and LSTM-AE model accounts for 77% (2691/3500). The percentage of correctly classified records of each model is shown in Recall column. CNN-BiLSTM model recall is 36% while LSTM-AE model recall is 94%. This large difference between both models' accuracy is mainly attributed to data imbalance, which in turn renders the overall engine accuracy to 81% (F1-score in the third row).

Table 10. Auto-Selective Model Prediction Engine classification report.

	Precision	Recall	F1-score	Support
CNN-BiLSTM	66%	36%	47%	809
LSTM-AE	83%	94%	88%	2691
Accuracy			81%	3500
Macro average	75%	65%	68%	3500
Weighted average	79%	81%	79%	3500

Figure 33 shows the feature importance using the Random Forest classifier method. Random Forest is used here not to make a prediction or eliminate features, but rather to provide insights about features ranking. The most important feature for classification is the solar zenith angle followed by DHI value of lag 1. The least important features are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

time-related features, such as DS, DC, MS, MC, and HS. In contrast, HS and HC are important features for forecasting (refer to Section 3.3).

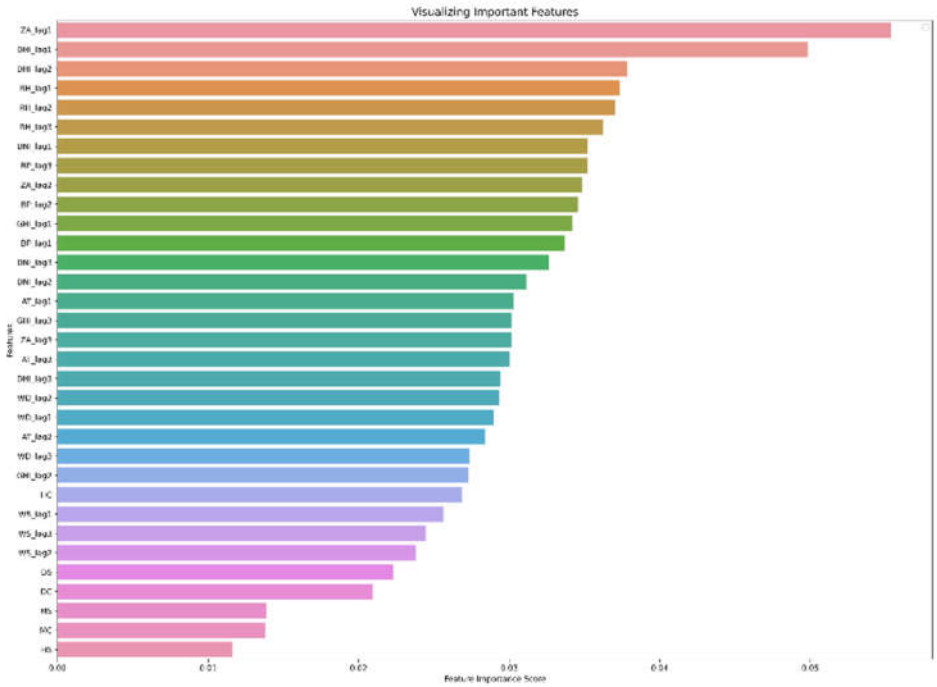


Figure 33. Feature importance using Random Forest classifier.

Our objective in this paper is to introduce our deep learning-based auto-selective approach to predicting the best performing machine learning model for GHI forecasting. We will investigate and improve the data balancing and other approaches in the future to improve the performance of the proposed auto-selective approach.

In the coming sections, the classification results are analyzed from three aspects: climate and location (4.2.1), sunny and cloudy weathers (4.2.2), and summer and winter seasons (4.2.3).

4.2.1. Model Prediction: Climate and Location

To further analyze the classification results, we first calculated the Auto-Selective Model Prediction Engine accuracy for each location as presented in Figure 34. The number of total records is also incorporated in the figure to see its effect on accuracy. The highest classification accuracy is 90% associated with Caracas and Toronto due to a large number of records for both locations while the lowest classification accuracy is 69% associated with Tabuk for which the low number of records plays a role in addition to the close forecasting performance between both forecasting models for this location (see Section 4.1.2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

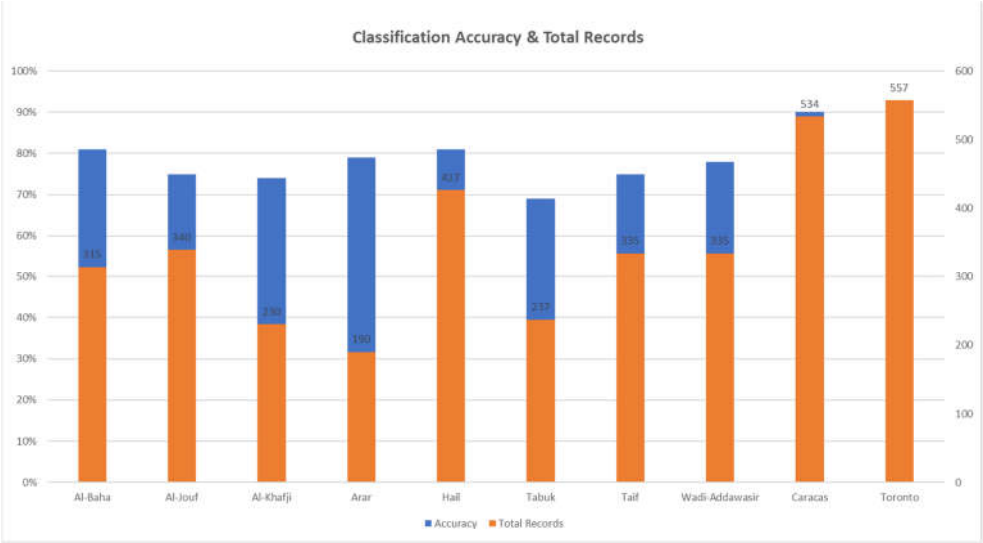


Figure 34. Classification accuracy of Model Prediction Engine based on location with total records.

We also calculated the recall of both models CNN-BiLSTM and LSTM-AE for each location as shown in Figure 35. The recall percentage for LSTM-AE model is 90% or higher for all locations except for Al-khafji, which equals 59%. On the other hand, the recall percentage for CNN-BiLSTM ranges from 9% to 44% except for Al-khafji, which equals to 86%. The reason for the superiority of CNN-BiLSTM model accuracy over LSTM-AE model associated with Al-khafji data is the imbalance in both models with 133 versus 97, unlike other datasets in which the total records of LSTM-AE is always higher than CNN-BiLSTM. This, in turn, is explained by the high variation in forecasting performance between CNN-BiLSTM model with MAPE of around 16% and LSTM-AE with MAPE around 35% for Al-khafji dataset (see Figure 25).

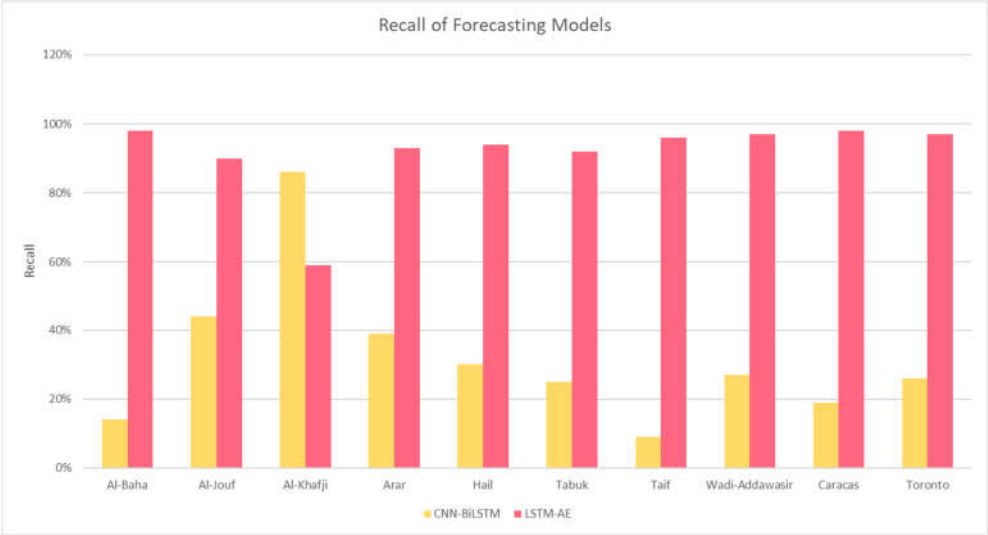


Figure 35. Recall of the two forecasting models in Model Prediction Engine.

4.2.2. Model Prediction: Sunny and Cloudy Weathers

The classification accuracy of the Model Prediction Engine on sunny days is 75% and 86% for cloudy days while the total record for sunny days is higher than cloudy by 28%. These results contradict forecasting results in which forecasting on sunny days is more

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

accurate than on cloudy days. The reason for that is the close prediction for both forecasting models CNN-BiLSTM and LSTM-AE in sunny weather, which makes it difficult for the classifier to pick one model. On the other hand, LSTM-AE shows superior performance on cloudy days, which the classifier learned from the data. Figure 36 shows the recall for each model in sunny versus cloudy weather. Notably, CNN-BiLSTM recall in cloudy weather is better than sunny with 38% while LSTM-AE recall in sunny weather is 2% better than cloudy. As explained earlier, on sunny days CNN-BiLSTM and LSTM-AE have a close prediction that causes the classifier to misclassify CNN-BiLSTM records as LSTM-AE.

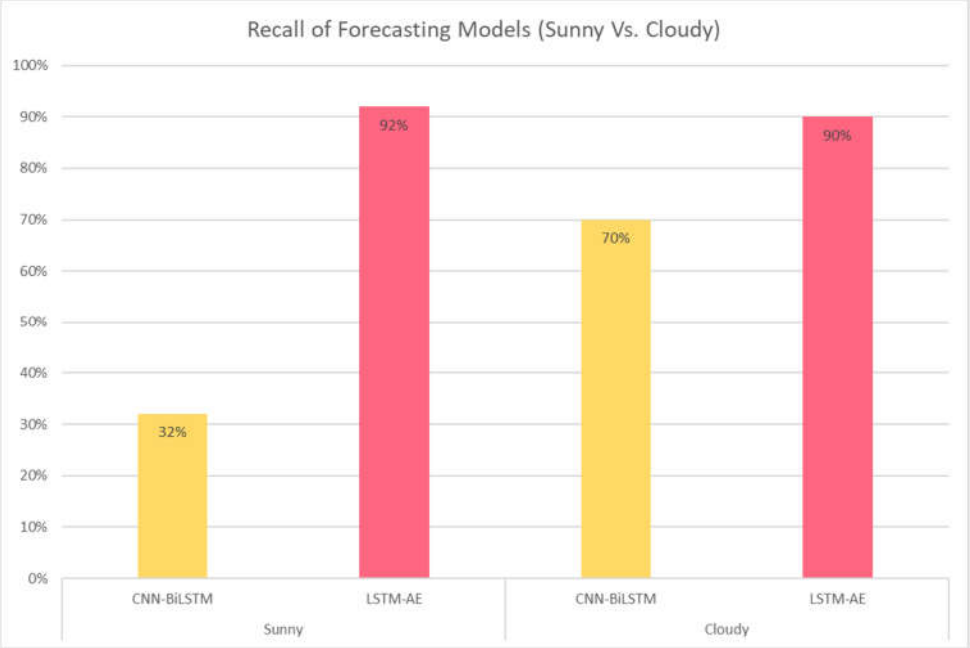


Figure 36. Recall of the two forecasting models (sunny Vs. cloudy) in Model Prediction Engine.

4.2.3. Model Prediction: Summer and Winter Seasons

The classification accuracy of the Model Prediction Engine in summer is 82% and 80% for winter even though the total number of records for summer is less than winter by 14%. This slight difference in the performance between seasons is aligned with the same trend found in the forecasting results. Figure 37 shows the recall for each model in summer versus winter. It is notable that CNN-BiLSTM recall in summer is better than in winter with a 7% difference while LSTM-AE recall is almost the same.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

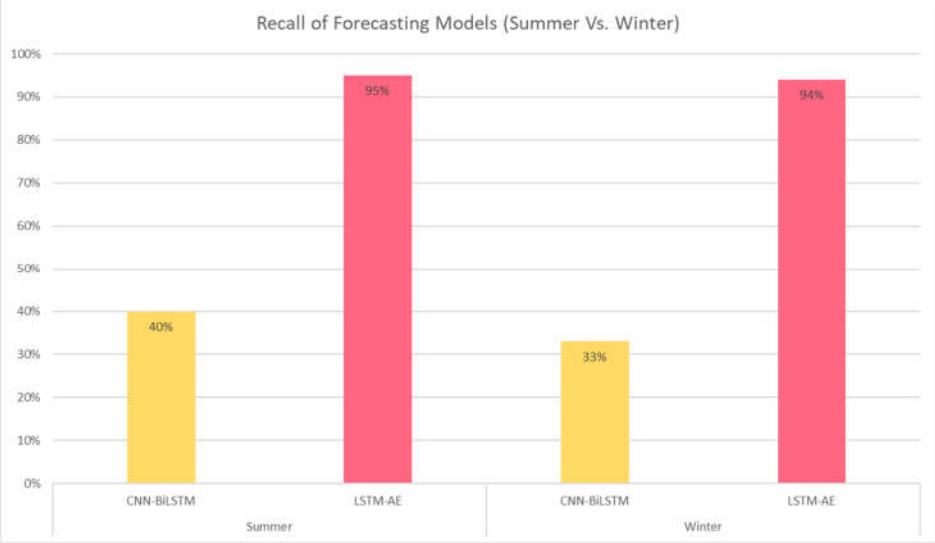


Figure 37. Recall of the two forecasting models (summer Vs. winter) in Model Prediction Engine.

4.3. SENERGY: Performance Gain and Loss

4.3.1. Actual Gains and Losses

To understand the benefit of using SENERGY, we calculated the performance gain (G) or loss (L) for the tool versus a model (m) as follows: the difference between the forecasting error of a model (CNN-BiLSTM, LSTM-AE) and the forecasting error of the model chosen by the tool.

$$\text{G or L} = \text{FE}_\text{m} - \text{FE}_t \tag{11}$$

A positive value indicates a gain, and a negative value indicates a loss. The gain or loss is calculated for each record in the testing set of Model Prediction Engine (total of 3500 records) using Equation (11). Table 11 shows an example of gain or loss calculation for three real records. As shown in the first row, the forecasting error of CNN-BiLSTM is 67.73 and for LSTM-AE is 4.83. The tool is able to choose the best model correctly for this record, thus, the achieved forecasting error is 4.83. To measure the gain over CNN-BiLSTM model, we calculate the difference between 67.73 and 4.83, which is 62.90. Therefore, we can say that the tool achieved gain in performance equals 62.90 over CNN-BiLSTM for this record. On the other hand, no gain was achieved for the tool over LSTM-AE model because it is the best anyway. Similarly, in the second record, the forecasting error of CNN-BiLSTM is 128.60 and for LSTM-AE is 44.41. The tool failed to choose the best model correctly for this record, thus, the achieved forecasting error is 128.60. There is no gain for the tool over CNN-BiLSTM model in this case. The difference between the forecasting error of LSTM-AE model and the forecasting error of the wrong best model chosen by the tool is 84.19. It is a negative number; thus, it is a loss in tool performance for LSTM-AE model for this record.

Table 11. An example of Gain/Loss of SENERGY over two models.

FE CNN-BiLSTM	FE LSTM-AE	FE Best Model	G/L CNN-BiLSTM	G/L LSTM-AE
67.73	4.83	4.83	62.90	0
128.60	44.41	128.60	0	-84.19

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

0.47	29.12	29.12	-28.65	0

Figure 38 shows the gain or loss of SENERGY versus CNN-BiLSTM model for each record. The gain is positive and hence above to zero line and the loss is negative and hence below the zero line. Locations' records are differentiated by colors. As noted, the gain is large in general because LSTM-AE model provides highly better forecasting than CNN-BiLSTM model. In contrast, the loss is small because when CNN-BiLSTM model achieves better forecasting than LSTM-AE model, the difference is small. Looking at gain or loss from a location perspective, the largest gains are achieved with Caracas and Toronto datasets while the smallest gains are achieved with Wadi-Addawasir, which is compatible with forecasting results discussed in Section 4.1.2.

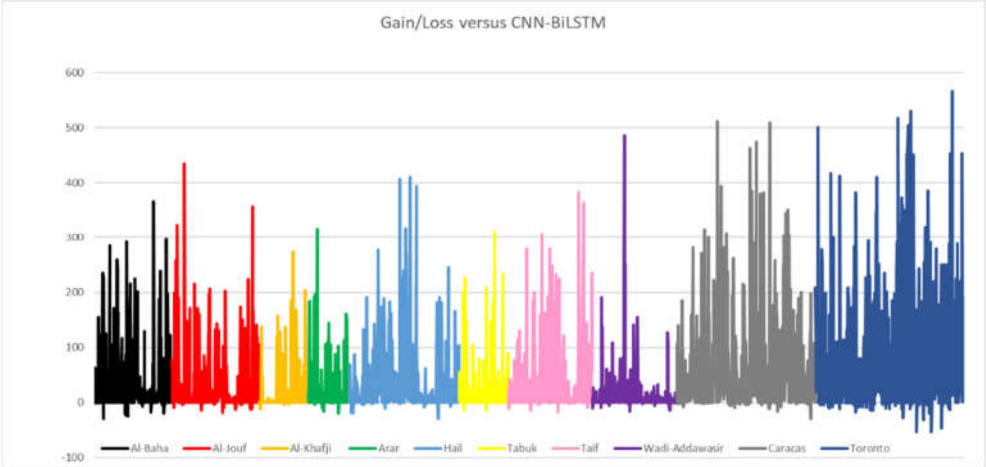


Figure 38. Gain/loss of SENERGY versus CNN-BiLSTM.

Figure 39 shows the gain or loss of SENERGY versus LSTM-AE model for each record. Locations' records are differentiated by colors. The gain is small in general because as we mentioned earlier when CNN-BiLSTM model achieves better forecasting than LSTM-AE model, the difference is small. In contrast, misclassifying records for which the best model is LSTM-AE as CNN-BiLSTM comes with a large loss because LSTM-AE accomplishes smaller forecasting error in general (refer to Table 11-second row for an example). Looking at gain or loss from a location perspective, the largest gains are achieved with Saudi datasets while the largest losses are achieved with Caracas and Toronto datasets, which is compatible with forecasting results discussed in Section 4.1.2.

Note that despite the large losses, the SENERGY tool still offers gains over the LSTM-AE method. These low gains and high losses are because the LSTM-AE method provides significantly better performance compared to any other method causing the LSTM-AE forecasting method to own most of the labels in the classification dataset (2691 out of 3500) as the best performing forecasting method, and this created a major data imbalance problem causing poor classification accuracy. Partly, to some extent, the performance of the LSTM-AE method could be attributed to the fact that we used a relatively optimized lagged feature for the LSTM-AE method giving LSTM-AE an advantage over the other four forecasting methods (see Section 4.1.1). It is possible to incorporate in SYNERGY a set of different features (e.g., Lag1, Lag2, Lag3) and treat each pair of a distinct feature (from this feature set) and a forecasting model as a separate forecasting engine (or model) and train the SYNERGY model prediction engine to predict a feature-model pair. This will allow SYNERGY to predict the best combination of a feature and model for a given GHI prediction instead of pre-defined fixed input features. The same approach can be

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

extended to hyperparameter optimisations and other parameters in the machine learning forecasting pipeline. These feature-related and parameter-related aspects of the SYNERGY approach should be investigated further before robust conclusions can be drawn. In addition, the use of additional meteorological datasets with high climate and data diversity and additional forecasting methods, coupled with solutions for data imbalance problems could create a more balanced classification dataset and allow improvements in the classification error leading to significantly better forecasting accuracies and gains.

The next section explains through graphical data what is potentially possible with the proposed SENERGY approach if the data imbalance problem can be solved. The exciting fact about the tool is that it would provide higher gains for higher diversity datasets while usually, the opposite is true for a single forecasting method. Also, as explained earlier, the approach allows selecting different models optimized for different climates rather than optimizing a model for multiple climates that may provide an optimally average performance for diverse climates. Moreover, further investigations into this approach could allow further understanding of optimal models for specific climates and weather leading to a better understanding of climates and forecasting methods and eventually developing better renewable energy forecasting approaches.

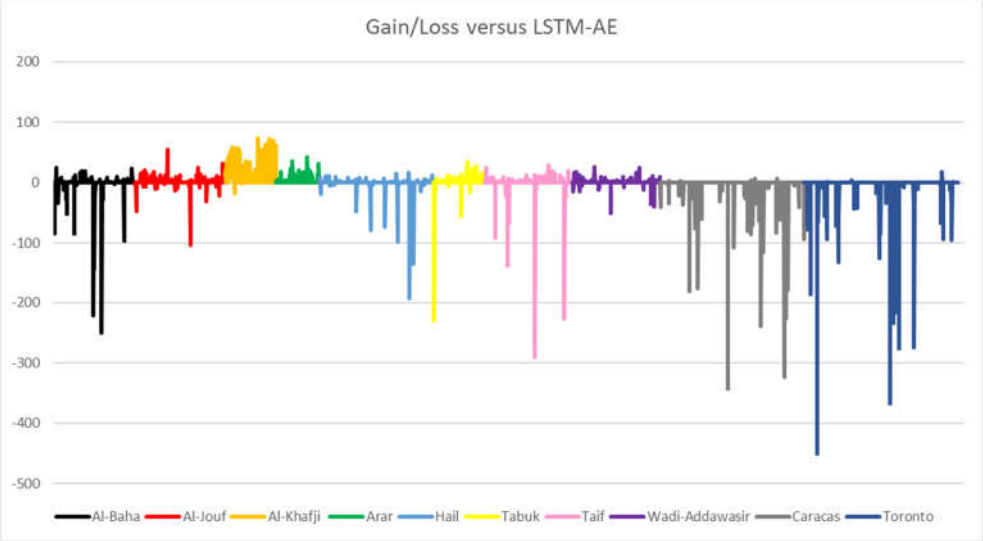


Figure 39. Gain or loss of SENERGY versus LSTM-AE.

Figure 40 displays the average gain or loss of the SENERGY tool for each record. Locations' records are differentiated by colors. The average is calculated by summing both models' gain/loss values and dividing the sum by 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

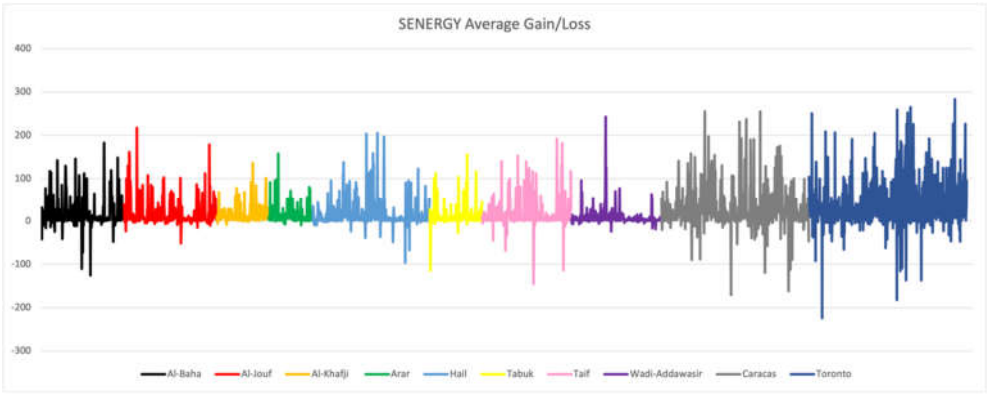
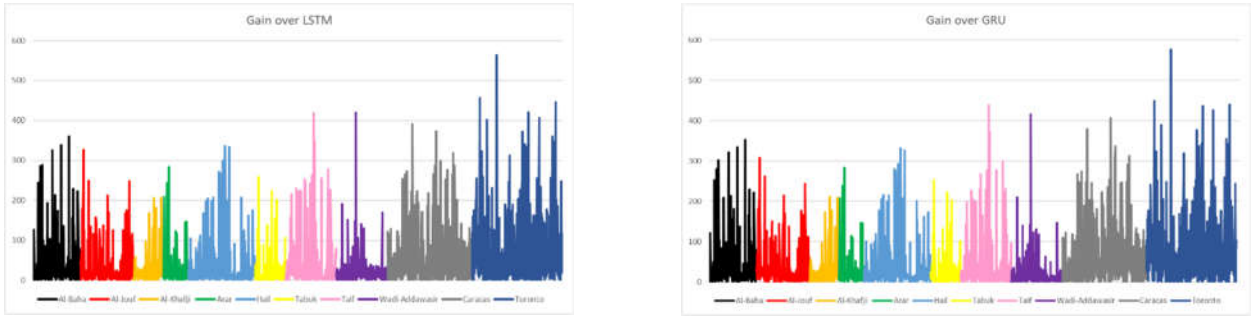


Figure 40. Average gain/loss of SENERGY.

4.3.2. Potential Performance

In the previous section, we demonstrated the gain and loss of SENERGY, which can choose the best forecasting model among the two models only. However, in an ideal situation, SENERGY would choose the best forecaster of the five models included in this work or even more models in the future. Therefore, the potential gain or loss is calculated here assuming that SENERGY can choose the best out of five forecasting models with 100% classification accuracy.

Figure 41 shows the gain of SENERGY over LSTM and GRU models in an ideal situation. There is no loss here because as mentioned before, the classification accuracy is 100%. It is noted that the gain over both models is almost the same for all locations because the forecasting performances of both models are convergent (refer to Section 4.1.2). location-wise, the largest gains of both models come with Caracas and Toronto datasets while the lowest gains are achieved with Al-Khafji and Wadi-Addawsir.

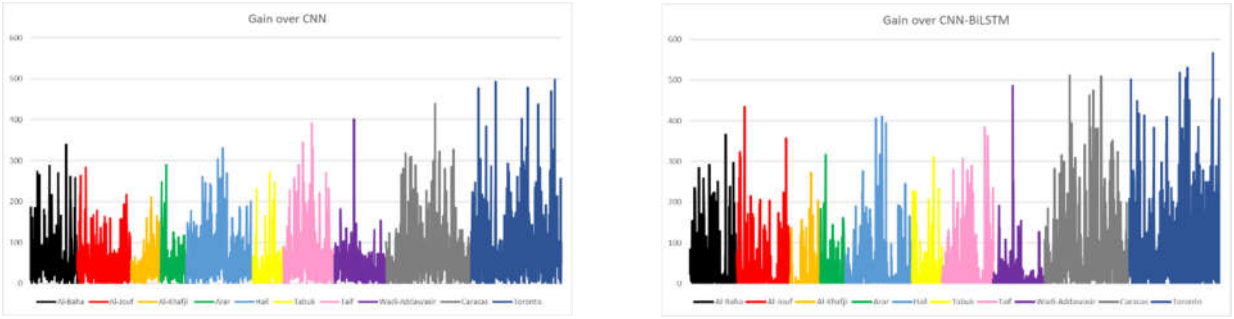


(**a**) (**b**)

Figure 41. Gain of SENERGY over:(a) LSTM; (b) GRU.

Figure 42 shows the gain of SENERGY over CNN and CNN-BiLSTM models in ideal situation. There is no loss here because as mentioned before, the classification accuracy is 100%. Gain over CNN model is like gain over CNN-BiLSTM model although the latter has more outliers. Looking at the gain from a location perspective, the largest gains are achieved with Caracas and Toronto datasets while the lowest gains are achieved with Al-Khafji and Wadi-Addawsir.

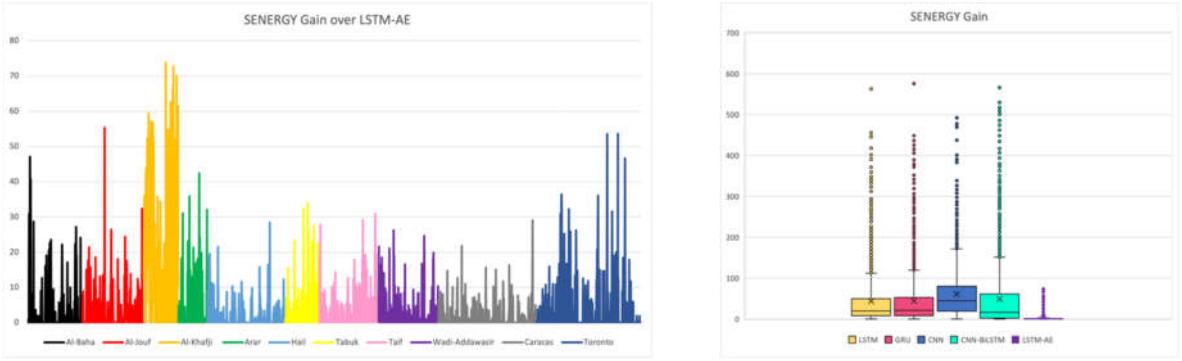
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1



(**a**) (**b**)

Figure 42. Gain of SENERGY over:(a) CNN; (b) CNN-BiLSTM.

Figure 43 (a) shows the gain of SENERGY over LSTM-AE model. Unlike other models, the largest gains are achieved with Al-Khafji. On the other hand, the lowest gains are related to Wadi-Addawsir like other models. Figure 43 (b) shows the gain of SENERGY over all the five models as a boxplot. It is obvious that the largest gain is achieved with CNN and CNN-BiLSTM models. Also, gain over LSTM and GRU models is similar while gain over LSTM-AE is very small since it is the best forecaster for most of the records anyway.



(**a**) (**b**)

Figure 43. Gain of SENERGY:(a) over LSTM-AE; (b) as a boxplot for the five base models.

4.4. SENERGY: Comparison with other Works

To the best of our knowledge, no work in the literature suggests a similar tool to combine deep learning for forecasting and classification to improve solar radiation forecasting performance and generalizability. Therefore, the comparison here will be mainly based on the forecasting results. The works that we have selected in this section for comparison with our tool SENERGY are based on two criteria. Firstly, these compared works propose models for forecasting next-hour GHI, and, secondly, they use multiple datasets from different climates. For the reasons explained, the works such as [23], [39]–[43], [74], [75] are excluded from the comparison because the datasets used are for one climate only. The works such as [20], [26], [44], [45], [69], [76]–[79] are also excluded from the comparison because they propose forecasting models for different time horizons such as day ahead or monthly GHI compared to the next-hour GHI that is the focus of our work in this paper.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

Comparison in this section includes MAE, RMSE, MAPE results, and their normalized values of all locations datasets used in each work whenever they are reported. The comparison also includes the forecasting skill metric FS_{MAE} and FS_{RMSE}. Equations of all these metrics are provided in Section 3.5. Moreover, GHI mean and standard deviation are added to the comparison to show the variation among locations.

Table 12 presents the comparison of SENERGY to six works, which met the aforementioned selection criteria. The comparison in the table is based first on information that shows the data variation aspect: data source location, GHI mean and standard deviation, climate classification of the location, and the use of weather parameters in addition to historical values of GHI in inputs. The second aspect of comparison is the model used for forecasting. For example, in reference [18], data from three locations in India are used, which represent three different climate classes (Cwa, Cwb, Bsh). GHI mean and SD are not provided. Weather data in addition to GHI historical values are used to develop the proposed ensemble model of XGBF-DNN. The third aspect of comparison is performance metric results, which are compared later in multiple figures.

Table 12. Comparison of SENERGY to related works.

Ref #	Location	GHI mean	GHI SD	Climate	Weather data	Model
[18]	Jaipur New Delhi Gangtok	NA	NA	• Cwa • Cwb • Bsh	✓	Ensemble model of XGBF-DNN
[21]	Los Angeles	217.37	291.73	• Csb, • BSk • Af • BWk	×	Hybrid model of CEEMDAN-CNN-LSTMv
	Denver	203.33	276.40			
	Hawaii's Big Island	220.12	307.79			
	Tamanrasset	269.98	361.83			
[22]	Ajaccio Tilos Odeillo	NA	NA	• Csa • Csb	×	ARMA RF
[24]	CA TX WA FL PA MN	NA	NA	• BSk • Cfa • Cfb • Am • Dfb • Dfa	✓	Generalized Random Forest
[25]	Bondville	398.04	284.66	• Cfa • BWk • BSk • Cfb • Dfa	×	68 machine learning algorithms (Cubist model is the best in most cases)
	Desert Rock	517.72	314.73			
	Fort Peck	368.17	277.33			
	Goodwin Creek	442.77	289			
	Penn. State Uni	384.31	277.24			
	Sioux Falls	406.94	277.55			
	Table Mountain	412.19	287.97			
[27]	Tucson Bermuda	532.5 417.1	NA	• Bsh • Cfa	×	ELM

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

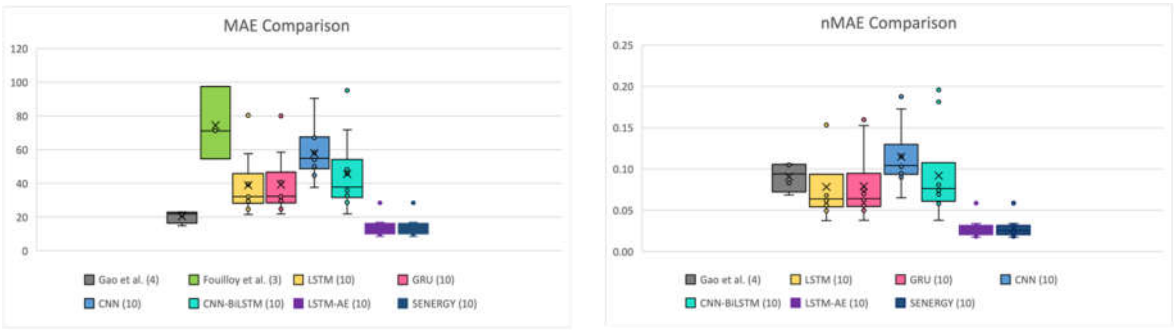
	Brasilia	475.6		• A		
	Sonnblick	347.2		• Dfc		
	Solar Village	580.9				
	Golden	459.4				
	Darwin	516.4				
	Ny-Alesund	184.3				
	Toravere	256.9				
	Lerwick	198.3				
	Al-Baha	582.09	311.16			
	Al-Jouf,	528.14	296.47			
	Al-Khafji	486.59	275.73			
	Arar	485.46	295.04			LSTM
This	Hail	543.77	303.82	• BWh		GRU
work	Tabuk	498.03	261.73	• A	✓	CNN
	Taif	567.62	308.47	• Dfb		CNN-BiLSTM
	Wadi-Adda-wasir	578.02	301.69			LSTM-AE
	Caracas	366.77	271.11			
	Toronto	524.82	297.12			

Ref [27] has 20 locations, we present data about 10 locations from various climates for simplicity.
NA: Not available.

A fair comparison of models' performances in the literature is a challenging task because first, there is great variation in the results reported by researchers. Also, it is difficult to find best-performing model by comparing various statistical measures at the same time, such as RMSE, MAE, MAPE, etc. For example, to compare six works included in this section, we need multiple figures (Figure 44, Figure 45, Figure 46, Figure 47). Some metrics are reported in these six papers and others are not. Sometimes normalized metrics' results are not given in a paper, but the GHI mean of each location is given. Therefore, we calculated normalized metrics in this case. Therefore, we could not include all the six works in these figures. This highlights the need to standardize the performance metrics used to report results. Each box plot in the following figures represents a performance metric result of several locations. Performance metrics results are averages calculated for a whole dataset. The desirable outcome is a low box to show small error and a short box to show small variation among different locations. The number of locations is ten for this work, and it is shown beside the authors' names in the legend for other works. SENERGY results reported in the next figures are calculated assuming it can choose the best out of five forecasting models with 100% classification accuracy. To elaborate, our proposed approach has the potential to provide better performance than any forecasting model alone, therefore, we reported the results for the ideal situation.

In Figure 44 (a), MAE results from Gao et al. [21] and Fouilloy et al. [22] are compared to MAE results of five forecasting models and SENERGY in this work. It is noted that reference [22] has the worst MAE results in terms of high value and large variation among the three locations. On the other hand, LSTM-AE model and SENERGY show the best performance in terms of both the lowest MAE values and low variations for ten locations. The work of Gao et al. [21] appears to show the next best performance, however, it is because the results are for four locations only. In Figure 44 (b), nMAE results are compared. Reference [22] is excluded because nMAE is not reported there. We see how normalization made the box of Gao et al. [21] bigger and thus, it is fair to say that LSTM, GRU, LSTM-AE, and SENERGY show better performance even with a larger number of locations. Both for MAE and nMAE, LSTM-AE model and SENERGY show the same performance because according to these metrics (averaged over each of the ten location datasets), the best model is always LSTM-AE for all locations (refer to Figure 23).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

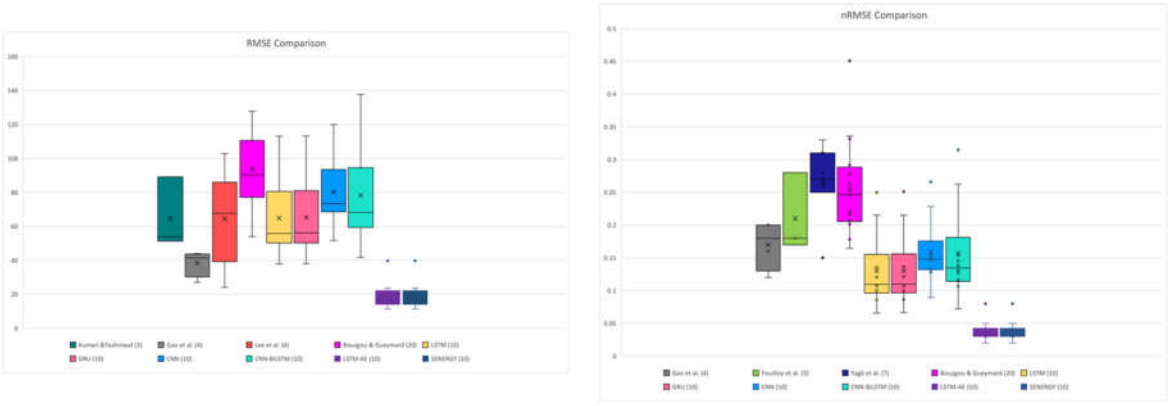


(**a**) (**b**)

Figure 44. Comparison of multiple works based on: (a) MAE; (b) nMAE.

In Figure 45 (a), RMSE results of four works: Kumari & Toshniwal [18], Gao et al. [21], Lee et al. [24], and Bouzgou & Gueymard [27] are compared to RMSE results of the five forecasting models and SENERGY results in this work. LSTM-AE model and SENERGY achieved the best performance in terms of lowest RMSE and smallest variation among ten locations. The work of Gao et al. reference [21] comes in the second place, however, it includes four locations only compared to six and twenty in other works. The worst performance in terms of value is associated with the work of Bouzgou & Gueymard reference [27], while the worst based on variation among locations is associated with the work of Lee et al. reference [24] with six locations. In Figure 45 (b), nRMSE results of this work are compared to four works. Since nRMSE is not reported in references [18] and [24], they are excluded in (b) and another two works are added: Fouilloy et al. reference [22] and Yagli et al. reference [25]. The best nRMSE results are achieved by LSTM-AE model and SENERGY while the worst are related to Yagli et al. reference [25] in terms of low value and Fouilloy et al. reference [22] in terms of large variation among locations. Comparing (a) and (b), we can see the benefit of normalization in providing a fair comparison. For example, in (a) Gao et al. reference [21] box is smaller and lower than LSTM, GRU, CNN, and CNN-BiLSTM models, but after normalization, it becomes higher than all of them. In both (a) and (b), LSTM-AE model and SENERGY achieved the best performance in terms of lowest value and smallest variation. Again, SENERGY shows the same performance as LSTM-AE model because according to RMSE and nRMSE results, the best model is always LSTM-AE for all locations (refer to Figure 24).

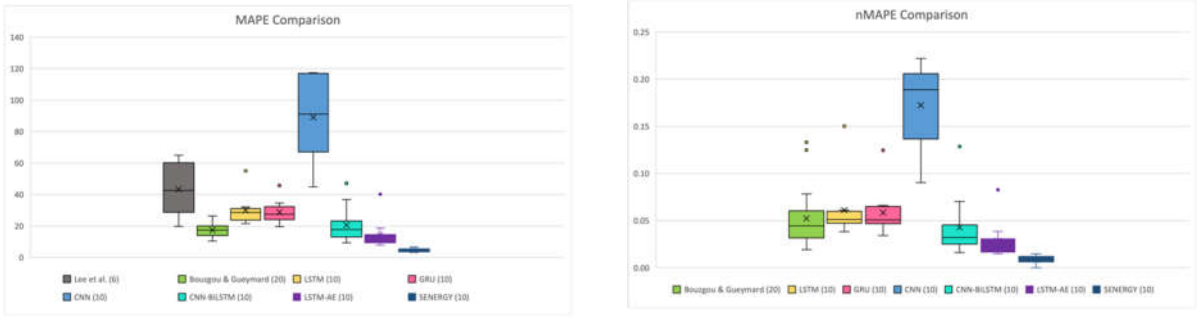
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1



(**a**) (**b**)

Figure 45. Comparison of multiple works based on: (a)RMSE; (b)nRMSE.

In Figure 46 (a), MAPE results of Lee et al. reference [24], and Bouzgou & Gueymard reference [27] are compared to the five forecasting models and SENERGY results in this work. SENERGY achieved the lowest error with the smallest variation among ten locations while CNN model is the worst. In Figure 46 (B), the comparison is based on nMAPE results and the same observation about the best and worst performance is true. The work of Lee et al. reference [24] is eliminated in (b) since GHI mean is not reported and thus nMAPE cannot be calculated. From (a) and (b), we can see the normalization effect on the work of Bouzgou & Gueymard reference [27]. In (a), it shows better performance than LSTM, GRU, and CNN-BiLSTM models while in (b) it becomes worse than all of them in value or variation among locations. Unlike MAE and RMSE results, SENERGY outperforms LSTM-AE model based on MAPE and nMAPE because the latter is not the best model for all locations according to these metrics (refer to Figure 25).



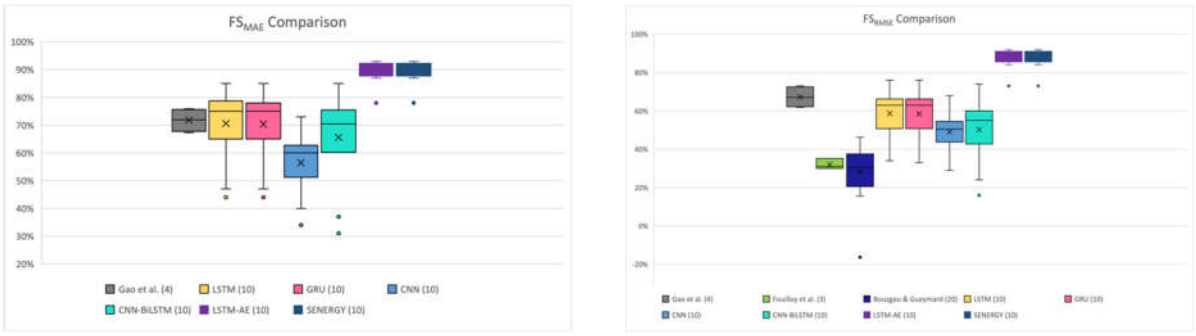
(**a**) (**b**)

Figure 46. Comparison of multiple works based on: (a) MAPE; (b) nMAPE.

In Figure 47 (a), FS_{MAE} results of Gao et al. [21] are compared to the five forecasting models and SENERGY results in this work. In this figure, the highest value is the best. It is noticed that LSTM-AE model and SENERGY have the highest value and the lowest variation among locations while CNN model is the worst in terms of value and CNN-BiLSTM is the worst in terms of variation among locations. In (b), FS_{RMSE} results of three works: Gao et al. [21], Fouilloy et al. [22], and Bouzgou & Gueymard [27] are compared to the five forecasting models and SENERGY results in this work. Again, LSTM-AE model and SENERGY have the highest value and the lowest variation among locations. The second best performance is achieved by the work of Gao et al. [21]. However, it only includes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2022 doi:10.20944/preprints202208.0345.v1

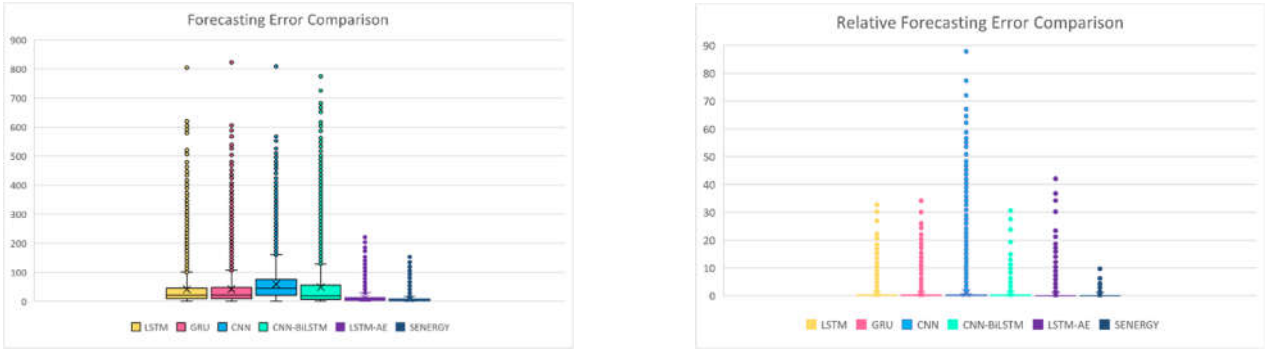
the results of four locations compared to ten and twenty in other works. On the other hand, Bouzgou & Gueymard reference [27] has the worst value and the largest variation among locations since it includes twenty results. Like MAE and RMSE results, SENERGY does not show better performance than LSTM-AE model in (a) or (b) because on both metrics, the latter is the best model for all locations (refer to Figure 26).



(**a**) (**b**)

Figure 47. Comparison of multiple works based on: (a) FSMAE; (b) FSRMSE.

Figure 48 compares the SENERGY performance with the five forecasting models in terms of the forecasting error (refer to Equation (4)). As we mentioned earlier, the comparison in this section is based on the assumption that SENERGY can choose the best among the five forecasting models with 100% accuracy. In this figure, the forecasting error is calculated and represented for each data item in the testing datasets of all locations together. (a total of 24676 records as shown in Table 6). Therefore, the number of outliers for each model is higher compared to the earlier figures in this section (those figures plot average statistics for each dataset). No other work is compared in this figure because we do not have results available at this precision from other researchers published works. From (a), we can see the improvement of SENERGY performance over the five models comes from the ability of the tool to choose one of the five models that achieves the least error for each data input. Similarly, in (b) the forecasting error is divided by actual GHI to get the relative error. Both in (a) and (b), SENERGY has the least error with fewer outliers and CNN model is the worst.



(**a**) (**b**)

Figure 48. Comparison of SENERGY to other models based on: (a) forecasting error; (b) relative forecasting error.

From all the figures shown in this section, it is evident how difficult it is to compare works when different metrics are reported and not all the needed information for a fair comparison is given. There is a need to improve, consolidate, and standardize

 doi:10.20944/preprints202208.0345.v1

international efforts on transparent and extensive testing of the proposed models for renewable energy forecasting [10]. One approach could be that researchers make the complete results data openly available for comparison purposes. The box plot used in this work provides the results at a higher granularity compared to the aggregate or average metrics. Particularly, the box plots for the forecasting error and the relative forecasting error provide a more detailed account of the performance because these results are plotted for each GHI prediction compared to the other performance metrics that show performance at a lower granularity of dataset levels.

5. Conclusions and Future work

This work introduced SENERGY, a novel tool for solar radiation forecasting. SENERGY utilizes the knowledge gained from the performance of deep learning-based forecasting models with different datasets collected from multiple locations and the meteorological data variables of these locations to recommend the best forecasting model suitable for data features. Using the recommended forecasting model by SENERGY with new data inputs would save time and effort in running experiments in addition to the gain in forecasting accuracy. To build the knowledge base of the models' performances, we trained and tested five forecasting models: LSTM, GRU, CNN, CNN-BiLSTM, and LSTM-AE with eight datasets collected from different locations in Saudi Arabia that have hot desert climate in addition to datasets from Toronto and Caracas, which have humid continental and tropical climate respectively. To provide the best forecasting model recommendation, an LSTM model was developed.

Future work would aim to make improvements in different aspects of the SENERGY tool design. One area is to improve the knowledge base. We used data from three different climates only. In the future, more datasets from different countries and climates would be used to enrich the knowledge base. This would allow SENERGY to provide more accurate recommendations to any meteorological data irrespective of the climate changes. Additionally, the SENERGY knowledge base contains only the performance of two forecasting models. Another way to improve it is to use more competitive forecasting models, specifically the models proven to provide high performance in the literature. Another aspect is to improve the model prediction engine performance. Currently, the classification accuracy is 81%, which should be enhanced in the future. One idea is to add a weather classification step (sunny, cloudy, etc.) to improve accuracy. Moreover, SENERGY predicts the best model based on models' performance in terms of forecasting accuracy only. Another performance option could be added to the tool, which is the model computation time. Model auto-selection would be provided based on both performance measures upon user preferences. A third aspect is to improve forecasting. Future work will also include improving the used forecasting models through a rigorous optimization process, such as hyperparameters tuning, and through inputs, such as using the forecast of meteorological variables or satellite data in addition to the historical measurements data to improve the GHI prediction.

The current performance of the SENERGY tool is limited because the LSTM-AE method outperforms all other methods, causing the LSTM-AE forecasting method to own the majority of the labels in the classification dataset as the best performing forecasting method, resulting in a major data imbalance problem and poor classification accuracy. The LSTM-AE method's performance could be attributed in part to the fact that we used a relatively optimized lagged feature for the LSTM-AE method, giving LSTM-AE an advantage over the other four forecasting methods. It is possible to include a set of different features (for example, Lag1, Lag2, Lag3) in SYNERGY and treat each pair of a distinct feature (from this feature set) and a forecasting model as a separate forecasting engine (or model) and train the SYNERGY model prediction engine to predict a feature-model pair. Instead of pre-defined fixed input features, SYNERGY will be able to predict the best

 doi:10.20944/preprints202208.0345.v1

combination of a feature and model for a given GHI prediction. The same approach can be used to optimize hyperparameters and other parameters in the machine learning forecasting pipeline. These feature- and parameter-related aspects of the SYNERGY approach should be investigated further before drawing firm conclusions. Furthermore, the use of additional meteorological datasets with high climate and data diversity, as well as additional forecasting methods, in conjunction with solutions to data imbalance problems, could result in a more balanced classification dataset and allow improvements in classification error, resulting in significantly better forecasting accuracies and gains.

Finally, in order to predict the best performing deep learning model for GHI forecasting, the proposed auto-selective approach currently considers minimum forecasting error. It can be extended to predict forecasting models based on additional criteria such as the amount of energy required or the speed with which the model is executed, different input features, different optimisations of the same models, or other user preferences. To improve the tool's performance and diversity, additional deep learning models for classification (to auto-select) or forecasting solar radiation can be incorporated. The method can be applied to other renewable energy sources and problems, such as wind energy forecasting.

Author Contributions: Conceptualization, G.K. and R.M.; methodology, G.K. and R.M.; software, G.K.; validation, G.K. and R.M.; formal analysis, G.K., R.M. and S.H.; investigation, G.K., R.M. and S.H.; resources, R.M. and S.H.; data curation, G.K.; writing—original draft preparation, G.K. and R.M.; writing—review and editing, R.M. and S.H.; visualization, G.K.; supervision, R.M. and S.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: We have provided details about the sources of data in the manuscript.

Acknowledgments: The work carried out in this paper is supported by the HPC Center at King Abdulaziz University. The authors would like to thank King Abdullah City for atomic and renewable energy (KACARE) for the supply of solar data.

References

[1] "Shining brightly | MIT News | Massachusetts Institute of Technology." https://news.mit.edu/2011/energy-scale-part3-1026 (accessed Aug. 14, 2022).

[2] E. Kabir, P. Kumar, S. Kumar, A. A. Adelodun, and K. H. Kim, "Solar energy: Potential and future prospects," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 894–900, Feb. 2018, doi: 10.1016/J.RSER.2017.09.094.

[3] Shell Global, "Global Energy Resources database." .

[4] A. Elrahmani, J. Hannun, F. Eljack, and M.-K. Kazi, "Status of renewable energy in the GCC region and future opportunities," *Curr. Opin. Chem. Eng.*, vol. 31, p. 100664, 2021.

[5] T. Peng, C. Zhang, J. Zhou, and M. S. Nazir, "An integrated framework of Bi-directional Long-Short Term Memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting," *Energy*, vol. 221, p. 119887, 2021.

[6] C. Voyant *et al.*, "Machine learning methods for solar radiation forecasting: A review," *Renew. Energy*, vol. 105, pp. 569–582, 2017.

[7] P. Kumari and D. Toshniwal, "Deep learning models for solar irradiance forecasting: A comprehensive review," *J. Clean. Prod.*, vol. 318, p. 128566, 2021.

[8] H. Wang *et al.*, "Taxonomy research of artificial intelligence for deterministic solar power forecasting," *Energy Convers. Manag.*, vol. 214, p. 112909, 2020.

[9] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, "Deep learning methods and applications for electrical power systems: A comprehensive review," *Int. J. Energy Res.*, 2020.

[10] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Convers. Manag.*, vol. 198, p. 111799, 2019.

[11] F. M. Reda, "Deep Learning an Overview," *neural networks*, vol. 12, no. 21, 2019.

[12] S. Shamshirband, T. Rabczuk, and K.-W. Chau, "A Survey of Deep Learning Techniques: Application in Wind and Solar Energy Resources," *IEEE Access*, vol. 7, pp. 164650–164666, 2019.

[13] I. Ahmad, F. Alqurashi, E. Abozinadah, and R. Mehmood, "Deep Journalism and DeepJournal V1.0: A Data-Driven Deep Learning Approach to Discover Parameters for Transportation," *Sustain.*, vol. 14, no. 9, p. 5711, May 2022, doi: 10.3390/SU14095711.

[14] N. Alahmari *et al.*, "Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia," *Sustain. 2022, Vol. 14, Page 3313*, vol. 14, no. 6, p. 3313, Mar. 2022, doi: 10.3390/SU14063313.

[15] S. Alswedani, R. Mehmood, and I. Katib, "Sustainable Participatory Governance: Data-Driven Discovery of Parameters for Planning Online and In-Class Education in Saudi Arabia During COVID-19," *Front. Sustain. Cities*, vol. 0, p. 97, Jul. 2022, doi: 10.3389/FRSC.2022.871171.

[16] N. Janbi *et al.*, "Imtidad: A Reference Architecture and a Case Study on Developing Distributed AI Services for Skin Disease Diagnosis over Cloud, Fog and Edge," *Sensors 2022, Vol. 22, Page 1854*, vol. 22, no. 5, p. 1854, Feb. 2022, doi: 10.3390/S22051854.

[17] G. Alkhayat and R. Mehmood, "A review and taxonomy of wind and solar energy forecasting methods based on deep learning," *Energy AI*, vol. 4, p. 100060, Jun. 2021, doi: 10.1016/j.egyai.2021.100060.

[18] P. Kumari and D. Toshniwal, "Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance," *J. Clean. Prod.*, vol. 279, p. 123285, 2021.

[19] M. A. F. B. Lima, P. C. M. Carvalho, L. M. Fernández-Ramírez, and A. P. S. Braga, "Improving solar forecasting using Deep Learning and Portfolio Theory integration," *Energy*, vol. 195, p. 117016, 2020.

[20] M. AlKandari and I. Ahmad, "Solar power generation forecasting using ensemble approach based on deep

learning and statistical methods," *Appl. Comput. Informatics*, vol. ahead-of-print, no. ahead-of-print, Jan. 2020, doi: 10.1016/j.aci.2019.11.002.

[21] 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," *Renew. Energy*, vol. 162, pp. 1665–1683, 2020.

[22] A. Fouilloy *et al.*, "Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability," *Energy*, vol. 165, pp. 620–629, 2018.

[23] J. Lago, K. De Brabandere, F. De Ridder, and B. De Schutter, "Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data," *Sol. Energy*, vol. 173, pp. 566–577, 2018.

[24] J. Lee, W. Wang, F. Harrou, and Y. Sun, "Reliable solar irradiance prediction using ensemble learning-based models: A comparative study," *Energy Convers. Manag.*, vol. 208, p. 112582, 2020.

[25] G. M. Yagli, D. Yang, and D. Srinivasan, "Automatic hourly solar forecasting using machine learning models," *Renew. Sustain. Energy Rev.*, vol. 105, pp. 487–498, 2019.

[26] S. Srivastava and S. Lessmann, "A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data," *Sol. Energy*, vol. 162, pp. 232–247, 2018.

[27] H. Bouzgou and C. A. Gueymard, "Minimum redundancy – Maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series," *Sol. Energy*, vol. 158, pp. 595–609, Dec. 2017, doi: 10.1016/j.solener.2017.10.035.

[28] M. Despotovic, V. Nedic, D. Despotovic, and S. Cvetanovic, "Review and statistical analysis of different global solar radiation sunshine models," *Renewable and Sustainable Energy Reviews*, vol. 52. Elsevier Ltd, pp. 1869–1880, Dec. 2015, doi: 10.1016/j.rser.2015.08.035.

[29] O. Behar, A. Khellaf, and K. Mohammedi, "Comparison of solar radiation models and their validation under Algerian climate - The case of direct irradiance," *Energy Convers. Manag.*, vol. 98, pp. 236–251, Jul. 2015, doi: 10.1016/j.enconman.2015.03.067.

[30] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond." Elsevier, 2016.

[31] T. Mohammed, A. Albeshri, I. Katib, and R. Mehmood, "DIESEL: A Novel Deep Learning based Tool for SpMV Computations and Solving Sparse Linear Equation Systems," *J. Supercomput.*, 2020, doi: https://doi.org/10.1007/s11227-020-03489-3.

[32] S. Usman, R. Mehmood, I. Katib, A. Albeshri, and S. M. Altowaijri, "ZAKI: A Smart Method and Tool for Automatic Performance Optimization of Parallel SpMV Computations on Distributed Memory Machines," *Mob. Networks Appl.*, 2019.

[33] S. Usman, R. Mehmood, I. Katib, and A. Albeshri, "ZAKI+: A Machine Learning Based Process Mapping Tool for SpMV Computations on Distributed Memory Architectures," *IEEE Access*, vol. 7, pp. 81279–81296, 2019, doi: 10.1109/ACCESS.2019.2923565.

[34] Y. Liu *et al.*, "Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network," *Appl. Energy*, vol. 253, p. 113596, 2019.

[35] J. Zheng *et al.*, "Time series prediction for output of multi-region solar power plants," *Appl. Energy*, vol. 257, Jan. 2020, doi: 10.1016/j.apenergy.2019.114001.

[36] X. Zhang, Y. Li, S. Lu, H. F. Hamann, B. M. Hodge, and B. Lehman, "A Solar Time Based Analog Ensemble Method for Regional Solar Power Forecasting," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 268–279, Jan. 2019, doi: 10.1109/TSTE.2018.2832634.

[37] J. Huertas-Tato, R. Aler, I. M. Galván, F. J. Rodríguez-Benítez, C. Arbizu-Barrena, and D. Pozo-Vázquez, "A short-

term solar radiation forecasting system for the Iberian Peninsula. Part 2: Model blending approaches based on machine learning," *Sol. Energy*, vol. 195, pp. 685–696, Jan. 2020, doi: 10.1016/J.SOLENER.2019.11.091.

[38] B. Brahma and R. Wadhvani, "Solar Irradiance Forecasting Based on Deep Learning Methodologies and Multi-Site Data," *Symmetry 2020, Vol. 12, Page 1830*, vol. 12, no. 11, p. 1830, Nov. 2020, doi: 10.3390/SYM12111830.

[39] W. Khan, S. Walker, and W. Zeiler, "Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach," *Energy*, vol. 240, Feb. 2022, doi: 10.1016/j.energy.2021.122812.

[40] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework," *Energy Convers. Manag.*, vol. 212, p. 112766, 2020.

[41] P. Singla, M. Duhan, and S. Saroha, "An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network," *Earth Sci. Informatics*, vol. 15, no. 1, pp. 291–306, 2022.

[42] C. Pan and J. Tan, "Day-ahead hourly forecasting of solar generation based on cluster analysis and ensemble model," *IEEE Access*, vol. 7, pp. 112921–112930, 2019.

[43] E.-S. M. El-Kenawy *et al.*, "Advanced Ensemble Model for Solar Radiation Forecasting Using Sine Cosine Algorithm and Newton's Laws," *IEEE Access*, vol. 9, pp. 115750–115765, 2021.

[44] K. Kaba, M. Sarıgül, M. Avcı, and H. M. Kandırmaz, "Estimation of daily global solar radiation using deep learning model," *Energy*, vol. 162, pp. 126–135, 2018.

[45] B. K. Jeon and E. J. Kim, "Next-Day Prediction of Hourly Solar Irradiance Using Local Weather Forecasts and LSTM Trained with Non-Local Data," *Energies 2020, Vol. 13, Page 5258*, vol. 13, no. 20, p. 5258, Oct. 2020, doi: 10.3390/EN13205258.

[46] K·A·CARE, "Renewable Resource Atlas, King Abdullah City for Atomic and Renewable Energy (K.A.CARE), Saudi Arabia," 2021. .

[47] L. Zepner, P. Karrasch, F. Wiemann, and L. Bernard, "ClimateCharts.net – an interactive climate analysis web platform," *International Journal of Digital Earth*, Mar. 2021. .

[48] M. Sengupta, A. Habte, Y. Xie, A. Lopez, and G. Buster, "National Solar Radiation Database (NSRDB)." United States, 2018, doi: https://doi.org/10.25984/1810289.

[49] F. Vignola, "GHI correlations with DHI and DNI and the effects of cloudiness on one-minute data," 2012.

[50] M. G. Yazdani, M. A. Salam, and Q. M. Rahman, "Investigation of the effect of weather conditions on solar radiation in Brunei Darussalam," *Int. J. Sustain. Energy*, vol. 35, no. 10, pp. 982–995, 2016.

[51] G. Petneházi, "Recurrent neural networks for time series forecasting," *arXiv Prepr. arXiv1901.00069*, 2019.

[52] F. Marchesoni-Acland and R. Alonso-Suárez, "Intra-day solar irradiation forecast using RLS filters and satellite images," *Renew. Energy*, vol. 161, pp. 1140–1154, 2020.

[53] G. M. S. Pereira, R. L. B. Stonoga, D. H. M. Detzel, K. K. Küster, R. A. P. Neto, and L. A. C. Paschoalotto, "Analysis and Evaluation of Gap Filling Procedures for Solar Radiation Data," in *2018 IEEE 9th Power, Instrumentation and Measurement Meeting (EPIM)*, 2018, pp. 1–6.

[54] N. B. Mohamad, A.-C. Lai, and B.-H. Lim, "A case study in the tropical region to evaluate univariate imputation methods for solar irradiance data with different weather types," *Sustain. Energy Technol. Assessments*, vol. 50, p. 101764, 2022.

[55] E. F. M. Abreu, P. Canhoto, V. Prior, and R. Melicio, "Solar resource assessment through long-term statistical analysis and typical data generation with different time resolutions using GHI measurements," *Renew. Energy*, vol. 127, pp. 398–411, 2018.

[56] KAPSARC, "KAPSARC Data Portal." .

[57] X. Tang, H. Yao, Y. Sun, C. Aggarwal, P. Mitra, and S. Wang, "Joint modeling of local and global temporal dynamics for multivariate time series forecasting with missing values," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 04, pp. 5956–5963.

[58] S. J. Hadeed, M. K. O'Rourke, J. L. Burgess, R. B. Harris, and R. A. Canales, "Imputation methods for addressing missing data in short-term monitoring of air pollutants," *Sci. Total Environ.*, vol. 730, p. 139140, 2020.

[59] B. Venkatesh and J. Anuradha, "A review of feature selection and its methods," *Cybern. Inf. Technol.*, vol. 19, no. 1, pp. 3–26, 2019.

[60] G. Memarzadeh and F. Keynia, "A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets," *Energy Convers. Manag.*, vol. 213, p. 112824, 2020.

[61] S. Shilaskar and A. Ghatol, "Feature selection for medical diagnosis: Evaluation for cardiovascular diseases," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4146–4153, 2013.

[62] V. Fonti and E. Belitser, "Feature selection using lasso," *VU Amsterdam Res. Pap. Bus. Anal.*, vol. 30, pp. 1–25, 2017.

[63] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063–78074, 2019.

[64] M. C. Sorkun, C. Paoli, and Ö. D. Incel, "Time series forecasting on solar irradiation using deep learning," in *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)*, 2017, pp. 151–155.

[65] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei, and G. Sun, "Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations," *Renew. Energy*, vol. 160, pp. 26–41, 2020.

[66] H. Zang *et al.*, "Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network," *IET Gener. Transm. Distrib.*, vol. 12, no. 20, pp. 4557–4567, 2018.

[67] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.

[68] A. Dolatabadi, H. Abdeltawab, and Y. A.-R. I. Mohamed, "Hybrid Deep Learning-Based Model for Wind Speed Forecasting Based on DWPT and Bidirectional LSTM Network," *IEEE Access*, vol. 8, pp. 229219–229232, 2020.

[69] S. Boubaker, M. Benghanem, A. Mellit, A. Lefza, O. Kahouli, and L. Kolsi, "Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia," *IEEE Access*, vol. 9, pp. 36719–36729, 2021.

[70] H. D. Nguyen, K. P. Tran, S. Thomassey, and M. Hamad, "Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management," *Int. J. Inf. Manage.*, vol. 57, p. 102282, 2021.

[71] A. Sagheer and M. Kotb, "Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems," *Sci. Rep.*, vol. 9, no. 1, pp. 1–16, 2019.

[72] G. Li, S. Xie, B. Wang, J. Xin, Y. Li, and S. Du, "Photovoltaic Power Forecasting With a Hybrid Deep Learning Approach," *IEEE Access*, vol. 8, pp. 175871–175880, 2020.

[73] M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast," *IEEE Access*, vol. 8, pp. 172524–172533, 2020.

[74] M. Alrashidi, M. Alrashidi, and S. Rahman, "Global solar radiation prediction: Application of novel hybrid data-driven model," *Appl. Soft Comput.*, vol. 112, p. 107768, 2021.

[75] C. Persson, P. Bacher, T. Shiga, and H. Madsen, "Multi-site solar power forecasting using gradient boosted regression trees," *Sol. Energy*, vol. 150, pp. 423–436, 2017, doi: 10.1016/j.solener.2017.04.066.

[76] R. Meenal and A. I. Selvakumar, "Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters," *Renew. Energy*, vol. 121, pp. 324–343, Jun. 2018, doi: 10.1016/j.renene.2017.12.005.

[77] L. Gigoni *et al.*, "Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 831–842, Apr. 2018, doi: 10.1109/TSTE.2017.2762435.

[78] R. C. Deo and M. Şahin, "Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland," *Renewable and Sustainable Energy Reviews*, vol. 72. Elsevier Ltd, pp. 828–848, 2017, doi: 10.1016/j.rser.2017.01.114.

[79] A. Marzo *et al.*, "Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation," *Renew. Energy*, vol. 113, pp. 303–311, 2017.