

Article

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

Ghadah Alkhatat ¹, Syed Hamid Hasan ¹ and Rashid Mehmood ^{2,*}

¹ Department of Computer Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

² High Performance Computing Centre, King Abdulaziz University, Jeddah 21589, Saudi Arabia

* Correspondence: rmehmood@kau.edu.sa

Abstract: Researchers have made great progress in developing cutting-edge solar energy forecasting methods. However, these methods are far from optimal in terms of their accuracy, generalizability, benchmarking, and other requirements. Particularly, no single method performs well across all climates and weather 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 generalizing a specific model for all climates, predicts the best performing deep learning model for global horizontal irradiance (GHI) forecasting in terms of forecasting error. The approach is based on carefully devised deep learning methods and feature sets created through an extensive analysis of deep learning forecasting and classification methods using ten meteorological datasets from three continents. We analyze the tool in great detail through a variety of metrics and means for performance analysis, visualization, and comparison of solar forecasting methods. SENERGY outperforms existing methods in all performance metrics including mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), the normalized versions of these three metrics (nMAE, nRMSE, nMAPE), forecast skill (FS), and relative forecasting error. The long short-term memory-autoencoder model (LSTM-AE) outperformed the other four forecasting models and achieved the best results (nMAE = nRMSE = nMAPE = 0.02). The LSTM-AE model is the most accurate in all weather conditions. Predictions for sunny days are more accurate than for cloudy days as well as for summer compared to winter. SENERGY can predict the best forecasting model with 81% accuracy. The proposed auto-selective approach can be extended to other research problems, such as wind energy forecasting, and to predict forecasting models based on different criteria such as the energy required or speed of model execution, different input features, different optimizations of the same models, or other user preferences.



Citation: Alkhatat, G.; Hasan, S.H.; Mehmood, R. SENERGY: A Novel Deep Learning-Based Auto-Selective Approach and Tool for Solar Energy Forecasting. *Energies* **2022**, *15*, 6659. <https://doi.org/10.3390/en15186659>

Academic Editors: Constantinos S. Psomopoulos, Helen C. Leligou, Ferdinanda Ponci, Josep M. Guerrero and Elisa Peñalvo-López

Received: 16 August 2022

Accepted: 8 September 2022

Published: 12 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: solar energy forecasting; generalizability; 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.

Solar energy could generate 10,000 times more than 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 linked to the national grid in November 2019. With a \$302 million investment, the plant will cover a six square kilometer area in Al-Jouf. This is the first in a series of projects under Saudi Arabia's national renewable energy program to generate 9.5 GW of renewable energy by 2023 [4].

The need for integrating solar energy into the electrical grid has motivated researchers across the globe 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 natural variability and uncertainty 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 greater performance, artificial intelligence methods such as machine and deep learning methods have grown in popularity [10,11]. 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 [12–14], transportation [15], healthcare [16], education [17], and smart cities [18]. 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 [12,19,20].

We have performed an extensive literature review (see Section 2 and [19]) on deep learning-based solar energy forecasting methods and have identified the key research gaps in this field. We explain the research gaps in Figure 1. The figure provides a performance comparison of different deep learning models. The compared works include Kumari and Toshniwal [21], Lima et al. [22], AlKandari and Ahmad [23], Gao et al. [24], Fouilloy et al. [25], Lago et al. [26], Lee et al. [27], Yagli et al. [28], Srivastava and Lessmann [29], and Bouzgou and Gueymard [30]. We will elaborate on the reasons for the selection of these methods in the later sections. In the figure, note 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 normalized MAE (nMAE). Each work is plotted as a boxplot, 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. [22] reported the performance of their proposed methods using MAPE with two datasets and is labelled as Lima et al. (MAPE, 2). Ideally, the boxplot should be closest to the *x*-axis to reflect a small value for the error metric. Additionally, the boxplot should be vertically small to reflect small variations in the error metrics for different datasets.

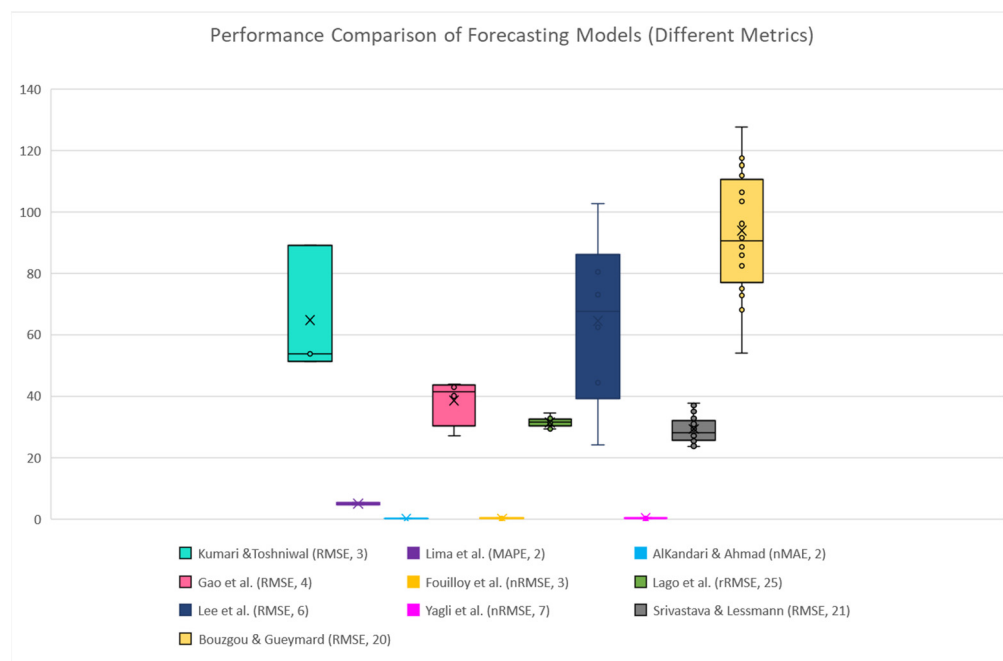


Figure 1. Performance comparison of solar forecasting models (different performance metrics).

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 the varying methods. A larger number of datasets may indicate better generalizability 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 using 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, etc. For example, although both Lima et al. [22], and AlKandari and Ahmad [23] used two datasets, it is hard to fairly compare them because the error metrics used by them are different (MAPE versus nMAE). Similarly, it is hard to compare the results reported by Kumari and Toshniwal [21] and Srivastava and Lessmann [29], 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 does 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 [12], 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 [31,32]. 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 [33]. These works and proposals demonstrate that the community has made significant progress in developing high-performance solar energy forecasting methods. However, 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 weather conditions.

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 predicts the best-performing deep learning model for GHI forecasting instead of generalizing a specific model for all climates. 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 (CNN-BiLSTM), and LSTM autoencoder (LSTM-AE). We analyze 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 mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the normalized versions of these metrics in addition to forecast skill (FS) and relative forecasting error. The LSTM-AE model outperformed the other four forecasting models and achieved the best results ($nMAE = nRMSE = nMAPE = 0.02$). The LSTM-AE model is the most accurate in all weather conditions. Prediction for sunny days is more accurate than for cloudy days and the same is true for the summer season compared to the winter season. SENERGY can predict the best forecasting model with 81% accuracy.

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 are 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.

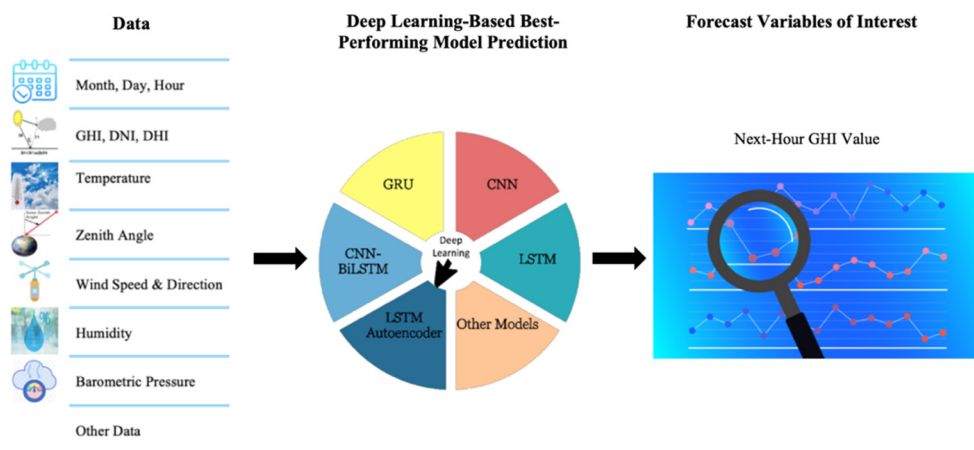


Figure 2. SENERGY: A high-level overview.

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 [34–36]. 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 the speed of model execution, different input features, different optimizations 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 summarized 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 standardization in performance evaluation of machine and deep learning modelling in solar forecasting by providing extensive analysis and visualization 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 2 reviews the related works and clarifies the research gap. Section 3 presents the methodology used in the work including subsections describing the SENERGY development process, the dataset's development process, feature importance, and the model's development process. Additionally, Section 3 includes a description of the performance evaluation metrics and tool implementation of the models. In Section 4, the results are discussed and analyzed in detail. Section 5 concludes and provides future directions.

2. Literature Review

Deep learning models' promising achievements have motivated 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 [19], we conducted an extensive review of solar and wind energy forecasting 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 first the hybrid models; next, recurrent neural network (RNN) models including the LSTM model and GRU model; and CNN models in the third place. Based on the numerous studies we reviewed, we found that deep learning-based forecasting models constantly outperform other machine learning models and statistical methods in accuracy and generalization ability, especially when they are merged with other methods in hybrid models. However, there will be no consensus regarding the best performing forecaster unless thorough testing is conducted using datasets that represent different climates and cover 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 regarding models' generalization and stability. More studies should focus on developing general forecasting models since developing a model for each location is infeasible. Few studies have suggested forecasting models for a whole region, such as [37–41]. However, general forecasting models should be able to provide forecasting for locations within different climatic zones rather than just similar regions.

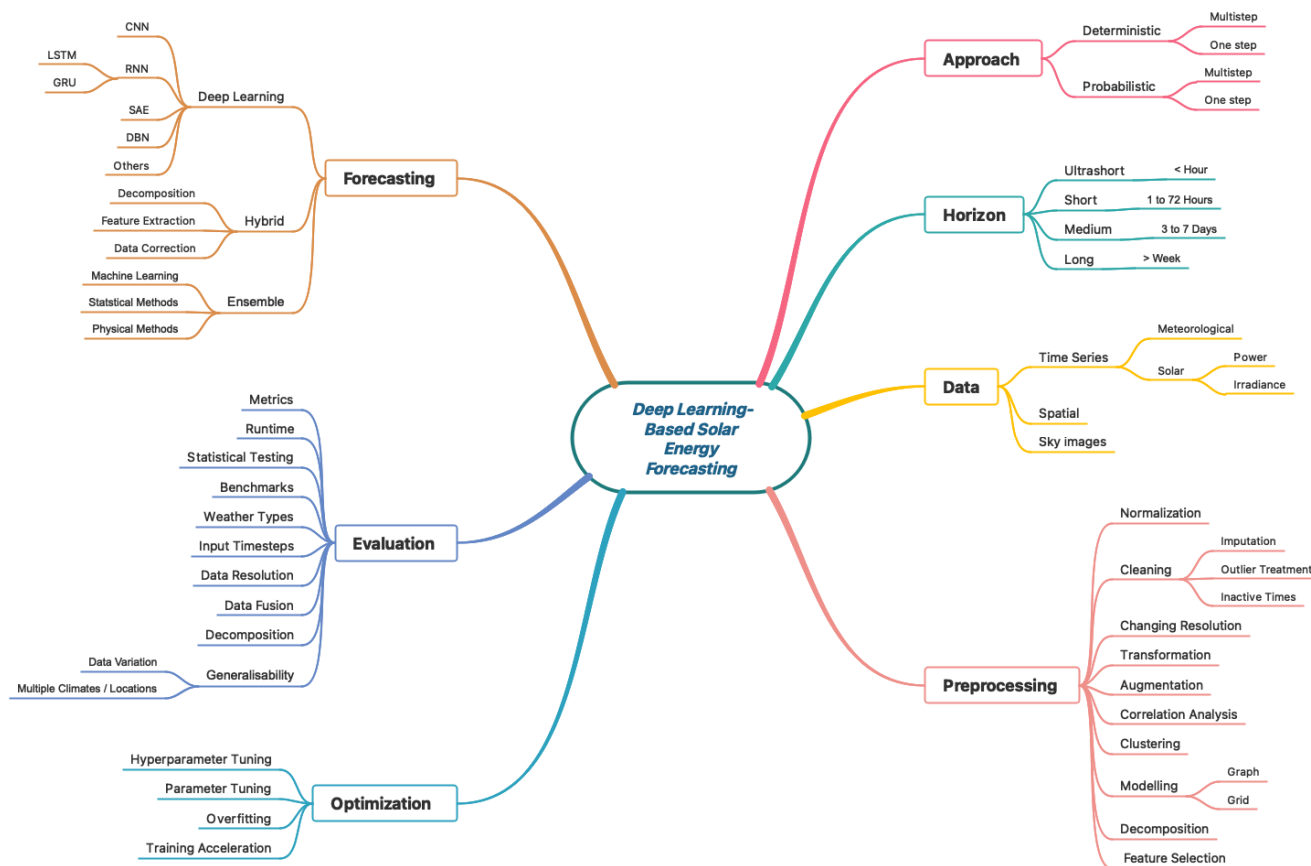


Figure 3. Deep learning-based solar energy forecasting taxonomy.

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. [22] 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 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 enables improved forecasting performance in terms of generalizability and accuracy. The authors also compared their model’s performance using datasets collected from Brazil and Spain. In addition, Khan et al. [42] combined shallow 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 and Ahmad [23] proposed a solar power forecasting model employing an ensemble approach, combining the GRU, LSTM, and theta models. They found that the ensemble technique involving 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. [43] utilized classification along with ensemble learning in their proposed PV power ensemble forecasting framework. The classification is used to identify the daily pattern label of the forecasting day to improve the forecasting accuracy performed by multiple LSTM models. Singla et al. [44] utilized wavelet transform (WT) to decompose the input time series data into different subseries. They then trained a bidirectional LSTM model for each subseries. The forecasted values of each subseries from the BiLSTM models are combined to deliver the final

24 h GHI forecast. Pan and Tan [45] performed cluster analysis on data to obtain weather regimes before employing random forests to acquire prediction from different weather regimes. El-Kenawy et al. [46] developed an ensemble model for solar radiation forecasting, which consisted of LSTM, NN, and support vector machine (SVM). This model's ensemble weights were optimized by advanced sine cosine algorithm that showed 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 found 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) to perform hourly GHI forecasting using data collected from three locations with humid subtropical, hot semi-arid, and subtropical climates of India [21]. Similarly, Gao et al. [24] tested the ability of their proposed CNN and LSTM hybrid model to provide hourly solar irradiance forecasting by using four datasets from locations with Mediterranean, semi-arid, rainforest, and desert climates. On a similar vein, Kapa et al. [47] compared the performance of a DNN model in relation to daily GHI forecasting on datasets collected from thirty-four cities in Turkey, situated in very wet, humid, semi-humid, and semi-dry climates. Fouilloy et al. [25] also compared eleven machine learning and statistical models for solar irradiance 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. [26] proposed a generalized model based on DNN for solar irradiance forecasting using data from twenty-five locations in the Netherlands, and Lee et al. [27] compared several ensemble models using datasets from six distinct locations in the USA. Yagli et al. [28] evaluated sixty-eight machine learning algorithms using data in five climate zones for hourly solar forecasting. Srivastava and Lessmann [29] compared LSTM performance in twenty-one locations in Europe and the USA with ten climate types to three other methods. Jeon and Kim [48] 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 Incheon. Bouzougou and Gueymard [30] trained an extreme learning machine (ELM) model using data from twenty locations with four different climates.

Table 1 highlights the approach used to improve forecasting performance generalization (ensemble model, multiple climates, or both) along with the results and main findings for each paper covered in this section. 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 in order to achieve greater accuracy. In this paper, we have 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.

Research Gap

The literature review presented in this section (also see [19]) 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 vary 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 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.

Table 1. Summary of the literature review.

Ref No.	Ensemble Model	Multiple Climates	Results	Main Findings
[21]	✓	✓	The ensemble model (XGBF-DNN) performed better than smart persistence, SVR, random forest (RF), XGBoost, and DNN models in hourly GHI prediction for all three locations in India and can be used for other locations.	The ensemble model (XGBF-DNN) attained RMSE = 53.79 for Jaipur, RMSE = 51.35 for New Delhi, and RMSE = 89.13 for Gangtok.
[22]	✓	✓	Integrating of LSTM, MLP, RBF, and SVR forecasting techniques provided better performance than the individual models for Brazil and Spain in 1 h ahead PV power forecasting.	The ensemble model of LSTM, MLP, RBF, and SVR achieved MAPE = 5.36% for Spain and 4.52% for Brazil.
[42]	✓		The ensemble model of ANN, LSTM, and XGBoost performed better than ANN and LSTM models alone in PV power 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 h data resolution.
[23]	✓	✓	The ensemble model of GRU, LSTM, and Theta achieved better performance with Shagaya dataset than with Cocoa because of the additional weather data and it achieved better accuracy than single ML algorithms and theta model in day-ahead solar power forecast for both locations.	The ensemble model of GRU, LSTM, and Theta achieved nMAE = 0.0317 for Shagaya in Kuwait while LSTM model alone achieved nMAE = 0.0739 for Cocoa in USA, which is slightly better than the ensemble model performance with nMAE = 0.0877.
[43]	✓		The ensemble model of LSTMs attained better performance than back propagation neural network (BPNN), SVM, and persistent models in day-ahead PV power forecasting.	The ensemble model of LSTMs attained RMSE = 5.68.
[44]	✓		The ensemble model of WT and bidirectional LSTM outperformed the naïve predictor, LSTM, BiLSTM, GRU and two different WT based BiLSTM models in 24 h ahead solar irradiance forecast.	The ensemble model of WT and bidirectional LSTM attained annual average RMSE = 45.61 and MAPE = 6.48%.
[45]	✓		The ensemble model of RF with cluster analysis for day-ahead solar forecasting performed better than RF alone and gradient boosted regression trees because weather classification improved the accuracy.	The ensemble model of RF with cluster analysis attained nRMSE = 8.8.
[46]	✓		The ensemble model of LSTM, NN, and SVM for solar radiation forecasting, optimized using advanced sine and cosine algorithm, outperformed all the reference models.	The ensemble model of LSTM, NN, and SVM achieved RMSE = 0.0018.
[24]		✓	The hybrid model of complete ensemble empirical mode decomposition adaptive noise (CEEMD), CNN, and LSTM to forecast hourly irradiance performed better compared to LSTM, BPNN, and SVM models as well as the hybrid CEEMDAN-LSTM, CEEMDAN-BPNN, and CEEMDAN-SVM models.	The hybrid model of CEEMD, CNN, and LSTM achieved annual RMSE = 42.84 for Tamanrasset, 43.98 for Hawaii's Big Island, 40.60 for Denver, and 27.09 for Los Angeles.
[47]		✓	The DNN model for daily GHI prediction showed good performance with 34 cities in Turkey using all inputs (extraterrestrial radiation, sunshine duration, cloud cover, and maximum and minimum temperature).	The DNN achieved RMSE ranges from 0.52 to 1.29 for 34 cities, which represent all climatic conditions in Turkey.

Table 1. Cont.

Ref No.	Ensemble Model	Multiple Climates	Results	Main Findings
[25]		✓	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, bagged regression tree and RF approaches performed better.	For a medium and low variability dataset (Tilos and Ajaccio), the best 1 h ahead forecasting is MAE = 71.27 and 54.58 achieved by SVR model, whereas for a high variability dataset (Odeillo), the best result is 97.48 achieved by RF.
[26]		✓	The 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 global DNN attained average relative RMSE = 31.31%, where the lowest relative RMSE = 29.24 for Hoek v. H. site and the highest relative RMSE = 34.55 for Deelen site.
[27]	✓	✓	The ensemble models (boosted trees, bagged trees, RF, and generalized RF) for short-term solar irradiance forecast outperformed SVR and Gaussian process regression.	The ensemble model 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).
[28]		✓	For hourly solar forecasting, tree-based methods were better in all-sky conditions, whereas variants of MLP and SVR were better in clear-sky and RF with quantile regression in overcast sky conditions.	Tree-based methods are superior for all-sky conditions with nRMSE ranges from 15.46% to 33.36% based on location.
[29]		✓	The LSTM model outperformed persistence, FFNN, and gradient boosting regression methods in day-ahead GHI forecasting.	The LSTM model achieved RMSE ranges from 23.6 to 37.78 for 21 locations.
[48]		✓	The global LSTM model, which was trained with international data for next-day GHI prediction, was able to predict GHI in Korea.	The global LSTM model achieved RMSE = 30 with Incheon in Korea.
[30]		✓	The ELM model, which was trained with data from 20 locations, has good performance for 15 min, 1 h, and 24 h ahead forecasting.	The ELM model achieved average RMSE = 93.82 for 20 locations for 1 h ahead forecast.

Using ten datasets from three continents, we conducted 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 analysis. None of the works reported forecasting based on five deep learning-based models with such many locations in Saudi Arabia, nor did they provide a comparison with locations elsewhere (Toronto and Caracas). We highlight the need for standardization 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 provide an overview of the SENERGY development process in Section 3.1, then we describe the steps in detail in later sections. In Section 3.2, we describe the datasets development process, 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 used in SENERGY are explained in Section 3.4. These are long short-term memory, gated recurrent unit, convolutional neural network in section, hybrid convolutional neural network 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 is discussed in Section 3.6.

3.1. Tool Development Process

The SENERGY development process is displayed in Figure 4. This starts with collecting datasets from multiple locations that have different climates and is 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. Next, preprocessed data are 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.

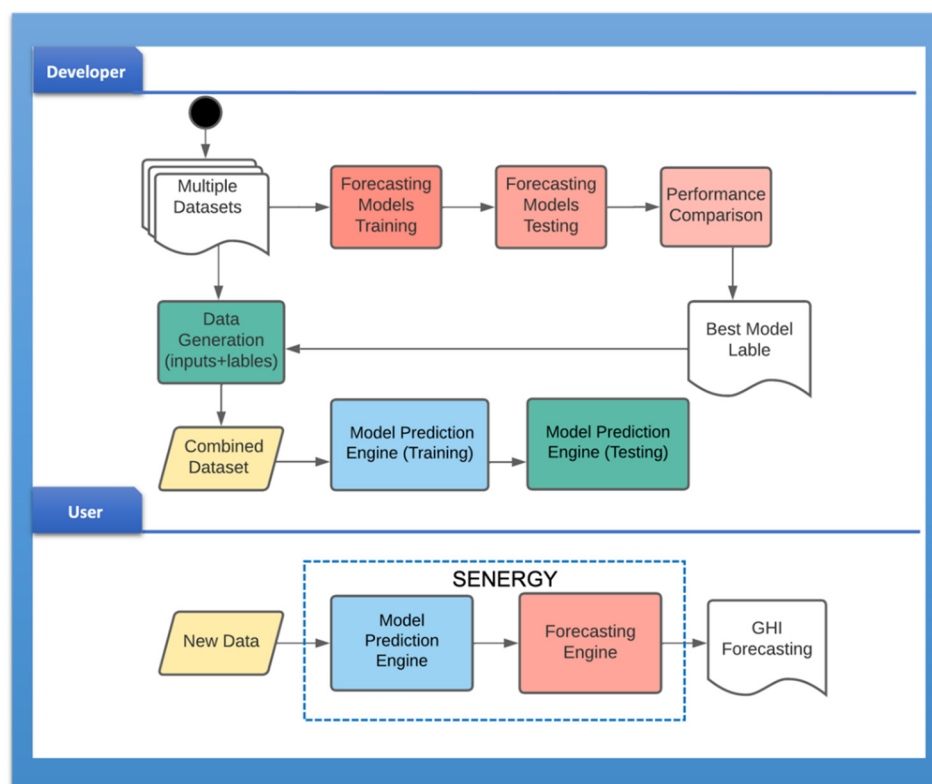


Figure 4. SENERGY development process.

3.2. Datasets Development

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

3.2.1. Data Collection

We used a total of ten datasets, eight of which were collected from solar monitoring stations in Saudi Arabia and the remaining two from a Toronto dataset and a Caracas dataset. The used datasets represent three different climates and contain records for five years, ensuring the inclusion of a variety of weather types, such as sunny, cloudy, and rainy, etc. The datasets from the Saudi Arabian locations were provided by the King Abdullah City for Atomic and Renewable Energy (K.A.CARE) [49]. They contain the measurements of three components of solar radiation: global horizontal irradiance (GHI), direct normal

irradiance (DNI), 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 min. Tier 1 stations provide the highest quality data, with an uncertainty of $\pm 2\%$ (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 the Köppen classification obtained from ClimateCharts.net [50]. Figure 5 shows the solar stations' locations on the Saudi Arabia map.

Table 2. Saudi Solar monitoring stations information.

Station No.	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 a map of Saudi Arabia.

The datasets of Toronto, Canada and Caracas, Venezuela were collected from the National Solar Radiation Database accessed through the National Renewable Energy Laboratory (NREL) website [51]. These datasets were gathered by geostationary satellites, unlike the Saudi datasets, which were collected from ground stations. The climate classification of

Toronto is humid continental (Dfb) and that of Caracas is tropical (A), according to Köppen classification. Table 3 provides the source information of both datasets and Figure 6 shows the Caracas and Toronto locations on a 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

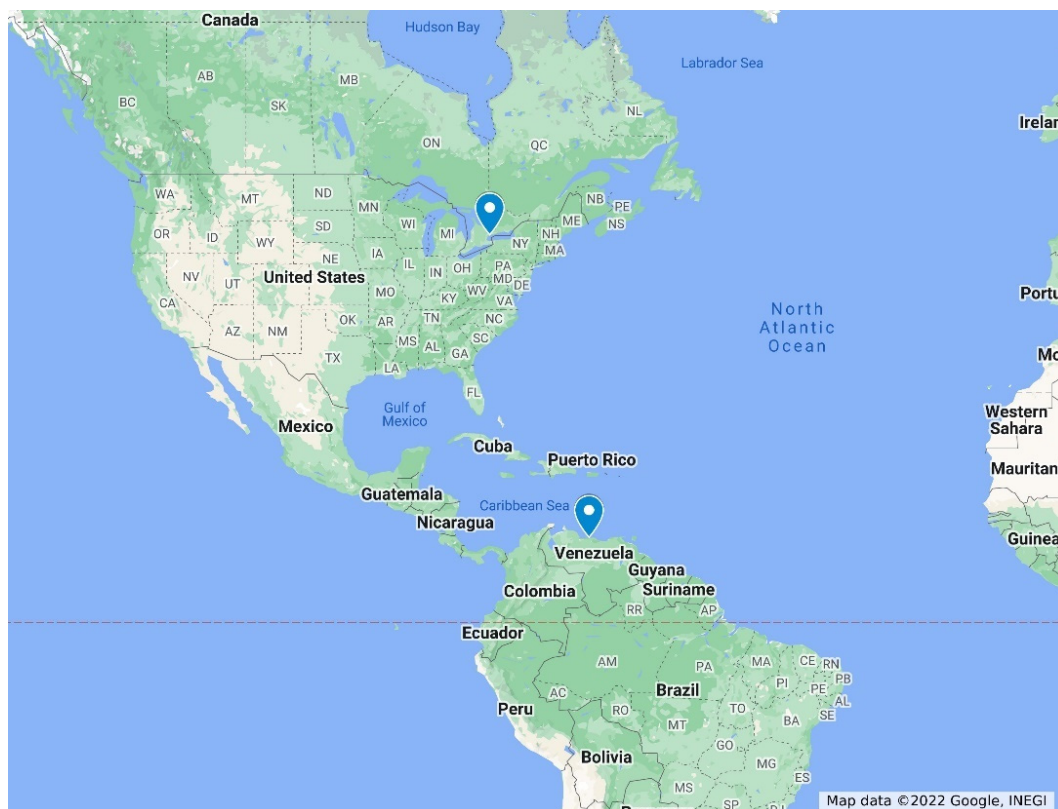


Figure 6. Toronto and Caracas locations on a map.

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 between them are clarified. Then, the steps for creating lagged features and temporal features steps are explained. Next, the steps for 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, such as numerical weather prediction (NWP) models' outputs, are also used as inputs [19]. In our work, the following nine measurements were 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).

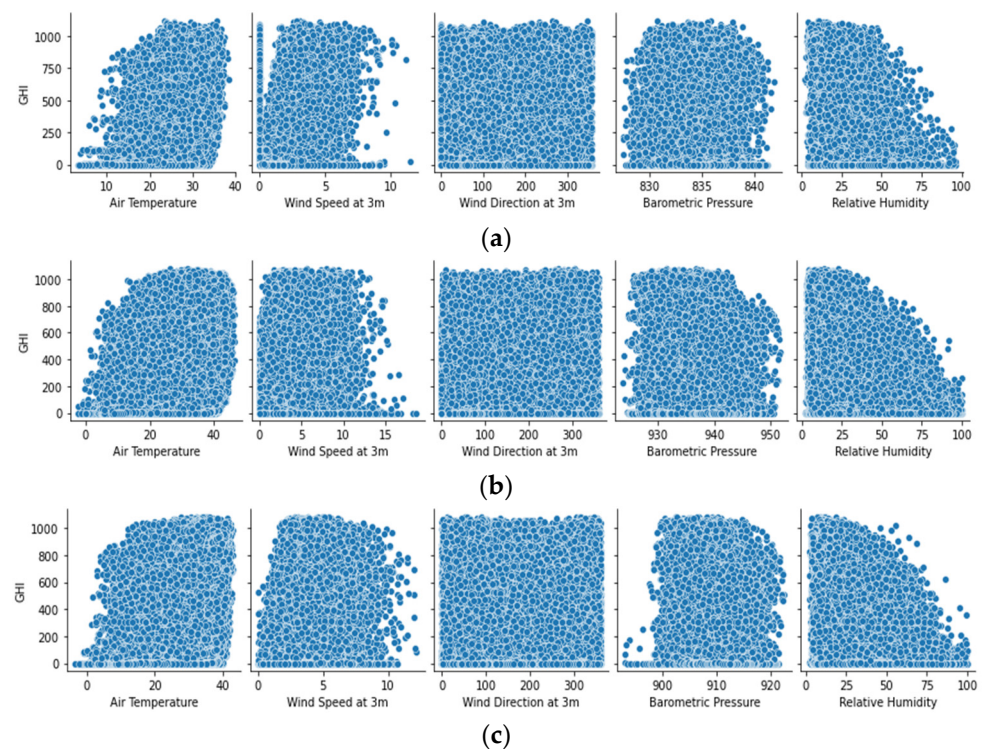


Figure 7. The relationship between GHI and the meteorological variables in: (a) Al-Baha; (b) Al-Jouf; and (c) Hail.

1. GHI: the total amount of shortwave radiation received from sun 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) [52];

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

2. DHI: solar radiation that does not come on a direct path from the sun, but has been spread by particles and molecules in the atmosphere and comes equally from all directions;
3. DNI: solar radiation that comes in a straight path from the direction of the sun at its current place in the sky. On a sunny day, GHI consists of 20% DHI and 80% DNI [52].
4. ZA: the angle between the sun's rays and a vertical line;
5. Air temperature (AT). This has a positive correlation with solar radiation [53] as can be seen in Figure 7;
6. Wind speed (WS) and wind direction (WD) at 3 m;
7. Barometric pressure (BP);
8. Relative humidity (RH). This has a negative correlation with solar radiation [53] as shown in Figure 7.

Using the previous three hours' measurements (lag = 3 h), 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 the library Pandas. 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 the partial autocorrelation function (PACF) for GHI as presented in Figure 8. The ACF shows a correlation of GHI with its 3 past values, whereas 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.

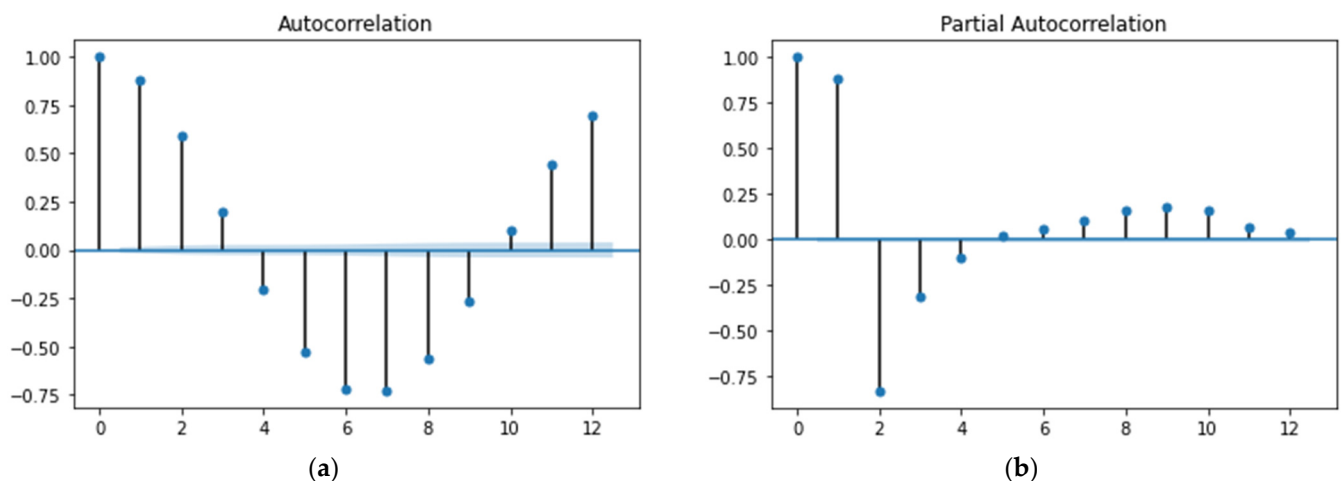
Table 4. Forecasting datasets features.

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/m ²
Hour_sin (HS)	DNI_lag1	DNI_lag2	DNI_lag3	Wh/m ²
Hour_cos (HC)	DHI_lag1	DHI_lag2	DHI_lag3	Wh/m ²
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.

Time Stamp	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

**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. These variables have a cyclical nature, for example, day 1 in a month is very close to day 30 of the previous month and the hour 12 pm is very close to the hour 1 am. Treating temporal variables as regular numbers would make day 1 of a month far from day 30 of the last month, even though the difference is one day, the same as the difference between days 1 and 2 of the same month. To avoid this problem that might affect the models' learning, we decided to eliminate the effect of the cyclical nature of time by encoding these variables into sine and cosine using the following equations [54]. The result of this transformation is an additional six features (hour sine, hour cosine, day sine, day cosine, month sine, and

month cosine). The total number of features used for training the forecasting models is thirty-three as shown in Table 4.

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

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

As mentioned earlier, there are missing records for many days in the Saudi datasets. To address this during input-output construction, we eliminated any hour record that does not have the previous three consecutive hours' records [55]. Records from the years 2016, 2017, and 2018 were used for training, whereas records from the years 2019 and 2020 were used for validation and testing, respectively. However, for the Arar, Al-Khafji, and Tabuk datasets, the number of missing days is large. Therefore, records from the year 2020 and the first four months of 2021 were used as testing sets of these locations. In the Wadi-Addawasir, Arar, and Al-Baha datasets, a few DHI values are missing, and they were filled by Equation (1). Many values of wind direction and wind speed are missing in the 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 a situation, researchers in this field usually either use a regression model to predict the missing values or use another source of data, such as a nearby station [56–58]. Since regression model accuracy might affect the data quality, we decided to use a nearby station's data to fill the missing wind speed/direction values in the Wadi-Addawasir and Taif datasets. The source of such data is the King Abdullah Petroleum Studies and Research Center (KAPSARC) [59]. The number of hourly records filled in Wadi-Addawasir dataset is 11,978 h, whereas it is 7630 in the Taif datasets. On the other hand, the Tabuk dataset only has 529 missing hours' records. Therefore, we decided to eliminate these records since records from the year 2021 had been added to the dataset to compensate for the shortage. Comparing methods for filling missing values and studying their impact on the forecasting results, as performed in [60,61], would be an opportunity for future work.

Preprocessing steps also included 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 scaling, 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 for the training, validation, and testing datasets. 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.

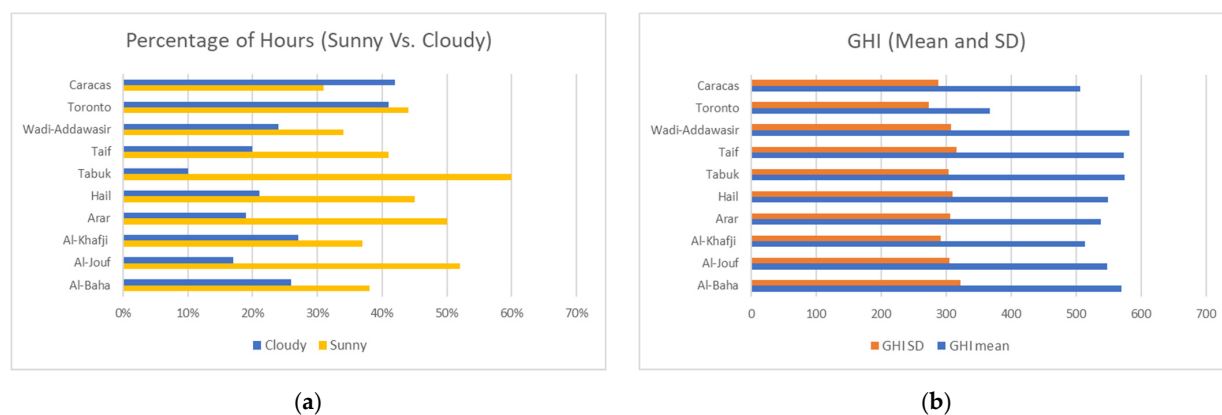


Figure 9. (a) Percentage of hours (sunny Vs. cloudy) of 10 datasets; (b) GHI (mean and SD) of 10 datasets.

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	104,896.29
	Val: 3056		552.10	325.90	106,176.30
	Test: 2247		582.09	311.16	96,780.11
Al-Jouf	Train: 8600	363 days	554.11	307.66	94,643.25
	Val: 2991		547.92	306.49	93,901.92
	Test: 2554		528.14	296.47	87,858.12
Al-Khafji	Train: 4618	970 days (Year 2019)	504.81	288.56	83,245.88
	Val: 2363		555.17	308.66	95,231.29
	Test: 2110		486.59	275.73	75,991.13
Arar	Train: 8339	575 days	546.71	310.06	96,128.23
	Val: 3589		537.73	300.20	90,097.23
	Test: 1357		485.46	295.04	86,983.40
Hail	Train: 8723	271 days	552.26	311.69	97,140.65
	Val: 3260		544.05	310.67	96,486.20
	Test: 2561		543.77	303.82	92,270.30
Tabuk	Train: 7576	542 days	593.27	310.35	96,307.42
	Val: 3100		579.62	303.93	92,342.88
	Test: 1937		498.03	261.73	68,465.05
Taif	Train: 8618	272 days	580.83	321.62	103,424.30
	Val: 3386		562.14	308.42	95,094.37
	Test: 2543		567.62	308.47	95,115.01
Wadi-Addawasir	Train: 9199	242 days	584.98	309.00	95,474.22
	Val: 3450		579.24	306.12	93,684.80
	Test: 2551		578.02	301.69	90,982.42
Caracas	Train: 10,112	0 days	499.28	284.48	80,922.07
	Val: 3428		505.95	288.71	83,327.90
	Test: 3428		524.82	297.12	88,255.24
Toronto	Train: 9892	0 days	381.15	273.39	74,732.91
	Val: 3392		336.74	266.95	71,242.70
	Test: 3388		366.77	278.11	77,322.36
All	Train: 81,904 Val: 32,015 Test: 24,676	3870 days	-	-	-

3.2.3. Datasets for Model Prediction

Preparing data for the auto-selective model prediction engine began by combining all ten datasets into one dataset and 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, we first 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 is the model that achieves the least forecasting error.

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

Figure 10 shows a snapshot of a few data records after adding the “Best model” feature to the thirty-three features used for forecasting. 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 24,576 (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.

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 (Section 3.3.1), mutual information (Section 3.3.2), forward feature selection and backward feature elimination (Section 3.3.3), and LASSO (Section 3.3.4). However, we did not eliminate any feature listed in Table 4 based on the results of these four methods since there was no agreement between them. For example, a feature that is considered insignificant by one 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 first training the models using only the first lagged features, and 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 is a measure of the linear relationship between two variables [62]. Figure 11 displays the correlation matrix for Al-Jouf and the same is displayed for Al-Khafji in Figure 12. The correlation matrices for Caracas and Toronto are shown in Figures 13 and 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, whereas the strongest negative correlation is between GHI and the hour cosine, except for the Toronto dataset, which is between GHI and the ZA of lag 1. From Table 7, it can be seen 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, the DNI of lag 2 is more important in the Toronto dataset than in other datasets.

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

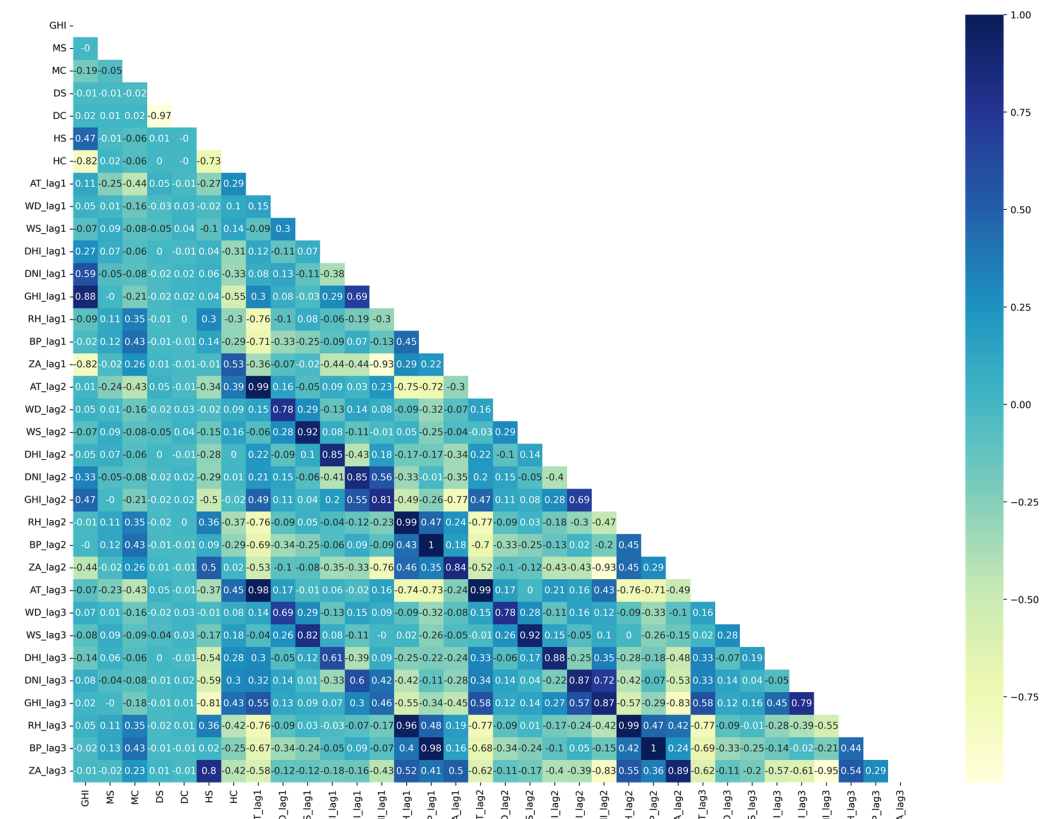


Figure 11. Al-Jouf dataset correlation matrix.

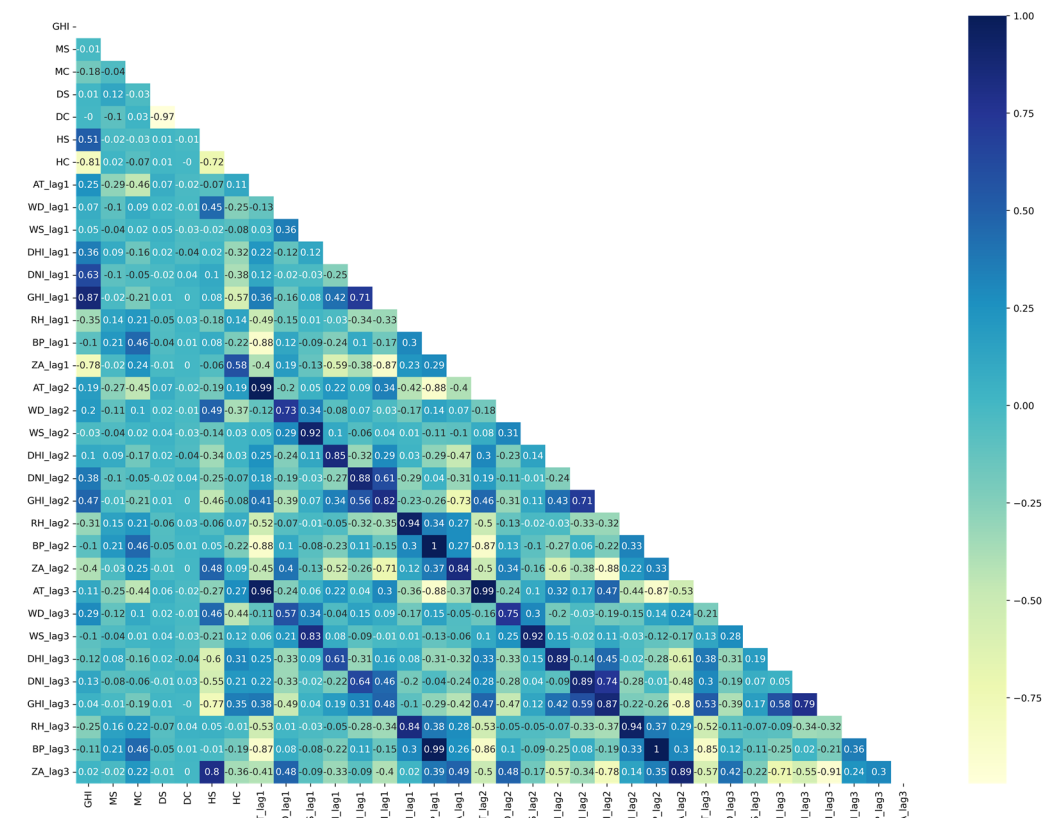


Figure 12. Al-Khaffi dataset correlation matrix.

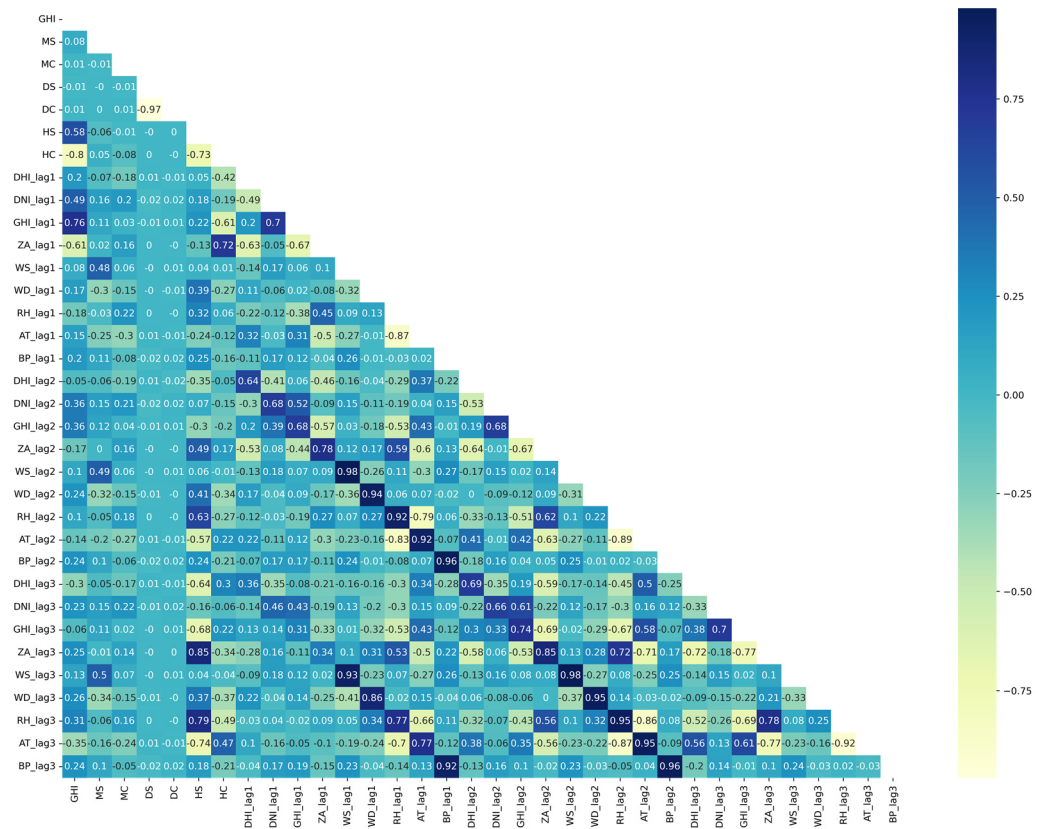


Figure 13. Caracas dataset correlation matrix.

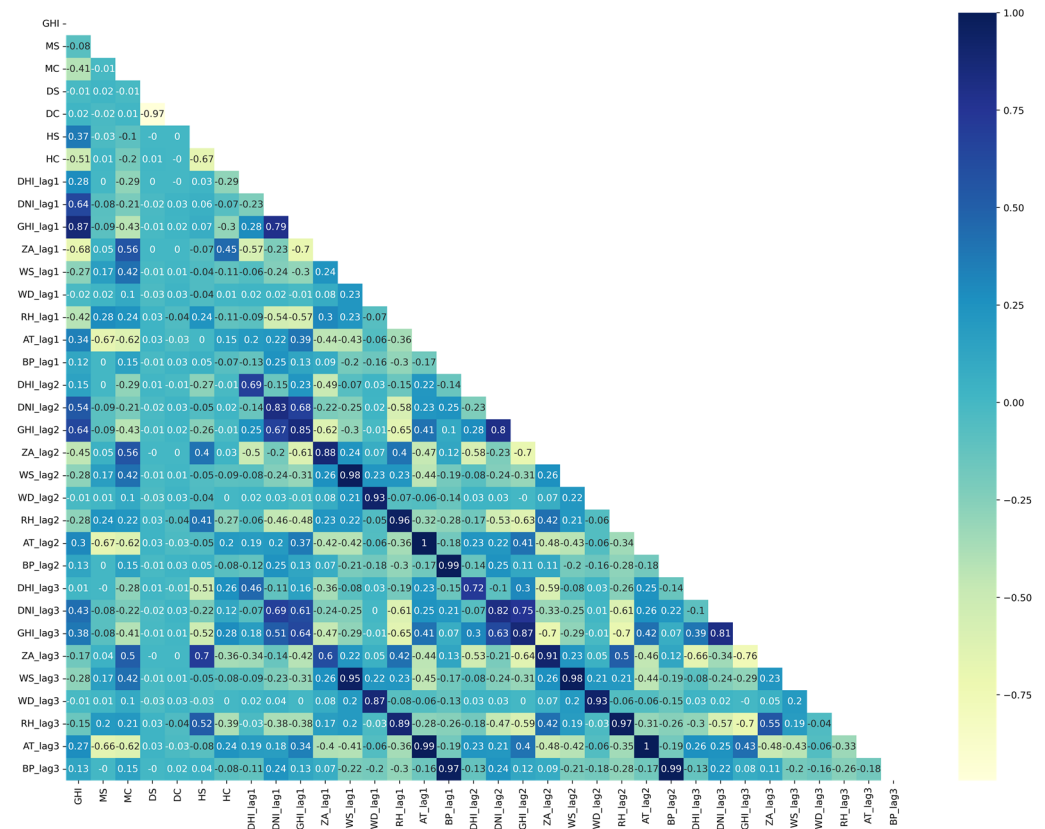


Figure 14. Toronto dataset correlation matrix.

3.3.2. Mutual Information

Mutual information (MI) measures the reduction in uncertainty for one variable given a known value of the other variable [63]. 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 and climate have 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 the Toronto dataset than in the others. GHI and Zenith Angle lagged observations are more important in the Saudi locations than in Caracas or Toronto.

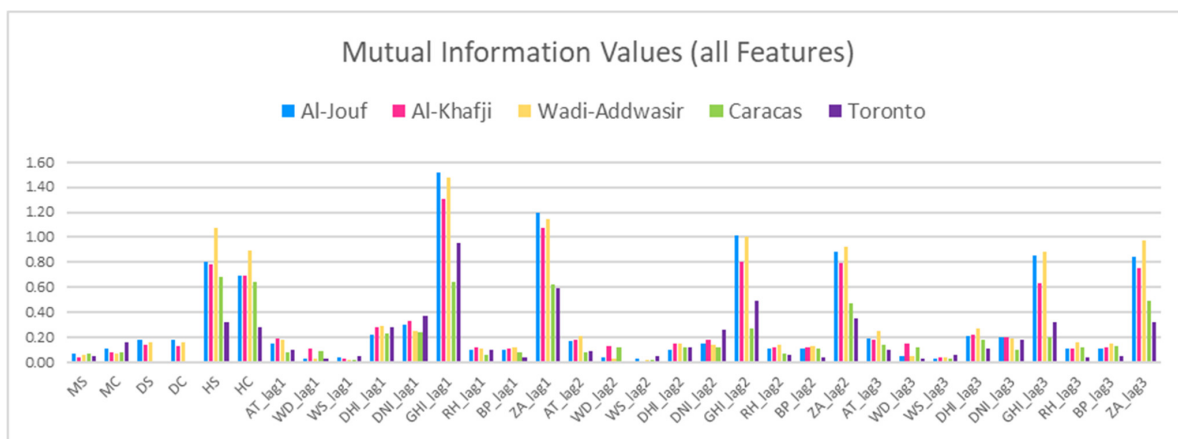


Figure 15. MI values of all features for Al-Jouf, Al-Khafji, Wadi-Addawasir, Caracas, and Toronto datasets.

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 best feature in improving the model is added until the addition of a new feature does not enhance the performance. On the other hand, backward feature elimination starts with all the features and eliminates the least important feature at each iteration. This process is repeated until no improvement is attained with feature elimination [64]. Table 8 shows ten selected features by FFS and 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 the datasets, such as the hour sine, DHI, DNI, and GHI lagged observations, whereas other features are rarely selected, such as the wind speed, relative humidity, barometric pressure, and air temperature.

Table 8. Selected features by FFS and 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 reducing some of the regression coefficients to zero. After the reduction, non-zero values are selected to be used as features in the model [65]. Figure 16 shows the selected features based on the LASSO method for five datasets (Al-Jouf, Al-Khafji, 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 the Caracas dataset. Surprisingly, time-related features are insignificant in all five datasets.

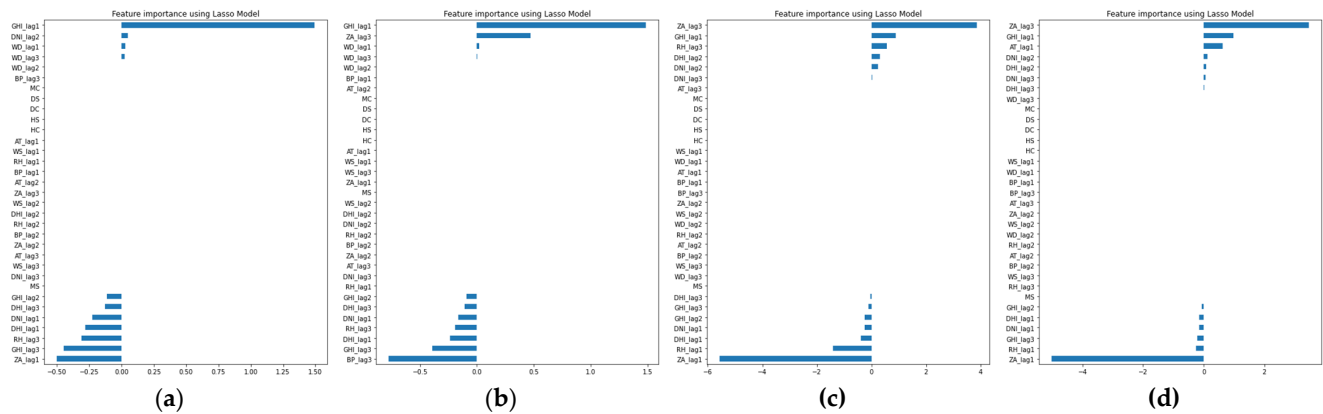


Figure 16. Selected features based on the LASSO method: (a) Al-Jouf; (b) Al-Khafji; (c) Caracas; (d) Toronto.

3.4. Models' Development

In this section, five deep learning models have been explained: long short-term memory (Section 3.4.1), gated recurrent unit (Section 3.4.2), convolutional neural network (Section 3.4.3), hybrid CNN and bidirectional LSTM (Section 3.4.4), and LSTM autoencoder (Section 3.4.5). All five models are used for next-hour GHI forecasting while only the LSTM model is also used for classification to serve as an auto-selective model prediction engine.

3.4.1. Long Short-Term Memory (LSTM)

LSTM is a special kind of RNN that has the ability of learning long-term dependencies. It performs better than traditional RNN in diverse tasks. Besides the hidden state, LSTMs contain the cell state that conveys important inputs from previous steps to later steps. Meanwhile, new inputs are added to or deleted from the cell state through input and forget gates. The output gate determines if the current memory cell will be output. More details about LSTM can be found in [5,66].

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 an auto-selective model prediction engine with two differences. First, two dense layers with eight and two neurons are used for classification instead of regression, respectively. Second, the criterion function is cross-entropy loss instead of mean squared error loss (MSE).

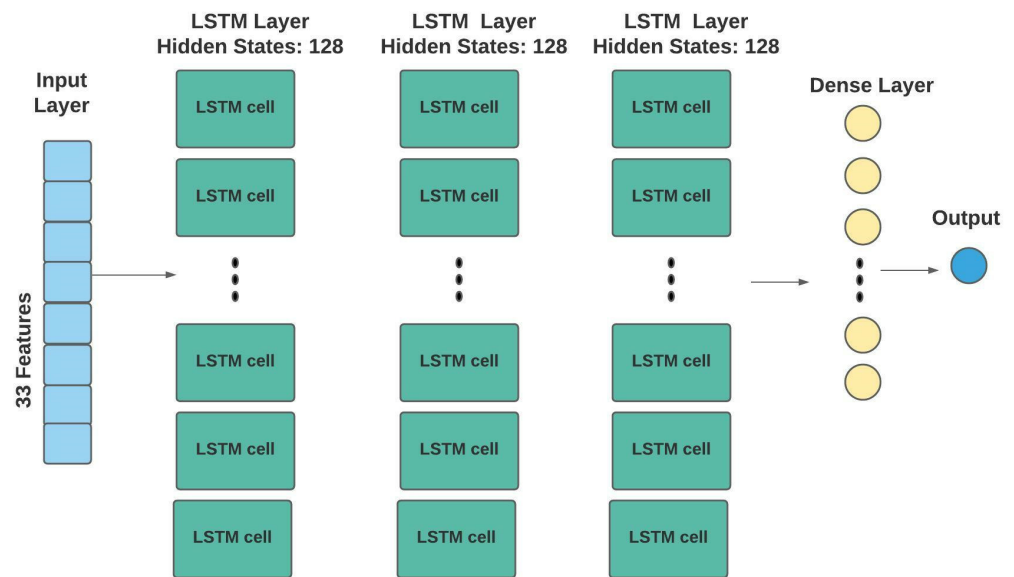


Figure 17. LSTM forecasting model.

3.4.2. Gated Recurrent Unit (GRU)

GRU is similar to LSTM in that it also captures long-term dependencies but does not contain the cell state. The update gate in GRU determines the amount of the past information needs to be kept, whereas the reset gate determines how much to forget. GRUs are usually faster and need less computation time and memory than LSTMs [67].

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.

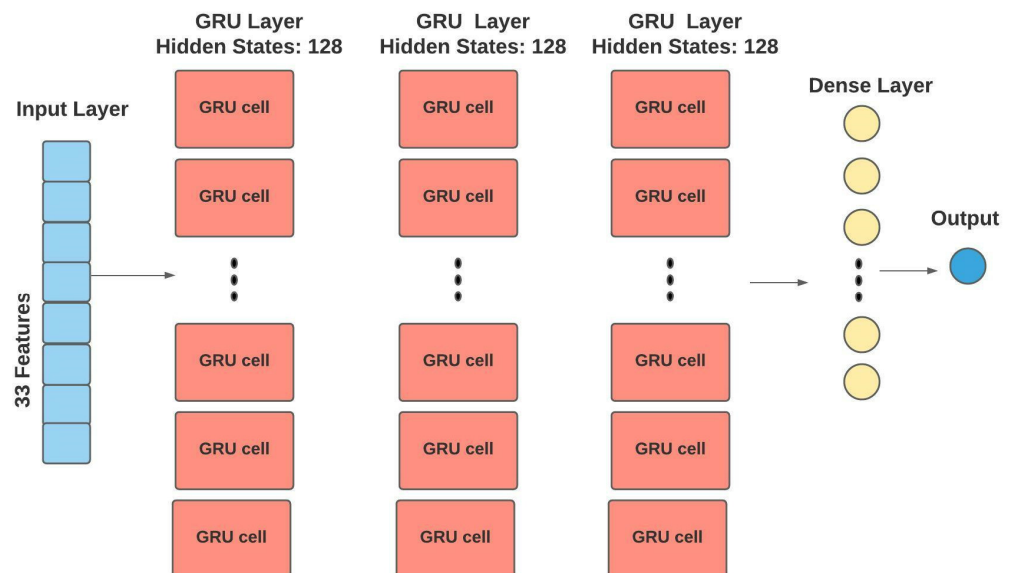


Figure 18. GRU forecasting model.

3.4.3. Convolutional Neural Network (CNN)

CNN is a kind 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. 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 on CNNs can be found in [68,69].

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 two and a stride of two, whereas in the second convolutional layer, five feature maps are created. The max-pooling layer uses a kernel of size two and a stride of one.

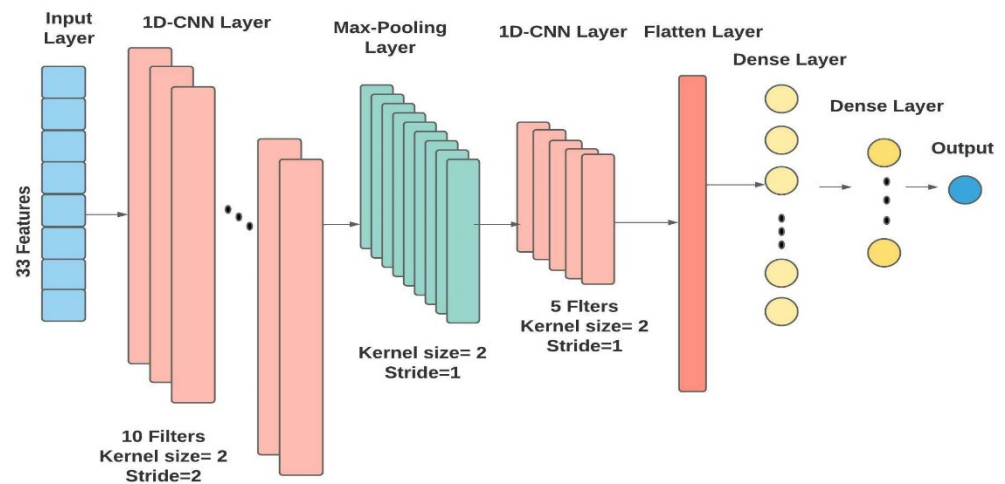


Figure 19. CNN forecasting model.

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 backwards direction. This structure allows learning from past and future information. More details on BiLSTMs can be found in [70,71].

In CNN and BiLSTM structure, convolutional and pooling layers are followed by BiLSTM layers, then one or more dense layers to generate the output [72].

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.

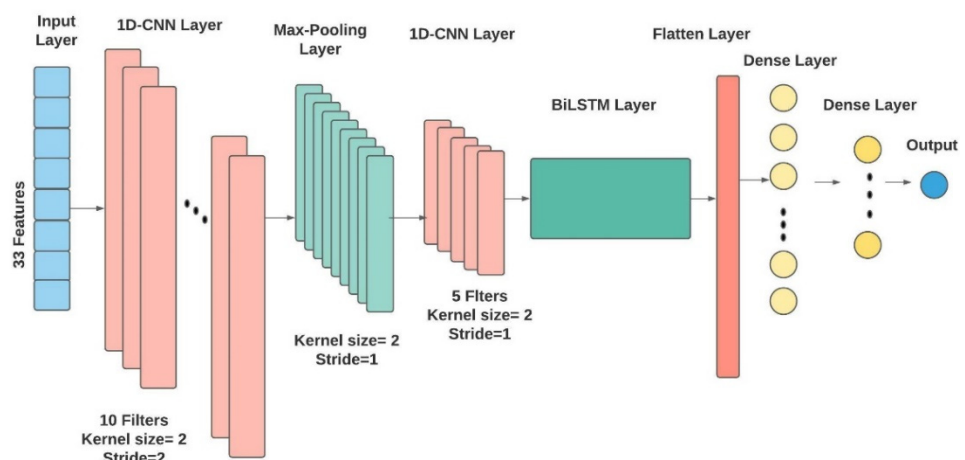


Figure 20. CNN-BiLSTM forecasting model.

3.4.5. LSTM Autoencoder (LSTM-AE)

Autoencoder is a neural network that comprises 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. The 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 [73,74].

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.

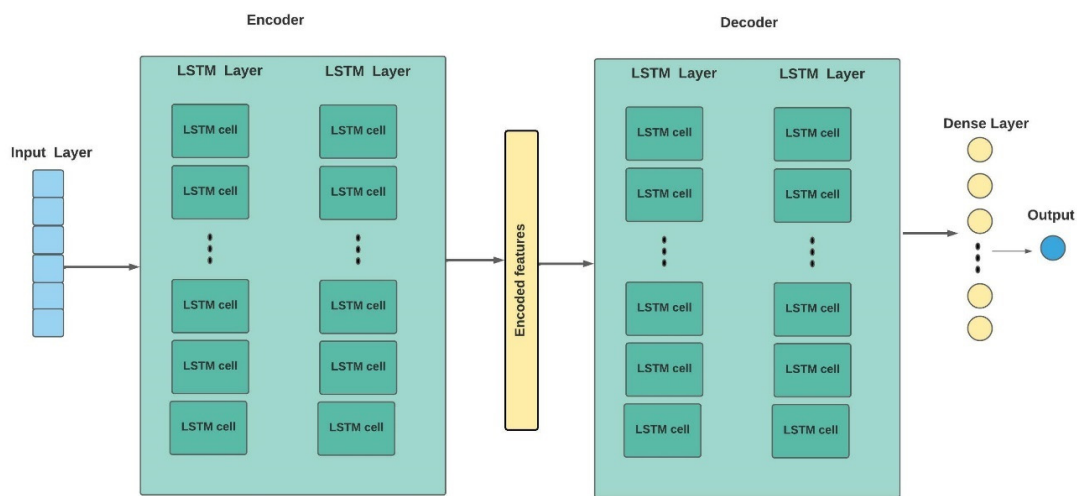


Figure 21. LSTM-AE forecasting model.

3.5. Performance Evaluation Metrics

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 [75].

$$MAE = \frac{1}{N} \sum_{i=1}^N |actual\ GHI_i - forecast\ GHI_i| \quad (5)$$

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 are concentrated around the line of regression. A lower value of RMSE is better. It is calculated as follows [75].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (actual\ GHI_i - forecast\ GHI_i)^2} \quad (6)$$

Coefficient of determination (R^2) is a statistical measure that defines the amount of variance in the dependent variable that can be explained by the independent variable. It displays the data's fit to 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 [75].

$$R^2 = 1 - \frac{\sum_{i=1}^N (actual\ GHI_i - forecast\ GHI_i)^2}{\sum_{i=1}^N (actual\ GHI_i - \overline{GHI})^2} \quad (7)$$

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 [76].

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\text{actual } GHI_i - \text{forecast } GHI_i}{\text{actual } GHI_i} \right| \times 100\% \quad (8)$$

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 more fair comparison [6]. Normalization could be applied to any metric, such as MAE, RMSE, and MAPE.

$$nMetric = \frac{Metric}{GHI} \quad (9)$$

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].

$$FS = 1 - \frac{Metric_{\text{proposed}}}{Metric_{\text{persistence}}} \times 100\% \quad (10)$$

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, 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 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' hyperparameters.

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 decay = 0.000001, Adam

4. SENERGY: Results and Evaluation

The performance of the forecasting engine and auto-selective model prediction engine components of SENERGY is evaluated in Sections 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 gains and losses in

terms of forecasting performance using SENERGY are discussed in Section 4.3. Finally, a comparison of SENERGY's 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 (Section 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 performed using four aspects: climate and location (Section 4.1.2), sunny and cloudy weather (Section 4.1.3), summer and winter seasons (Section 4.1.4), and forecasting error results (Section 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 a lag equal to three (the last three hours of observations). In this Section, we use the Toronto dataset to study the effect of using lags equal to one, two, and three 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 lags equal to one, two, and three with the Toronto dataset. With the LSTM, GRU, and CNN-BiLSTM models, it is apparent that using a lag of three made the results slightly worse. However, with MAPE, using a lag of three improved the results. In contrast, the LSTM-AE model achieved better results with a lag of two than a lag of one in all error metrics and achieved the best results with a lag of three. Given the fact that GHI is only highly correlated with GHI for a lag of one (see Table 7), using a lag equal to one would give satisfactory results, especially if dimensionality might affect the model efficiency. Otherwise, it is worth trying different lagged features to see if this 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.

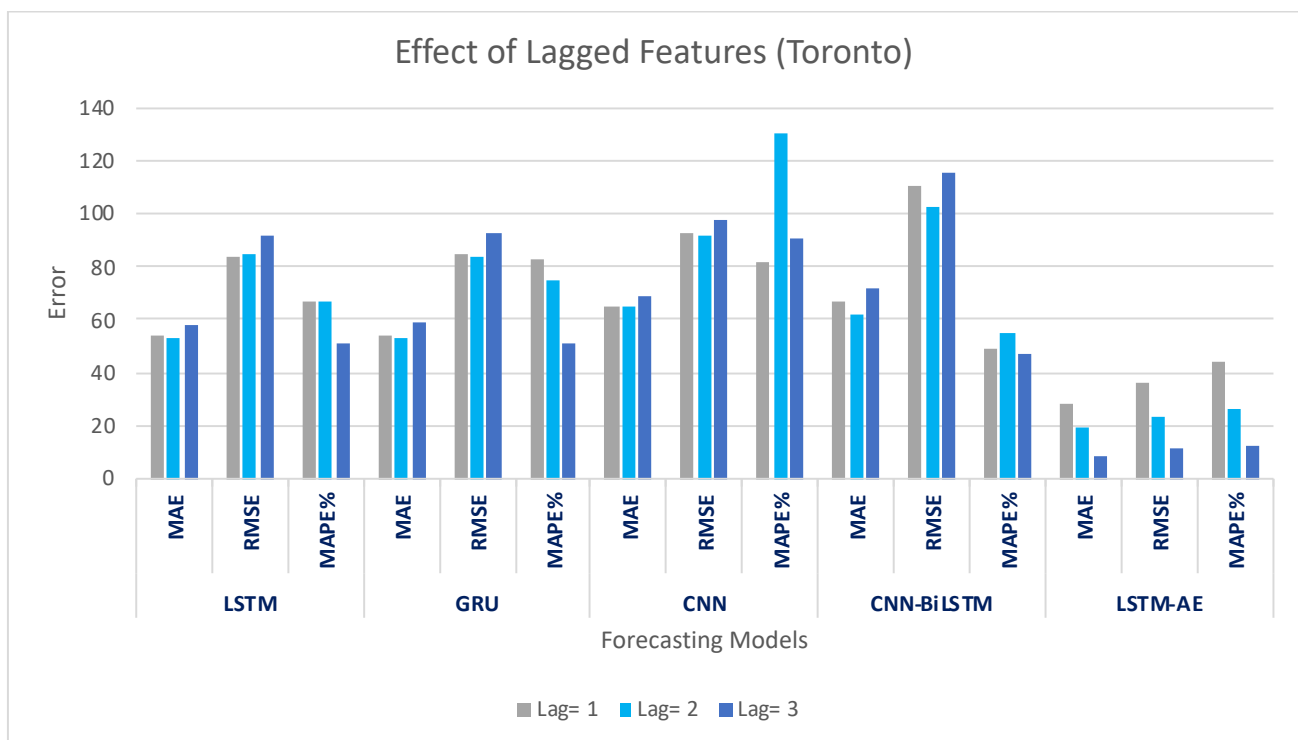


Figure 22. The effect of the lagged features on Toronto dataset.

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 the 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 whereas the worst values are associated with Caracas and Toronto, except for the LSTM-AE model. The high performance related to the 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 the Caracas and Toronto datasets might be attributed to the high percentage of cloudy hours (or unclear sky condition) compared to the Saudi locations. The best model according to MAE and nMAE values is the LSTM-AE model, which achieves nMAE equal to 0.02 with the 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 the CNN model with the Saudi datasets while the CNN-BiLSTM model is the worst for Caracas and Toronto. With time-series data, the temporal features are the most important ones, and these cannot be captured by the CNN model.

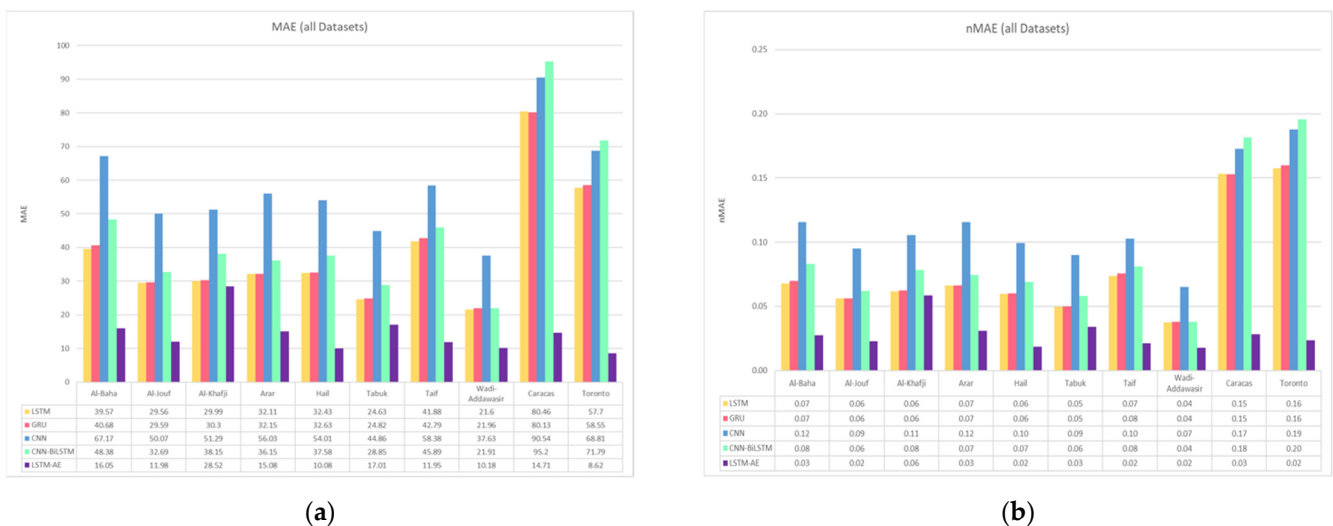


Figure 23. Forecasting results of 5 models for all datasets (a) MAE; (b) nMAE.

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 the 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 delivered the worst nRMSE value equal to 0.08 with the Al-Khafji dataset. We mentioned earlier the advantage of the Wadi-Addwasir dataset compared to other Saudi datasets and the disadvantage of Caracas and Toronto. Regarding the Al-Khafji dataset, it has the equivalent of a year’s missing data, which might explain the low performance of the LSTM-AE model here. However, the LSTM-AE model is the best model for all locations, whereas CNN is the worst with the Saudi datasets and the CNN-BiLSTM model is the worst with the Caracas and Toronto datasets. 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 for its superior performance.

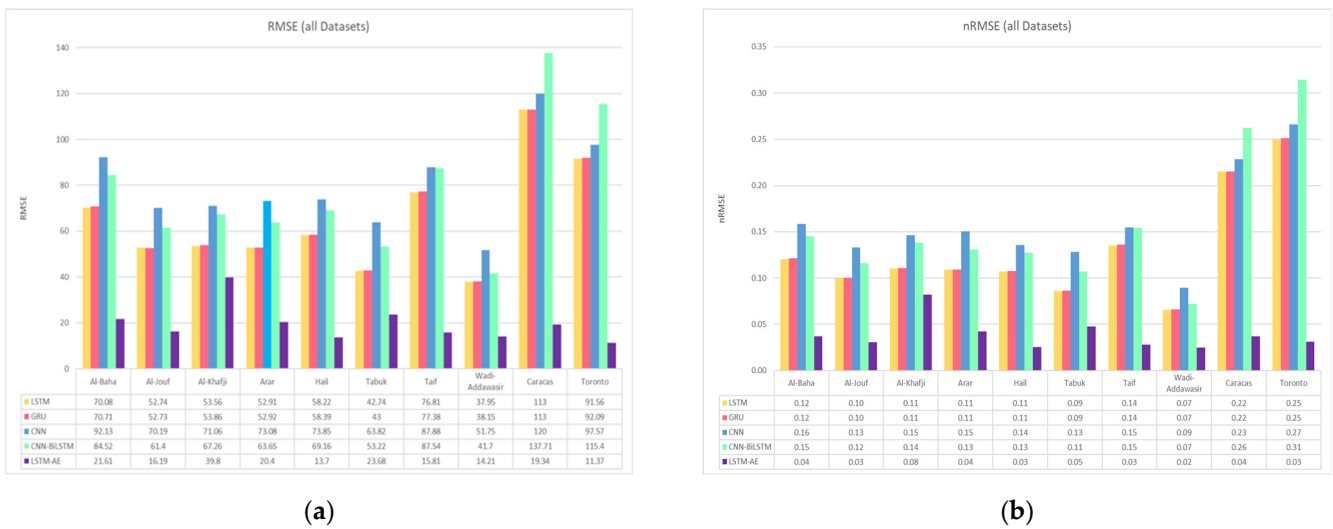


Figure 24. Forecasting results of 5 models for all datasets (a) RMSE; (b) nRMSE.

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

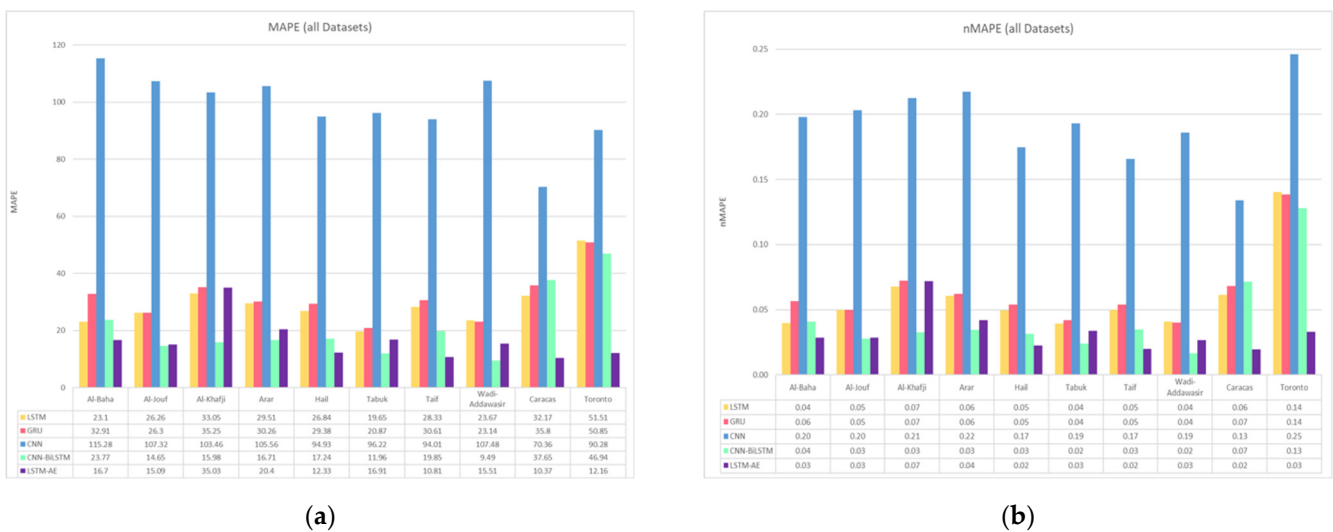


Figure 25. Forecasting results of 5 models for all datasets (a) MAPE; (b) nMAPE.

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.

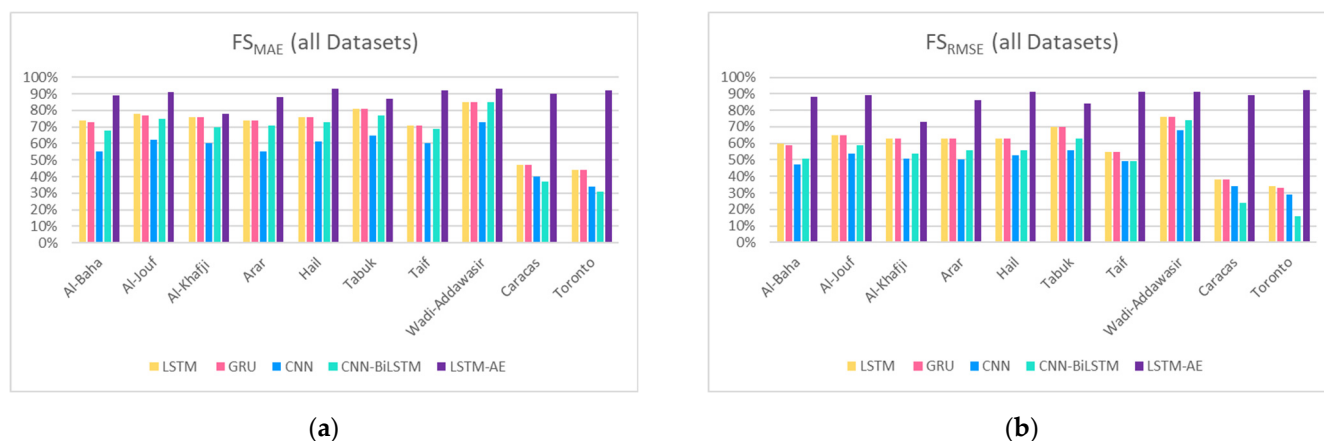


Figure 26. Forecasting results of 5 models for all datasets (a) FS_{MAE}; (b) FS_{RMSE}.

In summary, looking at the performance from the models' perspective (refer to Figures 23–25), 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 delivered the worst results. With time-series data, the temporal features are the most important features, and these cannot be captured by the CNN model. However, the CNN-BiLSTM model is the worst model for Caracas and Toronto according to the 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-Addawasir) and both models achieved the same value with Al-Jouf.

Looking at the performance from the location perspective (refer to Figures 23–25), we can see that the best nMAE, nRMSE, and nMAPE results for all models are mostly associated with the Wadi-Addawasir dataset. On the other hand, the worst results are linked with the Toronto and Caracas datasets. As mentioned earlier, the high performance related to the Wadi-Addawasir 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 the 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 the Toronto and Caracas datasets might be attributed to the high percentage of cloudy hours (or unclear sky condition) compared to the other Saudi locations (see Figure 9). This in turn means that GHI varies from time to time and is hard to predict. We can infer that the most important factor that affects the performance of the different models is the climate of 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 the performance of the different models, in Figure 27 we plot 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 hours 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's cloudy days, it is able to capture the trend line closely. In contrast, the CNN-BiLSTM model sometimes achieved a closer prediction than the LSTM-AE model, but it could not capture the trend line accurately as the LSTM-AE model as shown in the case of Toronto's cloudy days. On the other hand, the CNN model delivered the worst prediction, especially on cloudy days.

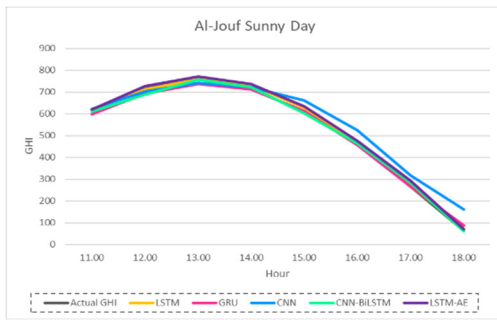
4.1.4. Effect of Summer and Winter Seasons on Forecasting

To examine the effect of seasons on the performance of the different models, we first show, in Figure 28, the actual vs. predicted GHI of the coldest and hottest months (January and August) for all models in five locations: Al-Jouf, Al-Khafji, Wadi-Addawasir, Caracas, and Toronto. Just as in the sunny and cloudy results, we found that the LSTM-AE model is the most accurate in January and August whereas the CNN model is the least accurate. It can be seen 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 was selected for no specific reason, we could have plotted using RMSE and other metrics, or all the metrics considered in this paper. However, we plotted with 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, and winter includes the remaining months. From Figure 29, we can see that the winter MAE is higher than the summer in all the datasets, except for Taif, Caracas, and Toronto, where MAE is higher in summer. Another observation is that the CNN model and the CNN-BiLSTM model have the largest difference in MAE from summer to winter whereas the LSTM-AE model has a very slight difference.

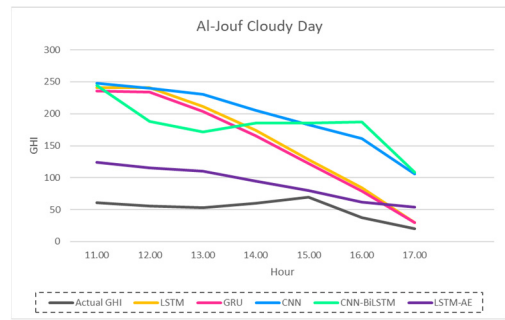
4.1.5. Digging Deeper into Forecasting Error for Each GHI Prediction

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 the 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 number of outliers, while dataset-wise, the Toronto and Taif datasets have the highest number of outliers.

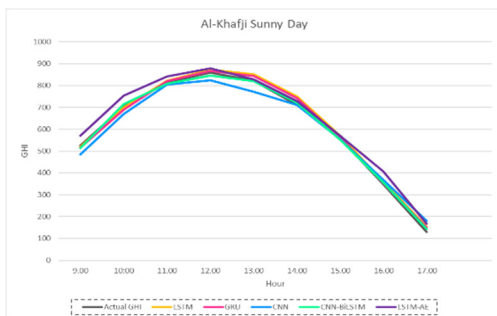
The forecasting error is used to determine the "best model" label of each record in the testing datasets of all locations. The best model is one that achieves the least forecasting error for each record. Figure 31 shows the achieved percentage of all the five models as the "best model" based on the forecasting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



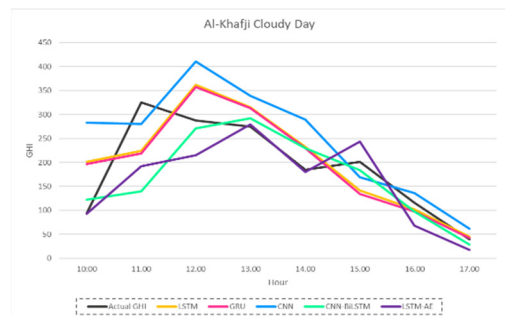
(a)



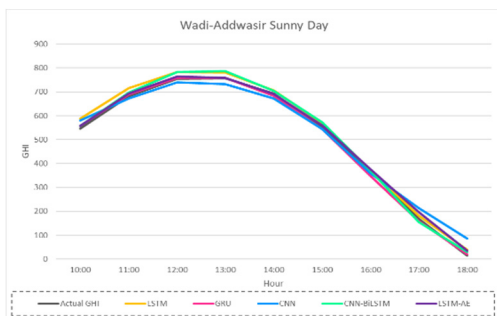
(b)



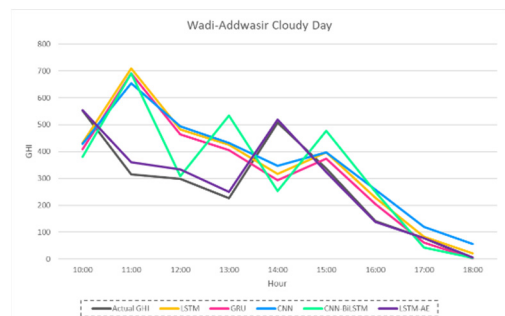
(c)



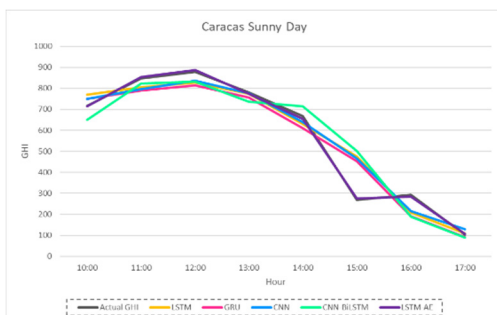
(d)



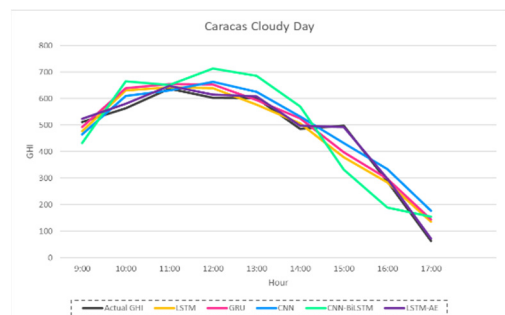
(e)



(f)

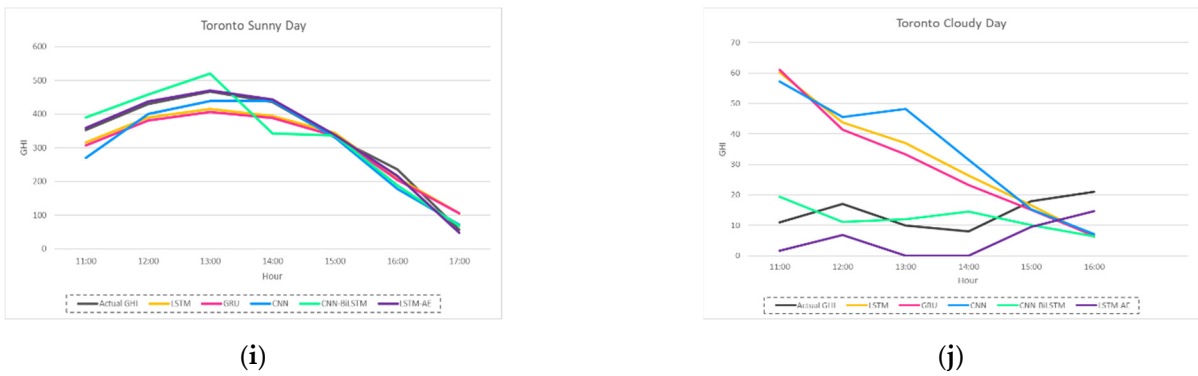


(g)



(h)

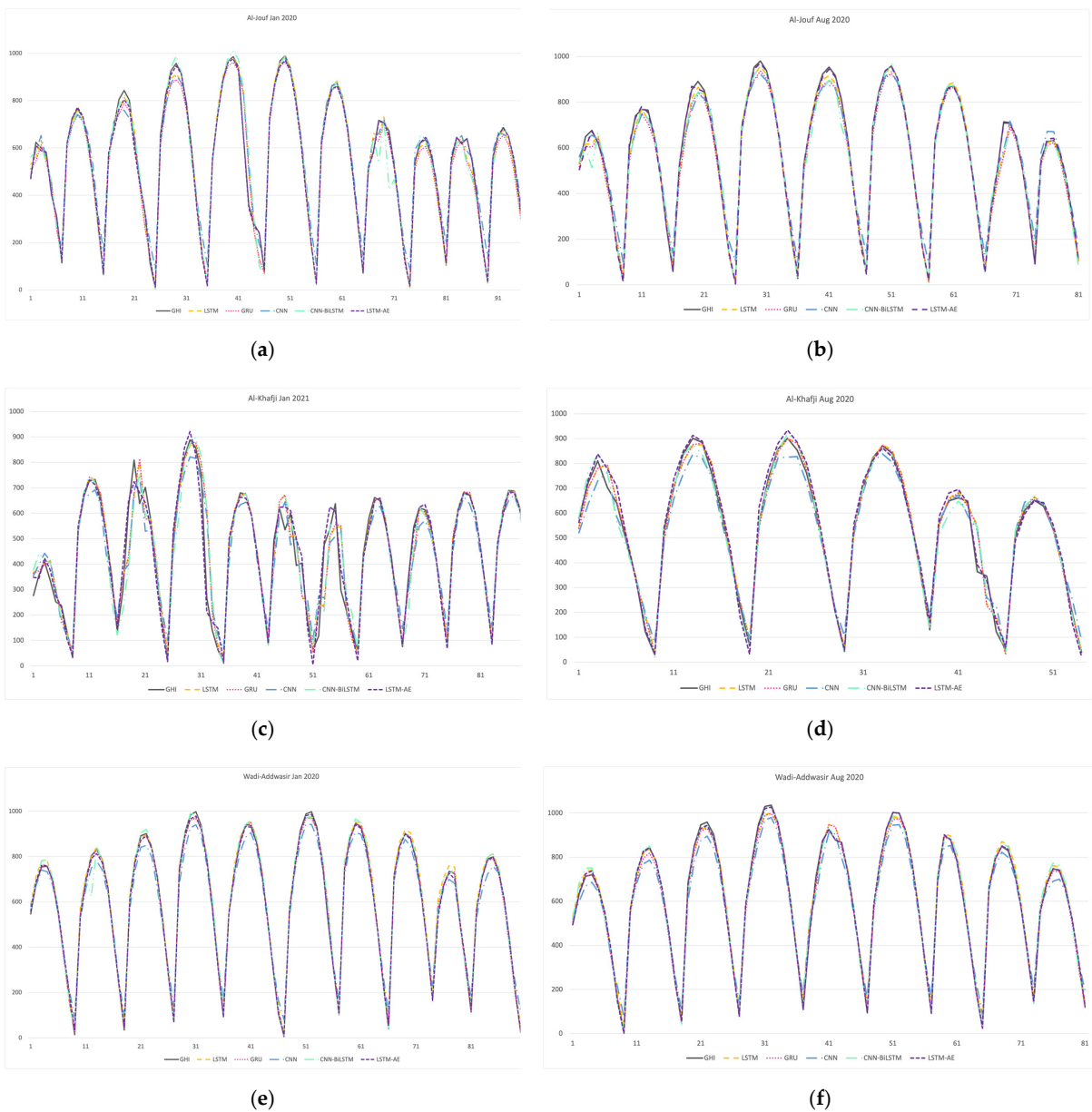
Figure 27. Cont.



(i)

(j)

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.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 28. Cont.

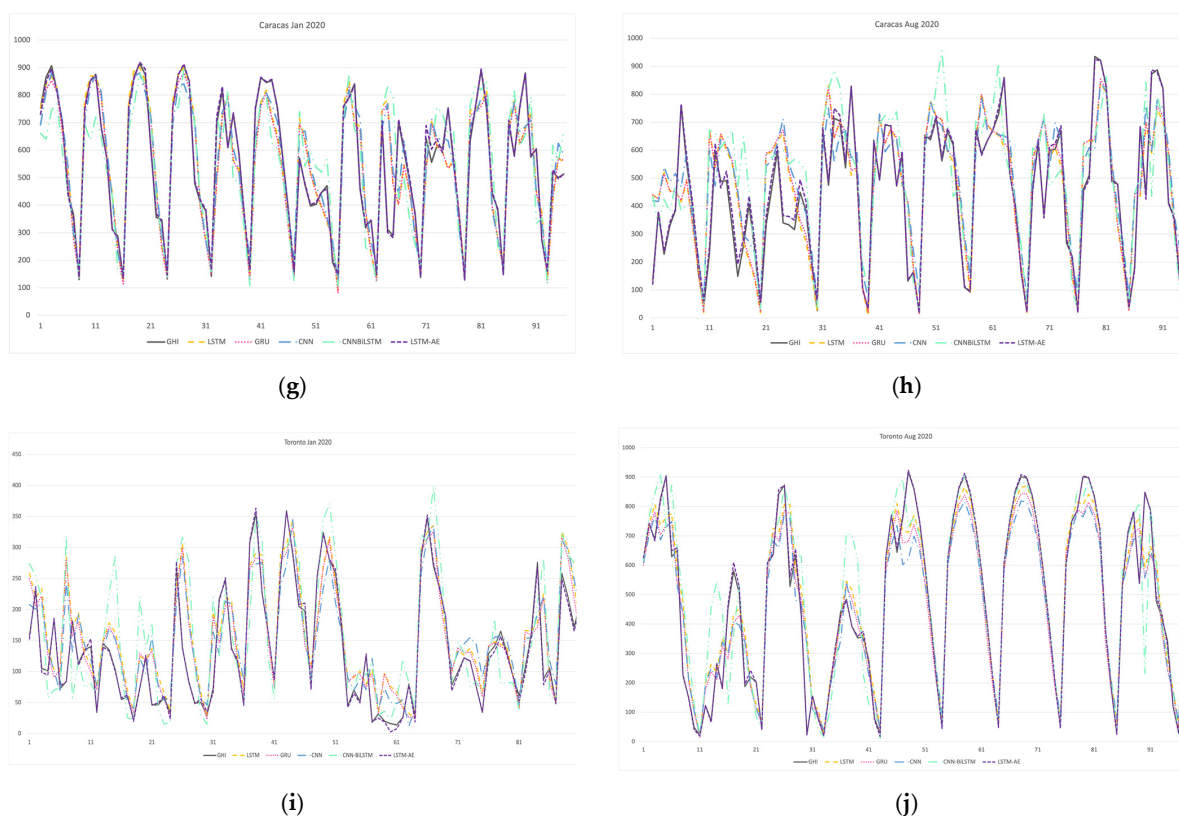


Figure 28. Summer vs. winter—actual vs. predicted GHI of 5 models for: (a) Al-Jouf Jan; (b) Al-Jouf Aug; (c) Al-Khafji Jan; (d) Al-Khafji Aug; (e) Wadi-Addwasir Jan; (f) Wadi-Addwasir Aug; (g) Caracas Jan; (h) Caracas Aug; (i) Toronto Jan; (j) Toronto Aug.

4.2. SENERGY: Auto-Selective Model Prediction Engine Performance

In Section 4.1, we compared the performances of the five forecasters 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 the CNN-BiLSTM model 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 the LSTM-AE model is the best model for 54% of the total records while the CNN-BiLSTM model is the best for 17%. The remaining models CNN, GRU, and LSTM together achieved only 29% (see Figure 31). This imbalance in the data that comes from the variation in the performance of the forecasting models affects the classifier training negatively. Considering both the overall performance of the forecasters represented in the MAPE metric and the item-wise performance represented in the forecasting error, we decided to use only two models LSTM-AE and CNN-BiLSTM in the 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. In the future, we will incorporate in the tool additional models for GHI forecasting. A description of the auto-selective model prediction engine's structure is given in Section 3.4.1. Figure 32 shows the confusion matrix of the engine. As shown by the matrix, correctly classified CNN-BiLSTM records account for 8.4% of the total records, while correctly classified LSTM-AE records account for 72.57% of the total records.

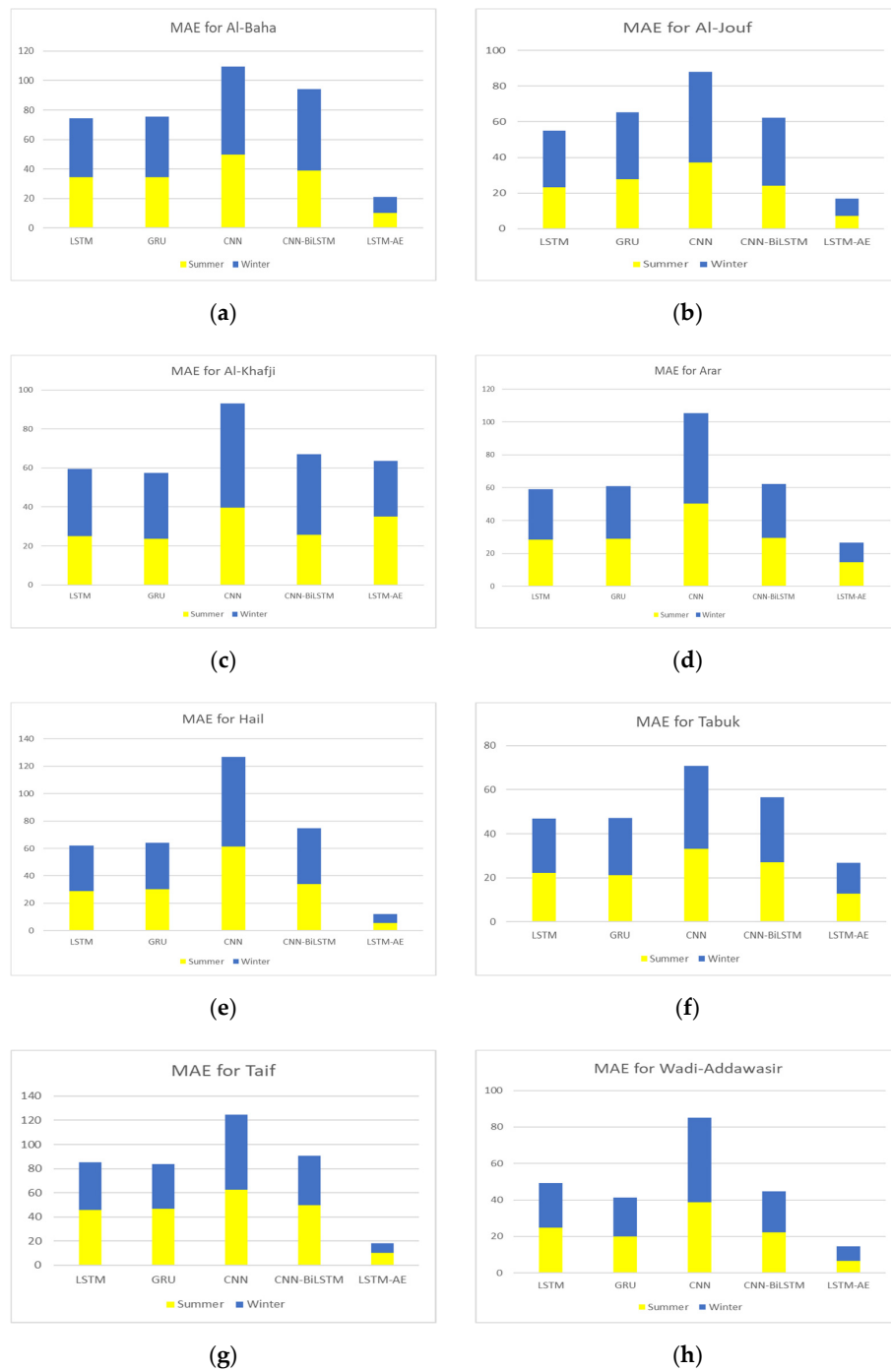


Figure 29. Cont.

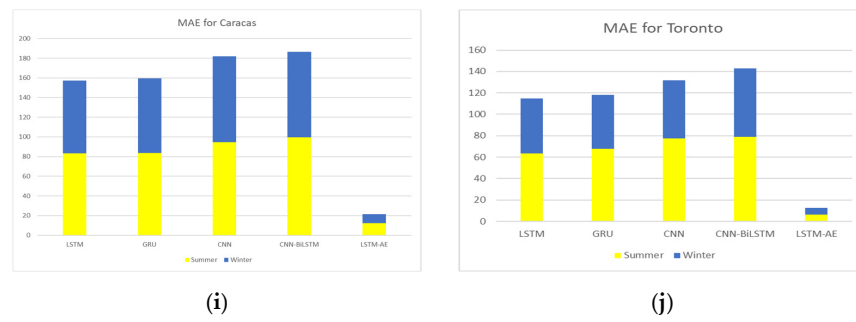


Figure 29. MAE of summer vs. winter for 5 models for: (a) Al-Baha; (b) Al-Jouf; (c) Al-Khafji; (d) Arar; (e) Hail; (f) Tabuk; (g) Taif; (h) Wadi-Addwasir; (i) Caracas; (j) Toronto.

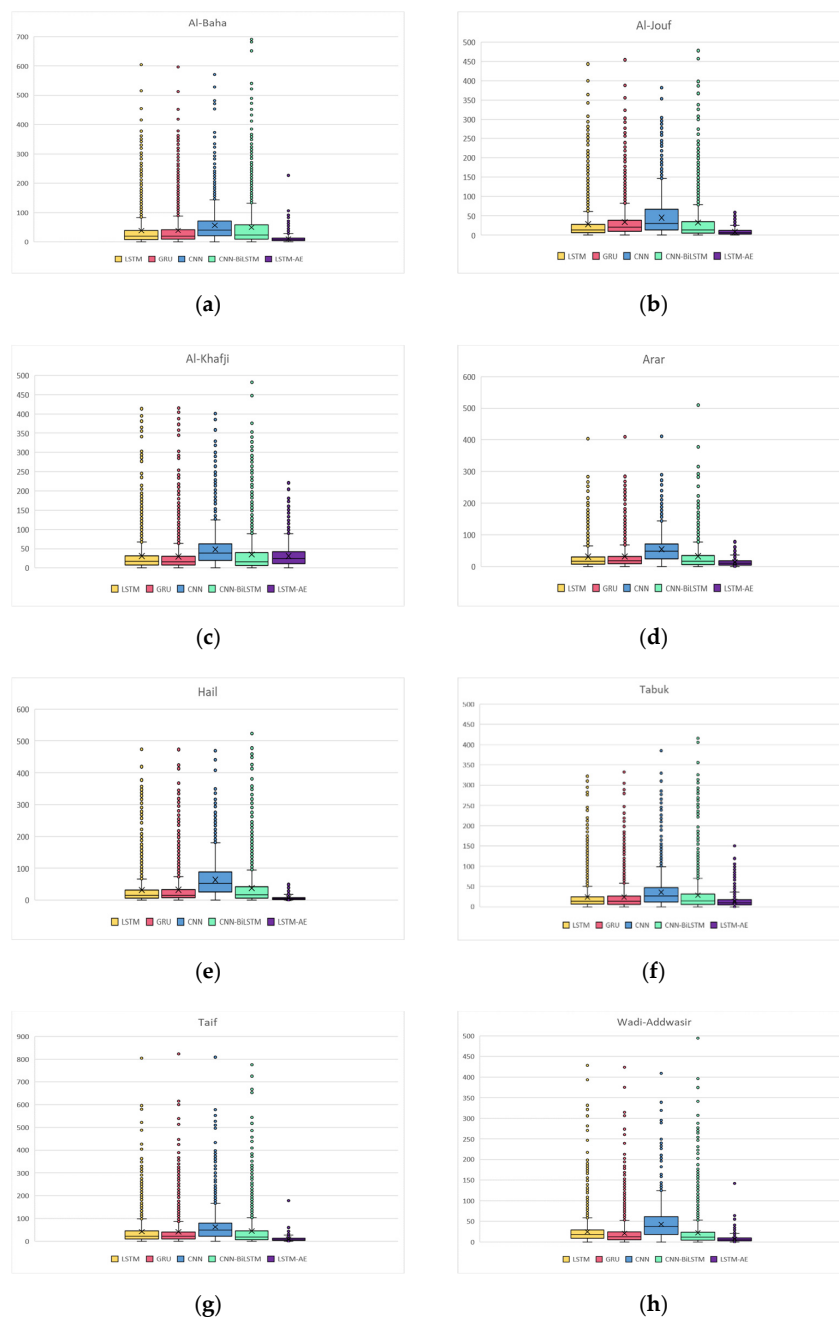


Figure 30. Cont.

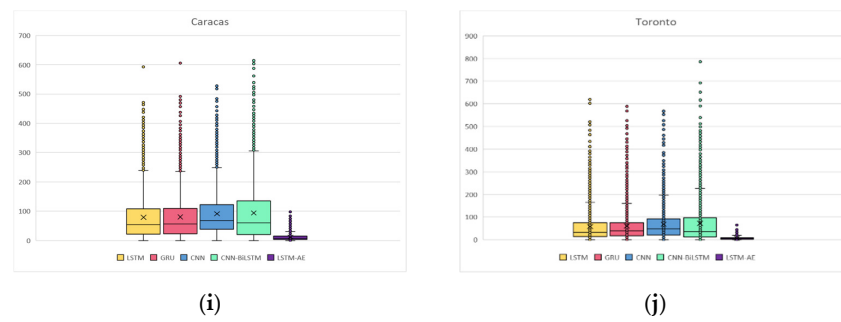


Figure 30. Boxplot of GHI forecasting error of 5 models for 5 models for: (a) Al-Baha; (b) Al-Jouf; (c) Al-Khafji; (d) Arar; (e) Hail; (f) Tabuk; (g) Taif; (h) Wadi-Addwasir; (i) Caracas; (j) Toronto.

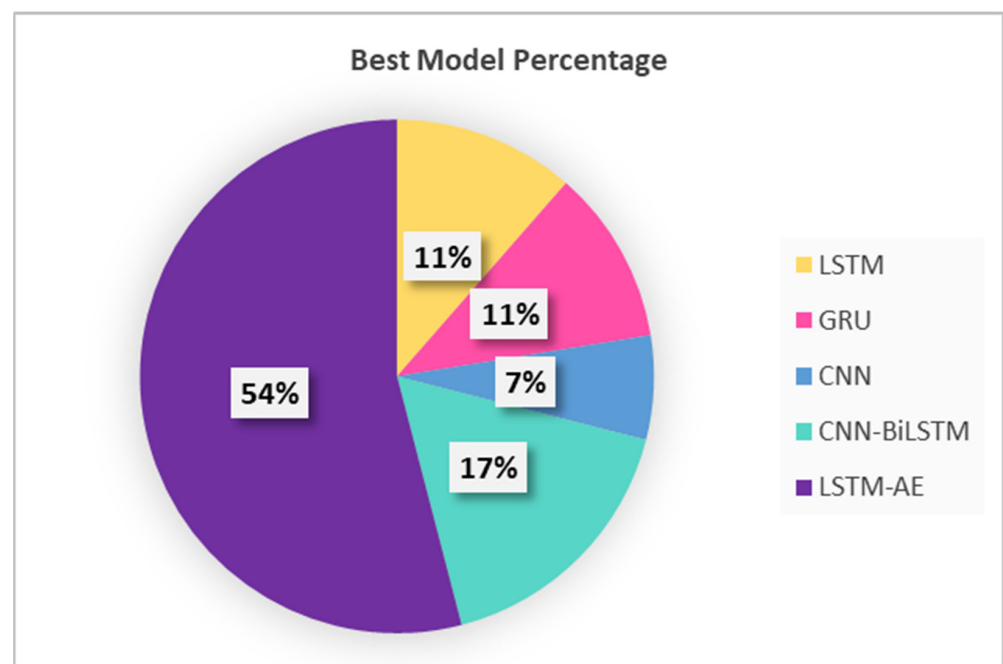


Figure 31. The achieved percentage of the models as “best model” based on the forecasting error.

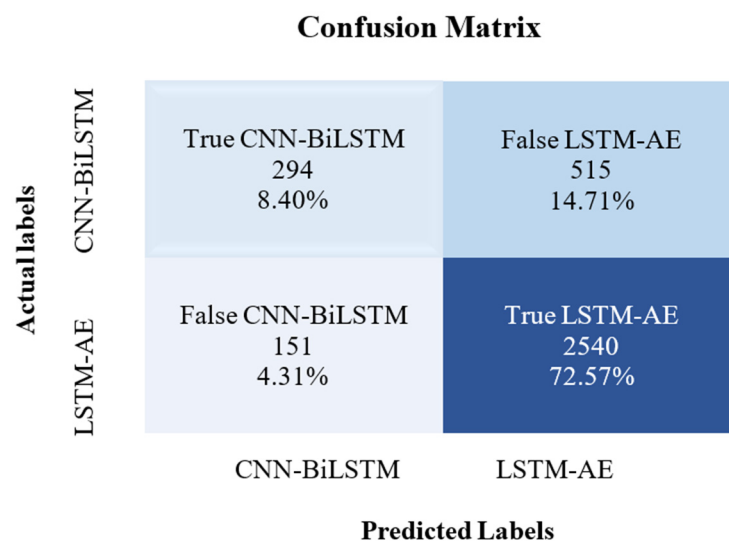


Figure 32. Auto-Selective Model Prediction Engine confusion matrix.

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 the CNN-BiLSTM and LSTM-AE models separately as well as the classification accuracy of the engine. The total number of records used for testing the engine is 3500 as shown in the Support column. Of these records, the CNN-BiLSTM model accounts for 23% (809/3500), and the LSTM-AE model accounts for 77% (2691/3500). The percentage of correctly classified records of each model is shown in the recall column. The CNN-BiLSTM model’s recall is 36% while the LSTM-AE model’s 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.

Model	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 the DHI value with a lag of one. The least important features are 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).

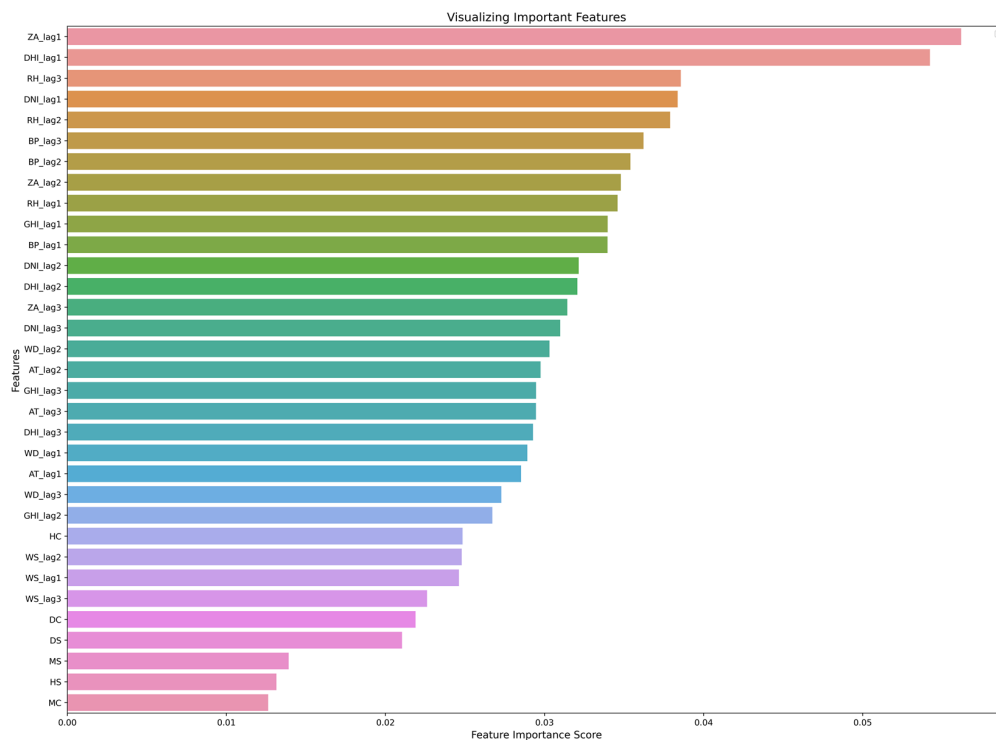


Figure 33. Feature importance using random forest classifier.

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 (Section 4.2.1), sunny and cloudy weathers (Section 4.2.2), and summer and winter seasons (Section 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%, and this is associated with Caracas and Toronto due to a large number of records for both locations. The lowest classification accuracy is 69%, associated with Tabuk, due to a low number of records and the close forecasting performance of both forecasting models for this location (see Section 4.1.2).

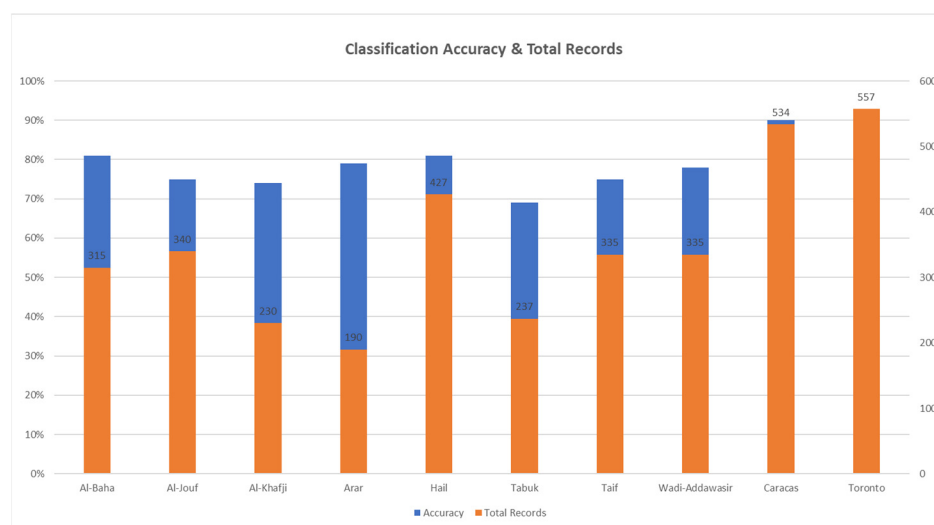


Figure 34. Classification accuracy of model prediction engine based on location with total records.

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

4.2.2. Model Prediction: Sunny and Cloudy Weathers

The classification accuracy of the model prediction engine is 75% for sunny days and 86% for cloudy days. The total number of records for sunny days is higher than for cloudy days by 28%. These results contradict forecasting results in which forecasting on sunny days is more accurate than on cloudy days. The reason for this is the close prediction for both the CNN-BiLSTM and LSTM-AE forecasting models in sunny weather, which makes it difficult for the classifier to pick one model. On the other hand, the LSTM-AE model shows superior forecasting performance on cloudy days, which makes it easy for the classifier to pick the best model on cloudy days. Figure 36 shows the recall for each model in sunny versus cloudy weather. Notably, the CNN-BiLSTM model's recall in cloudy weather is better than in sunny weather with 38%, whereas the LSTM-AE model's recall in sunny weather is 2% better than in cloudy weather. As explained earlier, on sunny days

the CNN-BiLSTM and LSTM-AE models make very similar predictions that causes the classifier to misclassify CNN-BiLSTM records as LSTM-AE.

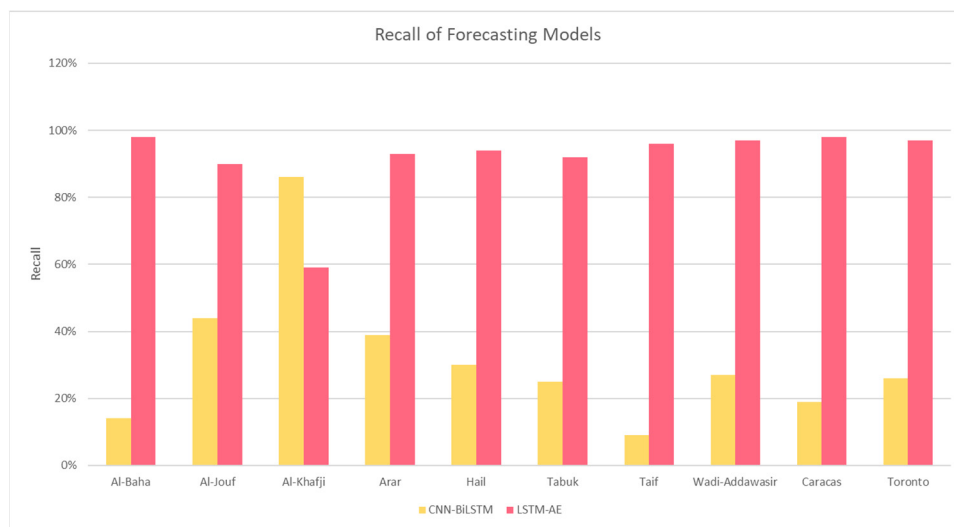


Figure 35. Recall of the two forecasting models in model prediction engine.

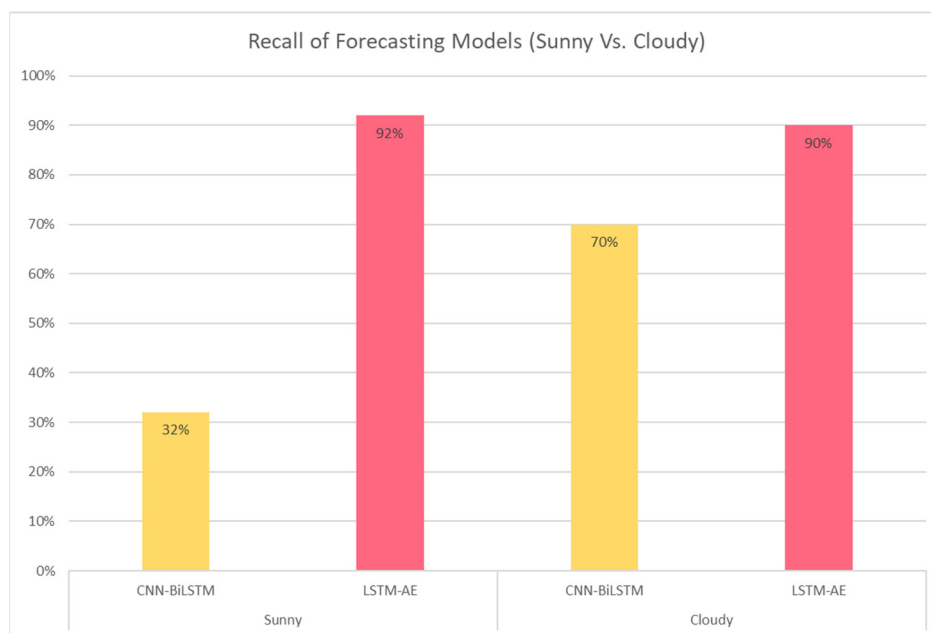


Figure 36. Recall of the two forecasting models (sunny vs. cloudy) in model prediction engine.

4.2.3. Model Prediction: Summer and Winter Seasons

The classification accuracy of the model prediction engine is 82% in summer and 80% in winter, even though the total number of records for summer is less than for 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 the CNN-BiLSTM model’s recall in summer is better than in winter, with a 7% difference while the LSTM-AE model’s recall is almost the same in both seasons.

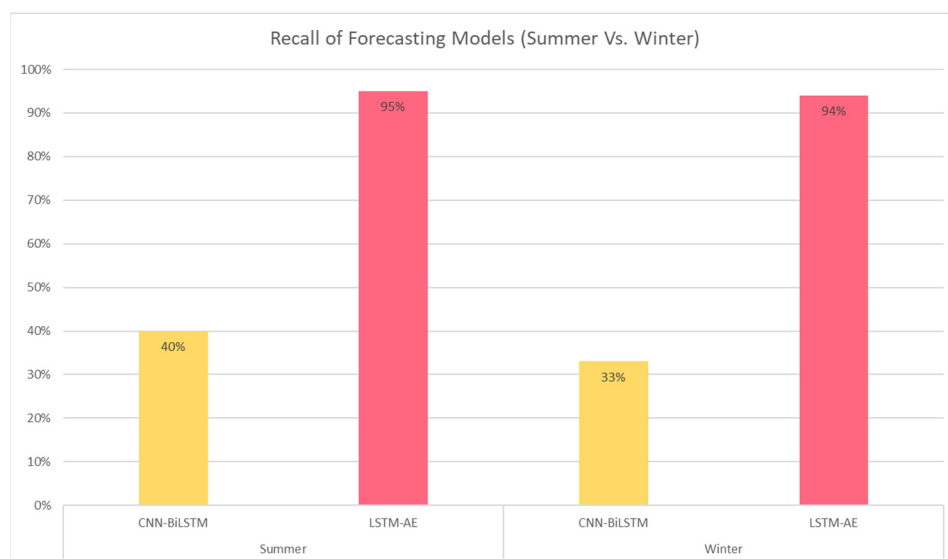


Figure 37. Recall of the two forecasting models (summer vs. winter) in model prediction engine.

4.3. SENERGY: Performance Gain and Loss

4.3.1. Actual Gains and Losses

To understand the benefits 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.

$$G \text{ or } L = FE_m - FE_t \tag{11}$$

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 the CNN-BiLSTM is 67.73 and that of the LSTM-AE model is 4.83. The tool was able to correctly choose the best model for this record. Thus, the achieved forecasting error is 4.83. To measure the gain over the 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 a gain in performance equal to 62.90 over the CNN-BiLSTM model for this record. On the other hand, no gain was achieved for the tool over the LSTM-AE model, because it is the best regardless. Similarly, in the second record, the forecasting error of the CNN-BiLSTM model is 128.60 and that of the LSTM-AE model 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 the CNN-BiLSTM model in this case. The difference between the forecasting error for the LSTM-AE model and the forecasting error for the wrong best model chosen by the tool is 84.19. It is a negative number, and thus it represents a loss in tool performance for the 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
0.47	29.12	29.12	−28.65	0

Figure 38 shows the gain or loss of SENERGY versus the CNN-BiLSTM model for each record. Gains are positive and thus above the zero line, and losses are negative and

thus below the zero line. Locations' records are differentiated by colors. As noted, the gain is large in general because the LSTM-AE model provides far better forecasting than the CNN-BiLSTM model. In contrast, the loss is small because when the CNN-BiLSTM model achieves better forecasting than the LSTM-AE model, the difference is small. Looking at gain or loss from a location perspective, the largest gains are achieved with the Caracas and Toronto datasets, whereas the smallest gains are achieved with the Wadi-Addawasir dataset, which is compatible with the forecasting results discussed in Section 4.1.2.

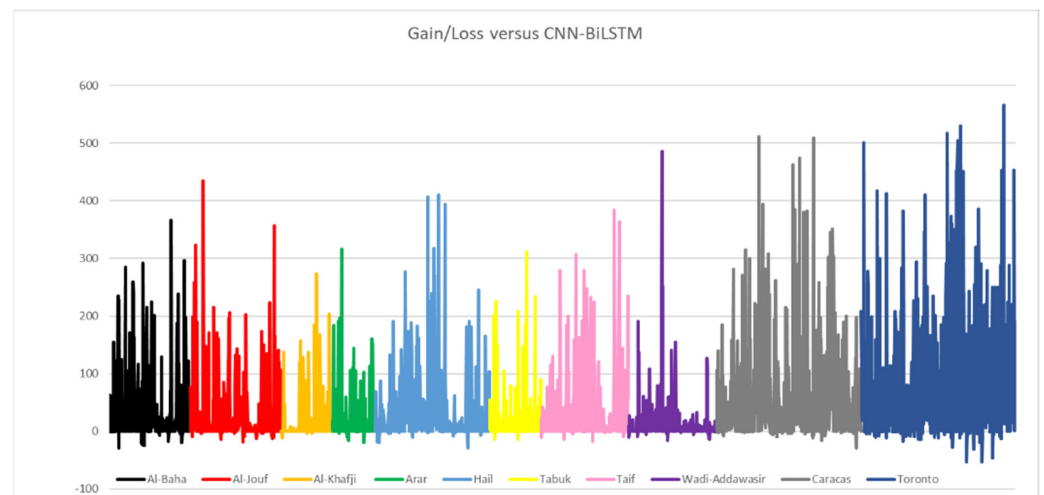


Figure 38. Gain/loss of SENERGY versus CNN-BiLSTM.

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

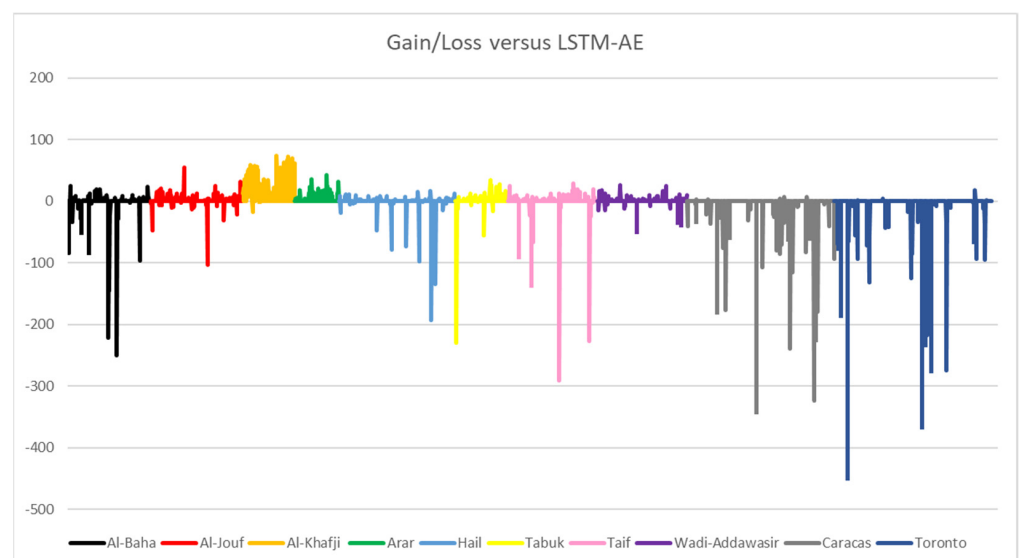


Figure 39. Gain or loss of SENERGY versus LSTM-AE.

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), 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 extended to hyperparameter optimizations 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 this tool is that it would provide higher gains for higher diversity datasets while usually, the opposite is true for a single forecasting method. Additionally, 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 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 two.

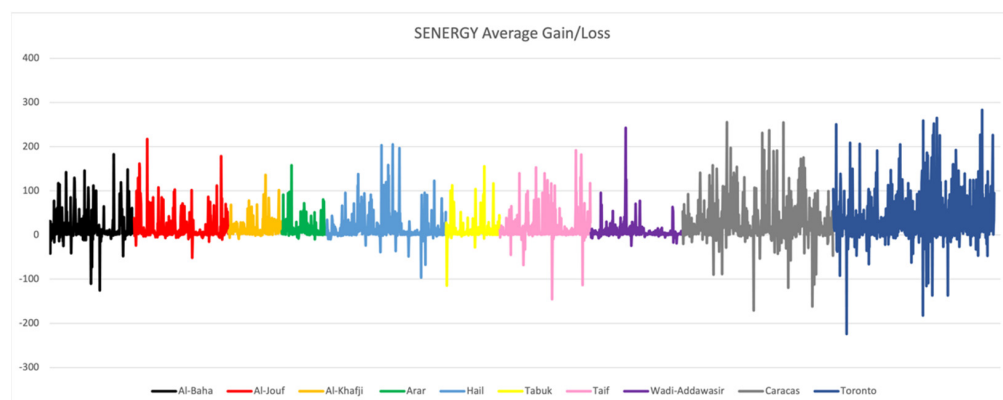


Figure 40. Average gain/loss of SENERGY.

4.3.2. Potential Performance

In the previous section, we demonstrated the gain and loss of SENERGY, which can choose the best forecasting model between the two models only. However, in an ideal situation, SENERGY would choose the best forecaster from 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 the 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 the Caracas and Toronto datasets, whereas the lowest gains occur with Al-Khafji and Wadi-Addawsir.

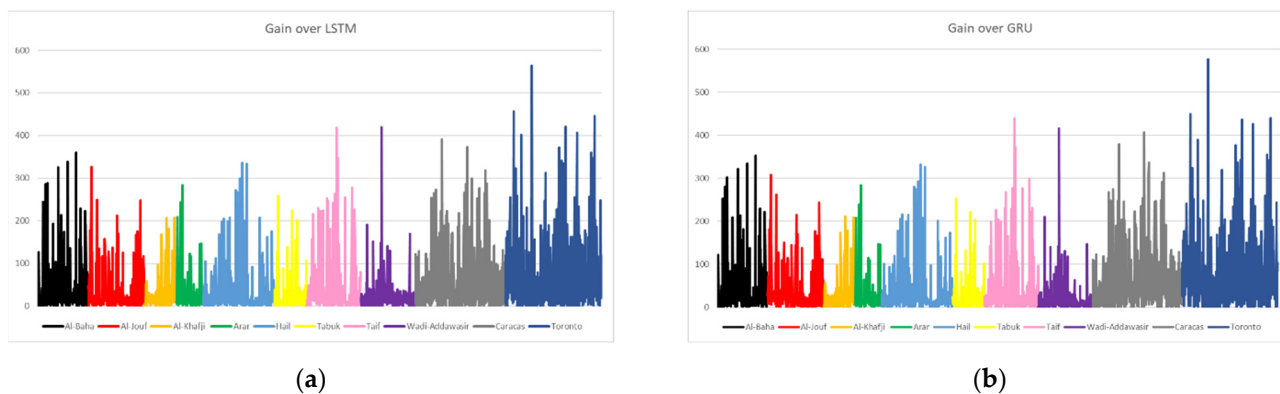


Figure 41. Gain of SENERGY over: (a) LSTM; (b) GRU.

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

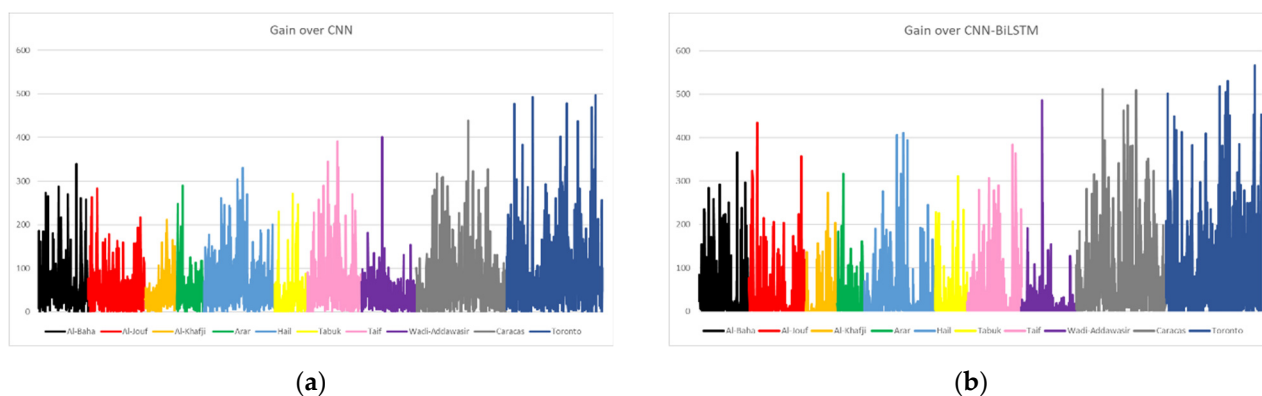


Figure 42. Gain of SENERGY over: (a) CNN; (b) CNN-BiLSTM.

Figure 43a shows the gain of SENERGY over the 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, as with other models. Figure 43b shows the gain of SENERGY over all the five models as a boxplot. It is obvious that the largest gain is achieved with the CNN and CNN-BiLSTM models. Additionally, gain over the LSTM and GRU models is similar, whereas gain over the LSTM-AE model is very small since it is the best forecaster for most of the records.

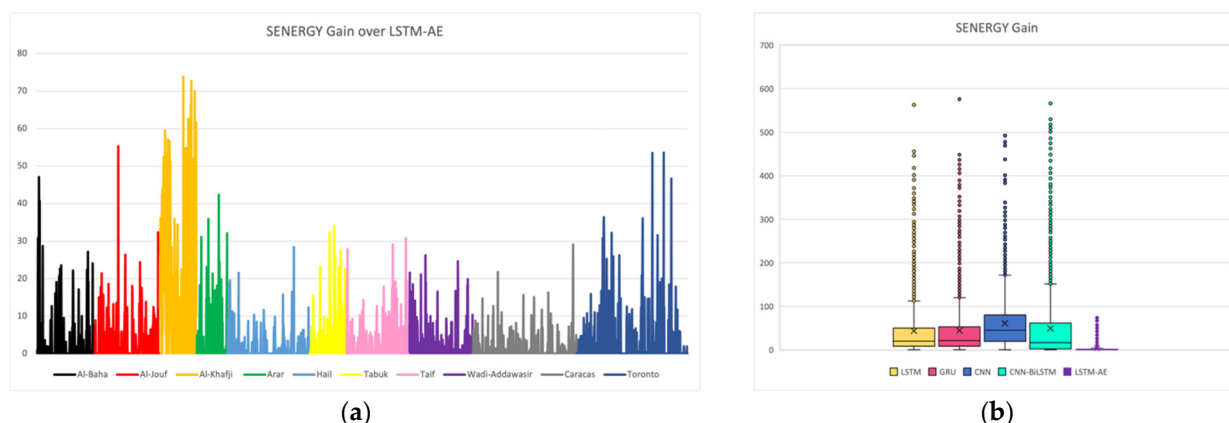


Figure 43. Gain of SENERGY: (a) over LSTM-AE; (b) as a boxplot for the five base models.

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. Secondly, they use multiple datasets from different climates. For the reasons explained, the works such as [26,42–46,77,78] are excluded from the comparison because the datasets used are for one climate only. Works such as [23,29,47,48,72,79–83] are also excluded from the comparison because they propose forecasting models for different time horizons, such as day-ahead or monthly GHI, whereas is the focus of our work in this paper is next-hour GHI.

The results 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 a comparison of SENERGY to six works that 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 the comparison is the model used for forecasting. For example, in reference [21], data from three locations in India, representing three different climate classes (Cwa, Cwb, Bsh) are used. 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 No.	Location	GHI Mean	GHI SD	Climate	Weather Data	Model
[21]	Jaipur New Delhi Gangtok	NA	NA	<ul style="list-style-type: none"> • Cwa • Cwb • Bsh 	✓	Ensemble model of XGBF-DNN
[24]	Los Angeles Denver Hawaii’s Big Island Tamanrasset	217.37 203.33 220.12 269.98	291.73 276.40 307.79 361.83	<ul style="list-style-type: none"> • Csb, • BSk • Af • BWk 	✗	Hybrid model of CEEMDAN-CNN-LSTM

Table 12. Cont.

Ref No.	Location	GHI Mean	GHI SD	Climate	Weather Data	Model
[25]	Ajaccio Tilos Odeillo	NA	NA	<ul style="list-style-type: none"> • Csa • Csb 	✗	ARMA RF
[27]	CA TX WA FL PA MN	NA	NA	<ul style="list-style-type: none"> • BSk • Cfa • Cfb • Am • Dfb • Dfa 	✓	Generalized random forest
[28]	Bondville Desert Rock Fort Peck Goodwin Creek Penn. State Uni Sioux Falls Table Mountain	398.04 517.72 368.17 442.77 384.31 406.94 412.19	284.66 314.73 277.33 289 277.24 277.55 287.97	<ul style="list-style-type: none"> • Cfa • BWk • BSk • Cfb • Dfa 	✗	68 machine learning algorithms (Cubist model is the best in most cases)
[30]	Tucson Bermuda Brasilia Sonnblick Solar Village Golden Darwin Ny-Alesund Toravere Lerwick	532.5 417.1 475.6 347.2 580.9 459.4 516.4 184.3 256.9 198.3	NA	<ul style="list-style-type: none"> • Bsh • Cfa • A • Dfc 	✗	ELM
This work	Al-Baha Al-Jouf Al-Khafji Arar Hail Tabuk Taif Wadi-Addawasir Caracas Toronto	582.09 528.14 486.59 485.46 543.77 498.03 567.62 578.02 366.77 524.82	311.16 296.47 275.73 295.04 303.82 261.73 308.47 301.69 271.11 297.12	<ul style="list-style-type: none"> • BWh • A • Dfb 	✓	LSTM GRU CNN CNN-BiLSTM LSTM-AE

Ref [30] has 20 locations, we present data from 10 locations from various climates for simplicity.
NA: Not available.

A fair comparison of the performance of the models in the literature is a challenging task because there is great variation in the results reported by researchers. Additionally, it is difficult to find the best-performing model by comparing various statistical measures, such as RMSE, MAE, and MAPE, etc., at the same time. For example, to compare six works included in this section, we need four figures. 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. Hence, 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 boxplot 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 that shows a small error and a short box that shows a small variation between the 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 have reported the results for the ideal situation.

In Figure 44a, MAE results from Gao et al. [24] and Fouilloy et al. [25] are compared to MAE results of five forecasting models and SENERGY in this work. It is noted that reference [25] has the worst MAE results in terms of high value and large variation between the three locations. On the other hand, the 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. [24] appears to show the next-best performance; however, this is because the results are for four locations only. In Figure 44b, the nMAE results are compared. Reference [25] is excluded because nMAE is not reported there. We see how normalization made the box of Gao et al. [24] bigger. Thus, it is fair to say that the LSTM, GRU, LSTM-AE models, and SENERGY show better performance, even with a larger number of locations. For both MAE and nMAE, the 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).

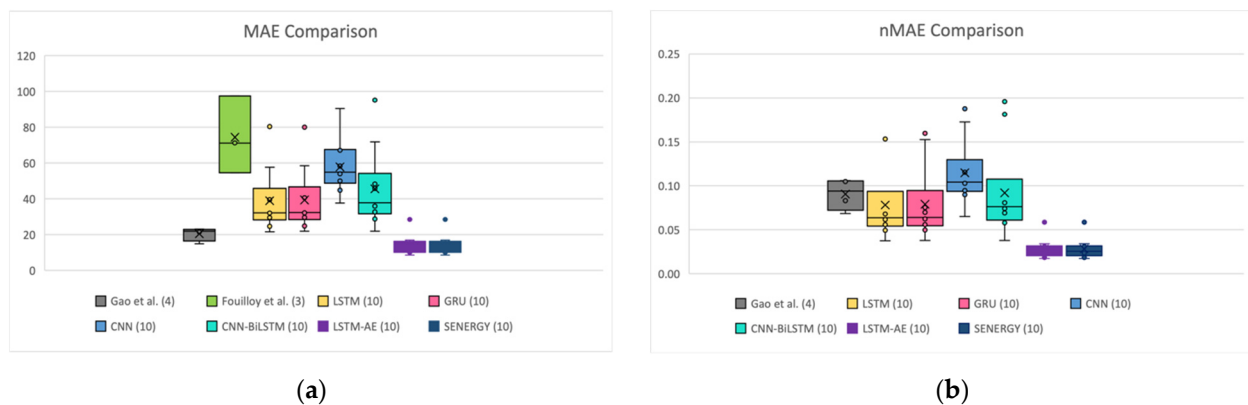


Figure 44. Comparison of multiple works based on: (a) MAE; (b) nMAE.

In Figure 45a, the RMSE results of four works (Kumari and Toshniwal [21], Gao et al. [24], Lee et al. [27], and Bouzguo and Gueymard [30]) are compared with the RMSE results of the five forecasting models and the SENERGY results in this work. The LSTM-AE model and SENERGY achieved the best performance in terms of the lowest RMSE and smallest variation among ten locations. The work of Gao et al. [24] comes next; 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 Bouzguo and Gueymard [30], while the worst based on variation among locations is associated with the work of Lee et al. [27] with six locations. In Figure 45b, the nRMSE results of this work are compared to four other works. Since nRMSE is not reported in references [21] and [27], they are excluded in (b) and another two works are added: Fouilloy et al. [25] and Yagli et al. [28]. The best nRMSE results are achieved by the LSTM-AE model and SENERGY, whereas the worst are related to Yagli et al. [28] in terms of low value and Fouilloy et al. [25] 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. [24], the box is smaller and lower than the LSTM, GRU, CNN, and CNN-BiLSTM models, but after normalization, it becomes higher than all of them. In both (a) and (b), the LSTM-AE model and SENERGY achieved the best performance in terms of lowest value and smallest variation. Again, SENERGY shows the same performance as the LSTM-AE model, because according to the RMSE and nRMSE results, the best model is always LSTM-AE for all locations (refer to Figure 24).

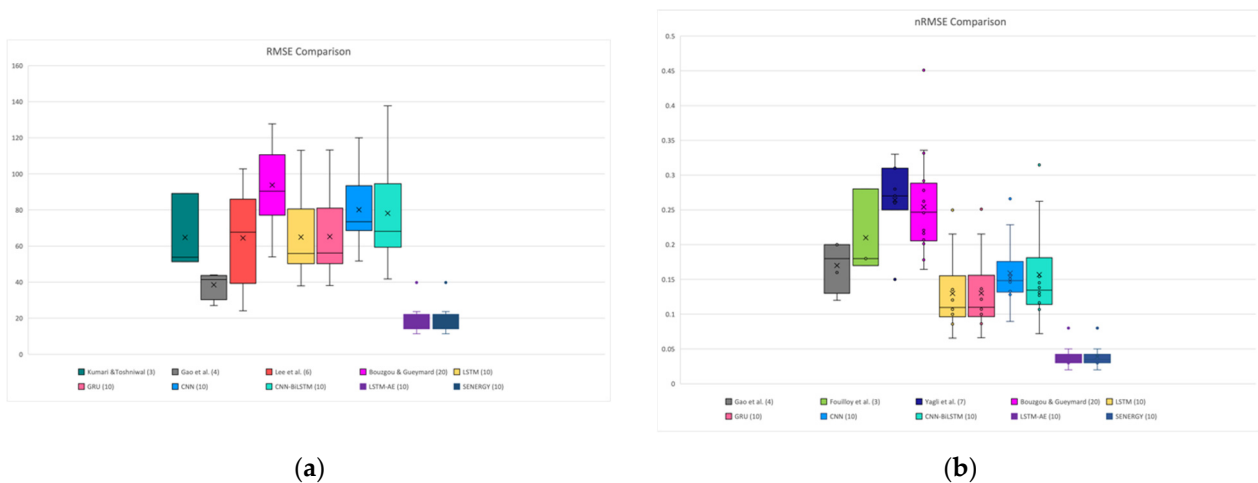


Figure 45. Comparison of multiple works based on: (a) RMSE; (b) nRMSE.

In Figure 46a, the MAPE results of Lee et al. [27], and Bouzguo and Gueymard [30] are compared to the five forecasting models and SENERGY results in this work. SENERGY achieves the lowest error with the smallest variation between ten locations, whereas the CNN model is the worst. In Figure 46b, the comparison is based on the nMAPE results and the same observation about the best and worst performance is true. The work of Lee et al. [27] 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 Bouzguo and Gueymard [30]. In (a), it shows better performance than the LSTM, GRU, and CNN-BiLSTM models, whereas in (b) it becomes worse than all of them in value or variation among locations. Unlike the MAE and RMSE results, SENERGY outperforms the 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).

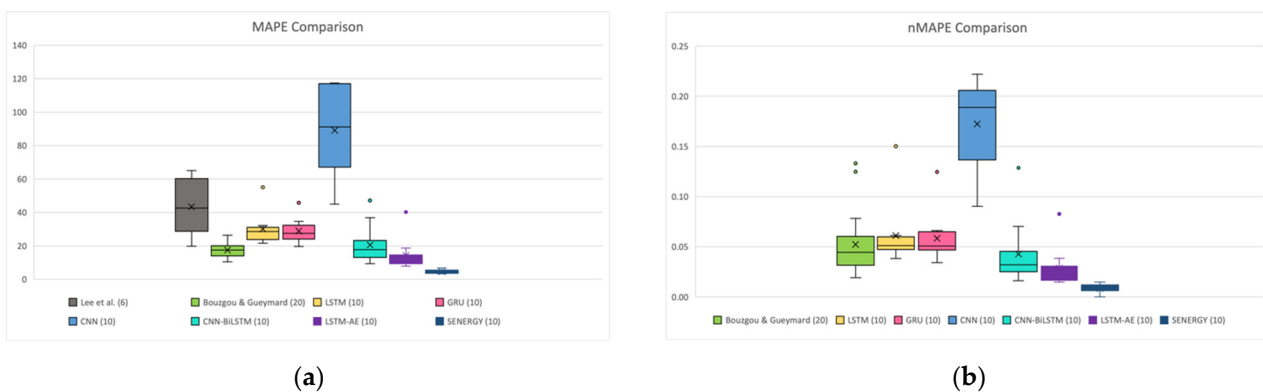


Figure 46. Comparison of multiple works based on: (a) MAPE; (b) nMAPE.

In Figure 47a, FS_{MAE} results of Gao et al. [24] are compared to the five forecasting models and the SENERGY results in this work. In this figure, the highest value is the best. It can be seen that the LSTM-AE model and SENERGY have the highest value and the lowest variation between locations, whereas the CNN model is the worst in terms of value and the CNN-BiLSTM model is the worst in terms of variation among locations. In (b), the FS_{RMSE} results of three works, Gao et al. [24], Fouilloy et al. [25], and Bouzguo and Gueymard [30], are compared to the five forecasting models and SENERGY results in this work. Again, the LSTM-AE model and SENERGY have the highest value and the lowest variation between locations. The second best performance is achieved by the work of Gao et al. [24]. However, they only include the results of four locations, compared to ten and

twenty in other works. On the other hand, Bouzgou and Gueymard [30] has the worst value and the largest variation between locations, since it includes twenty results. As in the MAE and RMSE results, SENERGY does not show better performance than the LSTM-AE model in (a) or (b) because on both metrics, the latter is the best model for all locations (refer to Figure 26).

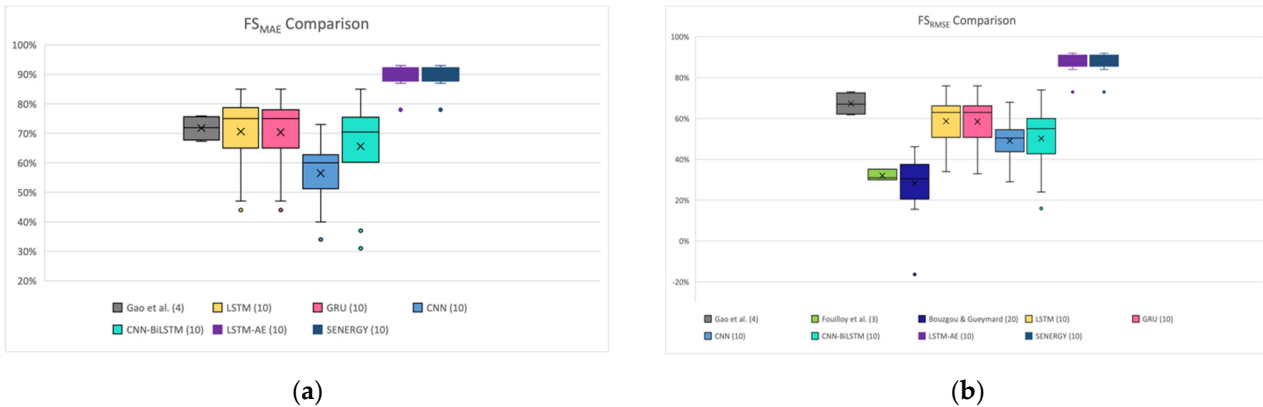


Figure 47. Comparison of multiple works based on: (a) FS_{MAE} ; (b) FS_{RMSE} .

Figure 48 compares SENERGY’s 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 24,676 record, 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 such precision results available from other researchers’ published works. From (a), we can see the improvement of SENERGY’s performance over the five models comes from the ability of the tool to choose the model that achieves the least error for each data input of the five models. 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 the CNN model is the worst.

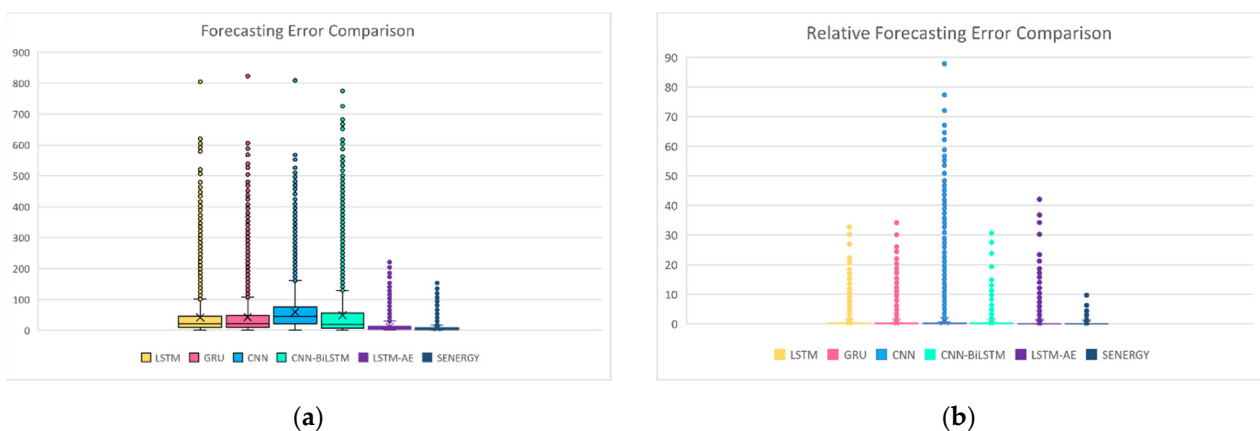


Figure 48. Comparison of SENERGY to other models based on (a) forecasting error; (b) relative forecasting error.

From all the figures shown in this section, the difficulty of comparing works when different metrics are reported and not all the needed information for a fair comparison is

given is evident. There is a need to improve, consolidate, and standardize international efforts on transparent and extensive testing of the proposed models for renewable energy forecasting [12]. One approach could be for researchers to make the complete results data openly available for comparison purposes. The boxplot used in this work provides the results at a higher granularity compared to the aggregate or average metrics. Particularly, the boxplots 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.

4.5. Results Summary

In Section 4.1, we evaluated the performance of the forecasting engine component of the SENERGY tool (we can also say that we evaluated the performance of the five forecasting models). Firstly, in Section 4.1.1, we studied the effect of lagged features on the forecasting results. Based on the MAE and RMSE results, a lag of two features have only improved the results of two models CNN-BiLSTM and LSTM-AE. Regarding MAPE results, a lag of three features only improved the results of GRU and LSTM-AE models. On the other hand, a lag of three features only improved the MAE and RMSE results of the LSTM-AE model as well as MAPE results of all models, except the CNN model. Next, in Section 4.1.2, we studied the effect of climate and location on the forecasting results. We found that the LSTM-AE model outperformed the other four forecasting models and achieved the best results ($nMAE = 0.02$, $nRMSE = 0.02$, and $nMAPE = 0.02$). 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. In addition, we found that the best forecasting results for all models are mostly associated with the Wadi-Addwasir dataset and the worst results are linked with the Toronto and Caracas datasets. From these results, we inferred that the most important factor that affects the performance of the different models is the climate of the dataset source (sunny or cloudy), followed by the completeness of the dataset. In Section 4.1.3, we studied the effect of sunny and cloudy weather on the forecasting results. We observed that predicting GHI on sunny days is more accurate than on cloudy days and the LSTM-AE model is the most accurate model among the five models in both conditions. In Section 4.1.4, we studied the effect of the summer and winter seasons on the forecasting results. We noticed that forecasting in the summer was generally more accurate than in the winter and the LSTM-AE model had a very slight difference in MAE results for both seasons. Finally, in Section 4.1.5, we examined the forecasting error of all five forecasting models (Section 4.1.5). We found that the LSTM-AE model achieved the least forecasting error for 54% of the records while the CNN-BiLSTM model achieved the least forecasting error for 17% of the records.

In Section 4.2, we evaluated the performance of the auto-selective model prediction engine of the SENERGY tool. First, in Section 4.2.1, we compared the classification accuracy of the model prediction engine based on climate and location. We found that the highest classification accuracy was 90%, and this was associated with Caracas and Toronto due to a relatively larger number of records for both locations compared to the other locations. Next, in Section 4.2.2, we compared the classification accuracy based on sunny and cloudy weather. We observed that the classification accuracy was 75% for sunny days and 86% for cloudy days. The reason for the lower accuracy in sunny weather was the close prediction for both the CNN-BiLSTM and LSTM-AE forecasting models, which made it difficult for the classifier to pick one of these models as the best model. Third, we compared the classification accuracy in the summer and winter seasons in Section 4.2.3. We found that the classification accuracy was 82% in summer and 80% in winter, which was in alignment with the trend found in the forecasting results.

In Section 4.3, we evaluated the performance of the SENERGY tool in terms of gain and loss in comparison to using a single forecasting model for forecasting. We first examined the actual gain and loss of SENERGY versus two forecasting models CNN-BiLSTM and LSTM-

AE in Section 4.3.1. We noted that the SENERGY gain versus CNN-BiLSTM was relatively large, because the tool would choose the LSTM-AE model for forecasting, which provided far better forecasting than the CNN-BiLSTM model. On the other hand, the SENERGY gain versus the LSTM-AE model was relatively small in general because when the CNN-BiLSTM model achieved better forecasting than the LSTM-AE model, the difference, and thus the gain, was small. Subsequently, we examined the potential gain and loss of SENERGY versus five forecasting models in Section 4.3.2. We found that assuming 100% accuracy of the classifier, the SENERGY tool would provide considerable gain over all the forecasting models with no loss.

Finally, in Section 4.4, we compared the performance of the SENERGY tool with other related works from the literature and we provided a performance comparison of models based on MAE, RMSE, MAPE results, their normalized versions, and F5MAE and F5RMSE values. The results establish that SENERGY provides superior performance overall based on all the different results.

5. Conclusions and Future Work

This paper introduced SENERGY, a novel deep learning-based auto-selective approach and a tool that predicts the best-performing deep learning model for GHI forecasting in terms of forecasting error rather than generalizing a specific model for all climates. The approach is based on carefully devised deep learning methods and feature sets created through an extensive analysis of deep learning forecasting and classification methods using ten meteorological datasets from three continents. We analyzed the tool in great detail through a variety of metrics and means for performance analysis, visualization, and comparison of solar forecasting methods. SENERGY outperforms existing methods in all performance metrics including MAE, RMSE, MAPE, nMAE, nRMSE, nMAPE, FS, and relative forecasting error. In all weather conditions, LSTM-AE is the most accurate. Prediction for sunny days is more accurate than the prediction for cloudy days; the same is true for the summer season versus the winter season. SYNERGY can predict the best forecasting model with 81% accuracy.

Future work would aim to make improvements in different aspects of the SENERGY tool design. For instance, the current performance of the SENERGY tool is suboptimal 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 predefined fixed input features, SYNERGY will be able to predict the best 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 for 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 optimizations 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.A. and R.M.; methodology, G.A. and R.M.; software, G.A.; validation, G.A. and R.M.; formal analysis, G.A., R.M. and S.H.H.; investigation, G.A., R.M. and S.H.H.; resources, R.M. and S.H.H.; data curation, G.A.; writing—original draft preparation, G.A. and R.M.; writing—review and editing, R.M. and S.H.H.; visualization, G.A.; supervision, R.M. and S.H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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 the King Abdullah City for Atomic and Renewable Energy (KACARE) for the supply of solar data.

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

Abbreviations

NWP	Numerical Weather Prediction
RNN	Recurrent Neural Network
ANN	Artificial neural network
AE	Autoencoder
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
MI	Mutual Information
MLP	Multilayer Perceptron Network
ARMA	Autoregressive Moving Average
WT	Wavelet Transform
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
SVM	Support Vector Machine
RMSE	Root Mean Square Error
nRMSE	Normalized Root Mean Square Error
MAPE	Mean Absolute Percentage Error
nMAPE	Normalized Mean Absolute Percentage Error
MAE	Mean Absolute Error
nMAE	normalized Mean Absolute Error
MSE	Mean Squared Error loss
WS	Wind Speed
AT	Air Temperature
RH	Relative Humidity
ML	Machine Learning
CNN	Convolutional Neural Network
PV	Photovoltaic
AI	Artificial Intelligence
SVR	Support Vector machine Regression
FFNN	Feed Forward Neural Network
GHI	Global Horizontal Irradiation
RF	Random Forest
DNN	Deep Neural Network
ELM	Extreme Learning Machine
BPNN	Back Propagation Neural Network
ReLU	Rectified Linear Unit

DHI	Diffuse Horizontal Irradiation
DNI	Direct Normal Irradiance
BiLSTM	Bidirectional LSTM
FS	Forecast Skill
XGBoost	eXtreme Gradient Boosting
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
WD	Wind Direction
BP	Barometric Pressure
ZA	Zenith Angle

References

- Shining Brightly | MIT News | Massachusetts Institute of Technology. Available online: <https://news.mit.edu/2011/energy-scale-part3-1026> (accessed on 14 August 2022).
- Kabir, E.; Kumar, P.; Kumar, S.; Adelodun, A.A.; Kim, K.-H. Solar energy: Potential and future prospects. *Renew. Sustain. Energy Rev.* **2018**, *82*, 894–900. [\[CrossRef\]](#)
- Shell Global, Global Energy Resources Database. Available online: <https://www.shell.com> (accessed on 26 June 2020).
- Elrahmani, A.; Hannun, J.; Eljack, F.; Kazi, M.-K. Status of renewable energy in the GCC region and future opportunities. *Curr. Opin. Chem. Eng.* **2021**, *31*, 100664. [\[CrossRef\]](#)
- Peng, T.; Zhang, C.; Zhou, J.; Nazir, M.S. An integrated framework of Bi-directional Long-Short Term Memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting. *Energy* **2021**, *221*, 119887. [\[CrossRef\]](#)
- Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [\[CrossRef\]](#)
- Kumari, P.; Toshniwal, D. Deep learning models for solar irradiance forecasting: A comprehensive review. *J. Clean. Prod.* **2021**, *318*, 128566. [\[CrossRef\]](#)
- Wang, H.; Liu, Y.; Zhou, B.; Li, C.; Cao, G.; Voropai, N.; Barakhtenko, E. Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Convers. Manag.* **2020**, *214*, 112909. [\[CrossRef\]](#)
- Ozcanli, A.K.; Yaprakdal, F.; Baysal, M. Deep learning methods and applications for electrical power systems: A comprehensive review. *Int. J. Energy Res.* **2020**, *44*, 7136–7157. [\[CrossRef\]](#)
- Said, Z.; Sharma, P.; Elavarasan, R.M.; Tiwari, A.K.; Rathod, M.K. Exploring the specific heat capacity of water-based hybrid nanofluids for solar energy applications: A comparative evaluation of modern ensemble machine learning techniques. *J. Energy Storage* **2022**, *54*, 105230. [\[CrossRef\]](#)
- Sharma, P.; Said, Z.; Kumar, A.; Nižetić, S.; Pandey, A.; Hoang, A.T.; Huang, Z.; Afzal, A.; Li, C.; Le, A.T. Recent advances in machine learning research for nanofluid-based heat transfer in renewable energy system. *Energy Fuels* **2022**, *36*, 6626–6658. [\[CrossRef\]](#)
- Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799. [\[CrossRef\]](#)
- Reda, F.M. Deep Learning an Overview. *Neural Netw.* **2019**, *12*, 14–18.
- Shamshirband, S.; Rabczuk, T.; Chau, K.-W. A Survey of Deep Learning Techniques: Application in Wind and Solar Energy Resources. *IEEE Access* **2019**, *7*, 164650–164666. [\[CrossRef\]](#)
- Ahmad, I.; Alqurashi, F.; Abozinadah, E.; Mehmood, R. Deep Journalism and DeepJournal V1.0: A Data-Driven Deep Learning Approach to Discover Parameters for Transportation. *Sustainability* **2022**, *14*, 5711. [\[CrossRef\]](#)
- Alahmari, N.; Alswedani, S.; Alzahrani, A.; Katib, I.; Albeshri, A.; Mehmood, R.; Sa, A.A. Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia. *Sustainability* **2022**, *14*, 3313. [\[CrossRef\]](#)
- Alswedani, S.; Mehmood, R.; Katib, I. Sustainable Participatory Governance: Data-Driven Discovery of Parameters for Planning Online and In-Class Education in Saudi Arabia During COVID-19. *Front. Sustain. Cities* **2022**, *4*, 97. [\[CrossRef\]](#)
- Janbi, N.; Mehmood, R.; Katib, I.; Albeshri, A.; Corchado, J.M.; Yigitcanlar, T.; Sa, A.A. Imtidad: A Reference Architecture and a Case Study on Developing Distributed AI Services for Skin Disease Diagnosis over Cloud, Fog and Edge. *Sensors* **2022**, *22*, 1854. [\[CrossRef\]](#) [\[PubMed\]](#)
- Alkhayat, G.; Mehmood, R. A review and taxonomy of wind and solar energy forecasting methods based on deep learning. *Energy AI* **2021**, *4*, 100060. [\[CrossRef\]](#)
- Abualigah, L.; Zitar, R.A.; Almotairi, K.H.; Hussein, A.M.; Elaziz, M.A.; Nikoo, M.R.; Gandomi, A.H. Wind, Solar, and Photovoltaic Renewable Energy Systems with and without Energy Storage Optimization: A Survey of Advanced Machine Learning and Deep Learning Techniques. *Energies* **2022**, *15*, 578. [\[CrossRef\]](#)
- Kumari, P.; Toshniwal, D. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *J. Clean. Prod.* **2021**, *279*, 123285. [\[CrossRef\]](#)
- Lima, M.A.F.B.; Carvalho, P.C.M.; Fernández-Ramírez, L.M.; Braga, A.P.S. Improving solar forecasting using Deep Learning and Portfolio Theory integration. *Energy* **2020**, *195*, 117016. [\[CrossRef\]](#)

23. AlKandari, M.; Ahmad, I. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Appl. Comput. Informatics* **2020**. *ahead-of-print*. [[CrossRef](#)]
24. Gao, B.; Huang, X.; Shi, J.; Tai, Y.; Zhang, J. Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. *Renew. Energy* **2020**, *162*, 1665–1683. [[CrossRef](#)]
25. Fouilloy, A.; Voyant, C.; Notton, G.; Motte, F.; Paoli, C.; Nivet, M.-L.; Guillot, E.; Duchaud, J.-L. Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. *Energy* **2018**, *165*, 620–629. [[CrossRef](#)]
26. Lago, J.; De Brabandere, K.; De Ridder, F.; De Schutter, B. Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data. *Sol. Energy* **2018**, *173*, 566–577. [[CrossRef](#)]
27. Lee, J.; Wang, W.; Harrou, F.; Sun, Y. Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. *Energy Convers. Manag.* **2020**, *208*, 112582. [[CrossRef](#)]
28. Yagli, G.M.; Yang, D.; Srinivasan, D. Automatic hourly solar forecasting using machine learning models. *Renew. Sustain. Energy Rev.* **2019**, *105*, 487–498. [[CrossRef](#)]
29. Srivastava, S.; Lessmann, S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Sol. Energy* **2018**, *162*, 232–247. [[CrossRef](#)]
30. Bouzgou, H.; Gueymard, C.A. Minimum redundancy–Maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series. *Sol. Energy* **2017**, *158*, 595–609. [[CrossRef](#)]
31. Despotovic, M.; Nedic, V.; Despotovic, D.; Cvetanovic, S. Review and statistical analysis of different global solar radiation sunshine models. *Renew. Sustain. Energy Rev.* **2015**, *52*, 1869–1880. [[CrossRef](#)]
32. Behar, O.; Khellaf, A.; Mohammedi, K. Comparison of solar radiation models and their validation under Algerian climate—The case of direct irradiance. *Energy Convers. Manag.* **2015**, *98*, 236–251. [[CrossRef](#)]
33. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *Int. J. Forecast.* **2016**, *32*, 896–913. [[CrossRef](#)]
34. Mohammed, T.; Albeshri, A.; Katib, I.; Mehmood, R. DIESEL: A Novel Deep Learning based Tool for SpMV Computations and Solving Sparse Linear Equation Systems. *J. Supercomput.* **2020**, *77*, 6313–6355. [[CrossRef](#)]
35. Usman, S.; Mehmood, R.; Katib, I.; Albeshri, A.; Altowaijri, S.M. ZAKI: A Smart Method and Tool for Automatic Performance Optimization of Parallel SpMV Computations on Distributed Memory Machines. *Mob. Networks Appl.* **2019**, 1–20. [[CrossRef](#)]
36. Usman, S.; Mehmood, R.; Katib, I.; Albeshri, A. ZAKI+: A Machine Learning Based Process Mapping Tool for SpMV Computations on Distributed Memory Architectures. *IEEE Access* **2019**, *7*, 81279–81296. [[CrossRef](#)]
37. Liu, Y.; Qin, H.; Zhang, Z.; Pei, S.; Wang, C.; Yu, X.; Jiang, Z.; Zhou, J. Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network. *Appl. Energy* **2019**, *253*, 113596. [[CrossRef](#)]
38. Zheng, J.; Zhang, H.; Dai, Y.; Wang, B.; Zheng, T.; Liao, Q.; Liang, Y.; Zhang, F.; Song, X. Time series prediction for output of multi-region solar power plants. *Appl. Energy* **2020**, *257*, 114001. [[CrossRef](#)]
39. Zhang, X.; Li, Y.; Lu, S.; Hamann, H.F.; Hodge, B.M.; Lehman, B. A Solar Time Based Analog Ensemble Method for Regional Solar Power Forecasting. *IEEE Trans. Sustain. Energy* **2019**, *10*, 268–279. [[CrossRef](#)]
40. Huertas-Tato, J.; Aler, R.; Galván, I.M.; Rodríguez-Benítez, F.J.; Arbizu-Barrena, C.; Pozo-Vázquez, D. A short-term solar radiation forecasting system for the Iberian Peninsula. Part 2: Model blending approaches based on machine learning. *Sol. Energy* **2020**, *195*, 685–696. [[CrossRef](#)]
41. Brahma, B.; Wadhvani, R. Solar irradiance forecasting based on deep learning methodologies and multi-site data. *Symmetry* **2020**, *12*, 1830. [[CrossRef](#)]
42. Khan, W.; Walker, S.; Zeiler, W. Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. *Energy* **2022**, *240*, 122812. [[CrossRef](#)]
43. Wang, F.; Xuan, Z.; Zhen, Z.; Li, K.; Wang, T.; Shi, M. 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.* **2020**, *212*, 112766. [[CrossRef](#)]
44. Singla, P.; Duhan, M.; Saroha, S. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. *Earth Sci. Informatics* **2022**, *15*, 291–306. [[CrossRef](#)] [[PubMed](#)]
45. Pan, C.; Tan, J. Day-ahead hourly forecasting of solar generation based on cluster analysis and ensemble model. *IEEE Access* **2019**, *7*, 112921–112930. [[CrossRef](#)]
46. El-Kenawy, E.-S.M.; Mirjalili, S.; Ghoneim, S.S.M.; Eid, M.M.; El-Said, M.; Khan, Z.S.; Ibrahim, A. Advanced Ensemble Model for Solar Radiation Forecasting Using Sine Cosine Algorithm and Newton’s Laws. *IEEE Access* **2021**, *9*, 115750–115765. [[CrossRef](#)]
47. Kaba, K.; Sarigül, M.; Avci, M.; Kandirmaz, H.M. Estimation of daily global solar radiation using deep learning model. *Energy* **2018**, *162*, 126–135. [[CrossRef](#)]
48. Jeon, B.K.; Kim, E.J. Next-Day Prediction of Hourly Solar Irradiance Using Local Weather Forecasts and LSTM Trained with Non-Local Data. *Energies* **2020**, *13*, 5258. [[CrossRef](#)]
49. Renewable Resource Atlas- King Abdullah City for Atomic and Renewable Energy. Available online: <https://rratlas.energy.gov.sa> (accessed on 1 December 2021).
50. Zepner, L.; Karrasch, P.; Wiemann, F.; Bernard, L. ClimateCharts.net—An interactive climate analysis web platform. *Int. J. Digit. Earth* **2021**, *14*, 338–356. [[CrossRef](#)]

51. Sengupta, M.; Habte, A.; Xie, Y.; Lopez, A.; Buster, G. National Solar Radiation Database (NSRDB). United States. *Renew. Sustain. Energy Rev.* **2018**, *89*, 51–60. [[CrossRef](#)]
52. Vignola, F. *GHI Correlations with DHI and DNI and the Effects of Cloudiness on One-Minute Data*; ASES: Schaumburg, IL, USA, 2012.
53. Yazdani, M.G.; Salam, M.A.; Rahman, Q.M. Investigation of the effect of weather conditions on solar radiation in Brunei Darussalam. *Int. J. Sustain. Energy* **2016**, *35*, 982–995. [[CrossRef](#)]
54. Petneházi, G. Recurrent neural networks for time series forecasting. *arXiv* **2019**, arXiv:1901.00069.
55. Marchesoni-Acland, F.; Alonso-Suárez, R. Intra-day solar irradiation forecast using RLS filters and satellite images. *Renew. Energy* **2020**, *161*, 1140–1154. [[CrossRef](#)]
56. Pereira, G.M.S.; Stonoga, R.L.B.; Detzel, D.H.M.; Küster, K.K.; Neto, R.A.P.; Paschoalotto, L.A.C. Analysis and Evaluation of Gap Filling Procedures for Solar Radiation Data. In Proceedings of the 2018 IEEE 9th Power, Instrumentation and Measurement Meeting (EPIM), Salto, Uruguay, 14–16 November 2018; IEEE: Salto, Uruguay, 2018; pp. 1–6.
57. Mohamad, N.B.; Lai, A.-C.; Lim, B.-H. A case study in the tropical region to evaluate univariate imputation methods for solar irradiance data with different weather types. *Sustain. Energy Technol. Assessments* **2022**, *50*, 101764. [[CrossRef](#)]
58. Abreu, E.F.M.; Canhoto, P.; Prior, V.; Melicio, R. Solar resource assessment through long-term statistical analysis and typical data generation with different time resolutions using GHI measurements. *Renew. Energy* **2018**, *127*, 398–411. [[CrossRef](#)]
59. KAPSARC Data Portal. Available online: <https://datasource.kapsarc.org/pages/home/> (accessed on 1 March 2022).
60. Tang, X.; Yao, H.; Sun, Y.; Aggarwal, C.; Mitra, P.; Wang, S. Joint modeling of local and global temporal dynamics for multivariate time series forecasting with missing values. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 5956–5963. [[CrossRef](#)]
61. Hadeed, S.J.; O'Rourke, M.K.; Burgess, J.L.; Harris, R.B.; Canales, R.A. Imputation methods for addressing missing data in short-term monitoring of air pollutants. *Sci. Total Environ.* **2020**, *730*, 139140. [[CrossRef](#)]
62. Venkatesh, B.; Anuradha, J. A review of feature selection and its methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26. [[CrossRef](#)]
63. Memarzadeh, G.; Keynia, F. A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets. *Energy Convers. Manag.* **2020**, *213*, 112824. [[CrossRef](#)]
64. Shilaskar, S.; Ghatol, A. Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. *Expert Syst. Appl.* **2013**, *40*, 4146–4153. [[CrossRef](#)]
65. Fonti, V.; Belitser, E. Feature selection using lasso. *VU Amsterdam Res. Pap. Bus. Anal.* **2017**, *30*, 1–25.
66. Zhou, H.; Zhang, Y.; Yang, L.; Liu, Q.; Yan, K.; Du, Y. Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. *IEEE Access* **2019**, *7*, 78063–78074. [[CrossRef](#)]
67. Sorkun, M.C.; Paoli, C.; Incel, Ö.D. Time series forecasting on solar irradiation using deep learning. In Proceedings of the 2017 10th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkey, 30 November–2 December 2017; IEEE: Bursa, Turkey, 2017; pp. 151–155.
68. Zang, H.; Liu, L.; Sun, L.; Cheng, L.; Wei, Z.; Sun, G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renew. Energy* **2020**, *160*, 26–41. [[CrossRef](#)]
69. Zang, H.; Cheng, L.; Ding, T.; Cheung, K.W.; Liang, Z.; Wei, Z.; Sun, G. Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Gener. Transm. Distrib.* **2018**, *12*, 4557–4567. [[CrossRef](#)]
70. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [[CrossRef](#)]
71. Dolatabadi, A.; Abdeltawab, H.; Mohamed, Y.A.-R.I. Hybrid Deep Learning-Based Model for Wind Speed Forecasting Based on DWPT and Bidirectional LSTM Network. *IEEE Access* **2020**, *8*, 229219–229232. [[CrossRef](#)]
72. Boubaker, S.; Benganem, M.; Mellit, A.; Lefza, A.; Kahouli, O.; Kolsi, L. Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia. *IEEE Access* **2021**, *9*, 36719–36729. [[CrossRef](#)]
73. Nguyen, H.D.; Tran, K.P.; Thomassey, S.; Hamad, M. Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *Int. J. Inf. Manage.* **2021**, *57*, 102282. [[CrossRef](#)]
74. Sagheer, A.; Kotb, M. Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems. *Sci. Rep.* **2019**, *9*, 19038. [[CrossRef](#)]
75. Li, G.; Xie, S.; Wang, B.; Xin, J.; Li, Y.; Du, S. Photovoltaic Power Forecasting With a Hybrid Deep Learning Approach. *IEEE Access* **2020**, *8*, 175871–175880. [[CrossRef](#)]
76. Hossain, M.S.; Mahmood, H. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* **2020**, *8*, 172524–172533. [[CrossRef](#)]
77. Alrashidi, M.; Alrashidi, M.; Rahman, S. Global solar radiation prediction: Application of novel hybrid data-driven model. *Appl. Soft Comput.* **2021**, *112*, 107768. [[CrossRef](#)]
78. Persson, C.; Bacher, P.; Shiga, T.; Madsen, H. Multi-site solar power forecasting using gradient boosted regression trees. *Sol. Energy* **2017**, *150*, 423–436. [[CrossRef](#)]
79. Meenal, R.; Selvakumar, A.I. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renew. Energy* **2018**, *121*, 324–343. [[CrossRef](#)]
80. Gigoni, L.; Betti, A.; Crisostomi, E.; Franco, A.; Tucci, M.; Bizzarri, F.; Mucci, D. Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants. *IEEE Trans. Sustain. Energy* **2018**, *9*, 831–842. [[CrossRef](#)]
81. Deo, R.C.; Şahin, M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew. Sustain. Energy Rev.* **2017**, *72*, 828–848. [[CrossRef](#)]

-
82. Marzo, A.; Trigo-Gonzalez, M.; Alonso-Montesinos, J.; Martínez-Durbán, M.; López, G.; Ferrada, P.; Fuentealba, E.; Cortés, M.; Batlles, F.J. Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. *Renew. Energy* **2017**, *113*, 303–311. [[CrossRef](#)]
 83. Ghimire, S.; Deo, R.C.; Wang, H.; Al-Musaylh, M.S.; Casillas-Pérez, D.; Salcedo-Sanz, S. Stacked LSTM Sequence-to-Sequence Autoencoder with Feature Selection for Daily Solar Radiation Prediction: A Review and New Modeling Results. *Energies* **2022**, *15*, 1061. [[CrossRef](#)]