



A Three-Stage Hierarchical Meta-Random Forest Stacking for One-Hour-Ahead Global Horizontal Irradiance Forecasting

D. Nelson Jayakumar*, J. Josalin Jemima, S. Charles Raja, M. Selva Vengatesh

Department of Electrical and Electronics Engineering, Thiagarajar College of Engineering, Madurai 625015, Tamilnadu, India.

Abstract: Accurate short-term solar irradiance forecasting is crucial for PV generation scheduling and energy management, however, hierarchical ensemble models can achieve optimistic performance if the temporal information flow is not maintained during meta-learning. This study presents a three-stage Meta-Random Forest (Meta-RF) stacking framework for one-hour-ahead global horizontal irradiance (GHI) forecasting. In the first stage, Lasso Regression, Ridge Regression, Support Vector Regression, and Decision Tree Regression are used; in the intermediate stage, Gradient Boosting Regression, XGBoost Regression, and AdaBoost Regression are employed; and in the final stage, Random Forest is used for meta-regression. A temporally ordered out-of-fold learning procedure is used to produce meta-features without direct propagation of in-sample fitted predictions across successive learning levels. The five years of hourly solar–meteorological data are divided chronologically into training (2016–2018), validation (2019), and independent testing (2020), where the hyperparameters are configured using the validation data. The complete Meta-RF framework outperforms the evaluated individual learners and intermediate stacking configurations in terms of MAE and RMSE, with an MAE of 1.13 W/m², RMSE of 3.41 W/m² and R² of 0.9998. Stage-wise analysis also shows the progressive improvement in forecasting with the hierarchical learning architecture, and measured–predicted characteristics and confidence-interval analysis provide evidence of consistency of the fitted relationship. The proposed framework is a temporally structured hierarchical learning approach for accurate one-hour-ahead GHI forecasting and short-term solar-energy applications.

Keywords: Global horizontal irradiance; Solar forecasting; Stacking ensemble; Meta-Random Forest; Machine learning

1. Introduction

The growing share of photovoltaic (PV) generation in modern power systems has made reliable short-term solar-resource forecasting more important than ever. The uncertainty of solar irradiance has a direct impact on PV generation and affects generation scheduling, energy management, reserve allocation, grid balancing, and coordinated operation of distributed energy resources [1, 2]. Global horizontal irradiance (GHI) is one of the main variables that characterize the solar resource for PV systems and is thus an important variable to forecast for operational energy applications [3, 4]. In terms of forecasting horizon, one hour ahead GHI forecast is a practical compromise between the advance information and short temporal resolution, which allows operational decisions to be taken in response to the changes of solar availability [5]. However, short-term forecasting is still difficult due to the nonlinear and time-varying nature of the interactions between solar geometry, cloudiness, temperature, humidity, and other atmospheric parameters that govern irradiance, and this has led to the development of forecasting methods that are able to represent complex solar–meteorological relationships.

In recent years, the research of solar forecasting has thus shifted from single machine-learning (ML) models to hybrid, ensemble, and hierarchical learning architectures. Comparative studies of hourly solar forecasting have shown the applicability of heterogeneous ML algorithms under different atmospheric conditions [6] and probabilistic XGBoost-based forecasting has proven the ability of gradient-boosted trees to capture the nonlinear behaviour of the irradiance [7]. The advantage of using complementary nonlinear representations has been further



demonstrated by hybrid CNN–SVR learning [8] and the potential of integrating multiple predictive mechanisms has been demonstrated by ridge-regression ensembles [9] and decomposition-assisted BiLSTM ensembles [10]. This evolution has been increasingly driving the stacking methods, which combine heterogeneous model predictions by an additional meta-learning mechanism. Ensemble stacking using deep learning has been used for PV generation forecasting [11] and stacking architectures have been explored for PV power prediction [12]. More recent advances have included stacking directly to GHI forecasting using boosting and bidirectional recurrent learning [13] and heterogeneous stacked ensembles for solar-irradiance forecasting [14]. Tree-based learning also offers additional support for heterogeneous ensemble construction, and Random Forest (RF) is shown to be effective for nonlinear solar-radiation prediction and irradiance-mapping [15, 16]. Optimized feature-selection strategies have also been explored for solar irradiance forecasting [17] and feature-selection-assisted recurrent learning for short-term PV prediction [18]. Recent stacking studies have also highlighted the importance of generalization across several PV systems [19] and wider studies of randomized ML methods have shown the benefits of predictive diversity in renewable energy forecasting [20]. Overall, the literature shows a clear evolution from single predictors to heterogeneous and hierarchical ensembles, and emphasizes the need for model diversity, information transfer and generalization in short-term solar forecasting.

Although these developments have been made, there are still significant methodological challenges in the design and assessment of multi-stage stacking models for short-term solar irradiance forecasting. The current methods mainly focus on the improvement of forecasting by hybridizing models, optimizing features, decomposing models, diversifying learners, or more and more complex ensemble architectures, and the predictive contribution added by the successive layers of a hierarchical stack is not always clearly separated. More important, the reliability of multi-stage stacking relies on the way predictive information is created and passed from one learning stage to the next. In-sample information can cascade through the hierarchy of higher-level learners, leading to an optimistic view of out-of-sample forecasting ability when the learners are developed with fitted predictions based on observations already seen when training the models. This is especially important for solar irradiance forecasting, since the observations are chronologically ordered and the chronological information availability is crucial for the operational forecasting. Therefore, there is a methodological gap for a hierarchical GHI forecasting framework that can simultaneously address the multi-stage learning of heterogeneous models, the generation of temporally ordered meta-features, the configuration and independent testing of models at different time points, and the stage-wise assessment of successive ensemble levels.

To fill this gap, this study proposes a three-stage Meta-Random Forest (Meta-RF) stacking model for one-hour-ahead GHI forecasting. The first stage consists of Lasso Regression, Ridge Regression, Support Vector Regression (SVR), and Decision Tree Regression (DT), which transform the input data in different ways. The first stage is the transformation of the input data by Lasso Regression, Ridge Regression, Support Vector Regression (SVR), and Decision Tree Regression (DT). The main methodological improvement is the incorporation of this hierarchical structure with an out-of-fold (OOF) learning procedure that is temporally ordered, where the meta-features provided to the various levels of the ensemble are derived from chronologically admissible observations instead of in-sample fitted predictions. The framework also distinguishes between model training, configuration based on validation data, and independent temporal testing, and stage-wise evaluation quantifies the predictive contribution of the successive learning stages. Therefore, the proposed approach integrates heterogeneous hierarchical integration, temporally valid meta-learning, independent chronological evaluation and stage-wise architectural verification in a single one-hour-ahead GHI forecasting framework.

The main findings of this study are as follows:

1. **Three-stage heterogeneous meta-stacking:** A Meta-RF architecture gradually combines regularized linear, kernel-based, tree-based and boosting representations to take advantage of complementary predictive properties for 1-hour-ahead GHI forecasting.



2. **Temporally ordered hierarchical learning:** An OOF meta-feature generation procedure is added at each successive stacking level to avoid direct propagation of in-sample fitted predictions and maintain chronological information flow in the hierarchical learning process.
3. **Chronologically independent evaluation:** Five years of hourly solar–meteorological data are divided into training, validation, and independent testing sets, with the model developed and configured during the training period, and the final assessment of the model's forecasting capability performed during the independent testing period, while keeping the training and independent testing periods temporally separated.
4. The performance of the individual base learners, intermediate GBR-, XGBoost-, and AdaBoost-based stacking configurations, and the overall Meta-RF framework are sequentially tested to measure the contribution of each learning stage.
5. **Multi-perspective forecasting evaluation:** Quantitative error measures, measured–predicted GHI characteristics, scatter-based agreement and confidence-interval analysis are used to examine forecasting performance, complementing the assessment of predictive accuracy and regression reliability.

2. Methodology

The proposed method uses a three-stage meta-stacking architecture for 1-hour-ahead global horizontal irradiance (GHI) forecasting. The framework gradually incorporates heterogeneous predictive representations by base learning, intermediate ensemble transformation and final meta-regression. The first learning stage is Lasso Regression, Ridge Regression, Support Vector Regression (SVR), and Decision Tree Regression (DT); the intermediate stage is Gradient Boosting Regression (GBR), XGBoost Regression, and AdaBoost Regression; and the final stage is Random Forest (RF) regression. To maintain the temporal integrity of the forecasting process, temporally ordered out-of-fold (OOF) prediction generation is included in the hierarchical model development, such that the meta-features provided to successive learning stages are not directly predicted out-of-sample from the corresponding observations. The resulting methodology integrates heterogeneous model diversity, progressive meta-feature transformation, leakage-safe temporal learning, and validation-guided model configuration in a single forecasting framework.

2.1. PROPOSED THREE-STAGE META-STACKING FRAMEWORK

The one-hour-ahead forecasting problem is represented by

$$D = \{(x_t, y_{t+1})\}_{t=1}^N \quad (1)$$

where $x_t \in R^p$ is the predictor vector, which is only based on solar and meteorological data available at or before the forecast-issuance time t , y_{t+1} is the measured GHI during the next hour, p is the number of predictors, and N is the number of forecasting samples. The measurements corresponding to the prediction hour $t + 1$ are excluded from x_t , thus preserving the temporal separation between the availability of the predictor and the prediction target. The variables that make up the predictor domain are described in **Section 3.1**, and the temporal alignment for one-hour-ahead forecasting is set in **Section 3.2**.

In the first learning stage, the predictor vector is provided separately to four different heterogeneous base regressors. Let $f_{m(\cdot)}$ denote the mapping learned by the m -th base learner. The corresponding forecasts are expressed as

$$\hat{y}_{t+1}^{(m)} = f_m(x_t), \quad m = 1, 2, 3, 4 \quad (2)$$

where $m = 1, 2, 3, 4$ corresponds to Lasso, Ridge, SVR, and DT, respectively. These four predictions from **equation (2)** are then fused together to form the first-stage meta-feature representation,

$$z_t^{(1)} = [\hat{y}_{t+1}^{(1)}, \hat{y}_{t+1}^{(2)}, \hat{y}_{t+1}^{(3)}, \hat{y}_{t+1}^{(4)}]^T \quad (3)$$

The representation in **equation (3)** combines predictive information from regularized linear, kernel and tree-based learning mechanisms. Instead of using a single model to capture the entire nonlinear solar–meteorological relationship, the first stage is to build a common prediction space in which complementary responses from the heterogeneous learners are preserved for the next stage of ensemble transformation.

In the second learning stage, the first-stage representation is separately learned by GBR, XGBoost, and AdaBoost. The second stage meta-feature representation is then defined as

$$z_t^{(2)} = [g_1(z_t^{(1)}), g_2(z_t^{(1)}), g_3(z_t^{(1)})]^T \quad (4)$$

where $k = 1, 2, 3$ corresponds to GBR, XGBoost, and AdaBoost, respectively. The intermediate stage is a transformation of the heterogeneous base-model responses using three different boosting mechanisms as expressed in **equation (4)**. This stage offers another layer of ensemble learning in which complementary patterns in the first stage prediction space are gradually improved prior to final aggregation.

The second stage representation is then fed to the RF meta-regressor. This one-hour-ahead GHI prediction is derived as

$$\hat{y}_{t+1} = h_{RF}(z_t^{(2)}) \quad (5)$$

where $h_{RF}(\cdot)$ denotes the final RF mapping. The entire hierarchical information flow of the proposed framework is described by **equations (2)–(5)**: heterogeneous base predictions are first aggregated into a common first-stage representation, then transformed by multiple boosting learners, and finally aggregated by RF meta-regression. This progressive organization allows information from various learning mechanisms to be stored and reworked at each ensemble level, with a small final prediction space.

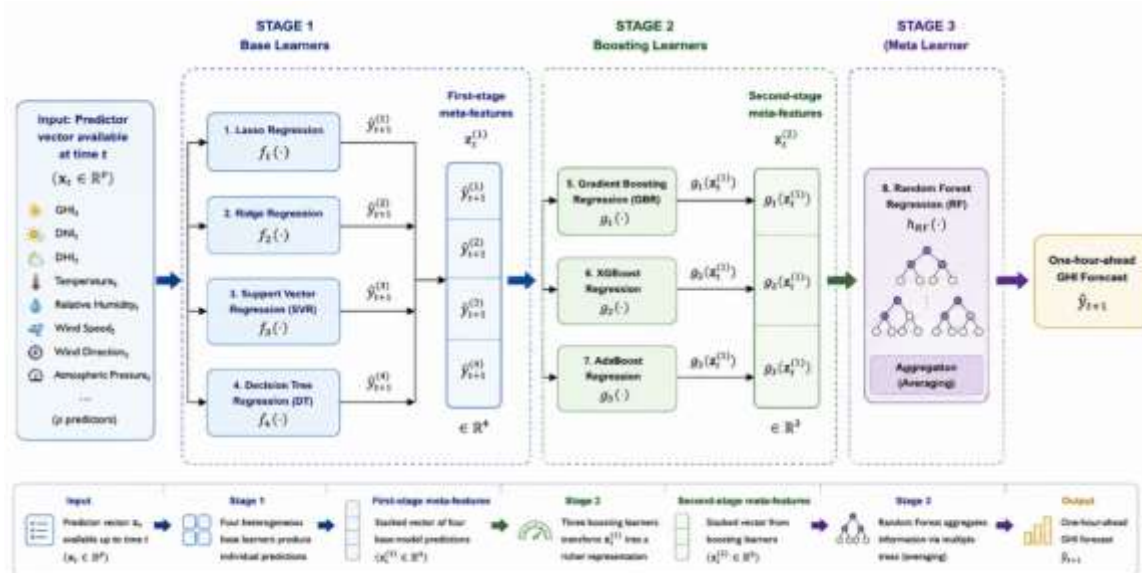


Fig. 1. Architecture of the proposed three-stage Meta-Random Forest stacking framework for one-hour-ahead GHI forecasting.



2.2. Leakage-Safe Hierarchical Learning And Forecast Generation

The hierarchical dependency among the learning stages requires careful construction of the meta-features used for model development. In-sample information can be cascaded through the stacking hierarchy by generating higher level training inputs from fitted predictions of the same observations used to estimate the previous learner, and can lead to overly optimistic estimates of generalization performance. This is especially important in the context of solar irradiance forecasting, as random mixing of observations over time can also introduce information after the prediction time. The proposed framework thus uses temporally ordered OOF meta-feature generation to maintain chronological causality across hierarchical learning.

The model-development sequence is broken down into a series of temporal blocks, but the measurements are kept in their original chronological order. The historical training set $D_{<r}$ for a temporal block $D^{(r)}$ consists of observations that are prior to the block. The necessary time restriction is given by

$$\max\{t: t \in D_{<r}\} < \min\{t: t \in D^{(r)}\} \quad (6)$$

By using **equation (6)**, predictions for each temporal block that is eligible are only made from models fitted with previous observations. The first admissible training window is defined in the first historical segment, and the history available grows gradually as more time blocks are traversed. This growing-history design maintains the natural flow of information and creates the OOF predictions needed for hierarchical model development.

The first-stage OOF representation for an observation i in temporal block $D^{(r)}$ is built as

$$z_{i,OOF}^{(1)} = [f_1^{(-r)}(x_i), f_2^{(-r)}(x_i), f_3^{(-r)}(x_i), f_4^{(-r)}(x_i)]^T, \quad i \in D^{(r)} \quad (7)$$

where $f_m^{((-r))^{(c)}}$ denotes the m -th base learner fitted using the admissible historical observations without using the temporal block being predicted. As a result, each meta-feature in **equation (7)** is an out-of-sample response with respect to the base learner.

The same temporal principle is carried over to the intermediate ensemble stage. The first-stage OOF representations of previous observations are admissible and used to fit the models, while the second-stage OOF representations are generated for future temporal blocks, with GBR, XGBoost, and AdaBoost being fitted. The training information provided to the RF meta-regressor is these temporally generated second stage meta-features. The hierarchical learning process therefore allows for the training of higher level models without having to train them directly on fitted predictions corresponding to the target observations, while maintaining the chronological order of the original measurement sequence.

After development and validation of the model, the developed learners are sequentially applied to the observations not seen before. The four base learners process the available predictor information according to **equations (2)–(3)**, the three boosting learners transform the first-stage representation according to **equation (4)**, and the final one-hour-ahead GHI prediction is produced by the RF meta-regressor according to **equation (5)**. This hierarchical information pathway is thus preserved in model development and independent forecasting, and the temporally ordered OOF procedure, which is controlled by **equations (6)–(7)**, ensures leakage-safe construction of meta-features.

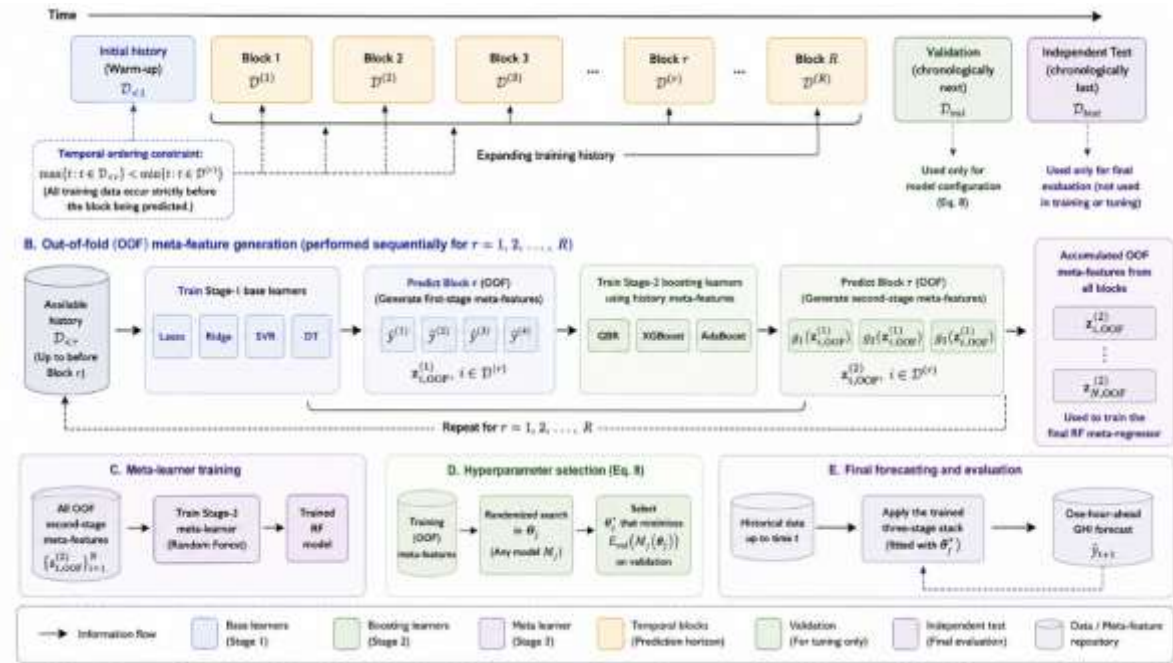


Fig. 2. Temporally ordered out-of-fold meta-feature generation and hierarchical learning procedure adopted for the proposed three-stage stacking framework.

2.3. Model Configuration And Hyperparameter Optimization

The constituent learning models have various hyperparameters related to regularization, nonlinear mapping, tree construction, boosting behaviour and ensemble aggregation. The independent test period was therefore kept separate from the parameter selection, and model configuration was carried out within the model-development boundary set in **Section 3.2**. The respective configuration spaces were explored under a common temporally ordered validation procedure by adopting randomized hyperparameter search.

For model M_j , let Θ_j denote the corresponding hyperparameter search space and $\theta_j \in \Theta_j$ an individual candidate configuration. The selected configuration is defined as

$$\theta_j^* = \arg \min_{\theta_j \in \Theta_j} E_{val} (M_j(\theta_j)) \quad (8)$$

where $E_{val}(\cdot)$ is the forecasting error from the temporally following validation data. As defined in equation (8), the choice of hyperparameters is based on the out-of-sample validation performance, not on the goodness of fit to the observations used for model estimation.

Model configuration is done in a hierarchical manner, according to the direction of information propagation. The base learners are configured first, and then the temporally ordered predictions from the base learners are used as the meta-features to configure GBR, XGBoost, and AdaBoost. The intermediate representations are then used for the configuration of the RF meta-regressor. This sequential process ensures consistency between model selection and the hierarchical information flow described in **Section 2.1** and preserves the temporal constraints described in **Section 2.2**. The hyperparameter search ranges and the chosen hyperparameters for the constituent models are summarized in **Table 5**.

Table 5. Hyperparameter search spaces and selected configurations of the constituent forecasting models.

Model	Hyperparameters considered	Search space	Selected configuration
Lasso Regression	Regularization coefficient (α)	$\alpha = 10^{-5} - 10^1$	$\alpha = 0.001$
Ridge Regression	Regularization coefficient (α)	$\alpha = 10^{-5} - 10^3$	$\alpha = 1.0$
Support Vector Regression	C, γ, ε , kernel	$C = 1 - 1000$; $\gamma = 10^{-4} - 1$; $\varepsilon = 0.01 - 10$; kernel = RBF	$C = 100$; $\gamma = 0.01$; $\varepsilon = 0.1$; RBF
Decision Tree Regression	Maximum depth, minimum samples for split, minimum samples per leaf	Depth = 3–30; split = 2–20; leaf = 1–10	Depth = 15; split = 4; leaf = 2
Gradient Boosting Regression	Number of estimators, learning rate, maximum depth	Estimators = 100–800; learning rate = 0.01–0.20; depth = 2–8	Estimators = 300; learning rate = 0.05; depth = 4
XGBoost Regression	Number of estimators, learning rate, maximum depth, subsample	Estimators = 100–1000; learning rate = 0.01–0.20; depth = 2–10; subsample = 0.6–1.0	Estimators = 500; learning rate = 0.05; depth = 6; subsample = 0.8
AdaBoost Regression	Number of estimators, learning rate	Estimators = 50–800; learning rate = 0.01–1.0	Estimators = 300; learning rate = 0.05
Random Forest Regression	Number of estimators, maximum depth, maximum features, minimum samples for split/leaf	Estimators = 100–1000; depth = 5–30; features = 2–7; split = 2–20; leaf = 1–10	Estimators = 500; depth = 23; features = 6; split = 2; leaf = 1

After the selection based on the validation, the selected configurations were kept without further adjustment based on the independent test results. The optimized configurations were then tested with the standard experimental setup outlined in **Section 3**.

3. Case Study

The proposed three-stage forecasting framework was tested with multi-year solar and meteorological data from Madurai, Tamil Nadu, India. The experimental investigation was designed to explore the one-hour-ahead GHI forecasting under temporally varying irradiance conditions and to give a solid ground to the model evaluation out of sample. This section introduces the measurement dataset and study site, sets the experimental configuration, and introduces the comparative evaluation framework used for performance assessment.

3.1. Study Site And Dataset

The experimental case study was set up at Thiagarajar College of Engineering (TCE), Madurai, Tamilnadu, India (9.9252° N, 78.1198° E) as shown in **Fig. 3**. The solar irradiance and related meteorological data were obtained from the weather-monitoring station at the study site, along with the data exported from the Elmeasure monitoring environment. The measurement records were exported at a regular hourly temporal resolution for the five-year period 2016–2020 for the present investigation. This setup will create a long-term data base of site-specific data for characterizing the temporal behaviour of solar irradiance under the meteorological conditions observed at the experimental site and for the development and evaluation of the proposed framework for solar irradiance forecasting.



Fig. 3. Geographical location of the experimental solar irradiance prediction site at Thiagarajar College of Engineering, Madurai, Tamil Nadu, India (9.9252° N, 78.1198° E).

The adopted data set combines the amount of solar radiation with meteorological and surface-related parameters related to the variability of solar radiation. Global horizontal irradiance (GHI) was chosen as the prediction target. The solar-radiation variables include GHI, direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), and clear-sky values of these variables. The following environmental parameters are included: cloud type, temperature, dew point, relative humidity, pressure, surface albedo, wind direction, and wind speed. Together, these measurements represent the solar availability in clear sky conditions, the conditions of the solar radiation actually received, the state of the atmosphere, moisture conditions, local wind behaviour, and surface reflectance. The summary of the overall configuration of the experimental data set is shown in **Table 2**.

Table 2. Configuration of the solar–meteorological dataset adopted for the case study.

Characteristic	Specification
Study location	Thiagarajar College of Engineering, Madurai, Tamil Nadu, India
Coordinates	9.9252° N, 78.1198° E
Data source	TCE weather-monitoring facility and Elmeasure environment
Observation period	2016–2020
Temporal resolution	1 h
Prediction target	Global horizontal irradiance (GHI)



Observed irradiance variables	GHI, DNI, DHI
Clear-sky irradiance variables	Clear-sky GHI, clear-sky DNI, clear-sky DHI
Meteorological variables	Cloud type, temperature, dew point, relative humidity, pressure, wind direction, wind speed
Surface-related variable	Surface albedo

The five-year measurement record shows that the statistical behaviour of the solar-radiation and meteorological variables are quite different, as shown in **Table 3**. The target GHI is 0 to 1075 W/m², mean 241.77 W/m², standard deviation (SD) 324.59 W/m², coefficient of variation (CV) 1.34. The other components of the irradiance are also quite variable. DNI ranges from 0 to 986 W/m², with a mean of 195.97 W/m², an SD of 276.29 W/m², and a CV of 1.40, whereas DHI ranges from 0 to 524 W/m², with a mean of 102.78 W/m², an SD of 139.91 W/m², and a CV of 1.36. The clear-sky GHI, DNI, and DHI have CV values of 1.27, 1.15, and 1.21, respectively. The figures show that the prediction target and the related components of the irradiance change significantly during the period of measurement considered.

Table 3. Descriptive statistics of the solar, meteorological, and surface variables in the 2016–2020 dataset.

Variable	Minimum	Maximum	Mean	SD	CV
Clear-sky DHI	0.00	518.00	70.72	86.07	1.21
Clear-sky DNI	0.00	986.00	301.52	347.66	1.15
Clear-sky GHI	0.00	1075.00	284.54	361.52	1.27
Cloud type	0.00	9.00	4.25	3.17	0.74
Temperature	15.50	43.70	28.18	4.81	0.17
Dew point	4.10	27.00	20.80	2.92	0.14
DHI	0.00	524.00	102.78	139.91	1.36
DNI	0.00	986.00	195.97	276.29	1.40
GHI	0.00	1075.00	241.77	324.59	1.34
Relative humidity	12.58	100.00	68.74	22.03	0.32
Surface albedo	0.13	0.23	0.18	0.01	0.05
Pressure	985.00	1004.00	993.72	2.95	0.00
Wind direction	0.00	360.00	173.25	104.90	0.60
Wind speed	0.00	7.80	2.10	1.24	0.59

The relative dispersion of the meteorological and surface variables is relatively small compared to the solar-radiation components. The CV values for temperature, dew point and relative humidity are 0.17, 0.14 and 0.32, respectively. The relative variation (CV) of surface albedo is the lowest (CV = 0.05), while wind direction and wind speed have CV values of 0.60 and 0.59, respectively. As a result, the experimental data set is composed of

strongly varying quantities of irradiance and meteorological variables with varying temporal dispersion. This is a heterogeneous measurement domain that offers complementary information to characterize the nonlinear variations of GHI as a function of solar and atmospheric conditions.

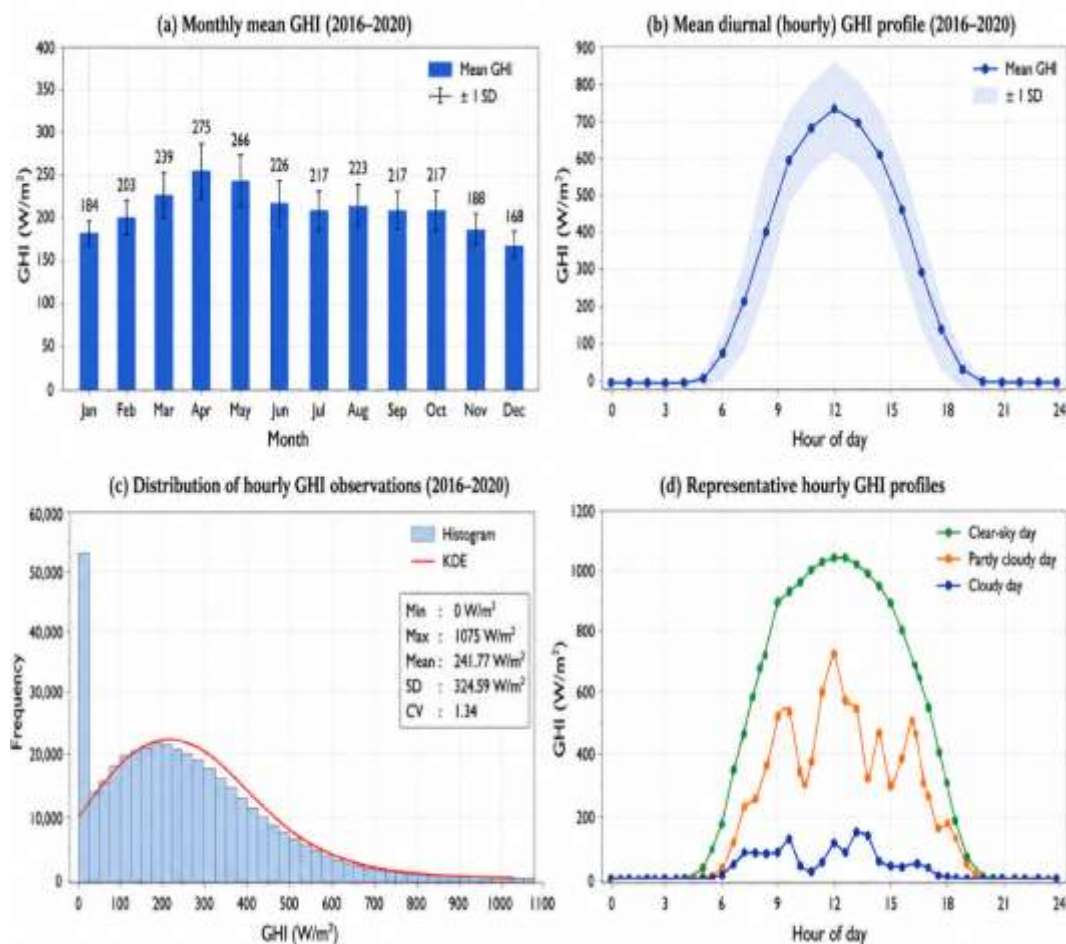


Fig. 4. Multi-year temporal profile of hourly global horizontal irradiance (GHI) at the Madurai case-study location during 2016–2020.

The temporal evolution of the prediction target was also analysed in the multi-year GHI profile shown in **Fig. 4**. The hourly measurements cover the range of irradiance observed from 0 to 1075 W/m² and repeated transitions between low, intermediate and high irradiance conditions are observed over the 5 year period. The measured GHI shows high temporal variability, which is consistent with the SD of 324.59 W/m² and CV of 1.34, and offers a data basis for developing and testing the proposed forecasting framework.

3.2. Forecasting And Experimental Configuration

The experimental analysis was set up for a 1-hour-ahead GHI forecast, which corresponds to the temporal resolution of the measurement records. In each forecasting instance, data up to the current time step was fed into the predictor to forecast GHI for the next hour, and the predictor and target sequences were aligned with a one-step temporal offset. To ensure that the available information was temporally separated from the forecasting target, measurements corresponding to the prediction hour were not included in the predictor set. The five-year record was divided chronologically instead of randomly to evaluate generalization under temporally unseen conditions.



The data from 2016–2018 were used for model training, the data from 2019 were used for model validation and configuration, and the data from 2020 were used for independent testing. The adopted partitioning strategy is summarized in **Table 4**.

Table 4. Chronological partition of the dataset for model development and evaluation.

Subset	Period	Role in experimental analysis
Training	2016–2018	Learning of constituent models and ensemble development
Validation	2019	Model configuration and hyperparameter selection
Independent test	2020	Final out-of-sample forecasting assessment

The hourly records were first checked for data consistency and then aligned in time prior to model development, and samples for one hour ahead forecasting were then created. Observations were chronologically ordered throughout the preprocessing, training, validation, and testing, and the preprocessing parameters needed by the individual learning algorithms were calculated from the model-development data, without utilizing the independent test period. The hierarchical stacking process was given special attention as predictions made at one learning stage are used as input to the next learning stage. Thus, the training information fed to the higher-level learners was not their own fitted predictions, but temporally ordered out-of-fold (OOF) predictions, which avoids the higher-level learners being trained directly on the fitted predictions of their previous learners.

The full hourly sequence includes nighttime hours where GHI is close to or equal to zero. These observations were kept in the temporally continuous data set, but the main forecasting evaluation was done during daylight hours to avoid having a large number of easily predictable zero-irradiance observations unduly affect the accuracy reported. The model and hyperparameter selection was done only within the training–validation boundary and then the selected models and hyperparameters were not adjusted further for independent testing. All the proposed framework and benchmark models were tested with the same forecasting horizon, chronological partition, preprocessing conditions, and test observations. Mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) were used to quantify the performance, offering complementary measures of average forecasting deviation, sensitivity to larger errors, and agreement with the observed GHI variability.

The entire experimental procedure is shown in **Fig. 5**. After temporal alignment of the hourly measurements, one-hour-ahead forecasting samples are created and then split into training, validation, and independent test sets. The model-development stage involves the generation of the proposed hierarchical ensemble in a temporally ordered manner and the configuration of the models based on the validation period. The finalized proposed and benchmark models are then tested on the independent 2020 dataset using the same experimental conditions, and the comparative, temporal, statistical, and uncertainty analyses are presented in **Section 4**.

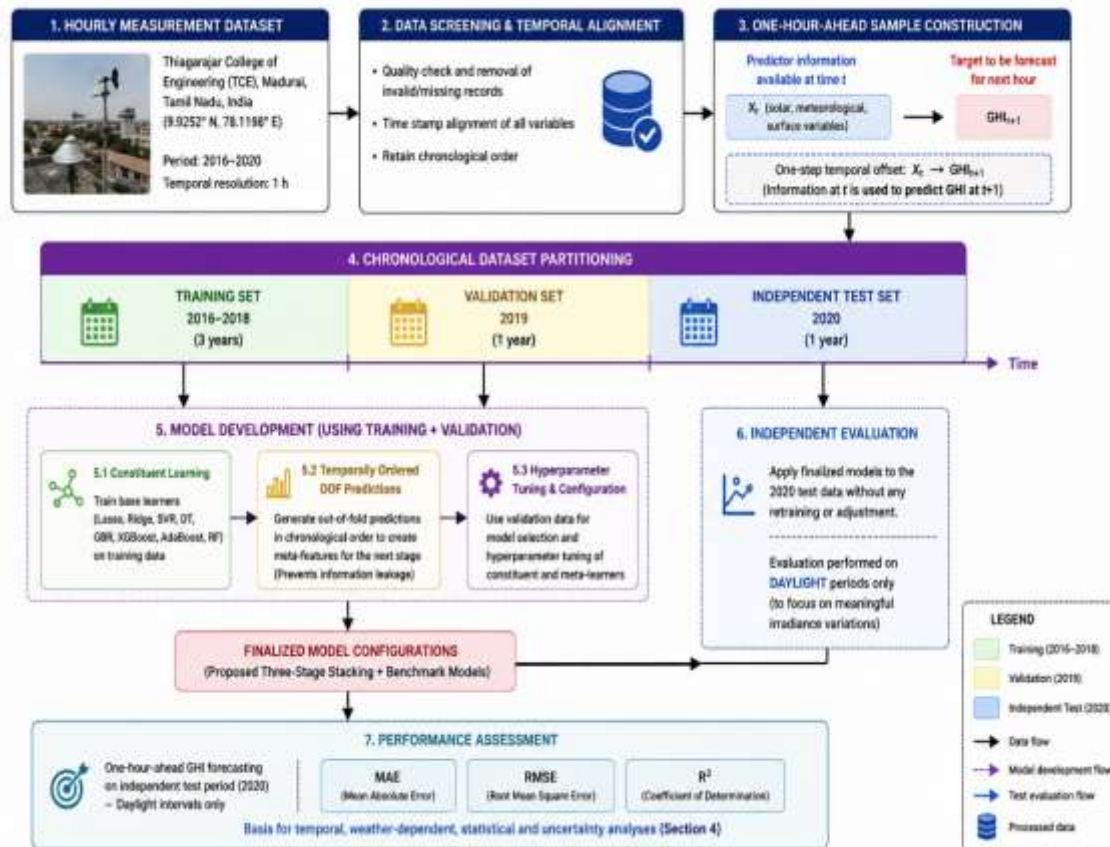


Fig. 5. Experimental protocol adopted for one-hour-ahead GHI forecasting, including temporal alignment, chronological dataset partitioning, hierarchical model development, validation, and independent performance evaluation.

3.3. Benchmark Models And Evaluation Framework

The comparative evaluation was designed to assess the forecasting capability of the proposed three-stage architecture against both standalone predictors and ensemble-based alternatives. Four different heterogeneous regressors were selected as individual learning benchmarks: Lasso Regression, Ridge Regression, Support Vector Regression (SVR), and Decision Tree Regression (DT), which are regularized linear, kernel-based, and nonlinear tree-based learning mechanisms, respectively. The intermediate ensemble configurations that were considered were Gradient Boosting Regression (GBR), XGBoost Regression and AdaBoost Regression, which were applied to the predictions of the first-stage learners. The final hierarchical configuration was then tested, the complete three-stage Meta-Random Forest (Meta-RF) framework. The common forecasting horizon, chronological data partition, preprocessing conditions, and independent test observations set in **Section 3.2** were used for all competing configurations.

A stage-wise comparison was added to quantify the contribution of the successive learning levels in the proposed architecture. First, the performance of the individual Stage-1 learners was investigated, then the three Stage-2 stacking configurations (GBR, XGBoost, and AdaBoost), and lastly the full Meta-RF architecture. This progressive comparison allows the forecasting enhancement due to the intermediate ensemble transformation and final RF meta-regression to be separated from the performance of the individual learners. The summary of the benchmark configurations and their role in the comparative assessment is given in **Table 5**.

Table 5. Models and comparison levels adopted for forecasting evaluation.

Comparison level	Model/configuration	Purpose
Individual learner	Lasso Regression	Regularized linear benchmark
	Ridge Regression	Regularized linear benchmark
	Support Vector Regression (SVR)	Kernel-based benchmark
	Decision Tree Regression (DT)	Nonlinear tree-based benchmark
Ensemble learner	Gradient Boosting Regression (GBR)	Boosting benchmark
	XGBoost Regression	Gradient-boosted tree benchmark
	AdaBoost Regression	Adaptive boosting benchmark
Intermediate ensemble	Stage-wise stacking configurations	Incremental contribution assessment
Proposed model	Three-stage Meta-Random Forest stacking	Final hierarchical forecasting framework

The evaluation is based on quantitative forecasting indices and on a graphical analysis of the prediction behaviour. The main performance metrics are mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2), which offer complementary information about the average forecasting error, the sensitivity to larger errors, and the agreement with the measured GHI variability. The MSE and MSLE values available were also kept in the overall comparative assessment for completeness. Measured and predicted GHI profiles were analyzed to determine the capability of the competing stacking configurations to reproduce the irradiance variations, and measured versus predicted scatter distributions were analyzed to determine the extent of agreement over the range of observed irradiances.

The reliability of the predictions was also analyzed using the adopted 95% confidence interval analysis for the individual, intermediate and final forecasting configurations. The interpretation of the confidence-interval assessment was not as a calibrated prediction interval for individual future observations but as additional information about the fitted regression relationship. The basis for the comparative evaluation presented in **Section 4** is the aggregate error measures, measured–predicted comparisons, stage-wise assessment, scatter distributions, and confidence-interval analysis.

4. Results And Discussion

The forecasting accuracy of the proposed three-stage Meta-Random Forest (Meta-RF) stacking framework was compared with the four base learners and three intermediate stacking combinations. The assessment considers the overall forecasting accuracy, agreement between measured and predicted irradiance, contribution of the successive hierarchical learning stages, and prediction reliability. All models were tested with the same experimental setup described in **Section 3**, so that the performance gain of the proposed hierarchical architecture could be consistently analyzed.

4.1. Overall Forecasting Performance

The forecasting performance of Lasso Regression, Ridge Regression, Support Vector Regression (SVR), Decision Tree Regression (DT), stacking with GBR, XGBoost and AdaBoost is compared with the proposed three-stage Meta-RF framework in **Table 6**. There are significant variations between the individual learners. The largest forecasting errors were obtained by SVR with an MAE of 18.54 W/m² and RMSE of 40.41 W/m², while Lasso

and Ridge obtained RMSE of 26.72 and 26.64 W/m², respectively. DT, on the other hand, significantly lowered the forecasting error, with an MAE of 2.25 W/m² and RMSE of 6.50 W/m². The difference between the linear, kernel-based, and tree-based learners illustrates the different capabilities of the individual models to capture the nonlinear relationships between the solar–meteorological predictors and GHI.

There are also unique predictive behaviours for the intermediate stacking configurations. The MAE and RMSE values for GBR were 2.62 W/m² and 6.65 W/m², respectively, and for XGBoost were 2.30 and 6.50 W/m², respectively. The errors obtained by AdaBoost were comparatively larger, MAE = 12.26 W/m² and RMSE = 14.49 W/m². The results presented here show that the addition of an intermediate ensemble stage does not necessarily lead to better forecasting accuracy, but rather the success of the stacking process relies on the transformation of the predictions of the ensemble members and the way they are stacked.

Table 6. Forecasting performance of the proposed three-stage Meta-RF framework and comparison models.

Forecasting model	MAE (W/m ²)	MSE	RMSE (W/m ²)	MSLE	R ²
Lasso Regression	15.53	714.00	26.72	1.4626	0.9931
Ridge Regression	15.88	709.76	26.64	1.6733	0.9932
SVR	18.54	1633.49	40.41	0.2124	0.9843
Decision Tree Regression	2.25	42.30	6.50	0.0003	0.9995
GBR stacking	2.62	44.34	6.65	0.0009	0.9996
XGBoost stacking	2.30	42.33	6.50	0.0003	0.9996
AdaBoost stacking	12.26	209.98	14.49	3.1807	0.9979
Proposed three-stage Meta-RF	1.13	11.64	3.41	9.9453	0.9998

The proposed three-stage Meta-RF configuration yielded the lowest MAE, MSE, and RMSE values of 1.13 W/m², 11.64, and 3.41 W/m², respectively, and the highest R² of 0.9998 among the evaluated configurations. The proposed framework outperformed GBR, XGBoost, and AdaBoost by 48.72%, 47.53%, and 76.33%, respectively, compared with the intermediate stacking configurations. The corresponding reductions in MAE were 56.87%, 50.86%, and 90.78%. The proposed model achieved the reduction of RMSE from 6.50 to 3.41 W/m² (about 47.5%) and MAE from 2.25 to 1.13 W/m² (about 49.8%) compared to the strongest individual learner DT.

The comparative performance is further illustrated in Fig. 6. The simultaneous decrease in both the MAE and RMSE is especially significant since the MAE reflects the average size of the error in the forecasts, while the RMSE gives more weight to relatively large errors. The increase in R² suggests that the decrease in absolute forecasting error is paralleled by a closer reproduction of the measured GHI variation. The proposed architecture therefore not only improves the overall goodness of fit but also the size of the individual forecasting errors.

There is one exception for MSLE. The proposed framework has the best performance in terms of MAE, MSE, RMSE, and R^2 , but the reported MSLE is higher than that of the comparison models. The proposed framework is thus not understood as being better in all reported metrics. The principal forecasting assessment outlined in **Section 3** is based on MAE, RMSE, and R^2 , all of which are easily understood.

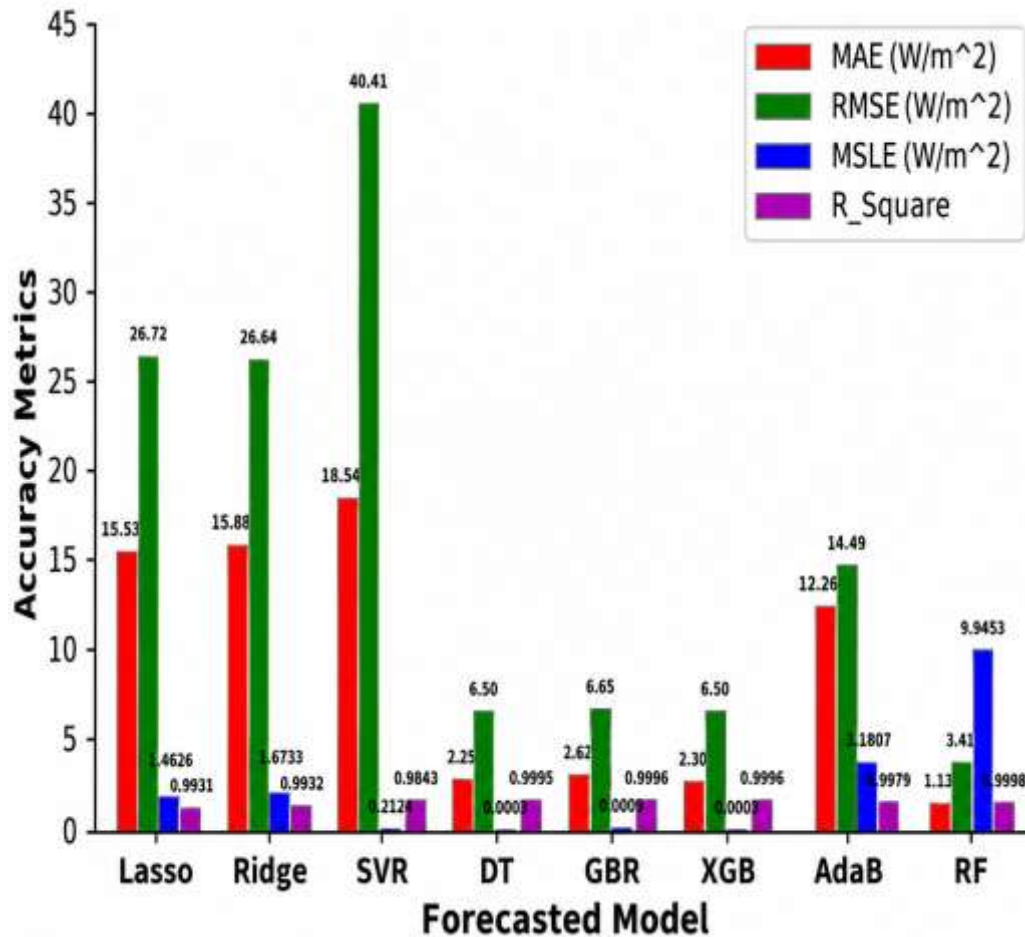


Fig. 6. Comparison of forecasting accuracy metrics for the individual learners, intermediate stacking configurations, and proposed three-stage Meta-RF framework.

4.2. Measured–Predicted Ghi Characteristics

The aggregate metrics were supplemented by an analysis of the measured and predicted irradiance profiles. The GHI estimates generated by the intermediate stacking configurations based on the GBR, XGBoost, and AdaBoost models are compared with the GHI estimates generated by the proposed Meta-RF model in **Fig. 7**. In all the observations analyzed, the proposed predictions are closer to the measured irradiance profile, as shown in **Table 6**, which is consistent with the lower MAE and RMSE. The improvement is especially noticeable compared to AdaBoost, which has a higher deviation of 14.49 W/m², as seen in its RMSE.

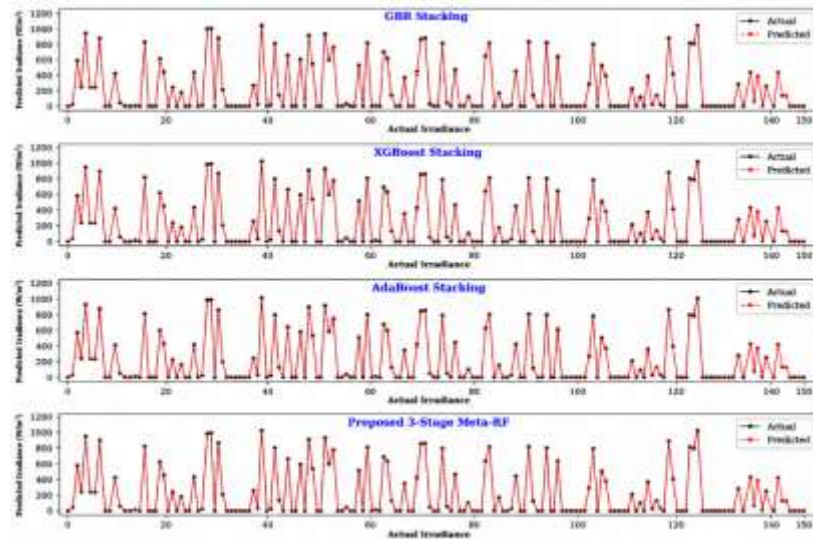


Fig. 7. Comparison of measured and predicted GHI for the GBR-, XGBoost-, and AdaBoost-based stacking configurations and the proposed three-stage Meta-RF framework.

The scatter distributions of measured versus predicted values in **Fig. 8** are another way of assessing the level of agreement in the forecasts. The closer the predictions are to the 1:1 line, the better the correspondence with the measured GHI, while the greater the dispersion, the greater the deviation of the predictions. The proposed Meta-RF predictions show the best overall agreement with the measured observations, as seen in the R^2 of 0.9998. DT, GBR, and XGBoost also have high R^2 values ranging from 0.9995 to 0.9996, but their higher MAE and RMSE values indicate that a high R^2 does not necessarily equate to the smallest point-wise forecasting error.

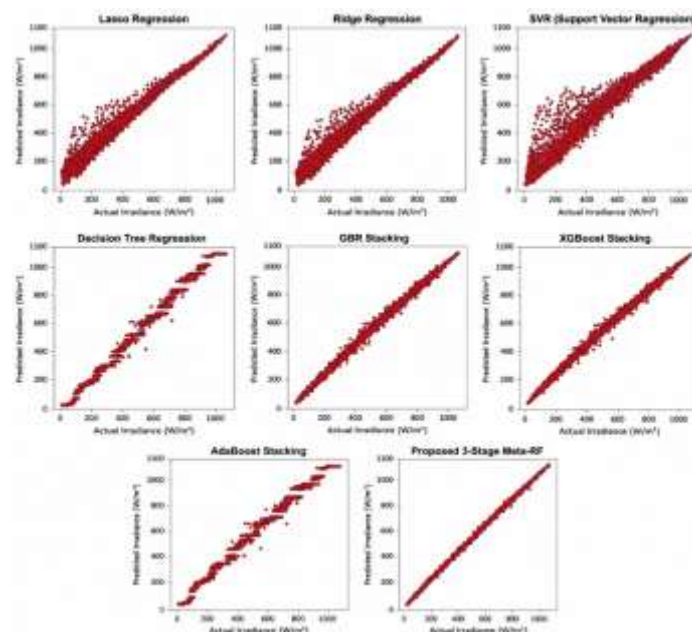


Fig. 8. Measured-versus-predicted GHI scatter distributions for the individual learners, intermediate stacking configurations, and proposed three-stage Meta-RF framework.

The interpretation of the temporal prediction profiles, scatter distributions, and quantitative error measures is therefore more comprehensive than the use of a single goodness-of-fit measure. The proposed framework is particularly characterized by the highest R^2 and the lowest MAE and RMSE, which show that the measured irradiance variation is well reproduced and that the magnitude of the corresponding forecasting error is minimized.

4.3. Stage-Wise Contribution Of The Hierarchical Architecture

The performance of the proposed framework was also investigated from a stage-wise point of view to see if the final RF meta-regression stage adds value to the constituent and intermediate models. DT was the best individual learner at the first learning stage with an MAE of 2.25 W/m² and RMSE of 6.50 W/m². For intermediate stacking configurations, XGBoost had the lowest MAE of 2.30 W/m², and XGBoost and GBR had RMSE values of 6.50 and 6.65 W/m², respectively. The full three-stage Meta-RF architecture then lowered the MAE to 1.13 W/m² and the RMSE to 3.41 W/m².

Table 7. Stage-wise performance of the proposed hierarchical forecasting architecture.

Learning level	Configuration	MAE (W/m ²)	RMSE (W/m ²)	R ²
Stage 1	Lasso	15.53	26.72	0.9931
	Ridge	15.88	26.64	0.9932
	SVR	18.54	40.41	0.9843
	Decision Tree	2.25	6.50	0.9995
Stage 2	Base predictions → GBR	2.62	6.65	0.9996
	Base predictions → XGBoost	2.30	6.50	0.9996
	Base predictions → AdaBoost	12.26	14.49	0.9979
Stage 3	Base → boosting → Meta-RF	1.13	3.41	0.9998

Compared to XGBoost, which achieved the best overall Stage-2 result, the final Meta-RF stage reduced the MAE by about 50.9% and the RMSE by about 47.5%. The improvement over each intermediate configuration shows that the final prediction is not only due to the best boosting learner. Instead, the three Stage-2 representations are passed to the RF meta-regressor, which uses the complementary responses of the three representations to get the final estimate.

This stage-wise behaviour provides direct support for the hierarchical organization introduced in **Section 2**. The base learners create heterogeneous predictive representations, the intermediate boosting models transform their respective outputs independently, and the RF meta-regressor integrates the outputs. Importantly, the comparison also reveals that higher hierarchy does not always lead to better performance, as is the case for the comparatively weaker AdaBoost configuration. The advantage of the full architecture is not just that there are more learning stages, but that the complementary intermediate representations are integrated at the end.

4.4. Forecast Reliability And Uncertainty Analysis

The deterministic forecasting assessment was complemented by the 95% confidence-interval analysis presented in **Fig. 9**. Confidence intervals were examined for the four constituent base learners, the three intermediate stacking configurations, and the proposed Meta-RF framework. The available analysis indicates closer regression agreement for the proposed framework, consistent with the lower deterministic errors and higher R^2 reported in **Table 6**.

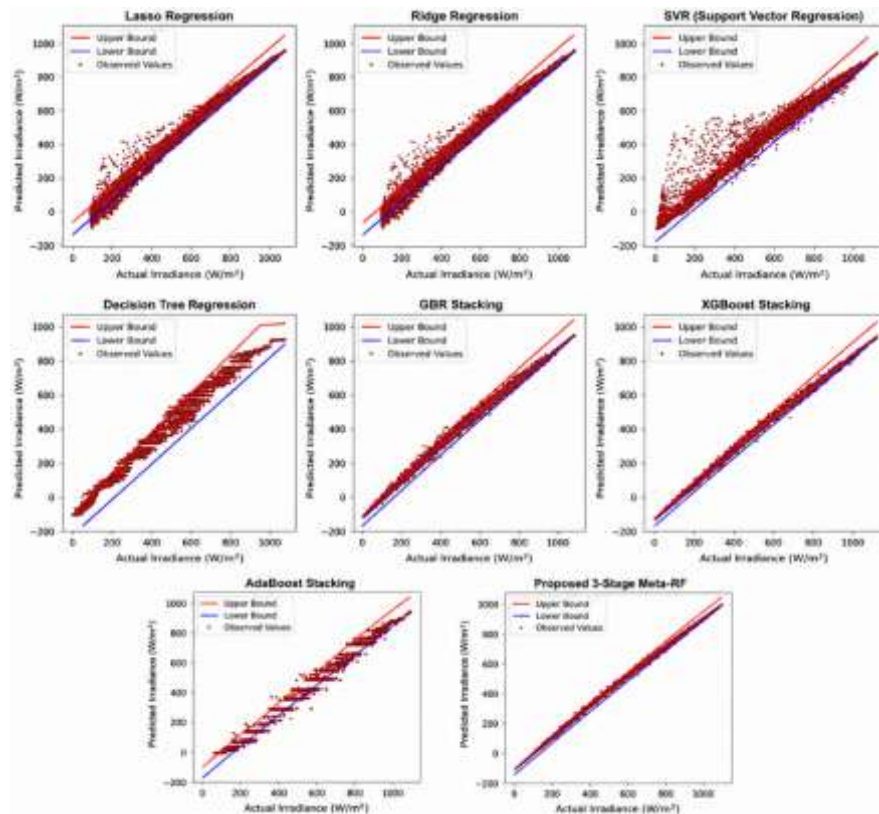


Fig. 9. 95% confidence-interval analysis for the individual learners, intermediate stacking configurations, and proposed three-stage Meta-RF framework

The confidence-interval results provide complementary information to the point-prediction metrics by illustrating the statistical range associated with the fitted relationship. Nevertheless, the confidence intervals should not be interpreted as calibrated prediction intervals for individual future GHI observations. The present uncertainty analysis therefore provides evidence concerning the reliability of the regression relationship, whereas the deterministic MAE, RMSE, and R^2 values remain the principal quantitative measures of one-hour-ahead forecasting performance.

This distinction is relevant when interpreting the practical reliability of the forecasting framework. A narrow confidence interval around an estimated regression relationship does not necessarily imply an equally narrow uncertainty range for every future irradiance observation. Consequently, future extension of the framework toward probabilistic forecasting could incorporate explicit prediction intervals and evaluate their empirical coverage and width. Within the present deterministic forecasting scope, however, the confidence-interval analysis provides supporting evidence consistent with the comparative performance trends observed across the evaluated models.

4.5. Practical Implications And Limitations

The proposed three-stage Meta-RF framework shows the potential of hierarchical integration of heterogeneous learning models for one-hour-ahead GHI forecasting, which results in lower forecasting errors than the evaluated individual and intermediate stacking configurations, and thus enables short-term PV generation estimation, energy scheduling, and renewable-energy management applications. The chronological evaluation and the temporally ordered OOF learning further reinforce the assessment by maintaining the time gap between the availability of the predictors and the following forecasting target. However, the present study is restricted to a single geographical



location in Madurai and a one-hour forecasting horizon and the performance of the study is dependent on the availability and quality of the solar–meteorological predictors used. In addition, the 95% confidence-interval analysis is not a measure of the uncertainty of calibrated prediction, and computational training and inference times were not quantified. Future investigations should therefore examine multi-site and multi-horizon generalization, calibrated probabilistic forecasting, and computational performance under operational deployment conditions.

5. Conclusion

This study proposed a three-stage Meta-Random Forest (Meta-RF) stacking model for global horizontal irradiance (GHI) forecasting at one-hour-ahead time horizon, which combines heterogeneous predictive models and temporally ordered hierarchical learning. The first stage is Lasso Regression, Ridge Regression, Support Vector Regression, and Decision Tree Regression; the intermediate stage is Gradient Boosting Regression, XGBoost Regression, and AdaBoost Regression; and the final stage is Random Forest, which is a meta-regression. To ensure the chronological information flow during the generation of the meta-features, a temporally ordered out-of-fold learning procedure was introduced, and the training, validation, and independent testing periods ensured temporal separation during the model development and evaluation process. The complete Meta-RF framework achieved an MAE of 1.13 W/m², MSE of 11.64 (W/m²)², RMSE of 3.41 W/m², and R² of 0.9998. The final framework achieved a reduction of RMSE of 48.72%, 47.53%, and 76.33% compared to the intermediate stacking based on GBR, XGBoost, and AdaBoost, respectively, and a reduction of MAE of 56.87%, 50.86%, and 90.78%, respectively, indicating the added predictive power of the final meta-regression stage. The consistency of the forecasting relationship was also confirmed by the measured–predicted characteristics, the scatter-based agreement, and the confidence-interval analysis. The results show the effectiveness of the proposed temporally structured meta-feature generation and heterogeneous hierarchical learning for accurate short-term solar forecasting. The present study is restricted to one geographical area, one-hour forecast period, and the confidence intervals are based on the regression reliability, and computational performance was not measured. Future research will thus be directed towards multi-site and multi-horizon validation, calibration of probabilistic forecasting, and computational assessment under operational conditions to further validate the generalizability and practical applicability of the framework for PV generation forecasting, energy scheduling, and renewable-energy management.

Reference

- [1] Yagli, G. M., Yang, D., & Srinivasan, D. (2019). Automatic hourly solar forecasting using machine learning models. *Renewable and Sustainable Energy Reviews*, 105, 487-498.
- [2] Li, X., Ma, L., Chen, P., Xu, H., Xing, Q., Yan, J., ... & Cheng, Y. (2022). Probabilistic solar irradiance forecasting based on XGBoost. *Energy Reports*, 8, 1087-1095.
- [3] Ghimire, S., Bhandari, B., Casillas-Perez, D., Deo, R. C., & Salcedo-Sanz, S. (2022). Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia. *Engineering Applications of Artificial Intelligence*, 112, 104860.
- [4] Carneiro, T. C., Rocha, P. A., Carvalho, P. C., & Fernández-Ramírez, L. M. (2022). Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain. *Applied Energy*, 314, 118936.
- [5] Singla, P., Duhan, M., & Saroha, S. (2022). An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. *Earth Science Informatics*, 15(1), 291-306.
- [6] Khan, W., Walker, S., & Zeiler, W. (2022). Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. *Energy*, 240, 122812.
- [7] Zhang, H., & Zhu, T. (2022). Stacking model for photovoltaic-power-generation prediction. *Sustainability*, 14(9), 5669.



-
- [8] Srivastava, R., Tiwari, A. N., & Giri, V. K. (2019). Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India. *Heliyon*, 5(10).
 - [9] Babar, B., Luppino, L. T., Boström, T., & Anfinsen, S. N. (2020). Random forest regression for improved mapping of solar irradiance at high latitudes. *Solar Energy*, 198, 81-92.
 - [10] Omar, N., Aly, H., & Little, T. (2022). Optimized feature selection based on a least-redundant and highest-relevant framework for a solar irradiance forecasting model. *IEEE Access*, 10, 48643-48659.
 - [11] Liu, R., Wei, J., Sun, G., Muyeen, S. M., Lin, S., & Li, F. (2022). A short-term probabilistic photovoltaic power prediction method based on feature selection and improved LSTM neural network. *Electric Power Systems Research*, 210, 108069.
 - [12] Del Ser, J., Casillas-Perez, D., Cornejo-Bueno, L., Prieto-Godino, L., Sanz-Justo, J., Casanova-Mateo, C., & Salcedo-Sanz, S. (2022). Randomization-based machine learning in renewable energy prediction problems: Critical literature review, new results and perspectives. *Applied Soft Computing*, 118, 108526.