International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 10, October-2021

A Novel Two Layer Stacking Ensemble for Improving Solar Irradiance Forecasting

Eric Nziyumva[1], Mathias Nsengimna[2], Jovial Niyogisubizo[1], Evariste Murwanashyaka[3], Emmanuel Nisingizwe[4], Alphonse Kwitonda[2]

[1]Fujian Key Lab for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou, China.

[2]African Centre of Excellence in Energy for sustainable Development, College of Science and Technology, University of Rwanda

[3]Institute of Rock and Soil Mechanics, University of Chinese Academy of Sciences, Wuhan, China

[4]Department of Electrical and Information Engineering, University of Nairobi, Kenya

***Abstract*: Solar irradiance forecasting plays a vital role in the reliable planning and efficient designing of solar energy systems. Moreover, solar power energy has gained significant importance as a clean, renewable, and alternative cheapest source of energy over the past few decades ago. However, the efficiency of solar power generation is strongly dependent on weather conditions and other natural intermittent parameters. Consequently, this leads to serious challenging issues during power grid management include non-stable operation and significant maintenance losses. To address these issues, accurate forecasting becomes an attractive solution to minimize the impact of uncertainty and energy costs. In this paper, we firstly built a novel computational framework based on stacking techniques to enhance the forecasting accuracy of solar irradiation. Then, the stacking-based ensemble is compared with the single models. The Adaptive Boosting (AdaBoost), Bootstrap aggregating (Bagging) regressor, Multi-Layer Perceptron (MLP), and its combination through stacking technique were compared. The working principle of the stacked AdaBoost-Bagging regressor-MLP model consists of combining the prediction of AdaBoost and Bagging regressor to generate final prediction using the MLP network. The dataset from the Philippines' government weather station especially located in Morong, Rizal province was used to validate the reliability of our study. We evaluate the forecasting performances via determination coefficient (R^2), mean absolute error (MAE), and root mean squared error (RMSE). The stacking-based ensemble learning performs better than any single model in terms of all three statistical indicators. This study contributes mainly to the development of reliable stacking ensemble-based model to minimize solar irradiance forecasting errors. Additionally, comparative assessment of the models leads to successful energy management.**

Keywords— Solar irradiation forecasting, machine learning, Stacking ensemble, Energy management, Multi-layer perceptron

I. INTRODUCTION

Solar-based energy becomes one of the most promising sources for generating power for residential, commercial, and industrial applications due to its characteristics of being environmental friendly[1]. However, the main difficulty with these resources is the uncertainty in their output power due to various uncontrollable and natural intermittent factors affecting solar energy. Consequently, this affects negatively to the overall power grid management. For instance, the power imbalance of photovoltaic system may cause significant losses, which compromises the development of any nation. In addition, the measurement process of those intermittent factors requires non-cheap sensor-based devises. Furthermore, it is also a complicated and time-consuming to install such measuring devices all over the world[2]. Hence, proper and accurate solar energy prediction is extremely important.

The variation of the temperature and irradiance have an extreme impact on the quality of solar-based electric power production[3]. Since solar irradiance and solar power output are highly related therefore solar irradiance forecasting is the best key factor to indicate the power production**.** Various models and algorithms have been widely explored to predict solar irradiance using different meteorological factors such as temperature and humidity. According to the literature, the development of solar power prediction is still an interested research topic as well as the desired prediction level is not yet reached for any electrical network.

Few decades ago, numerous models have been proposed for solar irradiance prediction issues. Some of them are based on mathematical formula and called empirical models[4]. The empirical became popular and widely used due to its ease of results interpretation. Among the various examples for solar irradiance prediction include cloudiness-based[5], sunshine-based[6], temperature-based[7],and meteorological parameters-based models[8]. However, these models are not capable to accurately predict the short-term solar irradiance due to the rapid changes in weather conditions. In addition, some researchers reported these models for not being able to reflect the complex and nonlinear relationships among both input and output variables in humid regions in which solar irradiation is strongly affected by heavy clouds throughout rainy days[9]. Previous studies reported also empirical models for presenting partially-unsatisfying forecasting results for daily global solar radiation data[10].

With the advancement of the technology, artificial intelligence (AI) became very popular and widely used for almost all engineering fields[11]. Lately, the AI algorithms have been reported as more accurate than empirical algorithms for solar irradiance prediction[9]. For instance, Quej et al. predicted daily global solar radiation data of six stations in Mexico by using support vector machine (SVM), artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). In the relevant study, the best results were achieved in SVM with RMSE = 2.578, MAE = 1.97 and R2 = 0.689[12].

Even if the AI algorithms are used to build the enhanced solar irradiance prediction models that have shown an outstanding advancement over empirical models, the performance of their models present various gaps of erroneous results due to variance, bias and noise. Moreover, high computational cost, instability issues, and less performance accuracy limit AI techniques while handling high dimensional and complex data[13]. These affect negatively to the solar irradiance prediction, which lead to significant losses and unsafe planning due to the bad management of power grid system. Consequently, AI algorithms became less competent for solar irradiance prediction.

A few years ago, ensemble-based machine learning became another alternative way for replying to the solar irradiance forecasting issues. Various tree-based ensemble methods have shown their significant role through not only their robust forecasting algorithms but also their stability and powerfulness[3]. In this paper, Adaptive Boosting (AdaBoost) and bootstrap aggregating (Bagging) regressor are combined using multi-layer perceptron (MLP) through stacking technique with the aim of investigate the capability of stacking ensemble over other ensemble learning. The proposed approach named stacked AdaBoost-Bagging regressor-MLP is firstly explored in solar irradiance forecasting. Then after, this new ensemble learning is compared with their benchmarks include AdaBoost, Bagging regressor and MLP. To the best of our knowledge, no comprehensive investigation using this method for solar irradiance forecasting has been reported yet.

The goal of this work is to save the significant losses by minimizing the aforementioned limitations. The contributions of this paper are summarized as follows:

- First, we introduce ensemble-learning models for improving solar irradiance prediction. Actually, the use of ensemble learning models is motivated by their characteristics of combining several weak learners to achieve an improved forecasting quality comparatively to conventional single learners. Moreover, they reduce the overall prediction error and with their ability of combining different models.
- Four machine learning models include AdaBoost, Bagging regressor, MLP and its stacking ensemble are compared each others. By considering all parameters for each models and using numerous evaluation metrics (MAE, RMSE, R^2), we obtain the acceptable results which leads to our target of reducing the significant losses. This enhances not only the power grid management but also the development of any nation.

The rest of the paper is arranged as follows. Section 2 presents dataset exploration and machine learning models. Section 3 contains evaluation criteria of models and comparative study. Lastly, section 4 concludes the paper and provides some recommendations of future research in this field.

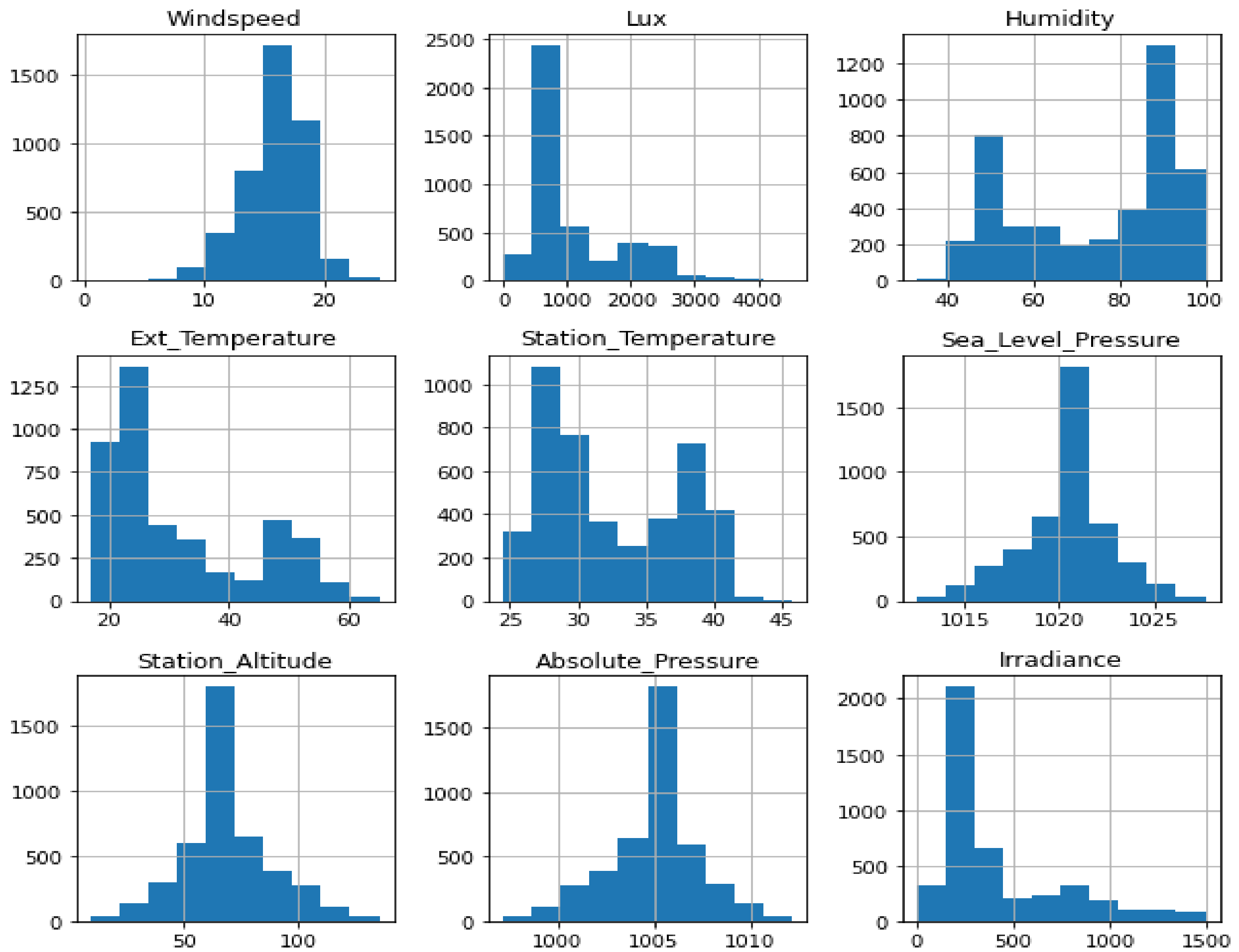

Fig. 1. Dataset attributes

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 10, October-2021

II. METHODOLOGY

This section is based on the four machine learning (ML) models used in this study. Fig. 1 summarizes the main steps of the proposed methodology. The proposed approach includes three key steps such as dataset exploration, data preprocessing and preliminaries on ML models.

A. Dataset Exploration

The dataset used in this study is provided by Philippines' government weather station especially located in Morong, Rizal province[14]. Data collection of nine weather-based attributes were recorded as comma separated values (.csv) format from September 2019. The raw data contains the information of 4330 samples with sampling frequency of one hour.

The solar irradiance is the dependent variable in this study. It is expressed as the intensity coming from the sun in the form of electromagnetic radiation. It is measured in terms of watt per square meter (W/m2). Since solar irradiation depends on weather conditions, thus the input elements are also almost weather-based parameters. These variables include absolute pressure, external temperature, humidity, Lux, sea level pressure, station altitude, station temperature and wind speed.

Fig. 1. presents the histogram of the dataset attributes. This histogram helps to check the normality of the dataset by assessing the shape of dataset distribution.

Fig. 2. presents the correlation heatmap between the variables.The strong inverse relationship is indicated by the darkest color. In other hand, the value between 0.7 and 1 indicates the strong direct relationship between two variables. The values at or close to zero imply a weak correlation .

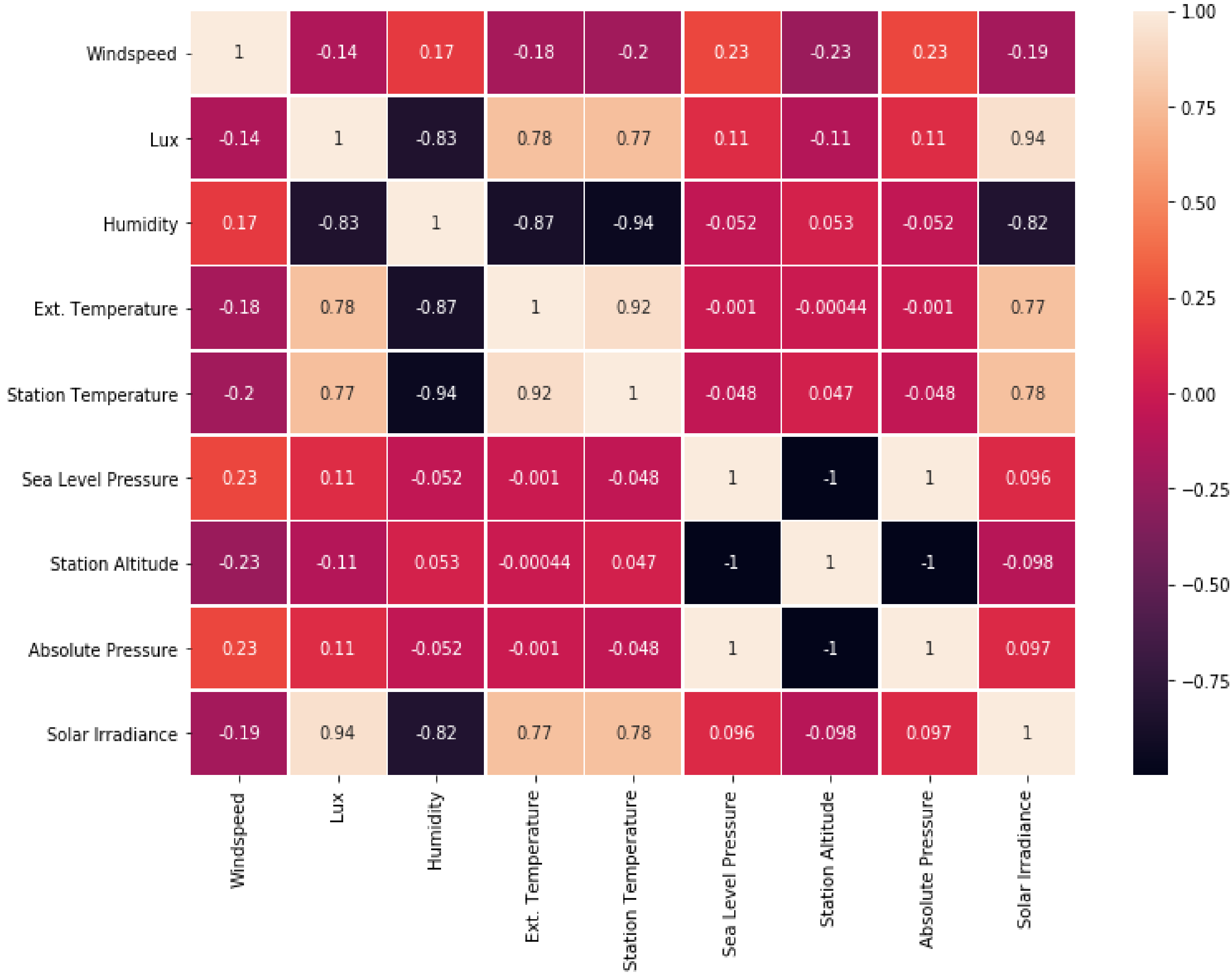

Fig. 2. Correlation Heatmap of the variables

B. Data preprocessing

The prediction system is improved by the quality of input variables and the forecasting engine. Moreover, the prediction errors are minimized by reliable data analysis and feature engineering. Therefore, the data should be cleaned to provide adequate quality in the dataset. Therefore, data preprocessing is required for ensuring the compatibility of the discussed dataset with regression models used in this study. Thus, data preprocessing is the process of transforming raw data into understandable format. Here, we have firstly imported

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 10, October-2021

necessary libraries and read data. Then, missing values and categorical data were checked. The missing values were dropped. Furthermore, data standardization and principal component analysis (PCA) transformation were done. Lastly, data-splitting phase contains two folds for training and testing data at a ratio of 80% and 20% respectively[15]. The input and output variables were fully identified into dataset exploration.

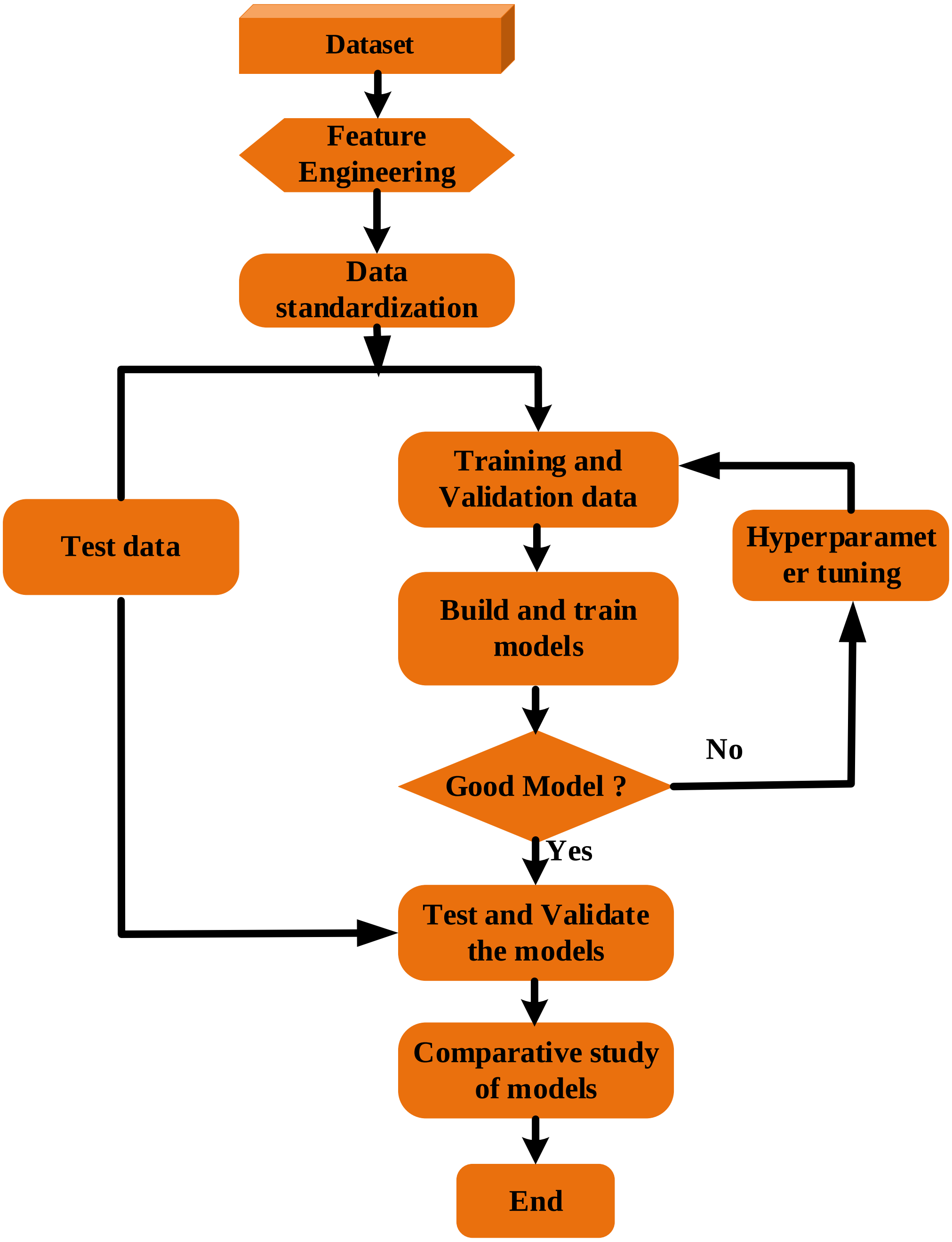

Fig. 3. Schematic block diagram of the study

Fig. 3. summarizes the main steps of the proposed methodology. This approach combines three key steps such as dataset exploration, data preprocessing and preliminaries on ML models.

C. Preliminaries on machine learning models

- **Adaptive Boosting (AdaBoost):** The AdaBoost is the first boosting-based algorithm developed by the joint of Freund and Schapire[16]. The boosting algorithm takes primarily its vital role as the machine learning meta-algorithm designed to enhance the forecasting accuracy. The boosting method expresses the sequential structure of base estimators in which one tries to minimize the bias and variance of the combined estimator[17]. Due to its advantages for handling regression and classification issues, adaptive boosting is widely used and applied in various engineering fields such as forecasting.

- **Bagging regressor:** Bagging (Bootstrap aggregating) method introduced by Breiman[18] is a

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 10, October-2021

ML ensemble meta-algorithm that primarily designed to improve the stability and the prediction performance of the model.

Bagging methods consist of several similar independent learners aggregated to compute the final prediction by averaging the outputs of all learners. They are widely used because they reduce the variance and avoids overfitting[19].

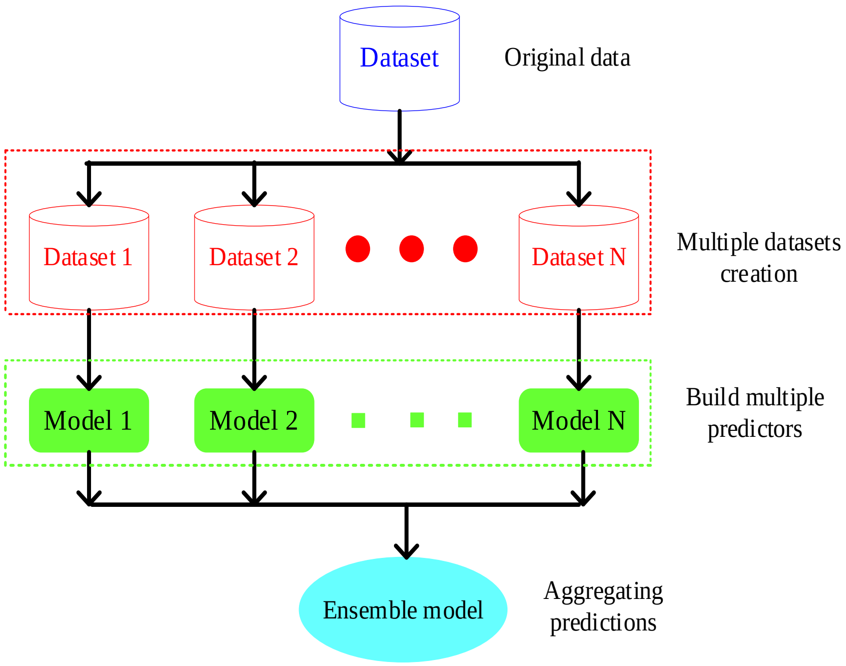

Fig. 4. Concept of bagging

Fig. 4. presents the bagging concept with the aim of minimizing prediction errors. N new datasets of the same size were firstly generated and used as input training data. By averaging all individual predictions, the final prediction is given by:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(x) \qquad (1)$$

Where each tree model f_1 is trained on bootstrap data i. Thus, the variance of prediction is decreased by 1/N compared to the variance of a standalone learner. By assuming that the error is unbiased and uncorrelated, the expected final error is defined by:

$$E_n = \frac{1}{n} E_1 \qquad (2)$$

Where E_n is the mean error while E_1 is individual model error.

- **Multi-Layer Perceptron (MLP):** Multi-Layer Perceptron MLP is a feed-forward neural networks (FFNN). It consists of sequential layers of neurons connected through synaptic weights[20]. A simple MLP consists of three connected layers arranged as follows: an input layer for receiving the input signals, a hidden layer, and an output layer that makes the final decisions about the input signals. The hidden layer performs the complex calculations and makes the MLP able of estimating any continuous function. Here, the MLP combines base learners and generates the final predictions. It is used due to its various advantages such as its simplicity and adaptive learning.

Fig. 5 presents the concept of simple MLP. The rectified linear unit (ReLU) is used as the activation function due to its characteristic of being the most efficient since it overcomes the vanishing gradient issues, allows the models to learn faster and perform better[21].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 10, October-2021

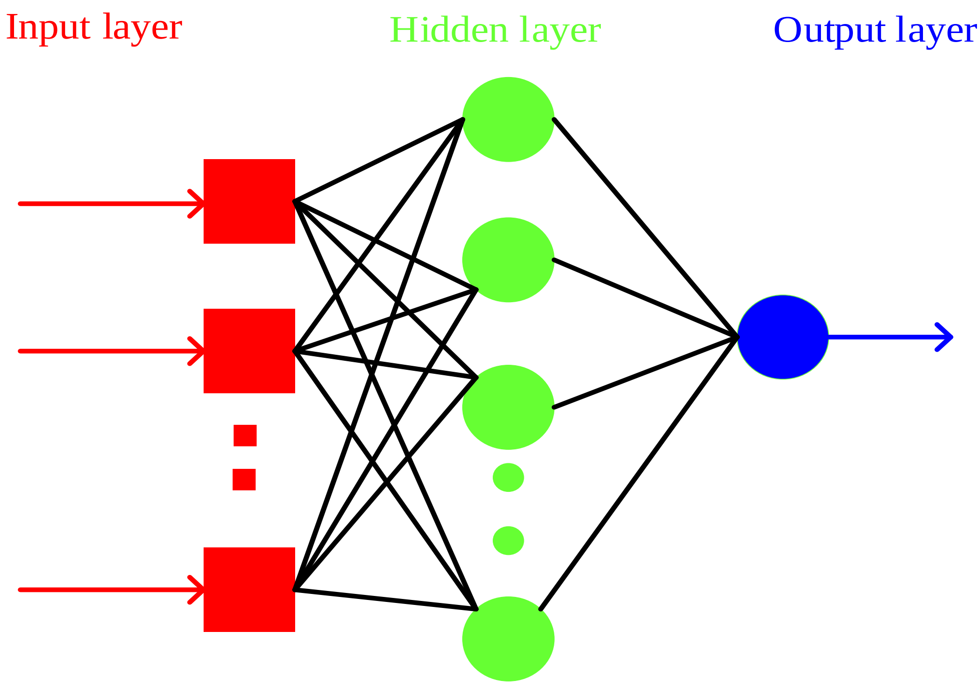

Fig. 5. Concept of simple MLP

- **Stacked adaboost-bagging regressor-MLP:** The working principle of ML ensembles leans to aggregate the outputs of numerous individual learners into a single output with the expectation of getting improved results compared to any individual learners. The combination technique of individual learners' outputs depends on problem's category to be handled. For instance, voting technique is reserved for classification while averaging technique is used for regression issues handling. Stacking based ML ensembles consist of combining the predictions of the base-learners to generate the input predictions of the next level learners and so on[22]. The base-learners are trained using the same training dataset. In this work, we briefly study the working principle of stacked AdaBoost-bagging regressor-MLP based on Fig. 6.

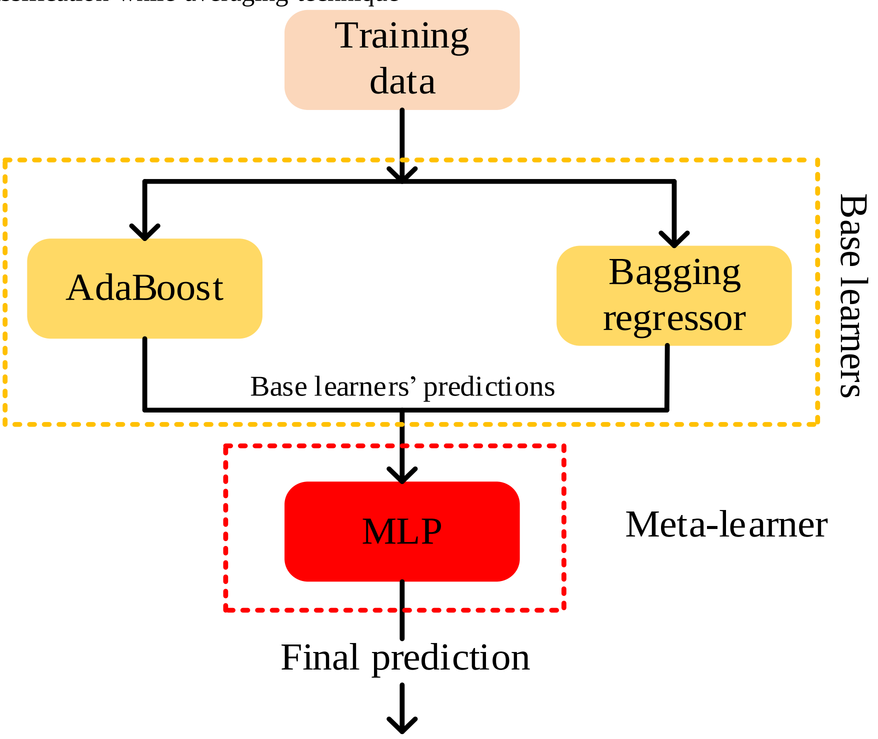

Fig. 5. Schematic diagram of stacking based ensemble

Fig. 6 presents the schematic diagram of stacked AdaBoost-Bagging regressor-MLP. All base-learners receive the same subset of data and trained in a parallel mode to make the forecast of solar irradiance. Afterwards, the aggregated of their output predictions is sent into meta-learner (MLP) using cross-validation technique. Then after, MLP analyzes the inputs and computes the final prediction.

III. RESULTS AND COMPARATIVE ANALYSIS

This section provides some insights of statistical metrics and the results' analysis of the models used in this study. Here, the described metrics are such as MAE, RMSE and R^2.

According to the results analysis of aforementioned metrics, the four machine learning models are assessed and compared. Those models are AdaBoost, Bagging regressor, MLP and its combination through stacking technique. In addition, there are various discussions, which leads to the best model.

A. *Model performance evaluation*

To analyze the forecasting performance, we compare some statistical indicators as follows:

TABLE I. A BRIEF SUMMARY OF THE STATISTICAL METRICS USED IN THE STUDY.

Metrics	Equation	Description
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	It gives us the measure of how far the predictions were from the actual output. However, they do not give us an idea of the direction of the error whether we are under predicting the data or over predicting the data.
RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	RMSE provides information on the short-term performance of the forecasting models. Its value is always positive and is desired to be close to zero[23]
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2}$	R^2 metric provides knowledge about how well a model can forecast a set of measured data. Its value varies between 0 and 1. The R^2 value approaching 1 indicates better performance[24]

Where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ expresses the mean ($\bar{y}$) of the actual values and n represents the total number of samples. While $\hat{y}_i$ and y_i are the i^{th} predicted values and the actual values respectively. The lower MAE and RMSE indicates prediction that is more accurate but in contrast, higher value of R^2 indicates better forecasting. Furthermore, for the model comparison, we also forecast the solar irradiance by using four machine-learning models. The simulation procedure was repeated to provide a high quality forecasting system. By using 10-fold cross-validation (CV) technique, the comparative study was made more authentic. Afterwards, the numerical results of statistical metrics for each k-fold cross-validation were presented in table II and table III.

B. *Results*

TABLE II. THE PERFORMANCE COMPARISON OF ADABOOST AND BAGGING REGRESSOR.

Model	AdaBoost			Bagging Regressor		
Fold number	MAE	RMSE	R^2	MAE	RMSE	R^2
1	69.414	94.176	0.912	67.348	151.572	0.774
2	80.658	105.587	0.896	74.407	156.630	0.772
3	73.780	100.851	0.908	74.360	153.887	0.786
4	69.344	93.893	0.928	88.948	180.880	0.733
5	75.757	98.592	0.892	65.929	136.966	0.792
6	67.437	91.039	0.908	71.156	143.103	0.773
7	76.526	103.209	0.902	77.130	168.839	0.738
8	74.460	99.035	0.913	79.753	158.409	0.777
9	72.641	95.936	0.921	84.283	171.559	0.748
10	77.723	104.101	0.921	76.080	168.717	0.742
Mean	73.774	98.749	0.902	75.939	159.575	0.764
SD	3.937	30.107	0.010	6.757	63.908	0.020
Time(s)	35.975			291.76		

TABLE III. THE PERFORMANCE COMPARISON OF MLP AND STACKING ENSEMBLE BASED MODEL.

Model	MLP			Stacking of AdaBoost-Bagging regressor-MLP		
Fold number	MAE	RMSE	R^2	MAE	RMSE	R^2
1	49.116	92.454	0.912	19.921	41.925	0.985
2	58.163	80.748	0.935	18.607	47.169	0.979
3	52.214	99.401	0.919	24.170	54.073	0.970
4	43.076	70.727	0.961	18.253	42.824	0.979
5	48.556	83.421	0.932	19.754	50.653	0.977
6	51.291	85.540	0.927	19.312	45.203	0.980
7	47.371	87.838	0.935	20.795	52.178	0.971
8	47.121	85.799	0.938	17.638	43.046	0.981
9	40.385	75.186	0.944	18.339	34.973	0.988
10	46.684	77.074	0.944	18.481	41.819	0.978
Mean	49.591	83.464	0.936	18.874	45.016	0.980
SD	10.050	34.304	0.022	1.343	22.587	0.004
Time(s)	2046.554			305.710		

The table II and table III summarize the numerical performance results of the models. The analysis show that stacked AdaBoost-bagging regressor-MLP generates the best prediction results in terms of the determination coefficient (R^2). Its (R^2) mean is 0.98 while AdaBoost, bagging regressor, and MLP have 0.90, 0.76 and 0.93 respectively. Moreover, stacked AdaBoost-bagging regressor-MLP presents the least mean absolute error (MAE) of 18.87 W/m^2 compared to its benchmarks. In addition, its root mean squared error of 45.01 W/m^2 confirms its high forecasting accuracy since AdaBoost, bagging regressor, and MLP generate 98.74 W/m^2, 159.57 W/m^2, and 83.46 W/m^2 respectively. Consequently, in this study, the stacked AdaBoost-bagging regressor-MLP outperformed the single models by generating the least values for both MAE and RMSE. Its high R^2 value shows also its potential for minimizing the forecasting error over the single models.

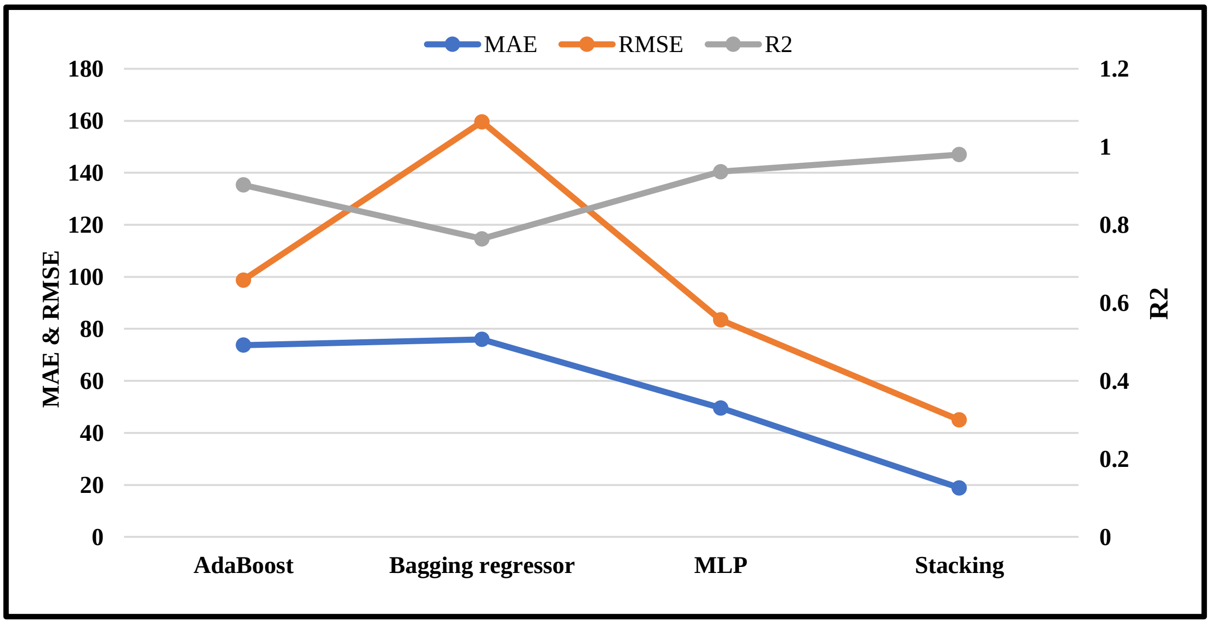

Fig. 7. Models performance comparison

By respecting to the model stability, the lowest relative standard deviation SD = 0.004 of the stacked AdaBoost-bagging regressor-MLP proves its effectiveness against random variations. The prediction results of this model is meaningful in terms of graphical assessment as shown in fig. 7. Therefore, this assessment motivate us also to apply stacking based ensemble in solar irradiance forecasting over single models.

IV. CONCLUSION

Solar power energy has gained significant importance as a clean, renewable, and alternative cheapest source of energy over the past few decades ago. Moreover, this source of energy enhances the economy of any nation because of its abundance and wide distribution. However, the efficiency of solar power generation is strongly dependent on weather conditions and other natural intermittent, uncertainty, uncontrollable parameters. Consequently, this leads to serious challenging issues during power grid management as it may imply non-stable operation and significant maintenance losses. To address these issues, accurate forecasting becomes an attractive solution to minimize the impact of uncertainty and energy costs and then enable suitable integration of photovoltaic (PV) systems in a smart grid.

In this paper, we firstly built a novel computational framework based on stacking techniques to enhance the forecasting accuracy of solar irradiation. Then, the stacking-based ensemble is compared with the single models. The AdaBoost, Bagging regressor, MLP, and its combination through stacking technique were compared. The working principle of the stacked AdaBoost-Bagging regressor-MLP model consists of combining the prediction of AdaBoost and Bagging regressor to generate final prediction using the MLP network. The dataset from the Philippines' government weather station especially located in Morong, Rizal province was used to validate the reliability of our study

We evaluate the forecasting performances via R^2, MAE, and RMSE. The stacking-based ensemble learning performs better than any single model in terms of all three statistical indicators. The analysis shows that stacked AdaBoost-bagging regressor-MLP generates the best prediction results in terms of the determination coefficient (R^2). Its (R^2) mean is 0.98 while AdaBoost, bagging regressor, and MLP have 0.90, 0.76, and 0.93 respectively. Moreover, stacked AdaBoost-bagging regressor-MLP presents the least mean absolute error (MAE) of 18.87 W/m^2 compared to its benchmarks. In addition, its RMSE of 45.01 W/m^2 confirms its high forecasting accuracy since AdaBoost, bagging regressor, and MLP generate 98.74 W/m^2, 159.57 W/m^2, and 83.46 W/m^2 respectively. Consequently, in this study, the stacked AdaBoost-bagging regressor-MLP outperformed the single models by generating the least values for both MAE and RMSE. Its high R^2 value shows also its potential for minimizing the forecasting error over the single models. The lowest relative standard deviation SD = 0.004 of the stacked AdaBoost-bagging regressor-MLP proves its effectiveness against instability

Even if the stacked AdaBoost-Bagging regressor-MLP model prooves its metrics over the single models, it has few limitations include longer running time compared to its benchmarks and its implementation process is slightly complex since it requires advanced skills and experience. However, these disadvantages have no meaningful effects compared to their various advantages. Therefore, this assessment motivates us to apply stacking-based ensemble in solar irradiance forecasting over single models. To further enhance solar irradiance forecasting, in future works, it is planned to develop ensemble ML methods that consider several independent variables especially spatiotemporal information.

REFERENCES

[1] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Conversion and Management*, vol. 156, pp. 459-497, 2018.

[2] Ü. Ağbulut, A. E. Gürel, and Y. Biçen, "Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison," *Renewable and Sustainable Energy Reviews*, vol. 135, pp. 110114, 2021.

[3] Y. Feng, W. Hao, H. Li, N. Cui, D. Gong, and L. Gao, "Machine learning models to quantify and map daily global solar radiation and photovoltaic power," *Renewable and Sustainable Energy Reviews*, vol. 118, pp. 109393, 2020.

[4] P. Kumari, and D. Toshniwal, "Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance," *Journal of Cleaner Production*, vol. 279, pp. 123285, 2021.

[5] M. Ustuner, and F. Balik Sanli, "Polarimetric target decompositions and light gradient boosting machine for crop classification: A comparative evaluation," *ISPRS International Journal of Geo-Information*, vol. 8, no. 2, pp. 97, 2019.

[6] S. Touzani, J. Granderson, and S. Fernandes, "Gradient boosting machine for modeling the energy consumption of commercial buildings," *Energy and Buildings*, vol. 158, pp. 1533-1543, 2018.

[7] G. E. Hassan, M. E. Youssef, Z. E. Mohamed, M. A. Ali, and A. A. Hanafy, "New temperature-based models for predicting global solar radiation," *Applied energy*, vol. 179, pp. 437-450, 2016.

[8] J. Fan, L. Wu, F. Zhang, H. Cai, X. Ma, and H. Bai, "Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 168-186, 2019.

[9] J. Fan, X. Wang, L. Wu, F. Zhang, H. Bai, X. Lu, and Y. Xiang, "New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: a case study in South China," *Energy Conversion and Management*, vol. 156, pp. 618-625, 2018.[1] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Conversion and Management*, vol. 156, pp. 459-497, 2018.

[2] Ü. Ağbulut, A. E. Gürel, and Y. Biçen, "Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison," *Renewable and Sustainable Energy Reviews*, vol. 135, pp. 110114, 2021.

[3] Y. Feng, W. Hao, H. Li, N. Cui, D. Gong, and L. Gao, "Machine learning models to quantify and map daily global solar radiation and photovoltaic power," *Renewable and Sustainable Energy Reviews*, vol. 118, pp. 109393, 2020.

[4] P. Kumari, and D. Toshniwal, "Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance," *Journal of Cleaner Production*, vol. 279, pp. 123285, 2021.

[5] M. Ustuner, and F. Balik Sanli, "Polarimetric target decompositions and light gradient boosting machine for crop classification: A comparative evaluation," *ISPRS International Journal of Geo-Information*, vol. 8, no. 2, pp. 97, 2019.

[6] S. Touzani, J. Granderson, and S. Fernandes, "Gradient boosting machine for modeling the energy consumption of commercial buildings," *Energy and Buildings*, vol. 158, pp. 1533-1543, 2018.

[7] G. E. Hassan, M. E. Youssef, Z. E. Mohamed, M. A. Ali, and A. A. Hanafy, "New temperature-based models for predicting global solar radiation," *Applied energy*, vol. 179, pp. 437-450, 2016.

[8] J. Fan, L. Wu, F. Zhang, H. Cai, X. Ma, and H. Bai, "Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 168-186, 2019.

[9] J. Fan, X. Wang, L. Wu, F. Zhang, H. Bai, X. Lu, and Y. Xiang, "New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: a case study in South China," *Energy Conversion and Management*, vol. 156, pp. 618-625, 2018.

[10] Y. Feng, D. Gong, Q. Zhang, S. Jiang, L. Zhao, and N. Cui, "Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation," *Energy Conversion and Management*, vol. 198, pp. 111780, 2019.

[11] H. Long, Z. Zhang, and Y. Su, "Analysis of daily solar power prediction with data-driven approaches," *Applied Energy*, vol. 126, pp. 29-37, 2014.

[12] V. H. Quej, J. Almorox, J. A. Arnaldo, and L. Saito, "ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 155, pp. 62-70, 2017.

[13] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression," *Energy*, vol. 164, pp. 465-474, 2018.

[14] D. Justin, R. S. Concepcion II, H. A. Calinao, R. R. Tobias, E. P. Dadios, and A. A. Bandala, "Solar Irradiance Prediction Based on Weather Patterns Using Bagging-Based Ensemble Learners with Principal Component Analysis." pp. 1-6.

[15] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation," 2018.

[16] R. Schapire, "Explaining adaboost. InEmpirical inference 2013 (pp. 37-52)," Springer Berlin Heidelberg.

[17] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin, "Advance and prospects of AdaBoost algorithm," *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745-758, 2013.

[18] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123-140, 1996.

[19] C. D. Sutton, "Classification and regression trees, bagging, and boosting," *Handbook of statistics*, vol. 24, pp. 303-329, 2005.

[20] A. Alfadda, S. Rahman, and M. Pipattanasomporn, "Solar irradiance forecast using aerosols measurements: A data driven approach," *Solar Energy*, vol. 170, pp. 924-939, 2018.

[21] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189-1232, 2001.

[22] C.-M. Lee, and C.-N. Ko, "Short-term load forecasting using lifting scheme and ARIMA models," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5902-5911, 2011.

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

[24] D. Shah, K. Patel, and M. Shah, "Prediction and estimation of solar radiation using artificial neural network (ANN) and fuzzy system: a comprehensive review," *International Journal of Energy and Water Resources*, pp. 1-15, 2021.