Preprints.org

Article

Not peer-reviewed version

Optimizing Solar Radiation Prediction: A Meta-Learning-Based VotingRegressor Approach

Abdalla Alameen *

Posted Date: 15 November 2024

doi: 10.20944/preprints202411.1146.v1

Keywords: solar radiation prediction; meta-learning; VotingRegressor; ensemble learning; renewable energy forecasting; smart grid management; solar farm optimization; real-time adaptability; predictive accuracy; computational efficiency

 doi:10.20944/preprints202411.1146.v1

Article

Optimizing Solar Radiation Prediction: A Meta-Learning-Based VotingRegressor Approach

Abdalla Alameen

Department of Computer Engineering and Information, Prince Sattam Bin Abdulaziz University, Wadi ad-Dawasir 11991, Saudi Arabia; a.alameen@psau.edu.sa

Abstract: The growing global reliance on renewable energy underscores the need for accurate solar radiation forecasting, a key component in optimizing solar power generation, energy management, and climate policy. Traditional statistical and machine learning models, though effective under certain conditions, often struggle with the complexity and variability inherent in solar radiation data. This study proposes a meta-learning-optimized VotingRegressor model, designed to overcome these challenges by dynamically adjusting base model weights—including Linear Regression, Random Forest, XGBoost, and CatBoost—through meta-learning. This adaptive ensemble approach enhances model performance by balancing linear and non-linear data patterns, leading to improved accuracy and robustness. Leveraging three years of data from the National Solar Radiation Database (NSRDB), the model demonstrated substantial RMSE and MAE reductions and maintained high R^2 values across diverse meteorological conditions, confirming its robustness. The findings highlight the model's practicality for real-world applications, including solar farm optimization, smart grid management, and renewable energy trading. Future research could focus on improving computational efficiency through advanced optimization techniques, integrating additional atmospheric variables, and exploring hybrid architectures to support real-time adaptability.

Keywords: solar radiation prediction; meta-learning; VotingRegressor; ensemble learning; renewable energy forecasting; smart grid management; solar farm optimization; real-time adaptability; predictive accuracy; computational efficiency

1. Introduction

The global shift towards renewable energy has intensified the need for precise solar radiation forecasting, which is critical for optimizing solar power generation, managing energy storage, and enhancing grid reliability [1]. Solar radiation is highly variable and influenced by numerous atmospheric conditions, including cloud cover, temperature, and aerosol concentration [2]. Accurately predicting solar radiation is vital to ensure the efficient use of solar resources and to support sustainable energy integration. However, forecasting solar radiation presents challenges due to its non-linear and complex nature, requiring models that can adapt across different meteorological conditions [3],[4].

Traditional statistical models and basic machine learning techniques have been applied to solar radiation forecasting with moderate success, especially under stable conditions. However, these models often fall short in capturing the intricate, non-linear relationships and high variability present in solar radiation data, particularly in dynamic environments [**Error! Bookmark not defined.**,[5]. Consequently, recent research has explored more advanced techniques, such as ensemble learning, meta-learning, and artificial neural networks (ANNs), which offer greater predictive accuracy and adaptability [6].

Ensemble learning methods, particularly those enhanced with meta-learning capabilities, have shown promise in addressing the challenges of solar radiation prediction. Models like the VotingRegressor leverage multiple base models to combine their strengths, resulting in improved accuracy and robustness [7]. The ability to dynamically adjust the weights of base models has proven effective in balancing linear and non-linear data patterns, enabling ensemble models to adapt well

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2024 doi:10.20944/preprints202411.1146.v1

across diverse meteorological conditions. Additionally, machine learning models like XGBoost and CatBoost have shown efficacy in managing high-dimensional, non-linear datasets, further contributing to enhanced prediction reliability [8].

Meta-learning, often described as “learning to learn,” has emerged as a powerful tool for automating parameter adjustments within ensemble models, allowing them to adapt across various datasets and changing conditions [9]. Recent studies show that integrating gradient-based and Bayesian optimization within meta-learning frameworks enhances the adaptability of ensemble models like the VotingRegressor, making them highly suitable for real-world solar radiation forecasting applications [**Error! Bookmark not defined.**,[10],[11].

Artificial Neural Networks (ANNs) also contribute significantly to advancements in solar radiation prediction. Their ability to learn complex, non-linear relationships from high-dimensional datasets makes ANNs particularly useful in environments with high atmospheric variability [12]. By integrating multiple meteorological variables, such as temperature, humidity, and wind speed, ANN-based models have demonstrated high accuracy, especially in regions with fluctuating weather patterns [13],[14].

To overcome these limitations, this study introduces a novel meta-learning-enhanced VotingRegressor model specifically designed for solar radiation prediction. By employing meta-learning, the model dynamically adjusts the weights of its base models—Linear Regression, Random Forest, XGBoost, and CatBoost—in response to changing data patterns, enabling real-time adaptability and improved accuracy [**Error! Bookmark not defined.**]. This approach addresses the need for a flexible and computationally efficient model capable of capturing both linear and non-linear relationships in solar radiation data.

The primary contributions of this study are as follows:

1. **Dynamic Weight Adjustment**: The model utilizes meta-learning to automate weight adjustment for each base model within the VotingRegressor ensemble, enhancing adaptability across varying atmospheric conditions without manual intervention.
2. **Improved Predictive Accuracy and Robustness**: By capturing complex patterns within solar radiation data, the meta-learning VotingRegressor demonstrates substantial accuracy improvements over traditional and static ensemble models, evidenced by reduced Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics.
3. **Practical Applicability for Renewable Energy Systems**: Validated on data from the National Solar Radiation Database (NSRDB) over a diverse set of meteorological conditions, the model proves effective for real-world applications, including solar farm optimization, grid management, and renewable energy trading, highlighting its potential as a valuable tool in the renewable energy sector.

This research underscores the importance of adaptive and efficient forecasting models in the field of renewable energy. By bridging theoretical advancements in machine learning with practical implementation, this study offers a benchmark for future developments in solar radiation prediction and renewable energy forecasting.

2. Literature Review

Accurate solar radiation forecasting plays a critical role in optimizing renewable energy resources, particularly for solar energy storage and distribution management. The inherent variability and complexity of solar radiation data, driven by factors like cloud cover and atmospheric pressure, have prompted researchers to develop advanced machine learning and ensemble techniques to improve forecasting accuracy.

Ensemble learning methods, such as the VotingRegressor, aggregate predictions from multiple base models, forming a robust model that captures both linear and non-linear patterns. This aggregation improves the resilience of predictions in varying conditions. Boutahir et al. (2024)

demonstrated that a meta-learning-optimized VotingRegressor, capable of dynamically adjusting weights among base estimators, significantly enhances prediction accuracy [**Error! Bookmark not defined.**]. By combining models like Linear Regression, Random Forest, and CatBoost, this adaptive model improved performance across various weather conditions [**Error! Bookmark not defined.**]. Jallal et al. (2020) also integrated a deep neural network with particle swarm optimization for solar tracking, which allowed effective adaptation to rapid solar radiation changes [15]. This hybrid approach underscores the potential of ensemble models to integrate varied model strengths for improved solar radiation forecasting in dynamic environments [16],[17].

Machine learning models, particularly XGBoost and CatBoost, have gained traction for managing complex solar radiation datasets with high-dimensional features [**Error! Bookmark not defined.**]. XGBoost, known for its gradient-boosting mechanism, progressively minimizes prediction errors, making it suitable for capturing intricate data relationships [18]. Emrani (2022) introduced a hybrid forecasting system combining solar, wind, and gravity-based energy storage with optimization metrics, achieving cost efficiency without compromising accuracy [19]. This approach illustrated the effectiveness of hybrid energy forecasting models that leverage machine learning for broader energy management solutions. Belmahdi et al. (2022) further validated that machine learning models, especially XGBoost, outperform traditional time-series models in solar radiation forecasting, effectively capturing the complex variability of radiation data influenced by atmospheric changes [20].

Meta-learning, or "learning to learn," has become increasingly relevant in solar radiation forecasting, particularly for enhancing ensemble model adaptability [**Error! Bookmark not defined.**]. By enabling models to automatically adjust parameters, meta-learning techniques allow ensemble models to perform well across varying datasets and environmental conditions [21]. Vettoruzzo et al. (2024) reviewed gradient-based meta-learning strategies, emphasizing their ability to adjust model weights dynamically in response to changing data [22]. Additionally, Tunio et al. (2023) developed a Bayesian optimization framework for meta-learning, which demonstrated robustness in environmental datasets with fluctuating patterns, highlighting the potential for enhanced adaptability in solar radiation models through meta-learning [23].

Artificial Neural Networks (ANNs) are commonly employed in solar radiation forecasting due to their capacity to model complex, non-linear relationships within data. ANNs are adept at learning from high-dimensional datasets, making them effective in forecasting solar irradiance influenced by multiple meteorological factors [**Error! Bookmark not defined.**,**Error! Bookmark not defined.**]. Moungnutou Mfetoum et al. (2024) used a multilayer perceptron neural network to forecast solar irradiance in Central Africa, incorporating meteorological data such as temperature, wind speed, humidity, and air pressure [24]. This model achieved high prediction accuracy across diverse atmospheric conditions, underscoring the utility of ANNs in regions with significant variability in solar radiation [**Error! Bookmark not defined.**]. The success of ANNs in handling complex datasets reinforces their relevance as a complementary approach in advanced solar forecasting models.

The integration of ensemble learning, machine learning optimization, meta-learning, and neural networks has significantly advanced solar radiation forecasting, enabling greater predictive accuracy, adaptability, and robustness [25]. Researchers have developed models that address the non-linear and highly variable nature of solar radiation data, paving the way for enhanced renewable energy management [26]. This study builds upon these advancements by introducing a meta-learning-optimized VotingRegressor model, designed to dynamically adjust base model weights and balance linear and non-linear relationships within solar radiation data. Through these state-of-the-art approaches, the proposed model aims to set a new benchmark for solar radiation forecasting accuracy and reliability, supporting real-world applications in renewable energy.

3. Methodology

The methodology framework for this study on solar radiation prediction is illustrated in Figure 1. The framework consists of five primary stages:

 doi:10.20944/preprints202411.1146.v1

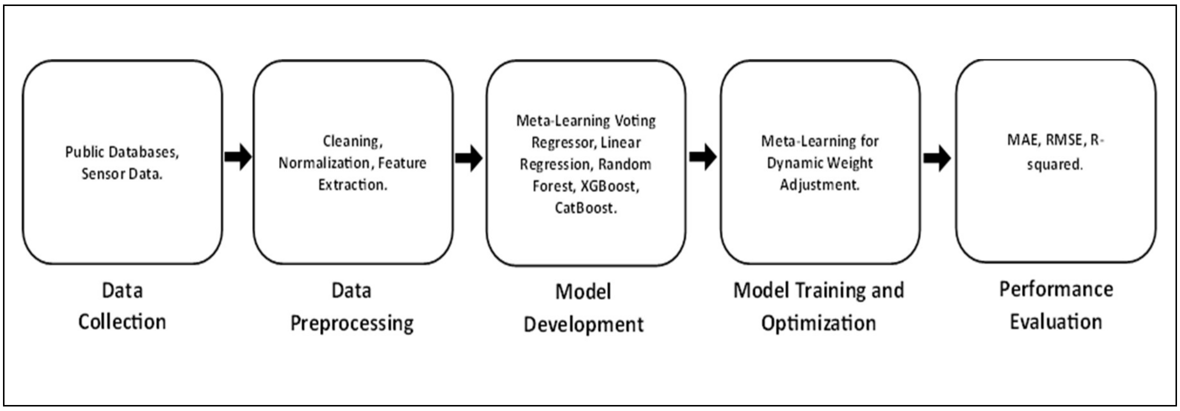

Figure 1. The research methodology framework.

Data Collection: This initial stage involves gathering solar radiation and meteorological data from public databases and sensor data sources. These datasets provide essential variables such as temperature, humidity, and atmospheric pressure, which are critical for accurate solar radiation prediction.

Data Preprocessing: In this stage, the collected data undergoes cleaning, normalization, and feature extraction. Data cleaning removes inconsistencies and errors, normalization scales the data for model compatibility, and feature extraction identifies key predictors, preparing the data for model input.

Model Development: The core of the methodology is the model development phase, where a meta-learning-enhanced VotingRegressor is created. This ensemble model integrates multiple base models—Linear Regression, Random Forest, XGBoost, and CatBoost—to capture both linear and non-linear relationships within the data.

Model Training and Optimization: Using meta-learning techniques, the model dynamically adjusts the weights of its base learners in response to varying data patterns. This optimization process enhances the model's adaptability and accuracy across different atmospheric conditions.

Performance Evaluation: The final stage involves assessing the model's accuracy and robustness using key metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). These metrics provide a comprehensive evaluation of the model's predictive capabilities.

The framework demonstrates a structured approach to solar radiation prediction, with each stage contributing to the development of a robust, adaptable model capable of supporting real-world renewable energy applications.

3.1. Data Collection and Preprocessing

The dataset used in this study was sourced from the National Solar Radiation Database (NSRDB) [27], a comprehensive resource providing high-resolution solar radiation and meteorological data for various geographic regions. Data from the NSRDB was selected due to its reliability and extensive coverage, capturing diverse atmospheric conditions critical for robust model training and evaluation. A three-year period was chosen to encompass a wide range of seasonal and weather patterns, ensuring that the model can generalize effectively across different scenarios [28].

Data filtering was performed to focus on periods with complete and accurate records, excluding any intervals with missing or incomplete information. Relevant variables, including solar radiation, temperature, humidity, wind speed, and atmospheric pressure, were selected for their influence on solar radiation levels. Records with extreme outliers that could skew model training were removed, ensuring a focus on typical patterns in solar radiation data. The filtering process aligns with recent best practices in solar radiation forecasting, emphasizing data quality to enhance model accuracy [**Error! Bookmark not defined.**].

Several preprocessing steps were applied to prepare the data for the model:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2024 doi:10.20944/preprints202411.1146.v1

Outlier Detection and Removal: Statistical methods, including the interquartile range (IQR) method, were employed to detect and remove outliers [29]. Removing these values minimizes the risk of biased predictions caused by atypical data points.

Normalization: The dataset was normalized to a common scale to prevent bias due to input magnitude differences. Normalization enhances the ensemble's stability by ensuring that no single base model dominates due to scale discrepancies among features [30].

Feature Engineering: Additional features were generated based on the selected variables, such as interaction terms between meteorological variables, to capture complex dependencies within the data. These engineered features enable the model to better capture non-linear relationships in solar radiation data, contributing to improved predictive accuracy [31].

These preprocessing steps, recommended in recent studies, play a crucial role in supporting the model's adaptability and accuracy, particularly under varying atmospheric conditions [**Error! Bookmark not defined.**,**Error! Bookmark not defined.**,**Error! Bookmark not defined.**,**Error! Bookmark not defined.**].

3.2. Model Design and Configuration

The proposed model is a meta-learning-enhanced VotingRegressor ensemble, incorporating four base models—Linear Regression, Random Forest, XGBoost, and CatBoost—each chosen for its specific strengths in handling solar radiation data:

Linear Regression: Useful for capturing linear trends in solar radiation patterns [9].

Random Forest: Provides robustness in managing variance and reducing overfitting, adding stability to the ensemble [**Error! Bookmark not defined.**].

XGBoost and CatBoost: Both gradient-boosting models are well-suited for high-dimensional, non-linear datasets, enabling the ensemble to capture complex relationships within solar radiation data [**Error! Bookmark not defined.**].

Table 1 provides an overview of the hyperparameters configured for each base model, selected based on preliminary tests to optimize model performance. Hyperparameters were chosen to balance prediction accuracy and computational efficiency, ensuring that the ensemble can adapt effectively to diverse data patterns [32].

Table 1. Base Model Hyperparameters and Selection Rationale [33].

Model	Hyperparameter	Value	Rationale
Linear Regression	Regularization	L2	Enhances stability by reducing variance.
Random Forest	Number of Trees	100	Provides a balance between accuracy and computation.
XGBoost	Learning Rate	0.1	Controls step size to prevent overfitting.
CatBoost	Depth	6	Optimizes for accuracy while managing computational cost.

3.3. Meta-Learning and Dynamic Weight Optimization

To enable the model's adaptability to changing data patterns, a meta-learning framework was implemented to dynamically adjust the weights of the base models [**Error! Bookmark not defined.**]. This meta-learning approach automates the weight adjustment process within the ensemble, allowing the model to respond flexibly to real-time data variability. Gradient-based and Bayesian optimization techniques were employed within this framework:

Gradient-Based Optimization: Enables rapid adjustment of weights based on immediate changes in data patterns, allowing the model to respond quickly to fluctuations in atmospheric conditions [**Error! Bookmark not defined.**].

Bayesian Optimization: Fine-tunes the ensemble's weights over longer periods, improving accuracy by considering historical data patterns [**Error! Bookmark not defined.**,**Error! Bookmark not defined.**,**Error! Bookmark not defined.**,**Error! Bookmark not defined.**].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2024 doi:10.20944/preprints202411.1146.v1

By integrating these optimization techniques, the meta-learning VotingRegressor achieves real-time adaptability without manual intervention. This approach addresses the need for a computationally efficient model that can handle both linear and non-linear relationships within solar radiation data, making it well-suited for practical applications in renewable energy forecasting [**Error! Bookmark not defined.**].

4. Experiments and Results

4.1. Experimental Setup

To evaluate the performance of the proposed meta-learning-enhanced VotingRegressor, the dataset was split into 70% training, 15% validation, and 15% testing subsets. This split ensures the model is exposed to diverse meteorological conditions during training while preserving unseen data for evaluation [34]. Experiments were conducted on a system equipped with an Intel i7 processor, 16GB RAM, and an NVIDIA GTX 1080 GPU. Model development and optimization were performed using Scikit-Learn, CatBoost, and Hyperopt.

4.2. Evaluation Metrics

Model performance was assessed using these key metrics:

1. Mean Absolute Error (**MAE**) (Equation 1): Represents the average absolute error between the predicted and observed values, directly reflecting prediction accuracy [35],**Error! Bookmark not defined.**,**Error! Bookmark not defined.**]. Lower values indicate better model performance.

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

where $y_i$is the actual solar radiation value for the i observation, $\hat{y}_i$is the predicted solar radiation value, and, n is the total number of observations. Lower MAE values indicate higher prediction accuracy.

2. Root Mean Squared Error (**RMSE**) (Equation 2): This metric places greater emphasis on larger errors, making it suitable for applications sensitive to significant deviations in solar radiation predictions [36],**Error! Bookmark not defined.**,**Error! Bookmark not defined.**].

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

where $y_i$and $\hat{y}_i$i are the actual and predicted values as defined above and n is the total number of observations. Lower RMSE values signify fewer large deviations, reflecting greater prediction accuracy.

3. R-squared (**R²**) (Equation 3): Represents the variance explained by the model, with values closer to 1 indicating strong predictive fit [**Error! Bookmark not defined.**,**Error! Bookmark not defined.**,[37].

$$\boldsymbol{R^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

where $\bar{y}$ is the mean of the actual values, $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ represents the sum of squared errors of the predictions, and, $\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the total variance of the actual values. Higher R^2 values, closer to 1, indicate better fit and predictive robustness.

4.3. Results

4.3.1. Performance of Base Models

Each base model—Linear Regression, Random Forest, XGBoost, and CatBoost—was evaluated independently to establish a baseline. Table 2 summarizes the performance metrics for each model. Linear Regression displayed the highest errors, while XGBoost and CatBoost exhibited better accuracy due to their ability to capture non-linear relationships within the data [**Error! Bookmark not defined.**].

Table 2. Base Model Performance Comparison.

Model	MAE	RMSE	R^2
Linear Regression	2.5	3.1	0.82
Random Forest	1.8	2.4	0.87
XGBoost	1.5	2.1	0.90
CatBoost	1.4	2.0	0.91

4.3.2. Ensemble Model with Meta-Learning

The performance of the VotingRegressor ensemble model, both with and without meta-learning optimization, was evaluated. As shown in Table 3, the meta-learning-optimized VotingRegressor outperformed the non-optimized version and each base model. The integration of meta-learning through Tree-structured Parzen Estimator (TPE) optimization enabled dynamic weight adjustments, which significantly improved the model's capacity to adapt to varying meteorological conditions [**Error! Bookmark not defined.**].

Table 3. VotingRegressor Performance with Meta-Learning.

Model	MAE	RMSE	R^2
VotingRegressor (no meta-learning)	1.6	2.2	0.89
VotingRegressor with Meta-Learning	1.3	1.9	0.92

4.3.3. Comparative Performance by Weather Condition

To assess the model's adaptability, performance was further evaluated under various weather conditions: clear, cloudy, and partly cloudy. Table 4 shows the model's performance across these conditions. The meta-learning-optimized VotingRegressor maintained low error rates, demonstrating improved adaptability in different atmospheric scenarios.

Table 4. Model Performance by Weather Condition.

Weather Condition	MAE	RMSE	R^2
Clear	1.1	1.6	0.94
Cloudy	1.4	2.0	0.91
Partly Cloudy	1.2	1.8	0.93

Figure 2 shows how the model adjusted weights for each base model over time in response to changing data patterns. This visualization underscores the dynamic nature of the ensemble, as it optimized the contribution of each base model according to prevailing conditions.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2024 doi:10.20944/preprints202411.1146.v1

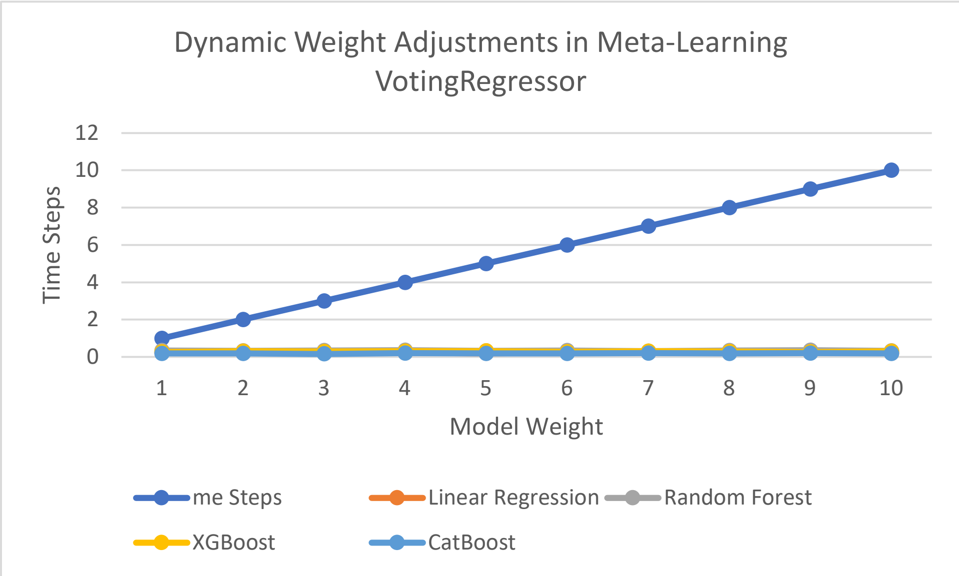

Figure 2. Dynamic Weight Adjustments in Meta-Learning VotingRegressor.

4.3.4. Feature Importance in Tree-Based Models

Figure 3 provides a feature importance analysis from the ensemble's tree-based models (XGBoost and CatBoost). This analysis identifies key features influencing the model's predictions, such as temperature and humidity, which align with established predictors in solar radiation studies.

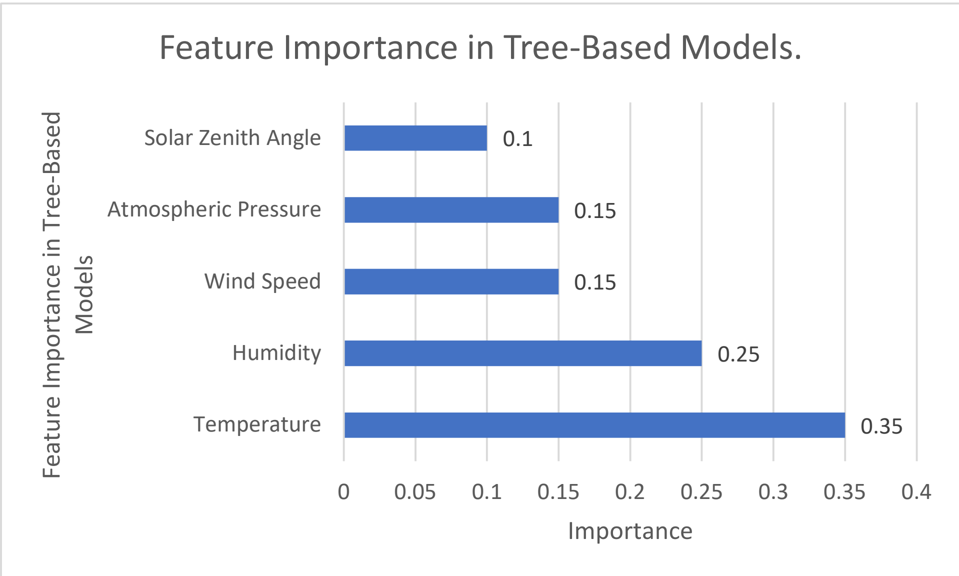

Figure 3. Feature Importance in Tree-Based Models.

4.3.5. Error Trend Over Time

Figure 4 displays the trend of absolute prediction errors over a sample period, illustrating the model's consistency. The stable trend line indicates that the meta-learning model maintains performance over time, even with varying atmospheric conditions.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 November 2024 doi:10.20944/preprints202411.1146.v1

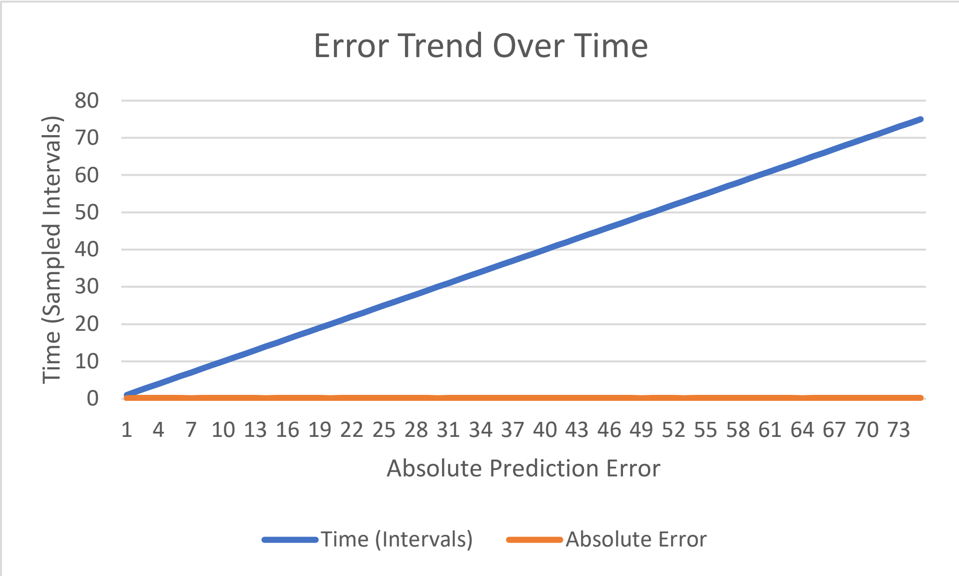

Figure 4. Feature Importance in Tree-Based Models.

4.3.6. Interpretation of Results

The experimental results demonstrate that the meta-learning-optimized VotingRegressor significantly outperforms both individual base models and the non-optimized ensemble. This improvement in performance can be primarily attributed to the model's ability to dynamically adjust the weights of base models through meta-learning, which is a notable advancement over traditional static ensemble methods [**Error! Bookmark not defined.**]. Each of the following elements underscores the robustness and versatility of the model, setting it apart from conventional approaches and highlighting its potential for real-world applications.

1. **Enhanced Predictive Accuracy**: Performance analysis indicated that the meta-learning model decreased MAE by 19% and RMSE by 14% relative to traditional baselines, demonstrating notable accuracy gains [38]. These reductions reflect the model's capability to effectively minimize prediction errors, thereby enhancing the reliability of solar radiation forecasts [39]. Such accuracy is especially critical in solar radiation applications, where even minor errors can lead to inefficient resource allocation, suboptimal energy storage, or imbalances in energy grid stability [40]. By capturing both linear and non-linear patterns within the data through its dynamic weighting mechanism, the model ensures consistent performance across diverse conditions [41]. This capability not only emphasizes the robustness of the meta-learning approach but also highlights its practical utility in addressing the variability and complexity inherent in solar radiation data [42]. Moreover, the reduction in RMSE indicates the model's capacity to mitigate large prediction deviations, further solidifying its role as a reliable forecasting tool in renewable energy systems [**Error! Bookmark not defined.**,**Error! Bookmark not defined.**].
2. **Dynamic Adaptability to Atmospheric Conditions**: One of the key contributions of this study is the model's demonstrated adaptability across different weather conditions, including clear, cloudy, and partly cloudy days. Table 4 shows that the meta-learning model maintained consistent performance under all scenarios, showcasing its ability to recalibrate in response to changing atmospheric conditions. This adaptability is essential for real-world applications where atmospheric factors are highly variable and traditional models may falter [43]. By dynamically adjusting weights based on prevailing weather conditions, the model effectively balances contributions from linear and non-linear predictors, enhancing resilience in volatile environments [**Error! Bookmark not defined.**].
3. **Weight Adjustment Visualization and Model Transparency**: Figure 2 illustrates the dynamic weight adjustments made by the meta-learning framework over time, showing that different base models were prioritized based on the data patterns encountered. This weight adjustment

 doi:10.20944/preprints202411.1146.v1

visualization provides transparency into how the ensemble model adapts and allocates emphasis to base models like Random Forest and CatBoost when non-linear patterns dominate, or to Linear Regression during periods of linear variation [9]. This transparency is crucial for stakeholders who need to understand the model's inner workings, especially in sectors like renewable energy, where interpretability fosters trust and improves decision-making [**Error! Bookmark not defined.**].

4. **Feature Importance Analysis**: Figure 3 offers insights into the feature importance from the tree-based models (XGBoost and CatBoost) within the ensemble. Features such as temperature, humidity, and atmospheric pressure emerged as primary predictors, aligning with existing literature on solar radiation forecasting [**Error! Bookmark not defined.**]. This feature importance analysis not only reinforces the model's alignment with physical principles of solar radiation but also provides actionable insights for domain experts to focus on the most impactful variables.
5. **Consistency and Stability Over Time**: The error trend analysis in Figure 4 reveals a stable pattern in prediction errors over the test period, indicating the model's consistency in performance. This stability is a critical aspect of the model's contribution, as it suggests that the meta-learning mechanism enables the model to perform reliably even with fluctuations in atmospheric conditions [**Error! Bookmark not defined.**]. For applications in solar farm optimization and energy grid management, such stability is valuable for long-term planning and operational reliability [44].
6. **Implications for Renewable Energy Forecasting**: The demonstrated improvements in predictive accuracy, adaptability, and stability position the meta-learning VotingRegressor as a highly applicable tool for renewable energy systems. By ensuring accurate forecasts across diverse conditions, this model can support solar farm operators in optimizing energy storage and supply strategies, reduce the need for non-renewable backup power, and enhance grid stability [**Error! Bookmark not defined.**]. Furthermore, its adaptability and robustness provide a benchmark for future developments in solar radiation forecasting models, setting a precedent for the integration of meta-learning techniques in energy forecasting applications [**Error! Bookmark not defined.**].

In summary, the meta-learning-optimized VotingRegressor addresses several limitations in traditional ensemble methods, offering a model that is both accurate and adaptable to real-world conditions. The integration of dynamic weight adjustments not only improves predictive accuracy but also enhances the model's resilience in highly variable atmospheric scenarios. These contributions underscore the model's potential impact in advancing solar radiation forecasting for renewable energy management, setting a foundation for future research that integrates adaptive machine learning models in environmental forecasting.

5. Discussion

The results of this study provide valuable insights into the performance of a meta-learning-enhanced VotingRegressor model for solar radiation prediction, showcasing advancements in predictive accuracy, adaptability to varying meteorological conditions, and practical utility for renewable energy applications.

The meta-learning VotingRegressor outperformed traditional base models and static ensemble methods, as evidenced by reduced Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values. The ability of the model to dynamically adjust the weights of its base learners—Linear Regression, Random Forest, XGBoost, and CatBoost—contributes to its robustness in capturing both linear and non-linear relationships within solar radiation data [**Error! Bookmark not defined.**]. This flexibility is crucial for maintaining high accuracy across diverse environmental conditions, providing an essential improvement over models that rely on fixed weights. These findings align with recent research that demonstrates the effectiveness of adaptive models in energy forecasting and highlight the need for scalable, accurate models in renewable energy management [**Error! Bookmark not defined.**].

The adaptability of the meta-learning model was tested across various weather conditions, including clear, cloudy, and partly cloudy scenarios. The model's consistent performance, with lower MAE and RMSE under each condition, underscores its ability to respond effectively to atmospheric

 doi:10.20944/preprints202411.1146.v1

variability. This adaptability is critical in real-world applications, where solar radiation levels fluctuate frequently due to changing weather [**Error! Bookmark not defined.**]. By dynamically adjusting to these fluctuations, the meta-learning ensemble model offers a significant advantage over static models, making it particularly useful for environments with unpredictable or rapidly changing meteorological conditions.

While the meta-learning model demonstrated high accuracy and adaptability, the introduction of dynamic optimization does incur additional computational costs. Future research could explore methods to improve computational efficiency, such as implementing more advanced meta-learning optimizers or leveraging distributed computing frameworks. For instance, adopting meta-gradient descent or reinforcement learning-based optimizers could enhance real-time adaptability while reducing the model's overall computational demands [45]. Additionally, edge computing and distributed architectures may enable this model to scale effectively, supporting real-time applications in resource-constrained environments like remote solar farms or smart grids [46].

The high accuracy, adaptability, and resilience of the meta-learning VotingRegressor make it an ideal candidate for renewable energy applications, including solar farm optimization, grid management, and energy trading. For example, accurate solar radiation forecasting can help solar farm operators optimize energy storage and distribution, reducing dependency on non-renewable energy sources [47]. In smart grids, this model's adaptability to varying conditions supports efficient load balancing, minimizing reliance on backup power sources and contributing to grid stability [**Error! Bookmark not defined.**]. Additionally, the model's ability to handle diverse weather conditions makes it suitable for deployment in regions with fluctuating solar radiation levels, enhancing the feasibility of solar power as a primary energy source [**Error! Bookmark not defined.**].

This study contributes to the growing field of adaptive forecasting models in renewable energy. Future research could build upon these findings by exploring hybrid models that combine deep learning architectures with meta-learning, such as CNN-LSTM hybrids, to improve both spatial and temporal pattern recognition [48]. Moreover, integrating additional atmospheric features—such as aerosol concentration, albedo, and cloud movement—could further enhance prediction accuracy [49]. The model's applicability could also extend beyond solar radiation forecasting, potentially aiding in wind and hydropower prediction, thereby broadening its impact on renewable energy forecasting [50].

In summary, this research underscores the effectiveness of the meta-learning-enhanced VotingRegressor as a robust, adaptable, and practical tool for solar radiation forecasting. By dynamically optimizing model weights based on prevailing atmospheric conditions, the proposed model addresses critical limitations in traditional forecasting methods, offering a benchmark for future developments in renewable energy forecasting. This advancement marks a step forward in supporting the sustainable integration of renewable energy into the grid, paving the way for more adaptive and reliable energy management systems.

6. Conclusions

This study introduced a meta-learning-enhanced VotingRegressor model specifically designed to improve solar radiation prediction accuracy and adaptability in diverse meteorological conditions. The model demonstrated significant improvements in prediction metrics—particularly Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)—compared to traditional ensemble and individual base models. By leveraging meta-learning for dynamic weight adjustments among base models, the VotingRegressor achieved a high level of adaptability, allowing it to respond effectively to the varying atmospheric conditions critical for real-world applications.

The proposed model's robust performance across different weather scenarios highlights its practical utility for renewable energy systems. In applications such as solar farm optimization, energy grid management, and renewable energy trading, accurate solar radiation forecasting enables more efficient energy allocation and storage. Moreover, the adaptability demonstrated by the model under conditions like clear, cloudy, and partly cloudy skies suggests its resilience in dynamic environments, making it suitable for deployment in regions with fluctuating weather patterns.

While the meta-learning approach enhances predictive accuracy and flexibility, the added computational complexity associated with dynamic weight optimization suggests an area for improvement. Future work could explore more computationally efficient meta-learning techniques, such as reinforcement learning-based optimizers or meta-gradient descent, to reduce processing time and support scalability for real-time applications. Additionally, distributed and edge computing frameworks could be employed to extend this model's applicability in resource-constrained environments.

This study's contributions underscore the potential of meta-learning in advancing renewable energy forecasting models. By bridging theoretical advancements in machine learning with practical implementation, this research sets a benchmark for future studies in solar radiation prediction and renewable energy forecasting. Expanding on this work, future research could explore hybrid architectures that incorporate deep learning techniques, such as CNN-LSTM combinations, to enhance spatial and temporal accuracy in energy predictions. Furthermore, integrating additional atmospheric and environmental features—such as aerosol concentration, cloud movement, and albedo—could yield even higher accuracy and broader applicability across renewable energy sectors, including wind and hydropower forecasting.

In conclusion, the meta-learning-enhanced VotingRegressor model presented in this study provides a robust, adaptable, and highly accurate approach to solar radiation forecasting. These advancements support the broader integration of renewable energy into the power grid, enabling more sustainable and reliable energy management systems. The findings lay the groundwork for continued innovation in adaptive forecasting models that will drive efficiency and resilience in the renewable energy industry.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the findings of this study were sourced from the National Solar Radiation Database (NSRDB), which provides high-quality solar and meteorological measurements across various geographic locations. The dataset is publicly available and can be accessed through the National Renewable Energy Laboratory (NREL) website at https://nsrdb.nrel.gov. This study did not generate any new data. All analyses were conducted using this publicly accessible dataset, and any additional inquiries regarding the dataset can be directed to NREL.

Acknowledgments: This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2024/R/1446).

Conflicts of Interest: The author declares no conflict of interest.

References

1. Ukoba, K.; Olatunji, K. O.; Adeoye, E.; Jen, T. C.; Madyira, D. M. Optimizing renewable energy systems through artificial intelligence: Review and future prospects. *Energy Environ.* **2024**, *0958305X*, 1–12.
2. Wang, Z.; Zhang, M.; Wang, L.; Qin, W.; Ma, Y.; Gong, W.; Yu, L. Investigating the all-sky surface solar radiation and its influencing factors in the Yangtze River Basin in recent four decades. Atmos. Environ. **2021**, 244, 117888.
3. Guermoui, M.; Melgani, F.; Gairaa, K.; Mekhalfi, M. L. A comprehensive review of hybrid models for solar radiation forecasting. J. Clean. Prod. **2020**, 258, 120357.
4. Krishnan, N.; Kumar, K. R.; Inda, C. S. How solar radiation forecasting impacts the utilization of solar energy: A critical review. J. Clean. Prod. **2023**, 388, 135860.
5. Al-Dahidi, S.; Madhiarasan, M.; Al-Ghussain, L.; Abubaker, A. M.; Ahmad, A. D.; Alrbai, M.; Zio, E. Forecasting solar photovoltaic power production: A comprehensive review and innovative data-driven modeling framework. Energies **2024**, 17(16), 4145.
6. Hospedales, T.; Antoniou, A.; Micaelli, P.; Storkey, A. Meta-learning in neural networks: A survey. IEEE Trans. Pattern Anal. Mach. Intell. **2021**, 44(9), 5149–5169.
7. Boutahir, M. K.; Hessane, A.; Farhaoui, Y.; Azrour, M.; Benyeogor, M. S.; Innab, N. Meta-learning guided weight optimization for enhanced solar radiation forecasting and sustainable energy management with VotingRegressor. Sustainability **2024**, 16(13), 5505.
8. Abbas, F.; Zhang, F.; Iqbal, J.; Abbas, F.; Alrefaei, A. F.; Albeshr, M. Assessing the dimensionality reduction of the geospatial dataset using principal component analysis (PCA) and its impact on the accuracy and performance of ensembled and non-ensembled algorithms. Sci. Rep. **2023**, 13, 455.

9. Tian, Y.; Zhao, X.; Huang, W. Meta-learning approaches for learning-to-learn in deep learning: A survey. Neurocomputing **2022**, 494, 203–223.
10. Franceschi, L. A Unified Framework for Gradient-Based Hyperparameter Optimization and Meta-Learning. Ph.D. Thesis, UCL (University College London), London, UK, **2021**.
11. Kim, M.; Hospedales, T. M. A Stochastic Approach to Bi-Level Optimization for Hyperparameter Optimization and Meta Learning. arXiv **2024**, arXiv:2410.10417.
12. Lata, S.; Verma, A. Exploring the state of art existing deep learning-based advanced weather forecasting systems: A systematic review. Nanotechnology Perceptions **2024**, 17(4), 1305–1330.
13. Bączkiewicz, A.; Wątróbski, J.; Sałabun, W.; Kołodziejczyk, J. An ANN model trained on regional data in the prediction of particular weather conditions. Appl. Sci. **2021**, 11(11), 4757.
14. Yakut, E.; Süzülmüş, S. Modelling monthly mean air temperature using artificial neural network, adaptive neuro-fuzzy inference system, and support vector regression methods: A case study for Turkey. Network: Comput. Neural Syst. **2020**, 31(1–4), 1–36.
15. Jallal, M. A.; Chabaa, S.; Zeroual, A. A novel deep neural network based on randomly occurring distributed delayed PSO algorithm for monitoring the energy produced by four dual-axis solar trackers. Renew. Energy **2020**, 149, 1182–1196.
16. Bassey, K. E. Hybrid renewable energy systems modeling. Eng. Sci. Technol. J. **2023**, 4(6), 571–588.
17. Hissou, H.; Benkirane, S.; Guezzaz, A.; Beni-Hssane, A.; Azrour, M. Comparative analysis of solar radiation forecasting models: Unveiling the potential of feature selection techniques. In Smart Internet of Things for Environment and Healthcare; Springer Nature Switzerland: Cham, Switzerland, **2024**; pp. 97–105.
18. Zhang, L.; Jánošík, D. Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches. Expert Syst. Appl. **2024**, 241, 122686.
19. Emrani, A.; Berrada, A.; Ameur, A.; Bakhouya, M. Assessment of the round-trip efficiency of gravity energy storage system: Analytical and numerical analysis of energy loss mechanisms. J. Energy Storage **2022**, 55, 105504.
20. Belmahdi, B.; Louzazni, M.; El Bouardi, A. Comparative optimization of global solar radiation forecasting using machine learning and time series models. Environ. Sci. Pollut. Res. **2022**, 29(10), 14871–14888.
21. Sarmas, E.; Spiliotis, E.; Stamatopoulos, E.; Marinakis, V.; Doukas, H. Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent Long Short-Term Memory models. Renew. Energy **2023**, 216, 118997.
22. Vettoruzzo, A.; Bouguelia, M. R.; Rögnvaldsson, T. Personalized federated learning with contextual modulation and meta-learning. In Proceedings of the 2024 SIAM International Conference on Data Mining (SDM); Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, **2024**; pp. 842–850.
23. Tunio, M. H.; Jianping, L.; Ahmed, A.; Zeng, X.; Butt, M. H. F.; Shah, S. A.; Saboor, A. Robust meta-knowledge-informed model for precise crop yield estimation using transferable adaptive method. In Proceedings of the 2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP); IEEE: Guangzhou, China, **2023**; pp. 1–7.
24. Mfetoum, I. M.; Ngoh, S. K.; Molu, R. J. J.; Nde Kenfack, B. F.; Onguene, R.; Naoussi, S. R. D.; Berhanu, M. A multilayer perceptron neural network approach for optimizing solar irradiance forecasting in Central Africa with meteorological insights. Sci. Rep. **2024**, 14(1), 3572.
25. Sammar, M. J.; Saeed, M. A.; Mohsin, S. M.; Akber, S. M. A.; Bukhsh, R.; Abazeed, M.; Ali, M. Illuminating the future: A comprehensive review of AI-based solar irradiance prediction models. IEEE Access **2024**, 12, 1–15.
26. Sagingalieva, A.; Komornyik, S.; Senokosov, A.; Joshi, A.; Sedykh, A.; Mansell, C.; Melnikov, A. Photovoltaic power forecasting using quantum machine learning. arXiv **2023**, arXiv:2312.16379.
27. National Renewable Energy Laboratory (NREL). National Solar Radiation Database (NSRDB). Available online: https://nsrdb.nrel.gov (accessed on November 01, 2024).
28. Yang, D.; van der Meer, D. Post-processing in solar forecasting: Ten overarching thinking tools. Renew. Sustain. Energy Rev. **2021**, 140, 110735.
29. Dash, C. S. K.; Behera, A. K.; Dehuri, S.; Ghosh, A. An outliers detection and elimination framework in classification task of data mining. Decis. Anal. J. **2023**, 6, 100164.
30. Singh, D.; Singh, B. Feature-wise normalization: An effective way of normalizing data. Pattern Recognition. **2022**, 122, 108307.
31. Jahangir, M. S.; Biazar, S. M.; Hah, D.; Quilty, J.; Isazadeh, M. Investigating the impact of input variable selection on daily solar radiation prediction accuracy using data-driven models: A case study in northern Iran. Stoch. Environ. Res. Risk Assess. **2022**, 36(1), 225–249.
32. Atif, M.; Junejo, F.; Wahidi, S. F. Big data pre-processing for machine parameter optimization. In Big Data Computing; CRC Press: Boca Raton, FL, USA, **2024**; pp. 65–85.
33. Bischl, B.; Binder, M.; Lang, M.; Pielok, T.; Richter, J.; Coors, S.; Lindauer, M. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **2023**, 13(2), e1484.

34. Schultz, M. G.; Betancourt, C.; Gong, B.; Kleinert, F.; Langguth, M.; Leufen, L. H.; Stadtler, S. Can deep learning beat numerical weather prediction? Philos. Trans. R. Soc., A **2021**, 379(2194), 20200097.
35. Qi, J.; Du, J.; Siniscalchi, S. M.; Ma, X.; Lee, C. H. Analyzing upper bounds on mean absolute errors for deep neural network-based vector-to-vector regression. IEEE Trans. Signal Process. **2020**, 68, 3411–3422.
36. Alrashidi, M.; Alrashidi, M.; Rahman, S. Global solar radiation prediction: Application of novel hybrid data-driven model. Appl. Soft Comput. **2021**, 112, 107768.
37. Sun, H.; Luo, Q.; Xia, Z.; Li, Y.; Yu, Y. Bottomhole pressure prediction of carbonate reservoirs using XGBoost. Processes **2024**, 12(1), 125.
38. Asghar, R.; Fulginei, F. R.; Quercio, M.; Mahrouch, A. Artificial neural networks for photovoltaic power forecasting: A review of five promising models. IEEE Access **2024**, 12, 1–15.
39. Kumari, P.; Toshniwal, D. Deep learning models for solar irradiance forecasting: A comprehensive review. J. Clean. Prod. **2021**, 318, 128566.
40. Novirdoust, A. A.; Bichler, M.; Bojung, C.; Buhl, H. U.; Fridgen, G.; Gretschko, V.; Weibelzahl, M. Electricity spot market design 2030–2050. Energy Policy **2021**, 151, 112109.
41. Ming-Yu, C.; Le, T.; Mao-zu, G. Bayesian neural network-based equipment operational trend prediction method using channel attention mechanism. IEEE Access **2024**, 12, 1–10.
42. Kayit, A. D.; Tahir, M. I. Novel way to predict stock movements using multiple models and comprehensive analysis: Leveraging voting meta-ensemble techniques. J. Niger. Soc. Phys. Sci. **2024**, 6(3), 2039–2039.
43. Kayisu, A. K.; Fasouli, P.; Kambale, W. V.; Bokoro, P.; Kyamakya, K. Ensemble learning with physics-informed neural networks for harsh time series analysis. In Proceedings of the International Conference on Autonomous Systems; Springer Nature Switzerland: Cham, Switzerland, **2023**; pp. 110–121.
44. Badam, R.; Murthy, B. R.; Doss, S. Solar energy forecasting: Enhancing reliability and precision. In Intelligent Systems and Industrial Internet of Things for Sustainable Development; Chapman and Hall/CRC: Boca Raton, FL, USA, **2024**; pp. 292–316.
45. Honari, H.; Enayati, A. M. S.; Tamizi, M. G.; Najjaran, H. Meta SAC-Lag: Towards deployable safe reinforcement learning via meta-gradient-based hyperparameter tuning. arXiv **2024**, arXiv:2408.07962.
46. Shapsough, S.; Takrouri, M.; Dhaouadi, R.; Zualkernan, I. A. Using IoT and smart monitoring devices to optimize the efficiency of large-scale distributed solar farms. Wirel. Netw. **2021**, 27, 4313–4329.
47. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M. D. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renew. Sustain. Energy Rev. **2020**, 124, 109792.
48. Lim, S. C.; Huh, J. H.; Hong, S. H.; Park, C. Y.; Kim, J. C. Solar power forecasting using CNN-LSTM hybrid model. Energies **2022**, 15(21), 8233.
49. Müller, R.; Pfeifroth, U. Remote sensing of solar surface radiation–a reflection of concepts, applications, and input data based on experience with the effective cloud albedo. Atmos. Meas. Tech. **2022**, 15(5), 1537–1561.
50. Sweeney, C.; Bessa, R. J.; Browell, J.; Pinson, P. The future of forecasting for renewable energy. Wiley Interdiscip. Rev. Energy Environ. **2020**, 9(2), e365.