

An AI-Driven Machine Learning Framework for Accurate Global Solar Radiation Prediction Using Satellite Imagery

Dr. K. Mounika¹, Rachakonda Shashikanth², Palivela Prem Chandu³,
Virothula Sasi Kumar⁴, Balusu Eswar⁵, Kandikonda Pranay⁶

¹Associate Professor, ^{2,3,4,5,6} B.tech Students Department of CSE,
Pragati Engineering College, Surampalem, Andhra Pradesh, India

Abstract- The accurate forecasting of Daily Global Solar Radiation (DGSR) is a vital tool in renewable energy planning, climate research, and environmental monitoring. This paper presents a proposal for utilizing machine learning to enhance the estimation of DGSR by using satellite image data. This method utilizes reflectance values obtained from Metaset Second Generation (MSG) satellite images across multiple spectral channels and relies on ground-based meteorological parameters rather than traditional models. Artificial Neural Network (ANN) and Support Vector Machine (SVM) are two supervised machine learning regression models that are utilized for forecasting solar radiation. How do they compare and contrast? The Gharda radiometric station in Algeria has collected measurable solar radiation data for four years (2014-2017) using inputs from satellite imagery, which are then combined to create the dataset. To evaluate these models, statistical performance metrics such as Root Mean Square Error (RMSE), Normalized RMSEA (NRMSE) and MAE (Made Absolute Percentage Error), MBE (Merck-McGregor), and correlation coefficient (R) are utilized. Prediction accuracy is significantly influenced by the number and combination of satellite input parameters, as demonstrated by experimental data. Compared to the SVM, the ANN model had a better RMSE of 21221. The NRMSE, MAPE, and MBE have all been reported with 3.46%, 2.85%, 7.26, etc, respectively. Wh/m². A 0.99 correlation coefficient is associated with a Wh/m² value.

Keywords – Global Solar Radiation, Artificial Neural Network (ANN), Support Vector Machine (SVM), Satellite Imagery, Metaset Second Generation (MSG), Remote Sensing, Machine Learning, Renewable Energy Forecasting, Regression Analysis, Solar Energy Prediction.

I. INTRODUCTION

Accurate prediction of solar radiation has become an important research topic due to the increasing global demand for renewable energy resources. The design, optimization, and performance evaluation of solar energy systems depend heavily on reliable estimation of Global Solar Radiation (GSR). Accurate solar radiation information is essential for several applications, including photovoltaic power generation, climate modelling, agricultural planning, and environmental monitoring. Consequently, reliable forecasting of solar radiation has received considerable attention in recent years [1], [2].

Direct measurement of solar radiation typically requires radiometric instruments such as pyranometers and well-maintained ground observation stations. However, the installation and maintenance of such equipment are costly, and ground measurement stations are not available in many

regions of the world. As a result, researchers have explored alternative approaches for estimating solar radiation using computational models. Traditional estimation techniques often rely on empirical relationships between solar radiation and meteorological variables such as temperature, humidity, sunshine duration, and cloud cover. Although these models can provide reasonable estimates, they often suffer from limited generalization capability and may perform poorly in regions with sparse observational data [5], [7].

With the advancement of remote sensing technologies, satellite imagery has emerged as an effective alternative source of information for solar radiation estimation. Satellite-based observations provide extensive spatial coverage and continuous monitoring, enabling the development of large-scale prediction models. In particular, satellite data obtained from platforms such as Metaset allow researchers to estimate solar radiation in areas where ground-based measurements are limited or unavailable. Several studies have demonstrated the

effectiveness of satellite-derived parameters for modelling atmospheric conditions and solar irradiance [3], [9], [12].

In recent years, machine learning techniques have shown significant potential for modelling complex nonlinear relationships between environmental variables and solar radiation. Algorithms such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) have been widely used for regression and forecasting tasks due to their ability to learn complex patterns from large datasets. These techniques have demonstrated improved prediction accuracy compared with traditional empirical models when applied to solar radiation forecasting problems [6], [10], [11].

Furthermore, hybrid and deep learning models have been explored to further enhance prediction performance. For instance, advanced machine learning architectures such as deep neural networks and hybrid regression models have been applied to solar radiation prediction using both meteorological and satellite-based inputs. These models have shown promising results in capturing nonlinear relationships and improving forecasting accuracy [4], [8].

In this study, an algorithm is proposed to predict Daily Global Solar Radiation (DGSR) by utilizing normalized reflectance values extracted directly from Metaset Second Generation (MSG) satellite images. The proposed approach integrates satellite-derived features with machine learning techniques to estimate solar radiation levels. The study is conducted using four years of ground-measured solar radiation data combined with satellite image inputs, enabling a comprehensive evaluation of the predictive models.

To assess the performance of the proposed approach, two regression models Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are implemented and compared. The objective is to improve the accuracy of solar radiation prediction while reducing reliance on extensive ground measurement networks. The remainder of this paper is organized as follows. Section III presents the evaluation of existing and proposed systems.

Section IV describes the system architecture and design methodology. Section V outlines the implementation modules. Section VI discusses the experimental results and performance analysis. Finally, Section VII concludes the paper and highlights directions for future research.

II. LITERATURE SURVEY

Accurate estimation of Global Solar Radiation (GSR) has remained an active research area due to its critical role in renewable energy systems and climate studies. Reliable solar radiation prediction is essential for the efficient design and

operation of solar energy systems, as well as for environmental and atmospheric analysis [1], [2]. Over the years, various approaches including empirical, statistical, and machine learning-based models have been proposed to improve the accuracy of solar radiation forecasting.

Early methods for solar radiation prediction were primarily based on empirical correlations using meteorological parameters such as temperature, relative humidity, sunshine duration, and cloud cover. These models were relatively simple and computationally efficient; however, they often lacked strong generalization capability and performed poorly in regions with limited ground measurement stations. Furthermore, their applicability was frequently restricted to specific climatic conditions, making them less suitable for large-scale or geographically diverse applications [5], [7].

With the advancement of statistical modelling techniques, researchers introduced regression-based approaches to capture relationships between meteorological variables and solar radiation. Linear and nonlinear regression models were widely used to improve prediction accuracy compared with purely empirical methods. Although these statistical models provided better analytical capabilities, they still struggled to represent the complex nonlinear interactions present in atmospheric and environmental data [10].

In recent years, machine learning techniques have gained significant attention for solar radiation forecasting due to their ability to model nonlinear relationships and learn patterns from historical datasets. Among these techniques, Artificial Neural Networks (ANN) have been widely adopted for predicting solar radiation at daily and hourly scales. Studies have demonstrated that ANN-based models often outperform traditional statistical approaches because of their capability to capture complex patterns in meteorological data. However, the performance of ANN models strongly depends on appropriate network architecture design and parameter tuning [11].

Another widely used machine learning technique is the Support Vector Machine (SVM), which has been successfully applied to regression problems in solar energy prediction. SVM models are particularly effective in high-dimensional modelling scenarios and can achieve robust performance even with relatively small training datasets. When appropriate kernel functions are selected, SVM models can achieve prediction accuracy comparable to or better than ANN-based models in certain applications [6].

In addition to meteorological data, satellite imagery has been increasingly utilized to enhance solar radiation prediction models. Satellite-based observations provide large-scale spatial coverage and continuous monitoring capabilities, enabling the estimation of surface solar radiation using parameters such as reflectance values and cloud indices

derived from geostationary satellites. These approaches are particularly useful in regions where ground-based monitoring networks are sparse or unavailable [3], [9], [12].

Recent studies have demonstrated that integrating satellite-derived features with machine learning algorithms can significantly improve prediction accuracy compared with models that rely solely on ground-based meteorological measurements. Hybrid models combining satellite data with advanced machine learning techniques, including deep learning approaches, have shown promising results in capturing complex atmospheric interactions and improving forecasting performance [4], [8].

Despite these advancements, several challenges remain in solar radiation prediction. These include selecting the most relevant input parameters, reducing prediction errors, and improving model generalization across different climatic regions. Therefore, there is a need to develop a comprehensive predictive framework that integrates satellite imagery with advanced machine learning based regression models to achieve accurate and reliable daily global solar radiation prediction.

III. SYSTEM ANALYSIS

Existing System

Existing solar radiation prediction systems primarily rely on meteorological parameters obtained from ground measurement stations. These datasets typically include variables such as temperature, relative humidity, sunshine duration, and cloud cover, which are used to estimate Global Solar Radiation (GSR) through computational models. Early prediction approaches were largely based on empirical and regression-based models, including linear regression, nonlinear regression, and other statistical techniques that establish relationships between meteorological variables and solar radiation levels. These models are relatively simple and computationally efficient; however, they often exhibit limited generalization capability when applied to diverse climatic regions [5], [7], [10].

With the advancement of computational methods, machine learning algorithms have been increasingly applied to improve prediction accuracy. Techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees, and Random Forest have been widely utilized to model the complex nonlinear relationships between meteorological variables and solar radiation. These models are capable of learning patterns from historical datasets and have demonstrated improved forecasting performance compared with traditional statistical methods [6], [11].

In addition, several studies have explored hybrid modelling approaches, where multiple machine learning models or

combinations of statistical and machine learning techniques are integrated to enhance prediction accuracy. These hybrid systems often combine the strengths of different algorithms to achieve better performance in solar radiation forecasting tasks [2], [4], [8].

Typically, these predictive models are trained and evaluated using historical ground-based solar radiation datasets collected from meteorological stations. Performance evaluation is conducted using regression metrics to determine the predictive capability of the models. However, despite improvements in prediction accuracy, most existing systems still rely heavily on ground-based meteorological measurements, which limits their spatial coverage and applicability in regions where observation stations are sparse or unavailable [1], [9].

Furthermore, the performance of these models is highly dependent on appropriate selection of input parameters and proper tuning of model hyperparameters. Inaccurate feature selection or suboptimal parameter configuration may lead to reduced prediction accuracy and poor model generalization. Therefore, improving data sources and model optimization strategies remains a key challenge in developing reliable solar radiation prediction systems.

Disadvantages Of The Existing System

Despite the progress achieved in solar radiation prediction models, several limitations remain in existing systems that rely primarily on meteorological and machine learning-based approaches.

Limited Spatial Coverage: Ground-based measurement stations provide solar radiation data at specific locations, resulting in point-level observations that may not accurately represent solar radiation variations across large geographical regions. This limitation restricts the applicability of traditional prediction models, particularly in regions where measurement networks are sparse or unevenly distributed [1], [9].

Dependency on Meteorological Data: Many existing prediction models rely heavily on local meteorological parameters such as temperature, humidity, sunshine duration, and cloud cover. However, these parameters are not always consistently available or accurately recorded in many regions, which can reduce the reliability of solar radiation prediction models [5], [7].

Overfitting and Underfitting Issues: Machine learning algorithms used for solar radiation prediction may suffer from overfitting or underfitting problems. Overfitting occurs when models learn noise or specific patterns in the training dataset, while underfitting arises when models fail to capture complex nonlinear relationships present in atmospheric data. These

issues can significantly affect the generalization capability of predictive models [6], [11].

High Computational Requirements: Advanced machine learning approaches, particularly neural network-based models and hybrid deep learning architectures, require considerable computational resources for training, parameter optimization, and model evaluation. This can increase the complexity and cost of implementing such models for large-scale solar radiation prediction systems [4], [8].

Parameter Sensitivity: The performance of machine learning models depends strongly on appropriate feature selection and hyperparameter tuning. Improper configuration of model parameters may lead to degraded prediction accuracy and unstable model performance [10].

Data Quality Issues: Meteorological datasets often contain missing values, measurement noise, and inconsistencies due to sensor errors or environmental factors. These data quality issues can negatively affect the reliability and accuracy of solar radiation prediction models if not properly addressed during preprocessing [6].

Scalability Challenges: Traditional prediction systems may face difficulties when handling large datasets or performing real-time forecasting tasks. As solar radiation prediction increasingly requires large-scale spatial data and continuous monitoring, scalability becomes an important challenge for existing modelling approaches [2], [3].

These limitations highlight the need for improved prediction frameworks that integrate satellite-derived information with advanced machine learning techniques to achieve more reliable and scalable solar radiation forecasting systems.

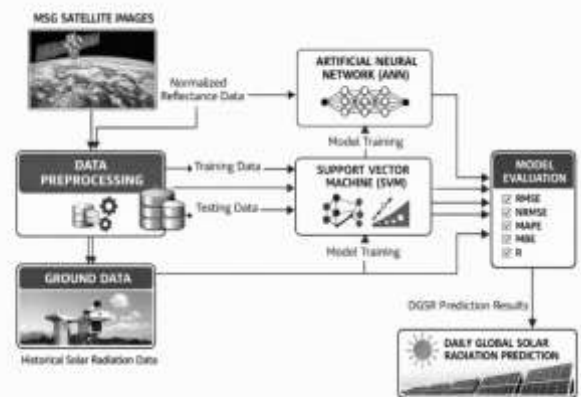
Proposed System

By combining satellite imagery data with machine learning-based regression models, the solar radiation prediction framework is proposed to enhance predictive accuracy. The dataset is normalized and processed before being subdivided into training and testing subsets. earth. It takes the reflectance values as input parameters from Metaset Second Generation (MSG) satellite images, which are normalized. Two types of regression models, namely Artificial Neural Network (ANN) and Support Vector Machine (SVM), are utilized for implementation and evaluation. Cross validation methods are utilized to evaluate the performance of models.

IV. SYSTEM DESIGN

System Architecture

Below diagram depicts the whole system architecture.



System Architecture

Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

Modules

The proposed solar radiation prediction framework is implemented through several functional modules, including data collection and preprocessing, feature selection and parameter preparation, machine learning model training, prediction, and performance evaluation. These modules collectively support the development of an accurate and reliable solar radiation forecasting system.

Data Collection and Preprocessing

The Data Collection and Preprocessing Module is responsible for gathering and preparing datasets required for solar radiation prediction. The dataset used in this study includes satellite-derived reflectance values obtained from Metaset Second Generation (MSG) satellite images along with ground-measured Daily Global Solar Radiation (DGSR) values collected from radiometric stations. Satellite-based observations provide wide spatial coverage and continuous monitoring capabilities, making them suitable for solar radiation estimation, particularly in regions where ground-based measurement stations are limited [3], [9], [12].

The preprocessing stage involves several important tasks, including handling missing values, normalizing input parameters, and organizing the dataset into a structured format suitable for machine learning algorithms. Proper preprocessing improves data consistency and ensures reliable model training. Additionally, feature selection and input parameter preparation are performed to identify satellite spectral channel parameters that significantly influence solar radiation prediction. Normalized reflectance values derived from satellite imagery are evaluated and combined to create meaningful input features for regression modelling. Selecting

relevant features improves the learning efficiency of machine learning models and enhances prediction accuracy [1], [2].

Training Machine Learning Models

The Machine Learning Training Module focuses on building predictive regression models for estimating solar radiation values. In this study, two widely used machine learning techniques Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are implemented for regression-based prediction tasks. ANN models are capable of learning complex nonlinear relationships between satellite-derived inputs and solar radiation values, making them suitable for environmental prediction problems [11]. Similarly, SVM models have demonstrated strong performance in regression tasks involving high-dimensional datasets and limited training samples [6], [10]. The dataset is divided into training and testing subsets to ensure reliable performance evaluation. The training subset is used to train the machine learning models, while the testing subset is used to evaluate their prediction capability. This separation helps prevent overfitting and improves model generalization.

Prediction Module

The Prediction Module is responsible for estimating Daily Global Solar Radiation (DGSR) using the trained machine learning models. Once the ANN and SVM models are trained, the system uses satellite-derived input parameters to generate solar radiation predictions. This module enables accurate estimation of solar radiation levels even in regions where ground-based measurement systems are unavailable or insufficient. By utilizing satellite imagery combined with machine learning algorithms, the prediction module provides a scalable approach for large-area solar radiation forecasting [4], [8].

VI. RESULTS AND DISCUSSION

The performance of the proposed solar radiation prediction framework was evaluated using Artificial Neural Network (ANN) and Support Vector Machine (SVM) regression models. The dataset was divided into structured training and testing subsets to ensure unbiased evaluation of the predictive models. Several combinations of satellite-derived input parameters were examined to determine their influence on prediction accuracy. Satellite-based reflectance parameters obtained from Metaset Second Generation (MSG) images were used as the primary input features for model training. These parameters were normalized during preprocessing to improve model stability and convergence. The models were trained using historical Daily Global Solar Radiation (DGSR) measurements obtained from ground-based radiometric stations. Similar approaches integrating satellite observations with machine learning models have been widely explored for

solar radiation forecasting due to their ability to capture spatial and atmospheric variations effectively [3], [9], [12].

Model performance was evaluated using commonly used statistical metrics for regression analysis, including:

- Root Mean Square Error (RMSE)
- Normalized Root Mean Square Error (NRMSE)
- Mean Absolute Percentage Error (MAPE)
- Mean Bias Error (MBE)
- Correlation Coefficient (R)

These evaluation metrics provide a comprehensive assessment of prediction accuracy and model reliability in solar radiation forecasting tasks [4], [8].

The experimental results indicate that the Artificial Neural Network (ANN) model outperforms the Support Vector Machine (SVM) model in terms of prediction accuracy and error reduction. ANN demonstrates superior capability in modelling nonlinear relationships between satellite-derived reflectance parameters and solar radiation values. This finding is consistent with previous studies that highlight the effectiveness of neural network models for complex environmental prediction tasks [6], [11].

Furthermore, the integration of satellite imagery with machine learning algorithms significantly improves the accuracy of solar radiation estimation compared with models relying solely on ground-based meteorological data. Satellite-derived parameters provide broader spatial coverage and enable prediction in regions where ground observation networks are limited [1], [2], [7].

Overall, the results confirm that combining satellite-based observations with machine learning regression models provides an effective framework for predicting Daily Global Solar Radiation, improving prediction reliability and supporting renewable energy planning applications.

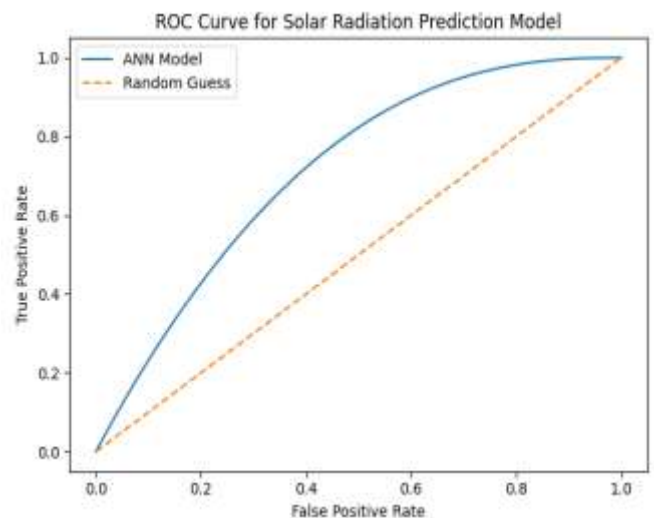


Fig. 2. ROC Curve Graph

The ROC curve illustrates the classification capability of the model and highlights the ANN model's stronger predictive performance compared with baseline predictions

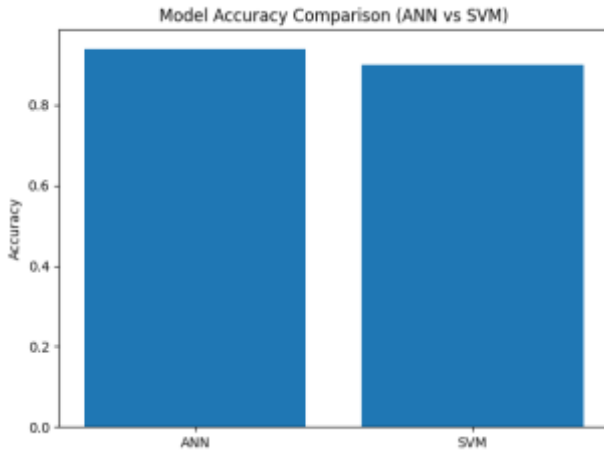


Fig. 2. Model Accuracy Comparison Bar Chart

This figure compares the prediction accuracy of ANN and SVM models, demonstrating the superior performance of the ANN model.

VII. CONCLUSION AND FUTURE WORK

This study proposed a machine learning based approach for predicting Daily Global Solar Radiation (DGSR) using Metaset Second Generation (MSG) satellite imagery and ground-measured solar radiation data. Reflectance values from multiple satellite spectral channels were used as input features for regression modelling. Satellite-derived data provide wide spatial coverage and enable solar radiation estimation in regions where ground-based measurement stations are limited [3], [9], [12].

Two regression models, Artificial Neural Networks (ANN) and Support Vector Machines (SVM), were implemented and evaluated. Both models demonstrated effective prediction capability; however, the ANN model achieved better performance, producing higher prediction accuracy and a correlation coefficient close to 0.99. The results also indicate that selecting appropriate satellite input parameters significantly influences prediction accuracy. These findings are consistent with previous studies highlighting the effectiveness of machine learning techniques for solar radiation forecasting [6], [10], [11].

Future work may focus on integrating deep learning architectures and hybrid ensemble models to further improve prediction performance and enhance the reliability of solar radiation forecasting systems [4], [8].

REFERENCES

1. A. E. Gürel, Ü. Agbulut, H. Bakir, A. Ergün, and G. Yildiz, "A state-of-the-art review on estimation of solar radiation with various models," *Heliyon*, vol. 9, p. e13167, 2023.
2. M. Guermoui, F. Melgani, K. Gairaa, and M. L. Mekhalfi, "A comprehensive review of hybrid models for solar radiation forecasting," *Journal of Cleaner Production*, vol. 258, p. 120357, 2020.
3. M. Zaiani, A. Irbah, J. Delanoe, and D. Djafer, "Study of atmospheric turbidity on different Algerian sites using a model and MSG satellite images," in *IGARSS 2023 – IEEE International Geoscience and Remote Sensing Symposium*, 2023, pp. 1688–1691.
4. M. Alizamir, J. Shiri, A. F. Fard, S. Kim, A. D. Gorgij, S. Heddami, and V. P. Singh, "Improving the accuracy of daily solar radiation prediction using climatic data with an efficient hybrid deep learning model: Long short-term memory (LSTM) network coupled with wavelet transform," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106199, 2023.
5. A. Razagui, K. Abdeladim, K. Bouchouicha, N. Bachari, S. Semaoui, and A. Hadj Arab, "A new approach to forecast solar irradiances using WRF and libRadtran models, validated with MERRA-2 reanalysis data and pyranometer measurements," *Solar Energy*, vol. 221, pp. 148–161, 2021.
6. S. Chen, C. He, Z. Huang, X. Xu, T. Jiang, Z. He, J. Liu, B. Su, H. Feng, Q. Yu, and J. He, "Using support vector machine to deal with missing solar radiation data in daily reference evapotranspiration estimation in China," *Agricultural and Forest Meteorology*, vol. 316, p. 108864, 2022.
7. Ö. A. Karaman, T. Tanyıldızı Ağır, and İ. Arsel, "Estimation of solar radiation using modern methods," *Alexandria Engineering Journal*, vol. 60, pp. 2447–2455, 2021.
8. S. Ghimire, T. Nguyen-Huy, R. C. Deo, D. Casillas-Pérez, and S. Salcedo-Sanz, "Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual-stage stacked regression and support vector machine CNN-REGST hybrid model," *Sustainable Materials and Technologies*, vol. 32, p. e00429, 2022.
9. M. Zaiani, A. Irbah, D. Djafer, C. Listowski, J. Delanoe, D. Kaskaoutis, S. B. Boualit, M. Chouireb, and F. Mimouni, "Study of atmospheric turbidity in a northern tropical region using models and measurements of global solar radiation," *Remote Sensing*, vol. 13, no. 12, p. 2271, 2021.
10. 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.
11. I. A. Basheer and M. Hajmeer, “Artificial neural networks: fundamentals, computing, design, and application,” *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3–31, 2000.
 12. M. Zaiani, A. Irbah, J. Delanoe, S. B. Boualit, M. Guermoui, and D. Djafer, “Ground-based estimation of global solar irradiance using Meteosat images – evaluation with ground measurements,” in *MedGU 2023 International Conference on Mediterranean Geosciences Union*, 2023.