PREDICTING SOLAR GENERATION FROM WEATHERFORECASTS BY USING MACHINE LEARNING

[1]D. DIVYA SRI, [2]G.LAVANYA

[1,2]Student of EEE department, GMRIT, Rajam, Srikakulam, Andhra Pradesh

ABSTRACT:

One of the primary goals of smart grid initiatives is to significantly boost the share of renewable energy sources in the grid's energy mix. Yet, a major challenge in incorporating renewables is their unpredictable and intermittent power generation. Accurate prediction of future renewable energy production is pivotal because it enables the grid to efficiently manage generators in response to the variable nature of renewable energy generation. While developing advanced prediction models for large-scale solar farms is feasible through manual efforts, extending such models to encompass the extensive distributed generation taking place at millions of homes across the grid poses a formidable challenge. To address this challenge, the research paper under consideration explores the automated development of site-specific prediction models for solar power generation. This is achieved by leveraging machine learning techniques and National Weather Service (NWS) weather forecasts. The daily average solar irradiance stands as a pivotal factor in determining the appropriate sizing for solar power generation installations. This metric, which represents the typical solar energy received at a specific location, serves as a crucial element for estimating the electricity output from solar panels. An accurate prediction of solar irradiation is instrumental in calculating the system's size, evaluating its return on investment (ROI), and assessing the system load requirements. The study delves into different regression methods, including linear least squares and support vector machines with multiple kernel functions, for constructing prediction models. The performance of these models is evaluated by comparing them to historical NWS forecasts and solar intensity measurements collected from a weather station over the course of nearly a year. The study's findings demonstrate that prediction models based on support vector machines (SVM) and incorporating seven distinct weather forecast metrics exhibit a 27% improvement in accuracy when compared to existing forecast-based models, specifically for the site in question.

Key Words: Grid, National Weather Service, Regression, Least Squares, SVM

INTRODUCTION:

Solar forecasting is a highly effective method for reducing solar power fluctuations due to atmospheric changes. This approach commonly uses ground or remote data to estimate current weather conditions and forecast future global solar irradiance or power, spanning from seconds to years ahead solar forecasting methods can be divided into three main categories: physical models, data-driven models, and hybrid models [1]. These categories depend on whether they rely on scientific principles or purely data analysis. Both ground- based and remote sensing data can be used in these models [1]. Satellite data is a valuable source of information and can be used in two ways: it can be processed by physical models like numerical weather prediction, or it can be directly used in statistical models to predict solar conditions. Ground-based sensors like pyranometers and pyrheliometer [1] is also useful, providing local solar irradiance measurements. Solar forecasting is particularly sensitive to cloud movements, which makes it unique

among energy forecasting tasks. Therefore, technologies like satellites and sky cameras play crucial roles in improving solar forecasting accuracy [7].

ROLE OF COMPUTER IN SOLAR FORECASTING:

The variation in solar irradiance throughout the day is mainly due to changes in cloud cover. To improve predictions of solar power fluctuations, computer vision techniques that use cloud observations have been developed. While cloud modeling involves some randomness, the predictable aspect of cloud movement can be leveraged to predict how clouds will change based on past observations. This ability to foresee cloud movements provides a significant advantage compared to forecasting methods that solely rely on local weather data, which are less reliable for longer-term solar predictions. This forecasting approach is essential for delivering accurate solar forecasts at various spatial and temporal scales.

The fluctuations in solar irradiance during the day are primarily due to changes in cloud cover. In order to better predict the resulting variations in solar power, advanced computer vision techniques have been developed based on cloud observations. While cloud modeling involves some randomness, it's possible to harness the partly predictable nature of cloud movements to project how cloud cover will change by analyzing past observations. This ability to forecast cloud movements offers a significant advantage over methods that rely solely on local weather data, which tend to be less reliable for making predictions beyond the short-term patterns seen in solar time series. As a result, this forecasting approach

is of utmost importance in delivering genuinely valuable solar predictions across different spatial and temporal scales.

Traditional computer vision methods used for analyzing sky or satellite images employ common techniques like cloud segmentation, localization, property estimation, tracking, and motion modeling. In sky images, cloud identification is often done by setting thresholds based on pixel statistics. Satellite image cloud segmentation can be improved by considering the effective cloud albedo or cloud index, which depends on various albedo factors. This continuous variable indicates how much light clouds allow through, and it can also be estimated using a pyrheliometer and a clear-sky model in sky camera applications. Cloud location is determined using multiple sky cameras in stereovision mode, followed by approximating cloud velocity vectors using techniques like block matching or optical flow. This helps predict the future impact of cloud dynamics on solar power generation, either deterministically or probabilistically. Future solar irradiance forecasts at a local or regional level can be derived from cloud properties estimations.

Recent advancements in deep learning and the availability of large datasets have led to data-driven approaches for this computer vision task. The typical approach involves training neural networks to extract spatiotemporal features from past sequences of sky or satellite images to predict future solar variability. Like traditional computer vision methods, these techniques can provide both local predictions [4] (e.g., photovoltaic power output) and regional

forecasts (e.g., cloud-index maps). Despite achieving high performance according to common metrics like root mean square error (RMSE) [4] or mean absolute error (MAE), deep learning models still have challenges, including difficulty in predicting critical events in a timely manner or accurately correlating images (especially satellite images) [1] with their corresponding solar values. Some models address these limitations by incorporating auxiliary data, such as ground-based measurements [1] or solar elevation information.

ROLE OF MACHINE LEARNING IN SOLAR PREDICTION:

Machine learning plays a crucial role in predicting solar generation by leveraging data-driven techniques to improve the accuracy and reliability of solar energy forecasts. Here are some key aspects of how machine learning is used in predicting solar generation [7]:

- **Feature Extraction by data analysis:**
  Machine learning models can process and analyze vast amounts of historical data related to solar generation, including solar irradiance [2], weather conditions, time of day, and geographical location. These models can identify important features and patterns that influence solar power production.

- **Solar Irradiance Prediction:**
  Solar irradiance is a critical factor in solar generation. Machine learning models can be trained to predict solar irradiance levels based on a variety of inputdata sources, such as satellite images, sky camera data [4], and historical weather information.

- **Time-Series Analysis:**
  Machine learning models can handle time-series data effectively, which is essential for making hourly, daily, or seasonal predictions of solar energy production. They can capture patterns and trends in the data to provide accurate forecasts.

- **Hybrid Models:**
  Some forecasting models combine physical modeling and machine learning techniques to leverage the strengths of both approaches. For example, machine learning can enhance the accuracy of physical models like numerical weather prediction (NWP) models.

- **Spatial and Temporal Prediction:**
  Machine learning can provide spatial and temporal predictions of solar generation. This is important for understanding how solar energy production varies across different locations and times.

- **Real-Time Updates:**
  Machine learning models can continuously update predictions in real time as new data becomes available, allowing for more accurate forecasts as conditions change throughout the day.

- **Predictive Maintenance:**
  Machine learning can also be used to predict maintenance needs for solar panels and associated equipment, ensuring that solar generation systems operate efficiently.

- **Forecasting Uncertainty:**
  Machine learning models can

provide estimates of uncertainty in solar generation predictions, helping energy grid operators and utilities make informed decisions and plan for contingencies.

- **Integrated Energy Management Systems:**

  Machine learning models can be integrated with energy management systems to optimize energy production and consumption in real-time, ensuring efficient use of solar power.

Machine learning has the potential to significantly improve the efficiency and effectiveness of solar energy generation and distribution, making it a valuable tool for the renewable energy sector [3]. These models can adapt to changing conditions and provide more accurate forecasts, which are essential for managing solar resources effectively and integrating them into the broader energy.

RELATED WORK:

Solar energy forecasts can be classified in a number of ways. The persistence or smart persistence model, which forecasts future electricity generation based on historical data.

The most basic way (2-3), over a short amount of time hours). whereby other forecasting approaches can be measured. In a forecast is usually completed in two steps. An NWP is intended for a certain time and location. first and foremost. The resulting NWP [1] is then used to Forecasting algorithms are used to forecast power generation. It is possible to use a physical model, a statistical model, or both. Approach, sometimes known as a machine learning methodology. For ML [5] techniques for prediction are compared to the Smart ML

models are used in the persistence (SP) technique.

Renewable energy sources are increasingly being incorporated into electric grids alongside non-renewable sources, which presents significant challenges due to their unpredictable and variable nature. To address these challenges, sophisticated computational solutions for predicting energy generation are crucial. We employ various data analysis techniques, such ashistorical load data preparation and the analysis of load time series. This is important because electricity consumption is closely linked to the use of other energy sources like natural gas and oil [1]. We have also examined and compared the trends in power consumption between renewable and non-renewable sources.

One innovative approach we've developed is a machine learning-based hybrid methodthat combines a multilayer perceptron (MLP) with support vector regression (SVR). Specifically, when it comes to solar power generation, using SVM [7] regression has yielded satisfactory results. However, it's worth noting that this approach has limitations, particularly when it comes to providing a detailed analysis of solar power generation and meteorological data. Therefore, its ability to accurately predict other datasets is constrained when relying solely on different SVM kernels [7] after basic statistical data processing.

To investigate the link between projected weather conditions and historical power output as a time series, artificial intelligence (AI) methods are employed. These AI techniques utilize algorithms capable of implicitly capturing the complex and non-linear relationship between input data, such as numerical

weather predictions (NWP) [1], and power output, instead of relying on formal statistical analysis. Artificial neural networks(ANN)[6]serve as models inspired by the structure of the human brain and are employed in various AI applications, including supervised, unsupervised, and reinforcement learning. In the supervised learning approach, ANNs [6] learn from data by training to approximate and estimate the underlying function or relationship.

Researchers have made advancements in their models to predict power generation from photovoltaic (PV) plants. Nevertheless, the inherent unpredictability in key factors, notably the diffuse component of solar irradiance, which comes from the entire sky hemisphere, remains a significant challenge. This unpredictability makes solar irradiance less predictable than temperature, especially when dealing with PV systems that consist of numerous individual panels spread over a large area.

Given the impracticality of examining all interconnected meteorological forecasts, alternative approaches have been devised. Some have considered weather forecasts from meteorological websites, while others have employed non-linear modeling methods like artificial neural networks to simplify solar forecasting (ANN) [6]. Two common types of neural networks used for forecasting solar radiation, including global solar radiation, solar radiation on inclined surfaces, daily solar radiation, and short-term solar radiation, are radial basis function (RBF) [6] and multilayer perceptron (MLP).

LITERATURE SURVEY:

Examining the prediction of solar power generation with the application of machinelearning methods:(1)

Solar power generation worldwide relies on photovoltaic (PV) systems. However, these systems produce alternating power outputs that are heavily influenced by various environmental factors, making solar power sources inherently unpredictable. Factors such as irradiance, humidity, PV surface temperature, and wind speed all play a role. Because of this unpredictability, it's essential to plan ahead for solar power generation, which necessitates solar power forecasting to support electric grids. Predicting solar power generation is challenging due to its dependency on weather [7] conditions. To tackle this complex forecasting task, Machine Learning (ML) algorithms have emerged as a valuable tool. These algorithms, including deep learning and artificial neural networks, have demonstrated impressive results in time series forecasting. They can anticipate power output by using weather conditions as input parameters. Various ML techniques, such as support vector machine regression [7], random forest regression, and linear regression, have been employed for solar power forecasting. Among these methods, the random forest regression [7] model has shown superior accuracy, outperforming the other two regression models.

Progress in solar forecasting: The use of deep learning and computer vision techniques:

Forecasting renewable energy is vital for efficiently integrating variable energy sources into the power grid. It helps power systems cope with the unpredictable nature of these energy supplies on different spatial and temporal scales. Traditional methods for predicting

the impact of cloud movements on solar energy production rely on numerical weather predictions or physical models. However, these methods struggle to incorporate real- time cloud information and correct [1] systematic errors. To overcome these challenges, combining computer vision with machine learning offers a solution. This approach blends up- to-the-minute cloud cover observations with data collected from various sources, such as ground-level sky cameras, satellites, weather stations, and sensor networks. This review highlights recent advancements in solar forecasting using a multi-sensor [1] Earth observation approach, with a particular emphasis on deep learning. Deep learning provides the essential foundation for creating models capable of extracting relevant insights from this diverse dataset. In conclusion, machine learning has the potential to enhance the precision and reliability of solar energy meteorology [1]. However, further research is needed to fullyharness this potential and address any associated limitations.

Utilizing large-scale data and machine learning to enhance practical weather predictions:

As we incorporate more renewable energy into our power grids, it becomes vital to predict the precise energy output from these sources. This is crucial for both long-term planning and immediate grid management. To achieve this precision, we must forecast how much energy the wind and solar resources will generate over different time periods. This task is essentially a complex data problem that involves collecting a wide range of data, using various specialized models tailored to different timeframes,

and applying advanced computational techniques to combine these models with real-time observations. The ultimate objective is to provide utility and grid operators with accurate and timely information. With the ongoing expansion of renewable energy capacity, there's also the challenge of continuously updating and maintaining the data used to train machine learning algorithms.

Using machine learning for practical weather forecasting:

The National Center for Atmospheric Research (NCAR) has a rich history of using machine learning for weather forecasting. They developed the Dynamic Integrated Forecasting (DICast®) System, one of the earliest automated weather prediction systems, which is now used in various industries for different purposes. NCAR also employs Diecast and other artificial intelligence technologies in applications such as renewable energy, surface transportation, and wildland fire forecasting. Weather forecasting has evolved from being heavily reliant on human efforts to being significantly empowered by computational methods. The initial major advancement came through numerical weather prediction (NWP)[1], involving the integration of motion equations over time with accurate initial conditions. However, more recent improvements have emerged by applying artificial intelligence techniques to enhance forecasting and generate large volumes of machine-generated predictions.

The importance of utilizing expert knowledge in assessing and predicting solarphotovoltaic power generation:

Accurate and reliable predictions

and forecasts of Photovoltaic (PV) power generation are critical for the successful integration of PV systems on a large scale. This critical importance is well-documented in scientific literature, where an increasing number of studies focus on creating models for estimating and predicting PV power. Notably, machine learning techniques have gained substantial attention in the last ten years.

However, a common oversight in these developments is the neglect of the importance of predictor variables. This oversight leads to models that don't fully recognize the value of incorporating physics-based models [2]. In our study, we set out to quantify the significance of these predictor variables in PV power estimation and forecasting. We identify the limitations in existing models and introduce several preprocessing steps to enhance the overall accuracy of these estimates and forecasts.

Additionally, we delve into optimizing the selection of predictor variables for PV power estimation and forecasting. Through a sensitivity analysis, we illustrate how the value of expert variables is influenced by the tilt angle of the PV system. To gain a deeper understanding of the importance of predictor variables, we conduct two case studies in different climate regions as part of our numerical evaluation.

Using machine learning to predict the output of a photovoltaic (PV) power generation system:

To reduce the carbon footprint of buildings, it's crucial to incorporate on-site renewable energy generation systems to power them independently of the national grid. However, renewable energy sources, which rely on weather conditions, can be unreliable as the sole energy source. Photovoltaic (PV) power generation is predicted using machine learning algorithms (MLA)[5]. Different MLA methods have varying levels of accuracy and training requirements, such as the need for more inputs or more data in general. No previous research has determined the optimal MLA for PV systems, but establishing this is essential for their effective use. To find the optimal MLA for a specific application, it's necessary to define the dataset and the desired outputs and assess how they affect the algorithm's performance. The objective of this study is to compare benchmark MLAs [5] in terms of accuracy and practicality for a university campus located in central Manchester, UK. The MLAs tested include random forest (RF) [1], neural networks (NN), support vector machines (SVM), and linear regression (LR) for forecasting PV power generation. Accurate PV power forecasting can enable a building management system (BMS) to optimize on-site renewable energy production.

In this study, a total of sixty-four MLA [5] models were developed for various forecasting horizons and dataset sizes, and they were validated using real-time data. The results show that RF algorithms had the lowest average error with a root mean squared error (RMSE) of 32, while SVM [7], LR, and NN had RMSE [1] values of 32.3, 36.5, and 38.9, respectively. Errors between the forecasted and actual results were measured using RMSE, and changes in error were depicted using mean actual percentage error (MAPE) [1] to show variations relative to the original value.

This research represents a novel attempt to evaluate the performance of different MLA [5] PV forecasting models across

datasets of various sizes, providing a critical analysis of the results. Comparing benchmark algorithms for PV generation [5] forecasting in a local system enables a detailed examination of their accuracy and associated characteristics.

Forecasting the solar energy variables for a major solar power plant using advanced triple deep learning models for predictive assessment:

The rapid expansion of large-scale solar power plants (LSSPs) has underscored the need for more precise models to forecast solar energy output. Solar energy generation remains unpredictable, posing a significant challenge to the solar industry. This study focuses on leveraging advanced deep learning techniques, specifically Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Convolutional Neural Network-Long Short- Term Memory, to address this issue. These techniques are applied to various solar energy variables such as power generation, soiling loss (%), and performance ratio (PR %) to identify the most effective forecasting model. What makes this research unique is that it's the first time key solar system parameters, including PR and soiling loss, have been examined to develop a practical forecasting model using a different deep learning approach. The real-time dataset for SEVs was obtained from Pakistan's largest solar plant, the "Quaid-e-Azam Solar Park" (QASP) [6]. The main significance of this study lies in the development of ANN [6], RNN, and CNN-LSTM-based models [6] within the deep learning framework. These models involve the creation of features, data scaling, training, and testing to determine the optimal model.

The forecasted values were compared against the actual values from the solar plant over the past 7 years, and predictions were made for the future trend over the next 20 years. The objective is to develop three models to thoroughly examine the accuracy of time-series forecasting for the SEVs [6] dataset and assess performance measurement errors to identify the most suitable model. Based on the forecasting results and error analysis, the study demonstrates that the CNN-LSTM [6] hybrid model outperforms the ANN and RNN [6] models. It offers more reliable predictions for power generation and PR values and is a more accurate predictor of future trends [6].

METHODOLOGY:

Predicting solar generation from weather forecasts using machine learning involves a detailed methodology that combines data preparation, model selection, training, evaluation, and deployment. Here's a step-by-step methodology for the task:

➢ **Define the Problem**:
Clearly define the problem you want to solve, such as predicting solar energy generation based on weather conditions.

➢ **Data Collection**:
Gather historical data on solar generation and weather conditions. Sources may include weather stations, satellite data, or online databases.

➢ **Data Preprocessing**:
Clean the data by handling missing values,

outliers, and inconsistencies. Convert timestamps to a consistent format. Merge weather and solar generation data based on timestamps.

➢ **Feature Engineering**:

Extract relevant features from weather data, such as temperature, humidity, cloud cover, wind speed, and solar irradiance. Create time-based features like hour of the day, day of the week, and season. Consider lag features that capture past weather conditions.

➢ **Data Splitting**:

Split the dataset into training, validation, and test sets, ensuring that the timeorder is maintained to simulate real-world forecasting.

➢ **Model Selection**:

Choose an appropriate machine learning model for regression. Common choices include:

- Linear Regression
- Random Forest
- Gradient Boosting (e.g., XGBoost, LightGBM)
- Neural Networks (e.g., LSTM for time series data)

➢ **Model Training**:

Train the selected model on the training data. Tune hyperparameters using the validation set to optimize performance.

➢ **Model Evaluation**:

Evaluate the model's performance on the test set using appropriate regression metrics (e.g., RMSE, MAE, R-squared). Examine residuals and residual plots to understand model errors.

➢ **Post-Processing**:

Apply any necessary post-processing techniques to improve predictions or interpretability. For example, you may want to scale predictions to match the units of solar generation.

➢ **Visualization**:

Visualize the model's predictions alongside actual solar generation data toassess accuracy and identify patterns or discrepancies.

➢ **Deployment**:

If the model performs well, deploy it in a production environment. Set up automated data pipelines to feed real-time weather forecasts into the model.

➢ **Monitoring**:

Continuously monitor the model's performance and retrain it periodically withnew data.

➢ **Reporting and Alerts**:

Implement reporting mechanisms to track the model's performance and send alerts if it deviates significantly from expected values.

➢ **Feedback Loop**:

Collect feedback from users and operators to improve the model and its predictions.

➢ **Optimization**:

Continuously optimize the model and the data pipeline to improve prediction accuracy and efficiency.

➢ **Scaling**:

If necessary, scale the infrastructure to handle increased data volume and model complexity.

➢ **Maintenance**:

Regularly update the model to account for changing weather patterns and improve forecasting accuracy.

➢ **Documentation**:

Document the entire methodology, including data sources, preprocessing steps, model details, and deployment procedures. This methodology combines data science and domain

expertise in solar energy generation and weather forecasting. It's essential to have access to quality data, stay up-to-date with weather forecasts, and keep the model and its infrastructure well- maintained for accurate and reliable predictions. Additionally, collaboration with experts in the field can help fine-tune the model for specific solar generation setups and regions.

CONCLUSION:

In conclusion, predicting solar generation from weather forecasts using machine learning represents a significant advancement in the renewable energy sector. This technology holds great promise for optimizing the integration of solar power into the energy grid, enhancing energy management, and reducing reliance on non-renewable energy sources. Some notablefindings are:

- **Improved Accuracy**: Machine learning models have demonstrated the ability toprovide more accurate and reliable solar generation forecasts compared to traditional methods. By leveraging historical weather data and advanced algorithms, these models can capture complex and non-linear relationships between weather conditions and solar output.

- **Enhanced Grid Integration**: Accurate solar generation forecasts enable grid operators to better plan and manage the integration of solar power into the electricity grid. This leads to improved grid stability and more efficient energy distribution.

- **Cost Savings**: Solar power prediction models help energy providers and consumers optimize energy production and consumption. This, in turn, can lead to cost savings by reducing the need for expensive energy storage solutions and backup power generation.

- **Environmental Benefits**: Solar power forecasting using machine learning contributes to a reduction in greenhouse gas emissions by promoting the use of clean, renewable energy sources. This aligns with the global effort to combat climate change and promote sustainability.

- **Challenges Remain**: While machine learning has made significant progress in this field, challenges persist. These challenges include the need for high-quality weather data, addressing the inherent variability of solar generation, and improving the interpretability of machine learning models for practical use.

- **Future Research**: Continued research and development in the field of solar generation prediction are essential. This includes exploring novel machine learning techniques, refining existing models, and integrating real-time data sources to further enhance forecasting accuracy.

In summary, the application of machine learning to predict solar generation from weather forecasts has the potential to revolutionize the renewable energy industry. It not only facilitates the

efficient use of solar power but also supports the transition to a more sustainable and environmentally friendly energy ecosystem. As technology and researchcontinue to advance, the accuracy and reliability of solar generation forecasts are expected to improve, contributing to a more sustainable and greener future.

Generation from Weather Forecasts Using Machine Learning
10.1109/PIICON56320.2022.10045237

REFRENCES :

1. Advances in solar forecasting: Computer vision with deep learning
https://doi.org/10.1016/j.adapen.2023.100150

2. Big Data and Machine Learning for AppliedWeather Forecasts
10.1109/SSCI.2015.79

3. Machine Learning for Applied Weather Prediction
10.1109/eScience.2018.00047

4. On the value of expert knowledge in estimation and forecasting of solar photovoltaicpower generation
https://doi.org/10.1016/j.solener.2023.01.019

5. Machine learning for forecasting a photovoltaic (PV) generation system
https://doi.org/10.1016/j.energy.2023.127807

6. Predictive evaluation of solar energy variables for a large-scale solar power plant based on triple deep learning forecast models
https://doi.org/10.1016/j.aej.2023.06.023

7. Predicting Solar