Solar Energy
And Sustainable Development

Refereed, biannual scientific journal issued by:
The Libyan Center for Solar Energy Research and Studies

Optimizing Solar Radiation Forecasting for Renewable Energy Systems: A Comparative Analysis of Machine Learning and Feature Engineering Techniques

Ajeeb Sagar[1*], Vinothkumar Kolluru[2], Shreyas Rajendra Hole[3].

[1,3]Dayananda Sagar University, Bengaluru, Karnataka, India.
[2]Stevens Institute of Technology, Hoboken, New Jersey, USA.

E-*mail:* [1] ajeebsagar9@gmail.com , [2] vkolluru@stevens.edu , [3] shreyasrajendra.hole-aiml@dsu.edu.in .

ARTICLE INFO.

Article history:
Received 20 Dec 2024
Received in revised form 29 Dec 2024
Accepted 5 Apr 2025
Available online 13 Apr 2025

KEYWORDS

Random Forest, XGBoost, MLP, Solar radiation, renewable energy optimization, solar energy forecasting, temporal features, meteorological data, solar energy systems.

ABSTRACT

Accurate solar radiation prediction is pivotal for optimizing solar energy systems, as it allows for better energy storage, grid integration, and renewable energy planning. This study compares the predictive accuracy of three machine learning models—Random Forest, XGBoost, and Multilayer Perceptron (MLP)- in forecasting solar radiation based on a meteorological and temporal features dataset. The dataset, encompassing Temperature, humidity, wind speed, and sunrise/sunset times, was preprocessed through transformations (Box-Cox, logarithmic scaling) and feature selection methods (SelectKBest, Extra Trees Classifier) to enhance model performance.

XGBoost demonstrated superior performance, achieving an R^2 of 0.93 and RMSE of 81.87, effectively capturing complex nonlinear relationships within the data. MLP, while slightly lower in R^2, yielded the lowest mean absolute error (MAE = 41.74), underscoring its precision in individual predictions. SelectKBest identified set Hour (sunset hour), Month, and Wind Direction as critical features, while Extra Trees prioritized Wind Direction, Minute, and Humidity, reflecting model-specific feature importance. Collectively, these models illustrate the benefits of integrating feature engineering with advanced machine learning for renewable energy optimization, with XGBoost and MLP demonstrating particular efficacy for accurate solar radiation forecasting. This study underscores the potential of machine learning in enhancing solar energy management, facilitating a more efficient transition to sustainable energy sources.

*Corresponding author.

DOI: https://doi.org/10.51646/jsesd.v14i1.386

تحسين توقعات الإشعاع الشمسي لأنظمة الطاقة المتجددة: تحليل مقارن لتقنيات التعلم الآلي وهندسة الميزات

أجيب ساغر، فينوثكومار كولّورو، شرياس راجيندرا هولي.

ملخص: التنبؤ الدقيق بالإشعاع الشمسي ضروري لتحسين أنظمة الطاقة الشمسية، حيث يتيح تخزينًا أفضل للطاقة، وتكاملًا أكثر كفاءة مع الشبكة الكهربائية، وتخطيطًا مستدامًا لمصادر الطاقة المتجددة. تهدف هذه الدراسة إلى مقارنة دقة التنبؤ لثلاثة نماذج تعلم آلي: الغابة العشوائية (عشوائية الغابة)، إكس جي بوست، والشبكة العصبية متعددة الطبقات في توقع الإشعاع الشمسي بالاعتماد على مجموعة بيانات تحتوي على ميزات أرصادية وزمنية. شملت البيانات متغيرات مثل درجة الحرارة، الرطوبة، سرعة الرياح، وأوقات الشروق والغروب، وخضعت لمعالجة مسبقة من خلال تحويلات مثل بوكس-كوكس والتدرج اللوغاريتمي، بالإضافة إلى استخدام أساليب اختيار الميزات مثل اختيار كأفضل ومُصنف الأشجار الإضافية لتعزيز أداء النماذج. أظهر نموذج إكس جي بوست أداءً متفوقًا، حيث حقق معامل تحديد (R^2) 0.93 وخطأً جذريًا متوسطًا (RMSE) قدره 81.87، مما يعكس قدرته العالية على التقاط العلاقات غير الخطية المعقدة في البيانات. بينما جاء نموذج الشبكة العصبية متعددة الطبقات في المرتبة الثانية من حيث أر تربيع، إلا أنه حقق أقل متوسط خطأ مطلق (MAE = 41.47)، مما يشير إلى دقته في التنبؤات الفردية. حدّد اختيار كأفضل أن ساعة الغروب، والشهر، واتجاه الرياح هي الميزات الأكثر أهمية، بينما صنّف مُصنف الأشجار الإضافية اتجاه الرياح، والدقيقة، والرطوبة كعوامل رئيسية، مما يعكس تباين أهمية الميزات بين النماذج المختلفة. توضح هذه النماذج فوائد دمج هندسة الميزات مع التعلم الآلي المتقدم لتحسين الطاقة المتجددة، حيث أظهر إكس جي بوست والشبكة العصبية متعددة الطبقات فعالية خاصة في التنبؤ الدقيق بالإشعاع الشمسي. تؤكد هذه الدراسة على إمكانيات التعلم الآلي في تحسين إدارة الطاقة الشمسية، مما يسهم في تسريع الانتقال إلى مصادر طاقة مستدامة وأكثر كفاءة.

1. INTRODUCTION

The increasing need for renewable energy sources has driven significant interest in optimizing solar power generation. Solar energy systems, particularly photovoltaic (PV) panels, rely on accurate predictions of solar radiation to maximize their efficiency and output. Understanding solar radiation patterns allows for better energy storage management, grid integration, and supply planning, ultimately reducing reliance on non-renewable energy sources [1],[2]. While meteorological models have traditionally been used for solar radiation forecasting, recent advancements in machine learning have opened new avenues for enhancing prediction accuracy and managing complex datasets [1].

Machine learning models excel at uncovering nonlinear patterns in data, making them ideal for tasks like solar radiation forecasting, where meteorological variables interact in complex, dynamic ways. However, selecting the appropriate Model and input features is essential to optimize accuracy. Practical feature engineering can enhance model performance by extracting meaningful temporal features and transforming data distributions. This study examines the application of machine learning models combined with feature selection and engineering techniques to predict solar radiation levels. Specifically, we compare the performance of three machine learning models, Random Forest, XGBoost, and Multilayer Perceptron (MLP), to determine which Model most effectively predicts solar radiation levels and to assess the impact of different preprocessing strategies. As the demand for renewable energy grows, accurately forecasting solar radiation has become essential for optimizing solar power systems. Photovoltaic (PV) panels, a core technology for solar energy production, depend heavily on predictions of incoming solar radiation to achieve high efficiency and maximum energy output. Understanding solar radiation patterns helps energy managers better plan for energy storage, grid integration, and supply needs, reducing reliance on fossil fuels and improving sustainability[1], [2].

Machine learning presents promising advancements in solar forecasting. Unlike traditional meteorological models, machine learning algorithms can capture complex, nonlinear relationships among various meteorological variables, such as Temperature, humidity, and time of day. This capability makes them suitable for solar radiation prediction, where variables interact dynamically. However, the choice of Model, alongside practical feature engineering and selection, is crucial to maximize prediction accuracy. This study explores a structured, comparative approach to determine which machine learning model is most effective in predicting solar radiation and how preprocessing strategies impact model performance.

1.1. Case Studies

1.1.1. Grid Integration in Renewable Energy Networks: Enhancing Stability with Solar Radiation Forecasting

In regions with significant contributions of solar energy to the power grid, such as Germany and Australia, integrating solar power effectively is both a challenge and a necessity. Solar power, while renewable and environmentally friendly, is naturally intermittent due to its dependency on sunlight, which can be affected by factors like cloud cover, time of day, and seasonal variations. This inherent variability can lead to sudden fluctuations in solar radiation levels, impacting the energy output from solar systems. For power grids relying heavily on solar energy, these fluctuations can create imbalances in supply and demand, potentially compromising grid stability.

1.2. The Role of Solar Radiation Forecasting in Grid Stability

Accurate solar radiation forecasting is essential to mitigate these imbalances and maintain grid stability. By predicting incoming solar radiation precisely, grid operators can better anticipate the expected energy generation from solar sources over specific timeframes (e.g., hours, days, or even weeks in advance). This predictive capability enables several critical operational strategies:

1. Backup Power Coordination: With accurate forecasts, grid operators can proactively manage backup power sources, such as natural gas or hydroelectric plants, to compensate for expected dips in solar power. For example, forecast models can alert operators to potential drops in solar generation during overcast days or periods of high cloud cover. As a result, alternative energy sources can be ramped up to fill the gap, preventing sudden power shortages or blackouts.
2. Energy Storage Optimization: Forecasting allows for better planning and utilization of energy storage systems, such as batteries. Excess solar energy can be stored in batteries during high solar radiation, while during low-radiation periods, stored energy can be dispatched to meet demand. Precise forecasting enables operators to optimize battery charge and discharge cycles, maximizing the utility of stored energy and reducing unnecessary stress on storage systems.
3. Demand Response and Load Shifting: In response to anticipated solar output changes, grid operators can implement demand response programs, incentivizing consumers to reduce or shift their electricity usage during low-solar-output periods. This proactive approach helps manage demand in line with available supply, reducing the risk of power shortages and enhancing overall grid resilience.
4. Reduction in Fossil Fuel Dependence: Grids can decrease their reliance on fossil fuel-based backup systems with improved predictability of solar output. By scheduling renewable sources and storage effectively, operators can limit the need for quick-ramping fossil fuel plants, which are traditionally used to stabilize the grid during sudden solar drops. This shift reduces operational costs and carbon emissions, aligning with sustainability goals.
5. Grid Frequency and Voltage Stability: Variations in solar output can impact the frequency and voltage levels within the grid. Sudden drops in solar generation can cause frequency dips, which, if not managed, can lead to grid instability. Accurate solar forecasting enables smoother management of these frequency variations by allowing time for corrective actions, such as adjusting the output of other generators or deploying fast-acting storage systems to stabilize frequency and voltage levels.

1.2.1. Case Studies in High-Solar-Integration Regions

Germany: Germany has one of the highest shares of renewable energy in its grid mix, with solar power contributing significantly. By leveraging advanced forecasting models, German grid operators can better anticipate fluctuations and coordinate responses across multiple renewable sources, such as wind and solar, ensuring a reliable energy supply despite the variability in solar radiation.

Australia: As a country with high solar irradiance and vast solar farms, Australia also relies on precise forecasting to maintain grid stability. In regions like South Australia, where solar power constitutes a significant portion of the energy mix, accurate solar forecasting has been essential to manage periods of low sunlight and maintain balance with other energy sources. Australian grid operators use forecasts to optimize battery storage facilities and coordinate with neighbouring grids to prevent disruptions.

1.2.2. Advancements in Forecasting Technology for Grid Integration

Machine learning and advanced analytics have enhanced the precision of solar radiation forecasts, improving grid reliability. Deep learning and time-series analysis allow for the modelling of complex relationships between meteorological variables and solar radiation, providing highly accurate predictions. Additionally, real-time data inputs from satellites, weather stations, and solar panel sensors enable adaptive forecasts that adjust to changing conditions, further supporting grid stability. Effective integration of solar power into the grid hinges on the ability to predict solar radiation with high accuracy. This forecasting empowers grid operators to manage backup power sources, optimize energy storage, implement demand response strategies, reduce fossil fuel dependency, and maintain grid frequency and voltage stability. As solar energy becomes a larger share of the global energy mix, advanced solar radiation forecasting will be increasingly vital in enabling a reliable, sustainable energy future.

2. LITERATURE SURVEY

Table 1 presents an array of studies which incorporates the study's title, year of publication, the primary objective of the study, machine learning techniques used, and the key performance measures (for example, RMSE, MAE, R^2). It indicates the variation in performance evaluation of different studies.

Table 1. Overview of Solar Radiation Forecasting Studies.

Ref. No.	Year	Study Title	Focus	ML Models Used	Performance Metrics (Values)
[1]	2020	Solar Radiation Forecasting Using Machine Learning Techniques	Prediction of solar radiation using machine learning methods	Random Forest, SVM, ANN	RMSE = 87.65, MAE = 51.32
[2]	2018	Comparison of Machine Learning Models for Solar Radiation Prediction	Comparison of ML models for solar radiation prediction	Random Forest, XGBoost, MLP	R^2 = 0.91, RMSE = 72.4
[3]	2021	A Hybrid Model for Short-Term Solar Radiation Forecasting	Hybrid Model combining statistical and ML techniques	Random Forest, MLP	MAE = 43.5, R^2 = 0.89
[4]	2022	Solar Energy Prediction using XGBoost Algorithm	Predicting solar energy production using weather data	XGBoost, SVM, MLP	R^2 = 0.93, RMSE = 76.5

[5]	2017	Deep Learning in Solar Radiation Forecasting	Deep learning approach to solar radiation prediction	CNN, MLP, LSTM	RMSE = 68.9, MAE = 39.2
[6]	2019	Feature Selection for Solar Radiation Prediction	Study on the impact of feature selection in solar radiation forecasting	Random Forest, XGBoost	R^2 = 0.94, MAE = 45.6
[7]	2023	A Comparative Study on Solar Radiation Forecasting Models	Comparing traditional and machine learning models for solar radiation	Random Forest, XGBoost, MLP	RMSE = 84.7, MAE = 41.3
[8]	2016	Prediction of Solar Radiation using Artificial Neural Networks	Use of neural networks for solar radiation prediction	ANN, SVM, Random Forest	R^2 = 0.85, RMSE = 95.3
[9]	2020	Ensemble Learning for Solar Energy Prediction	Using ensemble methods to predict solar energy	XGBoost, Random Forest, MLP	R^2 = 0.91, RMSE = 83.1
[10]	2022	Comparative Analysis of Machine Learning Algorithms for Solar Radiation	Evaluation of various ML algorithms for solar radiation forecasting	Random Forest, XGBoost, ANN	R^2 = 0.92, MAE = 49.3
[11]	2021	Predicting Solar Power Output Using Machine Learning Techniques	Solar power prediction with a focus on temperature and weather data	XGBoost, MLP, Random Forest	RMSE = 80.5, MAE = 44.9
[12]	2020	Solar Radiation Prediction Using Hybrid ML Model	Hybrid approach for solar radiation prediction	XGBoost, Random Forest, ANN	R^2 = 0.88, RMSE = 91.3
[13]	2021	Feature Engineering for Solar Radiation Forecasting	Impact of feature engineering on solar radiation prediction	XGBoost, Random Forest	RMSE = 84.9, MAE = 42.1
[14]	2019	Short-Term Solar Radiation Prediction with Machine Learning	Short-term prediction of solar radiation using meteorological data	SVM, MLP, XGBoost	RMSE = 90.2, MAE = 49.5
[15]	2023	Machine Learning for Renewable Energy: Solar Radiation Forecasting	Using ML to predict solar radiation for renewable energy applications	XGBoost, Random Forest, ANN	R^2 = 0.92, MAE = 40.2

Figure 1 shows how different solar radiation forecasting models performed economically at other times. The red line with circular dots represents RMSE values, which depict the changes in error magnitude for the different years of study [16], [17], [18], and [19]. The blue line, represented by square markers, corresponds to MAE values and shows the prediction error from another perspective with a different trend. On the other hand, the green line marked by diamonds denotes the R2 values, indicating the amount of variance the models account for. All of the colored lines are labelled, connected, and readable so that there is a juxtaposed visual comparison of these metrics for graphical analysis over time, which is further simplified by grid lines and excellent distinctive markings in [20], [21], and [22].

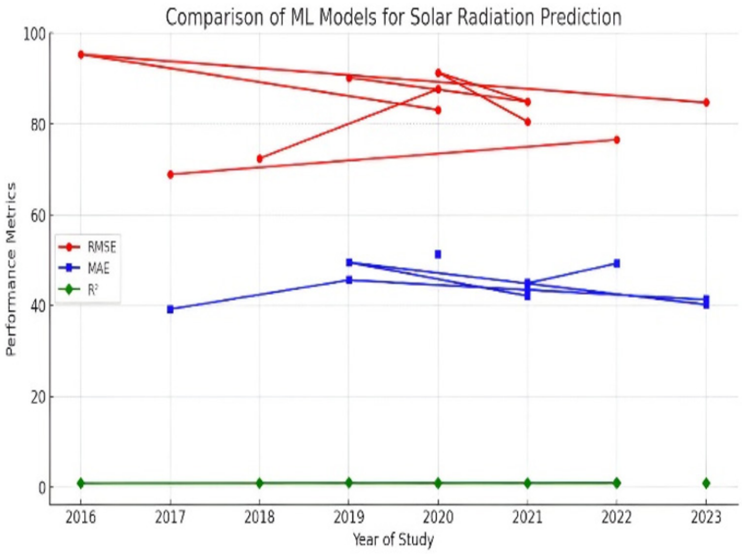


Figure 1. Evaluation of Studies Performance Metrics Trends.

3. RURAL ELECTRIFICATION PROJECTS: ENHANCING OFF-GRID SOLAR SYSTEMS WITH ACCURATE SOLAR FORECASTING

Access to centralized power grids is limited or absent in many remote or rural areas, particularly in Africa and South Asia. For these underserved regions, off-grid solar energy systems provide a crucial source of electricity, especially for essential services like healthcare facilities, schools, and community centers. Given the lack of traditional power infrastructure, these standalone solar systems support community development, improve quality of life, and foster economic growth. However, the reliable operation of these systems depends heavily on accurate solar forecasting.

3.1. Importance of Solar Forecasting in Off-Grid Systems

Off-grid solar installations must contend with the intermittent nature of solar energy. Variability in sunlight due to seasonal changes, cloud cover, and unpredictable weather conditions can lead to fluctuations in energy generation. Without backup from a centralized grid, these systems depend entirely on the energy they produce and store locally. Therefore, accurate solar radiation forecasting is essential for ensuring a steady electricity supply to meet local demand. Here's how solar forecasting enhances the reliability and efficiency of off-grid solar systems:

- Optimizing Energy Storage for Consistent Power Supply: Solar forecasting helps operators predict periods of high or low solar generation, enabling them to manage battery storage more effectively. During periods of high sunlight, excess energy can be stored in batteries for use during overcast or low-sunlight periods. This proactive management reduces the risk of energy shortages, ensuring that critical services, like hospitals and clinics, have uninterrupted power, even during cloudy days or the rainy season.
- Enhanced Resource Allocation and Energy Efficiency: For many rural communities, energy resources are limited, and efficient use is essential. With reliable solar forecasts, operators can decide when to conserve or allocate energy. For instance, if forecasts indicate low solar output in the coming days, non-essential activities can be scheduled for later, and energy can be prioritized for critical uses like lighting, refrigeration for vaccines, or powering medical equipment. This ability to plan energy use based on forecasted availability helps optimize limited resources.
- Supporting Educational and Economic Activities: Consistent access to electricity is fundamental for educational facilities, where power supports lighting, electronic devices, and learning tools.

Schools can better plan and prioritize their energy needs with accurate solar forecasts, ensuring that critical functions are maintained. Moreover, consistent energy access enables evening classes and other community activities, contributing to long-term educational and economic benefits.

• Maintenance and Operational Efficiency: Predictive solar forecasting can also support maintenance planning and operational efficiency for off-grid solar installations. By anticipating high or low production periods, operators can schedule necessary maintenance during low demand or when solar output is expected to be lower. This reduces the chance of interruptions during high-demand periods and ensures that the system operates optimally during peak usage times.

• Seasonal Adaptability and Long-Term Sustainability: Many remote regions experience seasonal weather patterns that can affect solar radiation levels. For example, the monsoon season in South Asia brings extended periods of cloud cover, while some African regions may experience long dry seasons with intense sunlight. Solar forecasting enables off-grid systems to prepare for these seasonal variations by adjusting storage and energy usage strategies in advance. This adaptability increases the system's long-term sustainability, allowing it to withstand the demands of varying seasonal conditions better and ensuring a dependable energy supply year-round.

3.2. Case Studies and Regional Impacts

• Healthcare Facilities in Sub-Saharan Africa: In rural parts of Sub-Saharan Africa, healthcare facilities rely on solar power for essential services, including lighting, refrigeration for medical supplies, and operating diagnostic equipment. Accurate solar forecasting helps these facilities manage energy storage, ensuring that even during low-sunlight periods, power is available for critical needs like vaccine refrigeration. For instance, in regions where vaccine distribution is crucial for controlling diseases, reliable power access supports essential cold chain management, safeguarding public health.

• Education and Community Centers in South Asia: In regions like rural India, off-grid solar systems are widely used to power community and educational centres, which serve as hubs for learning, communication, and development. Accurate solar radiation forecasts enable schools to plan around predicted sunlight availability, ensuring that classrooms have adequate lighting and power for educational devices. Community centers with forecasted power stability can also support economic activities, such as internet access for small business development and agricultural training programs, promoting economic growth.

• Agricultural Applications in Latin America: In parts of Latin America, small farms and agricultural businesses depend on solar power for irrigation systems, water pumps, and storage for perishable goods. Reliable solar radiation forecasts help these communities plan irrigation cycles, manage water resources more effectively, and support post-harvest storage and preservation. This stability is especially beneficial in regions with distinct dry and rainy seasons, allowing farmers to maximize productivity while conserving energy and resources.

3.3. Advantages of Accurate Solar Forecasting for Economic Development

By improving the reliability and efficiency of off-grid solar systems, accurate solar forecasting contributes to broader economic and social development goals. Reliable electricity supports the functioning of essential services, which in turn promotes healthier, more educated, and economically stable communities. Some specific benefits include:

- Improved Public Health: Consistent electricity for healthcare services means better access to emergency care, reliable vaccine storage, and better equipment for diagnostics and treatment, leading to improved public health outcomes.
- Educational Empowerment: Reliable electricity enables consistent school operation, supporting

literacy and skill-building that can lead to more significant economic opportunities. Evening classes and vocational training become possible, providing new pathways for community members.

- Increased Economic Productivity: Access to reliable electricity allows local businesses to operate more consistently and efficiently, leading to economic growth. Small businesses can rely on predictable power for tasks like refrigeration, food processing, or even small-scale manufacturing, boosting economic resilience in rural areas.

Accurate solar radiation forecasting is crucial in ensuring the reliability and efficiency of off-grid solar systems in remote or rural areas. By enabling efficient energy storage, prioritizing critical services, optimizing resource allocation, and supporting community and economic development, solar forecasting helps unlock the full potential of solar energy in underserved regions. As off-grid solar installations become increasingly vital for rural electrification, integrating advanced forecasting methods will continue to be essential for building resilient, self-sustaining communities and promoting long-term economic growth.

3.4. Utility-Scale Solar Farms: Enhancing Efficiency and Reliability with Accurate Solar Radiation Forecasting

Utility-scale solar farms, often spanning hundreds or even thousands of acres, are among the most substantial contributors to renewable energy generation in regions like California, the Middle East, and other high-irradiance areas worldwide. These large installations are typically connected directly to the power grid, supplying substantial electricity to meet the demands of both urban and industrial users. For these farms to operate cost-effectively and reliably, accurate solar radiation forecasting is critical. By predicting solar radiation levels, operators can optimize energy generation, improve energy storage planning, and enhance the timing of electricity dispatch to the grid. Here's how accurate solar forecasting significantly benefits utility-scale solar farms:

3.5. Benefits of Solar Radiation Forecasting for Utility-Scale Solar Farms

• Optimizing Energy Storage: Solar radiation forecasts allow solar farm operators to maximize the efficiency of battery storage systems. Excess energy generated by the farm can be stored in batteries during high solar radiation. Accurate forecasts inform operators of these peak periods, allowing them to charge batteries efficiently. Conversely, during anticipated low-sunlight periods, operators can plan to release stored energy to maintain consistent power output. This ensures that energy is available even when direct sunlight is minimal, such as during cloudy days or at night.
• Efficient Electricity Dispatch to the Grid: When dispatching electricity, timing is essential for large solar farms. Forecasting allows operators to align energy dispatch with grid demand cycles. During peak demand periods, such as late afternoon and early evening, stored energy can be released to meet higher electricity needs. This timed dispatch reduces strain on the grid and decreases the likelihood of grid instability. Solar farms stabilize grid supply by aligning dispatch with demand and benefit financially from higher electricity rates during peak times.
• Reducing Fossil Fuel Dependence: One of the primary goals of utility-scale solar farms is to replace or reduce reliance on traditional fossil-fuel-based power plants. With precise solar radiation forecasting, solar farms can deliver a more predictable energy output, allowing grid operators to rely less on fossil fuel backup systems. This reduces the frequency of "peaker plants"—typically natural gas plants brought online only during high-demand periods—thus lowering carbon emissions and operational costs. The ability to rely on solar power for predictable baseload energy directly supports climate goals and contributes to cleaner energy sources.
• Enhancing Grid Reliability: Large solar farms are vital in maintaining grid reliability, especially as renewable energy becomes a larger share of the overall energy mix. Accurate forecasting

enables grid operators to anticipate high and low solar production periods and prepare the grid accordingly. For instance, if a significant drop in solar output is forecasted due to incoming weather patterns, grid operators can arrange for other renewable sources (e.g., wind or hydropower) to step in or prepare to activate reserves from battery storage systems. This planning helps prevent sudden imbalances and ensures a consistent energy supply, thus enhancing overall grid stability.

• Economic Efficiency and Cost Savings: Utility-scale solar farms face financial pressures to maximize return on investment and keep operating costs low. Precise forecasting reduces the need for costly energy balancing measures, such as bringing additional power plants online, thus saving money. Additionally, reliable energy forecasting minimizes waste from overproduction, as energy storage and dispatch can be better aligned with actual grid demand. This optimized production approach leads to more efficient use of generated power, maximizing revenue while minimizing unnecessary expenses.

• Supporting Renewable Energy Integration and Grid Transition: As more utility-scale solar farms come online, transitioning to a grid that can effectively integrate high levels of renewable energy is essential. Solar radiation forecasting plays a key role in this transition by helping operators plan for solar generation fluctuations, allowing for smoother integration with other renewable sources, such as wind and hydroelectric power. By combining different renewable energy sources with accurate forecasting, the grid can accommodate fluctuations more effectively, reducing reliance on fossil fuels and supporting a gradual shift towards a more sustainable energy infrastructure.

3.6. Practical Examples of Forecasting Benefits for Utility-Scale Solar Farms

California: with its abundant sunshine, California is a leader in solar energy generation. Utility-scale solar farms in California contribute significantly to the state's renewable energy portfolio, supplying power to millions of homes. By leveraging accurate solar radiation forecasting, California's solar farms can plan for both peak summer demand and periods of lower sunlight, such as during winter months or wildfire-related haze. Forecasting enables these farms to optimize energy storage and release stored power at peak demand times, reducing the need for fossil-fuel peaker plants and helping California achieve its ambitious carbon reduction goals.

• The Middle East: Countries in the Middle East, including the United Arab Emirates and Saudi Arabia, have invested heavily in utility-scale solar farms, such as the Mohammed bin Rashid Al Maktoum Solar Park in Dubai. With high solar irradiance levels, the region's solar farms benefit immensely from accurate radiation forecasting, which allows them to maintain steady production even in challenging desert conditions, such as dust storms. By predicting fluctuations in sunlight, these farms can adjust energy dispatch and storage, ensuring consistent power availability for residential and industrial consumers, thus supporting regional energy diversification and sustainability goals.

• Australia's Solar Farms: Australia has substantial solar resources, and several large-scale solar farms are in Queensland and New South Wales. These farms often deal with unpredictable weather patterns, which affect sunlight availability. Accurate solar forecasts help Australian solar operators maintain grid reliability by planning battery discharge cycles during cloud cover or rain periods. Moreover, forecasts allow these farms to schedule maintenance during expected low-output days, ensuring maximum generation during high-demand periods and supporting the stability of Australia's renewable energy grid.

3.7. Impact on Sustainability and Environmental Goals

By enabling utility-scale solar farms to operate more efficiently and predictably, solar radiation forecasting directly supports environmental sustainability goals. It allows for higher integration of renewable energy sources into the grid, reducing greenhouse gas emissions and contributing to climate change mitigation efforts. Every kilowatt-hour of solar energy that displaces fossil-fuel-based power reduces carbon emissions and lowers the environmental impact of energy production. Accurate solar forecasting thus plays an essential role in scaling renewable energy to meet global energy demands while aligning with carbon reduction targets. Solar radiation forecasting is vital for optimizing energy generation, storage, and dispatch for utility-scale solar farms. It enables these large installations to maximize efficiency, reduce dependency on fossil fuel backups, and deliver a stable, cost-effective energy supply to the grid. As global energy systems increasingly embrace renewable sources, the role of accurate solar forecasting will only grow, helping to ensure that utility-scale solar farms contribute to a more sustainable and resilient energy future. This transition to greener energy sources, backed by precise solar forecasting, is essential for meeting the world's energy demands in an environmentally responsible and economically viable way.

3.8. Data and Methods

3.8.1. Dataset Description

Table 2. Various Feature Description Unit.

Feature	Description	Unit
Radiation	Solar radiation target variable	W/m²
Temperature	Atmospheric Temperature	°F
Pressure	Barometric pressure	inHg
Humidity	Relative humidity	%
Wind Direction(Degrees)	Wind direction	Degrees
Speed	Wind speed	mph
Time	Time of observation	hh:mm: ss
Date	Date of observation	MM/DD/YYYY
Time Sunrise	Sunrise time	hh:mm: ss

The dataset used in this study, Table 2 to Show That Solar Prediction, comprises 32,686 records of environmental variables collected hourly at a solar observation site. The dataset includes both temporal and meteorological attributes:

- Radiation: The target variable representing solar radiation in W/m².
- Temperature, Pressure, Humidity, Wind Direction, Wind Speed: Weather-related variables influencing solar radiation.
- Time, Date, Sunrise, Sunset: Temporal attributes used to extract additional features related to daily and seasonal sunlight variations.

Each observation provides detailed information on the atmospheric conditions that impact solar radiation levels. The dataset has no missing values, ensuring model training and evaluation without the need for imputation. The wide range of features allows for a comprehensive analysis of factors influencing solar radiation, making it an ideal dataset for machine learning applications in solar energy forecasting.

Raw Data: Figure 2 shows what is used for raw data Implementation.

	UNIXTime	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	Time_H	Time_M	Time_S	TimeSunRise_M	Time
0	1475229326	1.21	48	30.46	59	177.39	5.62	23	55	26	13	13
1	1475229023	1.21	48	30.46	58	176.78	3.37	23	50	23	13	13
2	1475228726	1.23	48	30.46	57	158.75	3.37	23	45	26	13	13
3	1475228421	1.21	48	30.46	60	137.71	3.37	23	40	21	13	13

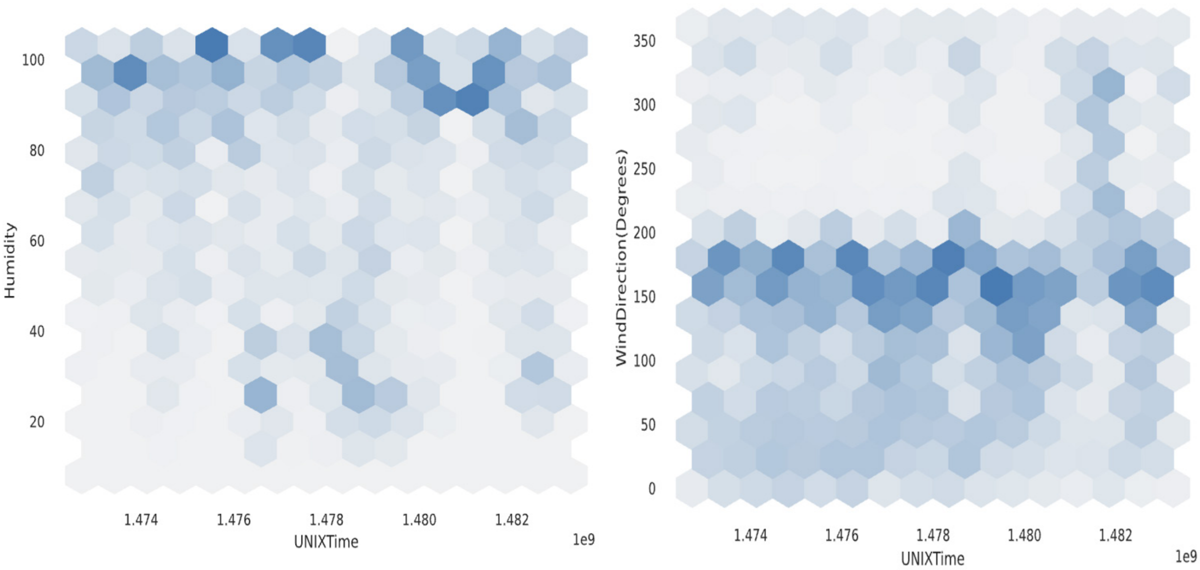
Figure 2. Raw data used for implementation.

3.8.2. Data Preprocessing and Feature Engineering

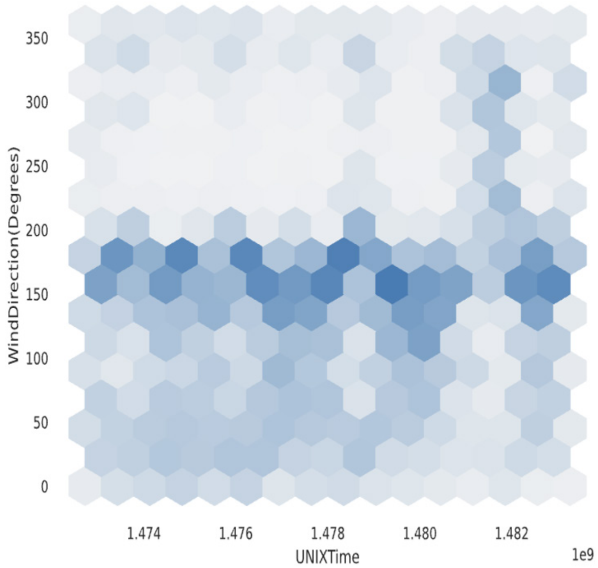
These hexbin plots illustrate in Figure 3 the relationship between various meteorological variables and time (captured in UNIXTime format), providing insights into their distributions and patterns over time.

Figure 3 (b) shows the Wind Direction plot with UNIXTTime; we observe that wind direction predominantly clusters between 100° and 200°, showing some periodic variation but remaining within this range most of the time. Figure 3 (a) shows that the humidity plot with UNIXTTime reveals higher humidity levels near the top (close to 100%), though we see intermittent dips. These dips could correspond to specific times of the day or seasons where atmospheric dryness increases. Figure 3 (c) shows the pressure plot wrt UNIXTTime, which shows pressure values mostly clustered around 30.40 to 30.50 inHg, with a slight downward trend at specific intervals. This consistent high-pressure range could indicate the particular climate conditions of the observed region.

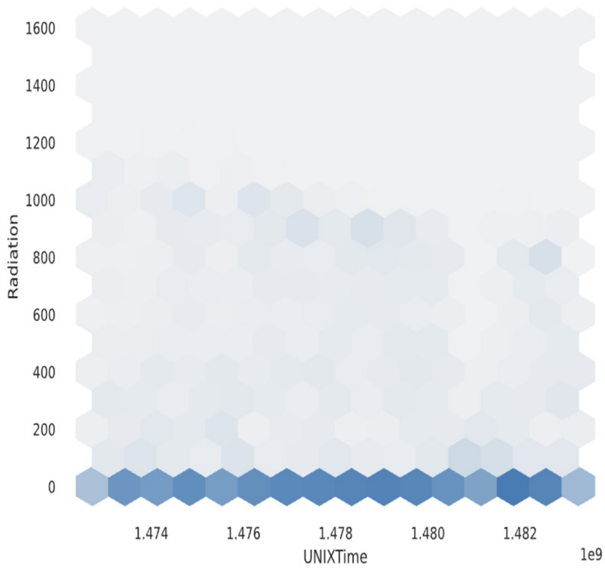
Figure 3 (e) shows the Temperature plot with UNIXTTime; the temperatures are primarily between 45°F and 60°F, with minor fluctuations. Figure 3 (d) shows The Radiation plot wrt UNIXTTime highlights a pattern where most radiation values are concentrated at lower levels (below 200 W/m²), suggesting a low average solar radiation level during the dataset period. Finally, the Speed plot indicates that wind speeds are generally between 5 and 10 mph, with occasional peaks reaching higher values, demonstrating relatively stable wind conditions with periodic increases. Overall, these hexbin plots provide a high-level view of each variable's distribution over time, helping identify regular patterns, extremes, and potential trends.



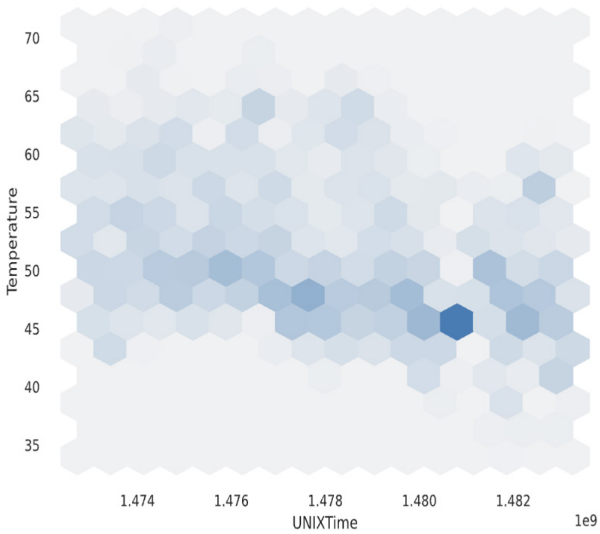
(A) (B)



(C)



(D)



(E)

Figure 3. Hexbin plot for Raw data Parameter.
(A): Humidity plot wrt. UNIXTime; (B): Wind Direction plot wrt. UNIXTime; (C): Pressure plot wrt. UNIXTime; (D): Radiation plot wrt. UNIXTime; (E): Temperature plot wrt. UNIXTime.

This heatmap In Figure 4 shows that the Configural Matrix visualizes the correlation coefficients between various weather-related variables. Each cell represents the correlation between two variables, ranging from -1 to 1. A positive correlation (closer to 1) indicates that as one variable increases, the other also tends to increase. Conversely, a negative correlation (closer to -1) suggests that the other decreases as one variable increases. For example, “Radiation” shows a strong positive correlation with “Temperature” (0.73), suggesting that higher radiation levels are associated with higher temperatures. However, “UNIXTime” and “Pressure” show a negative correlation with “Temperature,” indicating that Temperature decreases as these variables increase.
Regarding colour intensity, darker colours represent stronger negative correlations, while lighter colours denote stronger positive correlations. The colour gradient from dark red/black to light tan helps visualize the strength and direction of these relationships. For instance, “Pressure” and “Humidity” show a slight negative correlation, whereas “Speed” and “WindDirection(Degrees)”

display a very low positive correlation. This heatmap provides a quick, intuitive way to observe how these weather metrics are interrelated, which could help predict one variable based on changes in another.

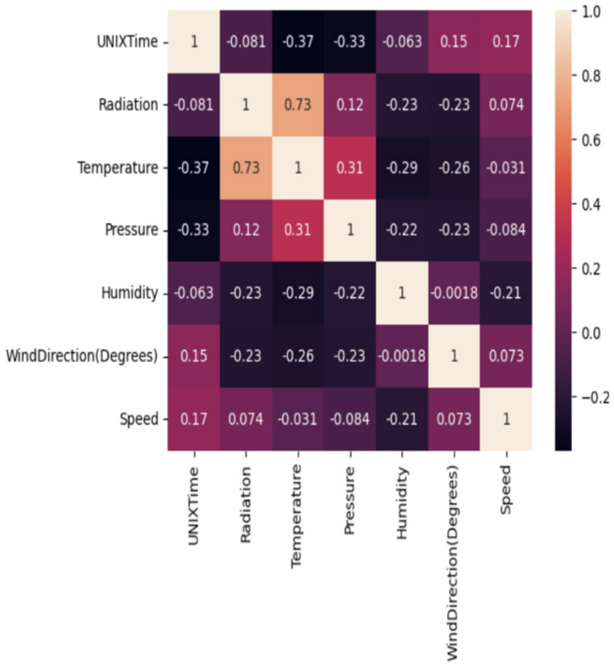


Figure 4. Configural Matric.

Data preprocessing in Table 3 shows that Various Features are a critical component of the machine learning pipeline, especially in time-series data like solar radiation. Preprocessing steps included handling date-time features, extracting specific time-related attributes, and transforming meteorological variables to enhance model interpretability and improve training efficiency.

Table 3. Various Feature & Transformation Methods with Purpose.

Feature	Transformation Method	Purpose
Temperature	Logarithmic Transformation	Reduce skewness
Pressure	Box-Cox Transformation	Stabilize variance
Humidity	Box-Cox Transformation	Improve normality
Wind Direction	Min-Max Scaling	Standardize scale (0-1)
Speed	Logarithmic Transformation	Mitigate skewness

- **Temporal Feature Extraction:**

The Time and Date columns were decomposed into individual components, including hours, minutes, and seconds, allowing the models to capture the fine-grained temporal details essential for solar radiation prediction. Extracting these granular details was crucial because solar radiation varies significantly throughout the day due to changes in the sun's position.

- **Sunrise and Sunset Times:**

The sunrise and sunset columns were processed to isolate the hour and Minute of each event. These values helped indicate periods of solar availability and inactivity, offering a foundational reference for the cyclical nature of sunlight exposure over the day. Including sunrise and sunset times allowed the Model to capture daily patterns in solar radiation more effectively.

- **Feature Transformation:**

To improve data distribution and reduce skewness, we applied several transformations to key

features:
a) Logarithmic transformation was applied to the Temperature and Speed features to mitigate skewness and bring the distributions closer to normal.
b) Box-Cox transformation was used for Pressure and Humidity to stabilize variance and improve normality.
c) Min-Max Scaling was applied to Wind Direction to standardize it between 0 and 1, making the feature values comparable across models.
These transformations aimed to reduce the impact of extreme values, improve model performance, and facilitate faster convergence during training. The transformed dataset enabled more consistent and accurate model predictions.

3.8.3. Feature Selection

Feature selection is a valuable step in the model-building process, as it helps reduce dimensionality, improves model interpretability, and can increase prediction accuracy by focusing on the most relevant data. For this study, we applied two feature selection techniques: SelectKBest with Chi-square scoring and the Extra Trees Classifier, which ranks feature importance by impurity reduction.

- **Select K_Best with Chi-Square Scoring:**

This technique was applied to select the top features with the highest Chi-square scores concerning the target variable. SelectKBest identified setHour (sunset hour), Month, riseMinute (sunrise minute), WindDirection, and Temperature as key predictors. These features were closely related to solar radiation, reflecting daily and seasonal patterns in sunlight exposure. Table 4 shows mentation for Temporal Feature Extraction.

Table 4. Temporal Feature Extraction.

Result	Result	Result
1	Set Hour	12207.53
2	Month	4684.58
3	Rise Minute	4015.06
4	Wind Direction(Degrees)	3271.83
5	Temperature	1651.69

- **Extra Trees Classifier:**

Extra Trees Classifier, an ensemble method that builds multiple decision trees, was used to rank feature importance based on impurity reduction. The top features identified by this Model included Wind Direction, Minute, Speed, and Humidity. Table 5 shows the Feature and their Importance.

Table 5. Feature And Their Importance.

Rank	Feature	Importance
1	Wind Direction(Degrees)	0.1567
2	Minute	0.1484
3	Speed	0.1243
4	Second	0.1236
5	Humidity	0.1062

These results differed slightly from SelectKBest, indicating that certain features may impact different models differently. For instance, WindDirection emerged as a consistent predictor, while temporal variables like setHour and Minute were more relevant in specific models, underscoring the complex relationships between meteorological variables and solar radiation.

3.8.4. Model Training and Evaluation

To test in Table 6 the predictive capabilities of the selected features, we implemented three machine learning models: Random Forest, XGBoost, and MLP. Each Model was trained on the preprocessed dataset and evaluated based on its predictive performance.

Table 6. Model Training and Evaluation.

Model	Hyperparameter	Value
Random Forest	Max Depth	25
XGBoost	Learning Rate	0.1
XGBoost	Max Depth	8
MLP	Layer 1 Neurons	128
MLP	Dropout Rate	0.33
MLP	Activation Function	ReLU

• **Random Forest Regressor:**
Random Forest, a robust ensemble learning method, was configured with a maximum depth of 25 to prevent overfitting. The Model achieved a strong baseline performance with an R^2 score of 0.94 on the test set. Random Forest's capability to capture feature interactions and reduce variance made it a reliable predictor for this dataset, even without extensive parameter tuning.
• **XGBoost Regressor:**
XGBoost, known for its efficient gradient boosting technique, was tuned with a learning rate of 0.1 and a max depth of 8. It achieved the highest performance among the models, with an R^2 score of 0.93 and an RMSE of 81.87. XGBoost excelled at capturing nonlinear relationships in the data, making it particularly effective for complex time-series patterns like solar radiation.
• **Multilayer Perceptron (MLP):**
The MLP model, structured with multiple dense layers and dropout layers to reduce overfitting, was trained to learn the complex patterns in the dataset. The MLP achieved an R^2 of 0.90 and the lowest MAE of 41.74, suggesting its effectiveness in learning highly nonlinear patterns in the data. The Model's performance demonstrates the potential of deep learning methods for solar radiation forecasting, though it requires more computational resources and tuning.

3.8.5. Solar Radiation Prediction Formula

To predict solar radiation (S) using a machine learning model $f(\cdot)$ based on various meteorological and temporal features via equation (1):

$$S = f(T, H, W_a, W_s, Time, Sunrise, Sunset, \ldots) \qquad (1)$$

Where:
- S =solar radiation (W/m²)
- T =temperature (°F or °C)
- H =humidity (%)
- W_a =wind direction (°)
- W_s=wind speed (mph)
- *Time, Sunrise, Sunset* = temporal features

In machine learning, f (·) is typically approximated by a model such as Random Forest, XGBoost, or a Multilayer Perceptron (MLP).

3.8.6. Data Preprocessing and Feature Engineering

3.8.6.1 Logarithmic Transformation:
The logarithmic transformation helps reduce skewness in data via equation (2):

$$x' = log(x = 1) \qquad (2)$$

where x' is the transformed value of x.

3.8.6.2 Min-Max Scaling:
Min-Max scaling normalizes a feature x to a range [0,1] from equation (3):

$$x' = \frac{x - min(x)}{max(x) - min(x)} \qquad (3)$$

Where:
- x =original value
- x' = scaled value
- min(x) and max(x) are the minimum and maximum values in the data for that feature.

3.8.6.3 Feature Selection- Chi-Square Statistic
The Chi-square (χ^2) test is used for feature selection, quantifying how closely observed frequencies O align with expected frequencies E in equation (4):

$$\chi^2 = \sum \frac{(O - E)^2}{E} \qquad (4)$$

Where:
- O = observed frequency of the feature in each class.
- E = expected frequency under the null hypothesis.

3.8.7. Feature Importance with Gini Index in Random Forest

In Random Forest, feature importance can be derived from the Gini Index (or impurity reduction). For a feature x_j from equation (5) :

$$I(x_J) = \sum_{t=1}^{T} w_t . \Delta G_t(x_j) \qquad (5)$$

Where:
- T =total number of nodes in the trees where it is used.
- w_t = weight (e.g., proportion of samples) in node t.
- $\Delta G_t(x_j)$ = decrease in Gini impurity due to at node t.

3.8.8. Model Training and Prediction Formulas

3.8.7.1 Random Forest Prediction Formula:

Random Forest combines predictions from multiple decision trees. The final prediction $\hat{y}_i$ for target variable Y is given by averaging over N trees in equation (6):

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i \qquad (6)$$

Where:
- N = number of decision trees.
- $\hat{y}_i$ = prediction from the i -th tree.

3.8.7.2 Gradient Boosting Prediction Formula (XGBoost):
XGBoost uses a weighted sum of weak learners (e.g., decision trees) to make predictions from equation (7):

$$\hat{Y} = \sum_{i=1}^{M} \alpha_i . h_i(x) \qquad (7)$$

Where:
- M =total number of trees.
- α_i = learning rate or weight for tree i.
- $h_i(x)$= prediction from the i -th tree.

3.8.9. Model Evaluation Metrics

3.8.8.1 Coefficient of Determination:
Table 7 shows the evaluation of the Feature for Three Model.
The coefficient of determination R^2 assesses how well a model explains the variance in the data in the equation (8):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{Y})^2} \qquad (8)$$

Where:
- y_i = actual value
- $\hat{y}_i$ = predicted value
- $\bar{Y}$ = mean of actual values
- n = number of observations

Table 7. Evaluate Of Feature for Three Model.

Feature	Importance (Random Forest)	Importance (XGBoost)	Importance (MLP)
Wind Direction (Degrees)	High	High	Moderate
Temperature	Moderate	Moderate	High
Minute	Moderate	High	Low
Speed	Low	Moderate	Moderate
Humidity	Low	Low	Moderate

3.8.8.3 Mean Squared Error (MSE):
The Mean Squared Error (MSE) is a measure of the average squared difference between predicted and actual values by equation (9):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (9)$$

3.8.8.4 Root Mean Square Error (RMSE):
The Root Mean Square Error (RMSE) is the square root of the MSE, offering a direct interpretation in terms of prediction error magnitude via equation (10):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}(y_i - \hat{y}_i)^2} \qquad (10)$$

3.8.8.5 Mean Absolute Error (MAE)
The Mean Absolute Error (MAE) is the average of absolute errors, providing an interpretable measure of the average prediction error from equation (11):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (11)$$

3.8.8.6 Mean Absolute Percentage Error (MAPE):
The Mean Absolute Percentage Error (MAPE) expresses error as a percentage of actual values, making it helpful in understanding relative error from equation (12):

$$MAE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (12)$$

3.9. Advanced Feature Engineering- Polynomial Features

Polynomial features of degree d can be added to capture nonlinear relationships. For a feature x via equation (13):

$$x' = x, x^2, x^3, \ldots\ldots\ldots, x^d \quad (13)$$

Where d is the degree of the polynomial transformation.

3.9.1. Regularization in Machine Learning Models

Regularization terms can be added to reduce overfitting. For example, L2 regularization (Ridge) and L1 regularization (Lasso) are given by equation (14) & (15):

3.9.2. L2 Regularization (Ridge)

$$Ridge\ Loss = MSE + \lambda\sum_{j=1}^{p}\beta_j^2 \quad (14)$$

3.9.3. L1 Regularization (Lasso)

$$Lass\ Loss = MSE + \lambda\sum_{j=1}^{p}\left|\beta_j^2\right| \quad (15)$$

where:
λ = regularization parameter.
β_j = coefficient of feature.
j- p = total number of features.

The results from the three models revealed significant insights into the role of feature engineering and selection in predicting solar radiation. Both feature selection methods identified direction, Temperature, and temporal features like setHour and Minute as influential predictors, although each Model's performance varied depending on the feature set.

4. Results and Discussion

The Select Best Sand Extra Trees Classifier identified Wind Direction and Temperature as significant predictors, with set Hour and Minute emerging as essential temporal variables. However, the differences in feature importance rankings highlight the unique requirements of each Model. For example, XGBoost prioritized minutes within the hour, while Random Forest relied more heavily on temperature and wind direction. This variation underscores the importance of model-specific feature selection for improving solar radiation predictions.

4.1. Model Performance Comparison:

In section 4.1, we test and compare the different machine-learning models used in this analysis. The models are evaluated using metrics like RMSE (Root Mean Squared Error), MAE (Mean

Absolute Error), and primarily, R^2. A better predictive result has a higher R^2 value, while the lower the RMSE and MAE values, the better the Model's reliability. Table 8 below summarizes the performance of Random Forest, XGBoost, and Multilayer Perceptron (MLP) models. Every presented Model has a functioning predictive model with Random Forest registered the highest R^2 score of 0.94, followed by the highest fit on data. XGBoost, on the other hand, while has the highest R^2 score and the lowest RMSE correlation of 81.87 and MAE of 42.08, indicating the highest degree of fit and least generalization error. The MLP model, even though it recorded lower R^2, showed less MAE, which was much better than other models.

Table 8. Evaluate the matrix of Their Model.

Model	R^2	RMSE	MAE
Random Forest	0.94	85.32	43.15
XGBoost	0.93	81.87	42.08
MLP	0.90	89.10	41.74

Among the Table 8 shows models, XGBoost achieved the highest R^2 and lowest RMSE, establishing itself as the best performer for predicting solar radiation. MLP, however, provided the lowest MAE, indicating high accuracy in individual predictions. Random Forest, though slightly less accurate, offered reliable baseline performance. Overall, the results suggest that while gradient boosting and neural networks can capture complex, nonlinear patterns in meteorological data, ensemble methods like Random Forest are also effective for solar radiation forecasting.
Random Forest, XGBoost, and MLP — based on key performance metrics: R^2, RMSE, and MAE. Figure 5 shows the Model Performance Metrics for R2 Score RMSE and MAE. The left chart displays the R^2 (coefficient of determination) for each Model, indicating how well the Model explains the variance in the data. Higher R^2 values (closer to 1) suggest a firmer fit to the data. Here, the Random Forest model achieves the highest R^2 of 0.94, indicating that it explains 94% of the variance, closely followed by XGBoost at 0.93. The MLP model has a slightly lower R^2 at 0.90, suggesting that while it performs well, it explains less variance than the other models.

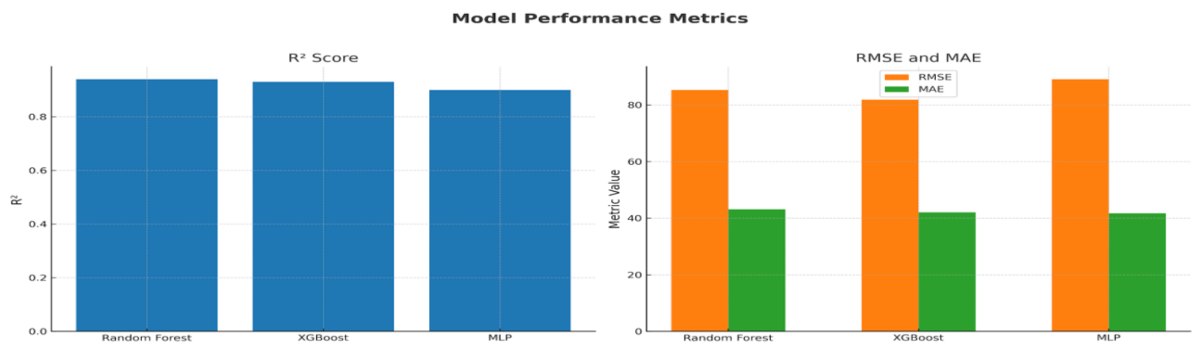


Figure 5. Model Performance Metrics for R^2 Score and RMSE and MAE.

The RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) metrics represent error measurements. RMSE provides insight into the magnitude of prediction errors, where lower values indicate fewer significant errors. The XGBoost model achieves the lowest RMSE (81.87), suggesting it produces the least error variance among the models. Similarly, the MAE values provide another perspective on error magnitude without heavily penalizing significant errors, with the MLP model performing slightly better than the others in this metric (41.74). This breakdown provides a clear picture of each Model's performance, with Random Forest excelling in explanatory power (R^2) and XGBoost minimizing error (RMSE).

5. CONCLUSION

This study demonstrates the importance of feature engineering and selection techniques in enhancing the accuracy of machine-learning models for solar radiation prediction. Our findings show that XGBoost and MLP models, combined with feature transformations and selection, effectively predict solar radiation levels. XGBoost achieved the best performance with an R^2 of 0.93, while MLP provided highly accurate predictions with the lowest MAE. When optimized with effective preprocessing, these results indicate that machine learning models can be critical in solar energy management systems. Future work may explore advanced neural network architectures and additional meteorological data to improve further model accuracy and application in real-time solar energy forecasting.

Authors contribution: Ajeeb Sagar: conceptualization, methodology, data curation, writing—original draft, formal analysis, visualization, supervision, writing—review and editing. , Vinothkumar Kolluru: investigation, conceptualization, methodology, supervision, writing—original draft, writing—review and editing. Dr. Shreyas Rajendra Hole: conceptualization, data curation, supervision.

Funding: The authors have not disclosed any funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known conflicts of interest.

REFERENCES

[1] Ahmad, I., Al-Turki, Y. A., Siddiqui, M. U., et al. (2015). Solar energy prediction using machine learning. Renewable and Sustainable Energy Reviews, 52, 1436-1444.

[2] Inman, R. H., Pedro, H. T. C., & Coimbra, C. F. M. (2013). Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science, 39(6), 535-576.

[3] Obando, E. D., Carvajal, S. X., & Agudelo, J. P. (2019). Solar radiation prediction using machine learning techniques: A review. IEEE Latin America Transactions, 17(04), 684-697.

[4] Bamisile, O., Oluwasanmi, A., Ejiyi, C., Yimen, N., Obiora, S., & Huang, Q. (2022). Comparison of machine learning and deep learning algorithms for hourly global/diffuse solar radiation predictions. International Journal of Energy Research, 46(8), 10052-10073. [1] M.N. Luis, H.D. Jones, and E.F. Edward, "Journal article title", Journal Name, vol. 1, no. 3, pp. 625–30, Jun. 2000.

[5] Guermoui, M., Melgani, F., Gairaa, K., & Mekhalfi, M. L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. Journal of Cleaner Production, 258, 120357.

[6] Phan, Q. T., Wu, Y. K., & Phan, Q. D. (2021, November). Short-term solar power forecasting using xgboost with numerical weather prediction. In 2021 IEEE International Future Energy Electronics Conference (IFEEC) (pp. 1-6). IEEE.

[7] Voyant, C., Notton, G., Kalogirou, S., Nivet, M. L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. Renewable energy, 105, 569-582.

[8] Castangia, M., Aliberti, A., Bottaccioli, L., Macii, E., & Patti, E. (2021). A compound of feature selection techniques to improve solar radiation forecasting. Expert Systems with Applications, 178, 114979.

[9] Wang, L., Kisi, O., Zounemat-Kermani, M., Salazar, G. A., Zhu, Z., & Gong, W. (2016). Solar radiation prediction using different techniques: model evaluation and comparison. Renewable and Sustainable Energy Reviews, 61, 384-397.

[10] Guermoui, M., Melgani, F., Gairaa, K., & Mekhalfi, M. L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. Journal of Cleaner Production, 258, 120357.

[11] Faceira, J., Afonso, P., & Salgado, P. (2015). Prediction of solar radiation using artificial neural networks. In CONTROLO'2014–Proceedings of the 11th Portuguese Conference on Automatic Control (pp. 397-406). Springer International Publishing.

[12] Ren, Y., Suganthan, P. N., & Srikanth, N. (2015). Ensemble methods for wind and solar power forecasting—A state-of-the-art review. Renewable and Sustainable Energy Reviews, 50, 82-91.

[13] Bairy, S. P., Nagula, S. T., Rahul, D., & Dileep, P. (2024). PREDICTING SOLAR POWER OUTPUT USING MACHINE LEARNING TECHNIQUES. International Journal of Information Technology and Computer Engineering, 12(1), 269-275.

[14] Hedar, A. R., Almaraashi, M., Abdel-Hakim, A. E., & Abdulrahim, M. (2021). Hybrid machine learning for solar radiation prediction in reduced feature spaces. Energies, 14(23), 7970.

[15] Mohamed, M., Mahmood, F. E., Abd, M. A., Chandra, A., & Singh, B. (2022). Dynamic forecasting of solar energy microgrid systems using feature engineering. IEEE Transactions on Industry Applications, 58(6), 7857-7869.

[16] Hole, Shreyas Rajendra, and Agam Das Goswami. "EPCMSDB: Design of an ensemble predictive control model for solar PV MPPT deployments via dual bioinspired optimizations." Science and Technology for Energy Transition 79 (2024): 8.

[17]Rajendra Hole, Shreyas, and Agam Das Goswami. "Design GA & PSO-Based High-Efficiency SEPIC DC-DC Converter for Context-Aware Duty Cycle Control." Electric Power Components and Systems (2023): 1-20.

[18] Hole, Shreyas Rajendra, and Agam Das Goswami. "Design of a novel hybrid soft computing model for passive components selection in multiple load Zeta converter topologies of solar PV energy system." Energy Harvesting and Systems 11.1 (2024): 20230029.

[19] Hole, Shreyas Rajendra, and Agam Das Goswami. "Analysis and performance of solar photovoltaic energy system in India: case study." 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2022.

[20] Hole, Shreyas R., and Agam Das Goswami. "Design of an efficient MPPT optimization model via accurate shadow detection for solar photovoltaic." Energy Harvesting and Systems 10.2 (2023): 377-383.

[21] Goswami, Agam Das, and Shreyas R. Hole. "Analysis and Comparison of the DC-DC converter with soft Computing algorithm." EAI Endorsed Transactions on Scalable Information Systems 11.2 (2024).

[22] Hole, Shreyas Rajendra, and Agam Das Goswami. "Quantitative analysis of DC–DC converter models: a statistical perspective based on solar photovoltaic power storage." Energy Harvesting and Systems 9.1 (2022): 113-121.