Volume 2 Issue 3, 2026

DOI: https://doi.org

International Journal of Advanced and Innovative Research
Journal homepage :https://scholarclub.org/index.php/IJAIR/login

Machine Learning-Based Prediction of Solar Energy Generation Using Weather and Environmental Data

Pawan Kumar Menghwar (Coresponding Author)
pawank2k1@gmail.com
National Research University Higher School of Economics, Saint Petersburg, Russia

ARTICLE INFO:

Received:
17 06 2026
Revised:
05 07 2026
Accepted:
25 07 2026

Keywords:
Machine learning; Solar-energy forecasting; PV power prediction; Weather variables; Environmental data;; Explainable artificial intelligence Renewable-energy management

ABSTRACT

Although solar is now playing a growing role in the move to low carbon electricity systems, PV electricity is highly variable, posing a constant challenge in energy planning, grid integration, energy storage and reliable electricity supply. The solar output is greatly affected by the variations in the atmospheric and environmental conditions like solar irradiance, temperature, humidity, cloud cover, wind speed, precipitation, seasonality and others which are location dependent. Therefore, it is important to understand the interaction between these variables and the generation from solar sources to enable more context-specific forecasting to be developed. There has been a recent surge of interest in machine learning (ML), given its data-driven, nonlinear, and time-dependent architectures, able to uncover complex relationships in meteorological and historical energy information (Alcañiz et al. 2023; Gaboitaolelwe et al. 2023). At the same time, the advancements of ML-based solar forecasting have raised concerns on the quality of the data, contextual dependence, complexity, interpretability and transferability of the models to different climatic and geographical conditions.This research study explores the current literature on the application of ML techniques to forecast solar-energy production based on weather and environmental information. The study is designed in an interpretivist, exploratory and qualitative research genre, using a documentary research approach that integrates peer-reviewed scholarly literature, technical publications and authoritative evidence related to the solar forecasting with ML. The study does not involve a new computational experiment, but rather uses a qualitative document analysis and thematic analysis to uncover patterns, conceptual relationships, methodological issues, and practical implications in the current research literature. The synthesis identifies five related observations. Weather and environmental variables are fundamental predictors since they directly affect solar resource availability and PV output, with variations in atmospheric conditions. Second, ML methods, such as traditional ML, ANNs, and deep-learning methods, have significant potential to identify nonlinear and complex patterns that are hard to capture using conventional methods (Alcañiz et al., 2023; Tian et al., 2023). Third, the usefulness of the predictions is critically dependent on the quality of the data, the temporal resolution, the geographical context, climate characteristics and environmental variability. Fourth, the greater the complexity of the model, the more important it becomes to be interpretable and explainable, especially when their forecasts feed operational and energy-management decisions. Recent studies also highlight the need for data-driven learning to enhance contextual robustness and generalization (de Oliveira Santos et al., 2024). Last but not least, it is important to *recognize that ML-based forecasting is not a standalone solution, but a decision support tool that can be integrated into renewable-energy planning.The study finds that, to be effective, solar-energy prediction must be more than just choosing complex algorithms. It needs variables that are environmentally relevant, reliable and context sensitive data, clear analytical processes and careful consideration of climatic and geographical variables. Therefore, future research should focus on approaches that are explainable, hybrid, physics-informed, and uncertainty-aware and can provide reliable and responsible management of renewable energies.*

INTRODUCTION

Solar energy has gained a prominent role in the global journey towards sustainable and low-carbon energy systems. Solar PVs have developed tremendously quickly because it can produce electricity without burning fossil fuels directly from the abundant solar resources. The share of solar energy is growing important in the context of the countries' efforts to reduce GHG emissions, to make their energy supply more secure and diversified. But unlike traditional generation methods which are often able to be dispatched upon demand, solar generation is intrinsically linked with the changing environmental and atmospheric conditions. The variability poses an important forecasting problem as electricity systems continuously have to match supply and demand. This makes the understanding and forecasting of solar-energy production fluctuations crucial for renewable-energy planning, grid integration, energy storage, and the operation of sustainable energy systems (International Energy Agency [IEA], 2024; International Renewable Energy Agency [IRENA], 2024).

The variability of the solar resource and environment has a strong link to variability in solar generation. Temperature, cloud conditions, humidity, wind, precipitation and atmospheric properties are all significant factors affecting solar irradiance, which affect the amount of solar radiation hitting the PV surface, and the efficiency of solar radiation being converted to electricity generation. These factors are not mutually independent. For instance, changing the irradiance can result from cloud movements and temperature can affect PV conversion efficiency. The environment, moisture, precipitation and aerosols, may also affect further the availability and quality of solar radiation. In addition, geographies vary in the way solar is generated, as the geographical location, seasonality, local climate and time trends differ. This does not only refer to the electricity generation, however, but also to the environmental context in which the electricity is produced, so it is important to take a wider environmental perspective (Inman et al., 2013; Yang et al., 2020).

These traditional forecasting techniques have proved beneficial to the understanding of renewable-energy variability but can be problematic if the link between environment and solar generation is highly nonlinear, non-stationary and context dependent. Conventional statistical methods are usually based on assumptions about the structure of the relationships in the historical observations. Physical forecasting techniques, however, can be used to model atmospheric processes and solar processes, and require detailed environmental information and considerable computing or modelling resources. This has inspired researchers to seek solutions based on information, which are able to learn complex relationships from the past historical and meteorological data. In this context, machine learning (ML) is becoming more and more popular as a tool for forecasting solar energy, since it can identify complex relationships between multiple variables without priorly establishing a relationship (Voyant et al., 2017; Alcañiz et al., 2023).

All algorithms (regression, decision trees, random forests, support vector machines, artificial neural network, gradient boosting, etc.) and any other more advanced deep learning architectures fall under the umbrella of machine-learning methods. The growing popularity of the applications in solar forecasting is an evidence of the understanding that solar generation is a complex interaction of temporal, meteorological, geographical and environmental factors. The use of a mixture of predictors can be captured by a combination of predictors and/or by patterns that are hard to model using simpler forecasting approaches, which can be done through ML-based approaches. This possibility has been further extended by deep-learning methods which allow to extract complex temporal and spatial representations from large and heterogeneous datasets. In recent years, a general trend in forecasting, moving away from traditional forecasting methods to ML and deep learning, and even hybrid approaches, has been described with a focus on meeting the complexity of renewable-energy systems (REs) (Alcañiz et al., 2023; Gaboitaolelwe et al., 2023).

However, the progress has not yet been reflected in providing reliable and representative solar-energy forecasts. One of the many issues is data quality. Machine-learning systems are dependent on the data that they are presented with, so missing observation, inconsistent measurements, too coarse a time scale, sensor errors, or underrepresentation of environmental conditions can affect the usefulness of the predictions that can be obtained from them. Therefor, the range of weather and environmental variables to choose from is not only a technical consideration, but also crucial to understanding the link between the atmosphere and solar generation. The explanatory importance of irradiance, temperature, humidity, cloud cover, wind speed, precipitation, air pressure, historical generation patterns, seasonality and geographical characteristics could vary in different climates. Therefore, the transfer of the effectiveness of an ML approach trained from one environment to another is not guaranteed (Tian et al., 2023).

The challenge of generalizability and transferability is also associated. Solar generation patterns are dependent upon latitude, climate, cloudiness, atmospheric conditions, seasonal changes, PV technology and local environmental conditions. Patterns learned by models in one geographical or climatic context may thus be difficult to apply to other contexts. This is particularly important for locations where there is limited quality or long-term PV generation or meteorological data is available. In such a context, a sophisticated algorithm might look good, but be difficult to operationalise because of a lack of or non-uniform environmental information required for reliable learning and validation. Therefore, the literature suggests that solar forecasting is not an algorithmic battle between different ML methods or approaches, but rather a problem that needs to be solved in a context-specific manner (Gaboitaolelwe et al., 2023; de Oliveira Santos et al., 2024).

A new challenge is interpretability/explainability. There are numerous sophisticated ML models that can model very complex relationships, and the complex decision making process within these models may not be easily discerned by practitioners. This poses an important conflict between the sophistication of the predictions and their believability in energy systems. Operators of the grid, renewable energy planners, policy makers and system managers need to forecast to inform generation scheduling, storage, grid balancing and investments decisions. However a prediction that is not interpretable may not be of practical use, especially if environmental conditions shift outside of the range of the data used to make the prediction. This has made explainable artificial intelligence (XAI) more and more pertinent as it aims to render more transparent and comprehensible complex computational prediction while maintaining the advantages of powerful learning techniques.

However, the literature includes a lot of empirical content directed towards algorithmic performance comparison and towards the development of new models. This study has contributed to the understanding of the technical aspects, but less work has been devoted to conceptual links between environmental variables and different ML methods, data quality, climatic variability, explainability and deployment of energy systems in the real world. Therefore the present study tries to fill this gap from the qualitative documentary point of view. It will instead review the type of weather and environmental information envisioned by ML-based solar forecasting, and the effects on its usefulness, based on the existing scholarly and authoritative literature. This is important because this is not about discovering new predictive success, but about developing an understanding of what is already known in the field in an interpretive manner.

Thus, the approaches taken to research in this study are interpretivist, exploratory and qualitative. It applies published studies, technical publications, energy reports and other authoritative documents, as sources of knowledge to be critically interpreted and thematically synthesized. The documentary study is organised under five related themes: weather and environmental variables as primary drivers; modelling complex patterns in solar generation using ML techniques; data quality, data variability, data contextual dependence; explainability, transparency and trust; practical applications in renewable-energy planning and sustainable management. The study not only pinpoints used algorithms and the "why" and "how" of the significance of environmental information, but also the influence of contextual factors on the forecast that is possible according to ML, as well as the challenges of responsible use of algorithms.

This study's research problem is not that there are no ML techniques that can be used for solar forecasting, but rather that there is a lack of understanding of the interplay between different environmental conditions, data characteristics, model complexity, interpretability, and geographical context that would enable us to understand how these can affect the usefulness of ML-based solar forecasting. The goal of this study is to critically review the current evidence on the use of weather and environmental data for solar-energy prediction when employing ML techniques and point to the key conceptual and practical issues.

What are the advantages and disadvantages of using ML to predict solar energy?(3) What are the weather and environment parameters that play a role in the prediction of solar energy? (3) What do recent studies reveal regarding the accuracy of pattern representation using ML in the context of complex and context-dependent pattern of solar generation? (3) What are the effects of data quality, environmental variability, seasonality and geographical context on the use and transferability of ML-based forecasting? It is emphasized that interpretability, explainability, and transparency are crucial to fostering trust in solar prediction using ML and for planning renewable energy and sustainable energy management.

The study is thus conceptual/interpretive in nature. It connects and contrasts various types of documentary evidence, and maps technical dimensions of ML forecasting to questions of data reliability, contextual adaptation, explainability and variability of the energy system and its decisions. The study implies that the usefulness of the ML-based solar forecasting is not based on the complexity of the algorithm. Instead, the environmental relevance of data, modelling in a context, transparency of the interpretation and an awareness of uncertainty are needed to be able to make something of it. This can provide a research and practical perspective of ML as a tool that can be used as a decision support aid or mechanism along with physical knowledge and professional judgment in increasingly complex and variable renewable-energy systems.

LITERATURE REVIEW

Solar energy is a clean, renewable, and abundant resource

Solar power has been an integral part of the global shift towards renewable and low-carbon energy sources for electricity production. While the PV generation expansion brings numerous environmental and energy security advantages, solar generation is weather sensitive, posing a unique forecasting challenge. Power generation from solar cannot be independently controlled like conventional power generation. Electricity generation may vary significantly over time due to changes in irradiance, cloud motion, temperature, atmospheric conditions, and seasonal changes. Therefore, accurate forecasting becomes crucial for maintaining electricity supply and demand balance, coordinating renewable-energy storage, managing grid operations, and planning for future renewable-energy capacity (International Energy Agency [IEA], 2024, International Renewable Energy Agency [IRENA], 2024).

The literature indicates that predicting solar generation is especially difficult because the underlying process is affected by multiple variables, not a single one, that are not necessarily easy to predict. Although traditional statistical and physical forecasting methods are not as effective as these newer techniques, the demand for representations of historical or atmospheric relationships that are interpretable remains. Statistical methods, however, can have difficulty with high levels of non-linearity in the relationship, and physical methods rely on detailed information about the environment and complex modelling assumptions. These constraints have led to the growing interest in data-driven machine-learning (ML) techniques that can learn relationships based on previous generation and meteorological data (Voyant et al., 2017; Alcañiz et al., 2023).

Noticeably, current studies have not indicated that ML can replace traditional forecasting. In contrast, the literature increasingly describes forecasting as a methodological continuum, where different methods, such as statistical, physical, ML, deep-learning and hybrid methods may deal with different aspects of solar variability. Qualitative examination is of particular relevance here, since different algorithms do not necessarily perform best on a universal basis, but rather how they perform under the environmental variability of solar generation.

Weather and Environmental Determinants of Solar Generation

Many solar forecasting methods rely on the weather and environmental data. Naturally, solar irradiance is one of the most important predictors as output from a PV system is directly linked to incoming solar radiation. However, irradiance exists beyond the vacuum of its own world. The efficiency of PV-system can be affected by temperature, and the radiation can change quickly with clouds. Other relevant contextual information that may be obtained from humidity, precipitation, atmospheric pressure, wind velocity and atmospheric composition can also be found. The temporal and geographical distribution of solar energy production is also affected by historical generation, seasonality, geographic location, and the local climatic characteristics.

The recent forecasting literature has begun to take on board the need to understand the meaning of environmental variables in the context in which they are applied. The relationship between a weather variable and PV output can vary depending on the climatic regime. Likewise, the dynamics of clouds could be more relevant for short-term forecasting while seasonal and geographical patterns could be more relevant for long-term forecasting. This context dependency raises the question whether a universal list of predictors would also be applicable to all solar-energy systems (Tian et al., 2023).

The literature then suggests a view of forecasting as wider than just the technical inputs of environmental data but as a way of representing the physical context of energy generation. The same point can also be made about the lack of attention that is sometimes paid to whether chosen environmental variables are a good representation of the atmosphere in the chosen area. Therefore, more complex models do not necessarily lead to improved prediction – sometimes more meaningful representation of the environment and careful consideration of the geographical and climatic context is necessary.

Solar Prediction with Machine-Learning Approaches

Given nonlinear relationships that could be hard to model using conventional statistical assumptions, machine learning has gained significant interest in the research community for solar-energy generation. Relatively simple regression models are available for modelling the relationships between environmental predictors and generation. Non-linear interactions are possible in decision trees and random forests, and they are relatively easy to interpret. Non-linear prediction can be achieved using SVM-based kernel learning and more complex relationship between meteorological and historical variables can be represented using ANNs.

Recent research has been more focused on ensemble learning, gradient boosting, and deep-learning techniques. The methods are appealing, because they can handle complex relationships and in some cases time and dependency in large sets. Given the time-series nature of solar generation, deep-learning architectures, especially those based on recurrent networks and other sequence-oriented methods, have emerged as key players in forecasting. The research on solar forecasting continues to evolve toward ML and deep-learning techniques, and hybrid models that are hybrids of several modelling techniques (Alcañiz et al., 2023; Gaboitaolelwe et al., 2023).

But, there are also warnings against seeing the sophistication of the algorithms as a sign of practical superiority. More complex models may need larger amounts of high quality data, more computational resources and more complex validation procedures. They can also be hard to understand. In cases where transparency is key, for computational efficiency, or where data is not amply available, simpler methods can be desirable. So, there is a recurring tension between modelling complexity, predictive power, interpretability, and feasibility between the literature.

Hybrids of these are especially relevant: they try to combine complementary strengths. Hybrid methods and physics-informed methods aim to bring information about the physical behavior of solar-energy systems into computation models, instead of using physical knowledge or data-driven learning exclusively. This is a trend that shows how the field of solar forecasting has evolved over the years with more and more understanding that it is an inherently physical problem, and therefore data-driven approaches can help, but not at the expense of losing domain knowledge.

The learning objectives for this unit are to study data quality, feature selection, temporal and spatial variability

In the literature, data quality is one of the most critical underlying aspects of solar prediction using ML that are often overlooked. Machine-learning systems learn from the information available; therefore missing observations, measurement errors, inconsistencies in the data, inappropriate temporal resolution, and insufficient past coverage can affect the quality and reliability of learned relationships. Good quality environmental information is especially relevant at high speed changing atmospheric conditions.

Interpretation is necessary in the context of the feature selection. The more variables you add, the better a forecasting framework won't necessarily be. Variables should be meaningful and represent relevant physical or environmental processes, and should have a meaningful relationship with solar generation. Irradiance, temperature, cloud characteristics, humidity, wind, precipitation, and past generation may also be useful but their relative importance can differ from place to place, season to season and day to day.

This places an extra challenge into the mix because of the temporal and spatial variation. The output of solar generation fluctuates during the day and from one season to another, and the weather system is different in various geographical locations. A climate or geographic model may thus not be easily translatable to another climate or geography. This generalizability issue is especially critical in regions that have various solar-resource characteristics, environmental conditions, or data infrastructures. Recent studies have shown that model transferability cannot be taken for granted, just because an algorithm has proven successful in another context (Gaboitaolelwe et al., 2023; Tian et al., 2023).

This literature also reveals a methodologically missing link. A considerable number of studies focus on comparisons within given data sets and environments and make it hard to assess whether the findings for a specific ML method would also apply if the environment were different. A qualitative synthesis can be helpful here as it can help combine these disparate findings and consider the conditions in the context that impact the usefulness of the model, not as performance independent from place and context.

Explainability, Sustainability, Practical Deployment and Future Directions

Complication of the ML-based solar forecasting has led to the growing interest in explainability and transparency. Forecasting is only useful if it yields an output for renewable-energy planners and system operators, but if it does yield an output it is also useful to the stakeholders if they understand the environmental and operational factors associated with that output. Explainable AI can therefore help foster trust by making predictions and their relationships with the predictors more understandable. This is particularly significant for use in scenarios where models are used in an operational setting with decisions regarding grid balancing, storage, maintenance, and/or energy management relying on forecasting information.

However, the practical situation in which the ML systems are deployed must also be taken into consideration for sustainability. The feasibility of implementing advanced forecasting methods outside of research settings can depend on the computational needs, data infrastructure, maintenance requirements, transferability, and accessibility. In certain areas of the planet, such as those undergoing development and climate change, having access to reliable and historical PV data could prove to be a more basic problem than which algorithm to choose.

Increasingly, the literature suggests a path of explainable, hybrid, physics-informed and uncertainty-aware approaches as promising directions. These methods recognize that there is inherent uncertainty in solar forecasting due to the environment and that no computational model is applicable in every case. Instead, forecasting should be a blend of data-driven learning, physical understanding, and decision support transparency, as described by de Oliveira Santos et al. (2024).

In summary, a clear shift from traditional forecasting to increasingly advanced ML, deep-learning, and hybrid approaches is evident in the literature. However, there are some critical gaps yet to be filled. The existing research is mostly technically oriented, geographically biased and is rather model development/performance comparison oriented. Less focus has been placed on weaving together the various aspects of the weather variables, environmental uncertainty, data quality, geographical context, model interpretability, and practical deployment as a composite problem. To fill this gap, the current qualitative study uses an interpretivist documentary analysis of the scholarly and authoritative evidences. Rather than test a specific algorithm, it summarizes common trends in the literature to discuss how environmental conditions affect the utility of ML-based solar prediction and why the usefulness of a model varies from context to context and how responsible deployment and interpretability are crucial to the sustainable management of renewable energy. This is an approach that gives more context to solar forecasting through a lens of machine learning as a decision-support tool that is part of a larger environmental and energy system and not just an algorithm.

RESEARCH METHODOLOGY

Research Design and Philosophical Orientation

The current study uses qualitative, interpretivist and exploratory research design to explore the existing scholarship on the machine-learning-based prediction of solar-energy generation using weather and environmental information. The study is explicitly of a documentary type and not of an experimental one. It doesn't train and test a new machine-learning model, gather numerical prediction data, perform statistical analysis, or test the accuracy of algorithms. Rather, it decodes and combines information from authoritative and scholarly documents. An interpretivist approach is suitable when the study aims to gain an understanding of how researchers perceive the relationship between the environmental conditions, machine-learning techniques, data quality, prediction uncertainty, explainability, and renewable-energy applications. Qualitative research is especially well suited for the goal of developing contextual and interpretive understanding, not establishing statistical relationships, (Creswell & Poth, 2018).

Documentary Data Sources

The documentary evidence includes peer-reviewed journal articles, systematic and scholarly reviews, authoritative renewable-energy reports, technical publications, policy documents and other institutional sources. Literature review focusing on solar-energy forecasting, PV generation, meteorological and environmental predictors, machine learning, deep learning, hybrid approaches, explainable AI, data quality, and renewable-energy planning. Recent literature (since 2020) is emphasized given the quick evolution of the forecasting challenges in the fields of renewable energy and ML. Important conceptual and/or methodological foundations are retained in seminal publications.

The documentary method enables evidence from various kinds of sources to be taken together; it also acknowledges that the purpose and form of authority of different kinds of sources, such as academic articles, technical reports, and institutional publications, may differ. The analysis does not treat all documents as being on equal footing, but takes into account their relevance, scholarly credibility, transparency in methodology and contribution to the research questions.

The selection process and its criteria

Documents are chosen by purposive selection, which involves identifying sources that are intentionally selected due to their substantive information that fits with the research goals. Preference is placed on papers that directly address or critically discuss ML-based solar forecasting, as well as weather or environmental variables. In the context of solar energy, the temperature, relative humidity, cloud cover, wind speed, precipitation, and atmosphere are taken into account, along with seasonality, geographical variation and historical environmental patterns.

The inclusion criteria include: (a) scholarly or authoritative publications relevant to solar-energy generation forecasting, or forecasting of solar-energy generation; (b) studies that mention or discuss ML, deep learning, statistical, physical or hybrid forecasting approaches; (c) literature pertaining to meteorological or environmental predictors; (d) publications that discuss data quality, generalizability, interpretability, uncertainty or practical implementation; and (e) recent literature (supplemented with seminal literature when necessary). Documents which are not relevant to solar forecasting, are not significant enough to the research questions, or do not significantly contribute to understanding the environmental prediction factors are filtered out. Sources that have no known source or the source is not verifiable with the bibliographic data are also excluded.

Document Analysis Procedure

The analysis adopts the same sequence as in the structured qualitative document analysis. Firstly, the information from the selected literature is read many times to familiarise yourself with the evidence and to establish patterns in the concepts. A few passages are then analysed with respect to the research questions. Special attention is paid to similarities, differences, conflicting findings, methodological issues and conditions of the studies throughout the literature.

The analysis makes the distinction between direct evidence that was reported by previous researchers, and interpretations that were developed by the present study. For instance, if the currently available research suggests a variable such as irradiance or cloudiness to be an important predictor, it is taken into account along with other variables that are discussed, including temperature, humidity, wind, precipitation, seasonality and geographical characteristics. Likewise, statements about the applicability of specific ML methods are understood with reference to the datasets, context and timing of the data, forecasting goals and procedure conditions reported in the original study.

Thematic Analysis and Coding

The principles of Braun and Clarke's (2006) thematic analysis are used to analyse the documentary evidence thematically. The process is one of familiarization with the documents, initial coding, pattern finding, theme review, theme articulation, and final synthesis of the documents' meaning. Much of the coding is conceptual and descriptive, and emphasizes the ideas that are repeated, including those of environmental predictors, nonlinear relationships, data quality, contextual variability, model generalizability, interpretability, uncertainty, and practical deployment.

Codes are iteratively compared to form larger overarching analytical themes related to concepts. Thematic structure aims to reflect the complexity of the solar forecasting science by not considering algorithms and environmental variables as individual technical challenges. This method is appropriate for the exploratory nature of the study since the qualitative nature of the data allows for important patterns and relationships to emerge, rather than being developed on a priori quantitative assumptions.

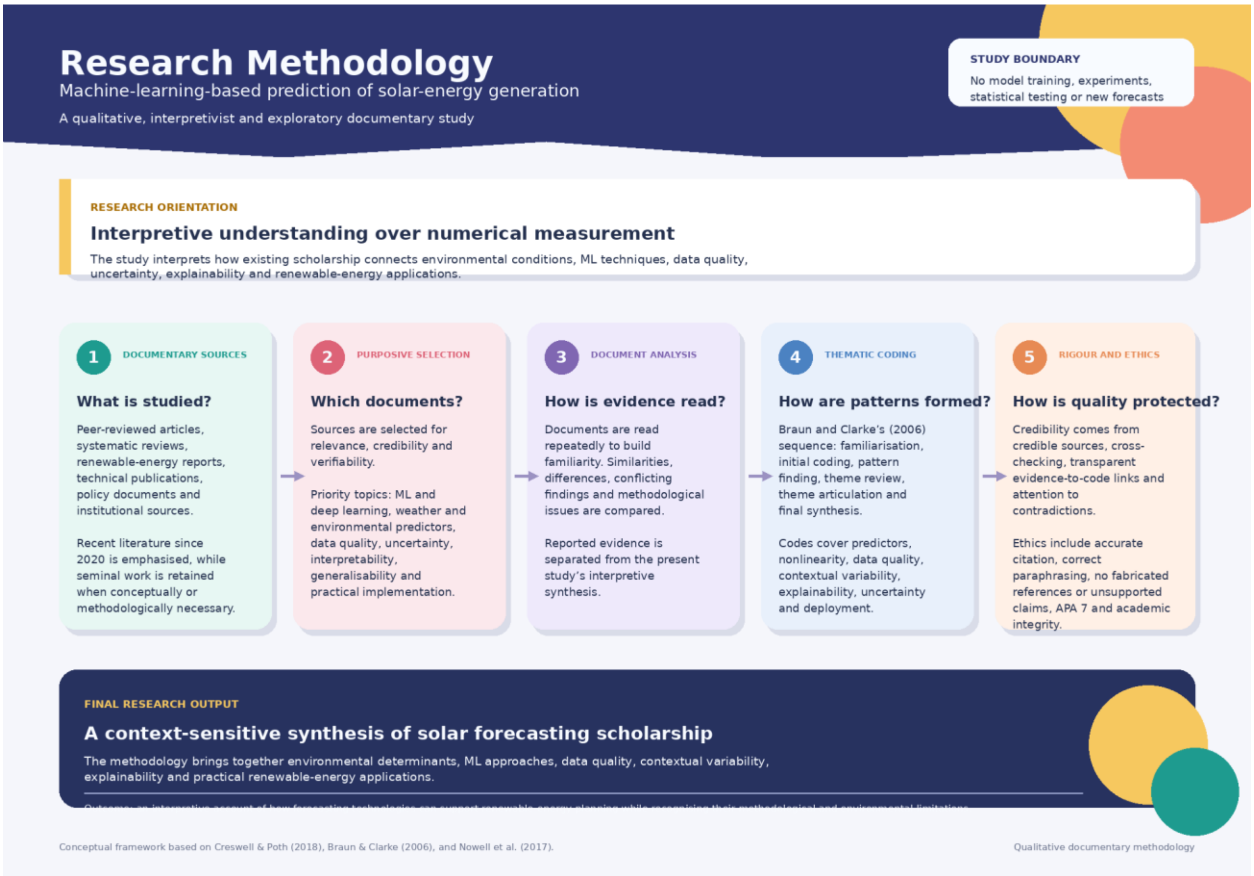
Students will develop analytical rigour and credibility

To enhance credibility, transparency and analytical depth several strategies are employed. First, relevant and credible sources are used based on explicit relevance and credibility criteria. Second, interpretations are constantly cross-checked between multiple documents and not isolated to a single document. Third, the distinction between the interpretation of the source reported evidence and the researcher's interpretive synthesis is clearly drawn. Fourth, contradictory or divergent findings are given attention rather than selective attention to evidence supporting a pre-determined conclusion. Analytical links that are made explicit from documents to codes and themes also improve transparency and dependability (Nowell et al., 2017).

Ethical Considerations

The study does not include human participants, interviews, questionnaires or private personal information as it is based solely on publicly available scholarly and institutional documents. However, it is still possible to conduct ethical research. Sources are appropriately and accurately cited; ideas are paraphrased and cited correctly; conclusions are not attributed to authors that are not documented in their publications. Care is taken to avoid fabricated references, unsupported claims, or invented findings. The study also addresses the issue of intellectual-property rights and academic integrity, offering proper APA 7th edition citations in the entire document.

In general, this approach allows a systematic and interpreted study of the existing knowledge on the ML-based prediction of solar energy. The study combines evidence of the environmental determinants, machine learning methodologies, data quality, contextual variability, explainability and practical applications to create the qualitative understanding of how forecasting technologies can help in renewables planning, and acknowledge their environmental and methodological limitations.



Findings and Discussion

After conducting the literature review, a documentary analysis process identified five themes related to the prediction of solar energy generation using machine learning (ML) and weather and environmental data, which are interconnected. The ability to make useful forecasts is not a purely technical issue; the synthesis indicates that the match between meteorological conditions, algorithmic capabilities, data quality and context variability, as well as the interpretability and applicability of forecasts to the practical requirements of the energy systems is important for making useful forecasts. The results thus indicate that ML is not just a computational forecasting tool, but a data-based approach to managing renewable energy in general.

Theme 1 is the core predictors of weather and environmental variables

Weather and environmental data are key to understanding solar-energy variability – one of the most consistent findings in the literature. The natural variations in solar irradiance are naturally related to atmospheric and environmental conditions, and other parameters like temperature, humidity, cloud cover, wind speed, atmospheric pressure, and the history of weather information are also indicative of variations in the photovoltaic (PV) output. The observations of meteorological variables are one of the main sources of information needed for solar-radiation prediction as repeatedly shown in the reviews of solar-radiation prediction (El-Hendawy et al., 2021). A more recent study also shows that some variables, such as ambient temperature, pressure and humidity, were also used with solar variables in the deep learning forecasting studies. (ScienceDirect)

Qualitative synthesis suggests, however, that these variables are not to be viewed as equally influential in each context. They rely on the local atmospheric processes, geographical properties, seasonality and the forecasting period. Cloud motion is another dynamic that can cause rapid changes in the amount of radiation reaching the ground, and is not well covered by the standard weather observations. Therefore, it is crucial to combine solar sky images, satellite data, a weather station, and other sensors to capture the changes of clouds, which can be seen from the recent research on computer vision-based solar forecasting (Huang et al., 2023). (ScienceDirect)

This suggests that for a good solar prediction, information about the context must be taken into account, rather than fixed variables which are generic to all contexts. As such, it is recommended from the literature that the quality and relevance of the meteorological inputs can be as significant as that of the sophistication of the ML algorithm itself.

Theme 2;The Solar-Generation Pattern Catching via Machine Learning pattern is explored

The second theme is the ability of ML to learn non-linear and context dependent relationships between the environment and solar generation. When the dependency of weather conditions and PV outputs changes over time, seasons, space or atmosphere, the traditional forecasting methods may be questioned. ML approaches can be an alternative since they can learn complex patterns from multi-dimensional datasets without having to prescribe all the relationships beforehand.

The literature review suggests that there is a shift of interest towards neural networks, ensemble learning, recurrent networks, convolutional methods, and hybrid models. Some deep-learning research also references LSTM and CNN architectures, which are able to represent temporal dependencies and spatial or visual information, respectively (Huang et al., 2025). Concurrently, there is no evidence in the literature that the more complex the model is, the more useful it will be. A new recent review has highlighted the trend towards hybrid learning models, where physical knowledge is combined with data-based learning. According to the authors, these models can be applied without violating the principles of solar energy known, and they can model the nonlinear behaviors. (ScienceDirect)

But from the current interpretative point of view, the main benefit of ML is not just the higher predictive accuracy. Its bigger contribution is its ability to accommodate not easily representable relationships. The greater complexity of the models introduces new challenges including computational complexity, model interpretability, model transferability and a need for rich data sets. Based on the results found, algorithm selection should not only be influenced by the algorithms' technical aspects but also by the context-based forecasting and practical considerations.

Themes 3: Data Quality, Variability and Contextual Dependence

A third, and particularly important, theme is that the quality and context-relevance of the data available for prediction by ML methods is important. It is found that the following issues are important ones discussed in reviewed literature, which are frequently occurring: missing observations, error in measurement, varying sampling rates, limitations of sensor, noisy weather data, and variations in data sources. The researchers make use of both surface observations and satellite data and numerical-weather or reanalysis datasets in this review of the in-depth solar radiation forecast worldwide and each of these datasets has its merits and demerits (El-Hendawy et al., 2021). (ScienceDirect)

The synthesis also suggests that a successful model in one geographical or climatic area cannot be extrapolated to other areas. Solar generation is dependent on the solar conditions, which vary by latitude, cloud regimes, atmospheric conditions, terrain, seasonality, and other local environmental characteristics. This is a huge issue of contextuality. The forecasting relationship acquired from another area might need modification before being applied to an alternate area.

The latest reviews also identify the persistence of the issues in the data quality and generalization and the geographical variability of the solar forecasting research. This is a particular point of interest in regions where the meteorological data is not freely available for use and in developing areas where long-term observations of high-resolution meteorological data may be limited. The qualitative interpretation is thus that ML is not data neutral; that is, the reliability of intelligent prediction is intimately bound up with the representativeness, continuity and contextual appropriateness of the data used.

This theme aims to develop understanding of explainability, transparency and trust in ML-based solar prediction.The theme's goal is to develop understanding of explainability, transparency and trust in ML-based solar prediction.

Theme 4 the growing importance of explainability

A very complicated ML system may give an accurate prediction, but provide little information about why a given prediction was made. It creates problems for energy planners and technical users who must determine whether there is any mismatch between the results of the model and physical expectations.

Recent research on interpretable ML emphasizes the importance of not relying on black-box ML models, as they may reduce the trustworthiness of users and make it difficult to explain the meteorological processes behind the prediction (Yang et al., 2024). Solar PV forecasting using Explainable AI (SHAP, feature-importance, attention visualization and others used for explanation) is an example of a new field of study that's gaining in importance and is specifically devoted to systematic analysis for the purpose of EAI. Note that that review also points out current problems with explanation stability, correlated variables, computational cost, and explanations actually reflecting model reasoning. (ScienceDirect)

For this, the explainability is called more than a technical addition in the current synthesis. A need for a meaningful human interaction with forecasting using ML. If planners and operators can articulate what aspects of the environment they believe impact on a prediction, they might be more able to think about the unusual cases that can be output from their algorithms and to consider the algorithms as a complement to the professional knowledge. The literature then suggests moving away from the question of whether a model is a good predictor to the question of whether the model is a model that is understandable, physically plausible and operationally trustworthy.

Theme 5: Practical Integration for Renewable-Energy Planning and Sustainable Management

Finally, the fifth theme demonstrates the link of solar forecasting to planning and management of renewable energy. The forecast information can be utilized for decision making regarding grid operation, energy balancing, storage, demand management, PV system operation and other renewable energy planning. As VRE becomes an increasing share of electricity systems, reviews emphasize the need for accurate forecasts in solar power. (IEEE Xplore)

But literature indicates that technical forecasting ability does not necessarily lead to real value added. The models to deploy need to be computationally tractable, sensitive to local conditions, transparent to a certain extent, and compatible with existing energy-management models. The relevance of hybrid approaches is especially highlighted as they aim to integrate the flexibility of ML with the incorporation of physical knowledge, which can enhance generalization and interpretability (Santos et al., 2024). (ScienceDirect)

This is particularly important in the context of developing countries. The significant complexity of the forecasting systems can be difficult to implement due to the few infrastructure facilities for monitoring and reporting, the lack of a coherent data set, climatic diversity, and insufficient computing power. The interpretive synthesis thus lends itself to a context sensitive thinking in technological uptake: sometimes simpler, easier-to-understand and locally adapted models will be more useful than sophisticated systems that need extensive data and computational facilities,Machine learning is a socio-technical process which can't be separated from its social and technical aspects. The basis is weather and environmental variables, the ability of ML to find complex relationships, the data quality and geographical context which makes it reliable, the ability to explain the results to gain trust and the ability to integrate it into a usable application that creates operational value. The literature thus suggests the future of successful solar forecasting will not only be about how sophisticated the algorithms are, but also about how environmental data is of good quality, how the algorithms are contextually smart, how physical knowledge is involved, how transparent they are, and how they are related to the practical energy-management needs.

Table: Synthesis of Findings on Machine-Learning-Based Solar Energy Forecasting

Theme	Main Finding	Key Variables, Methods and Evidence	Challenges and Limitations	Interpretation	Practical Implications
Theme 1: Core Weather and Environmental Predictors	Weather and environmental variables are fundamental for predicting solar-energy generation because photovoltaic output changes according to atmospheric and environmental conditions.	Important predictors include solar irradiance, temperature, humidity, cloud cover, wind speed, atmospheric pressure and historical weather data. Weather stations, satellite data, sky images and other sensors can help capture changes in solar radiation and cloud movement. Relevant evidence includes El-Hendawy et al. (2021) and Huang et al. (2023).	The influence of each variable differs according to geographical location, climate, season, atmospheric conditions, terrain and forecasting horizon. Standard weather observations may not capture rapid cloud movement accurately.	Effective forecasting requires context-relevant meteorological information rather than a fixed set of variables applied universally. Data quality and relevance may be as important as algorithmic sophistication.	Forecasting systems should combine multiple data sources, such as weather stations, satellite imagery, sky cameras and environmental sensors, where available. Feature selection should be adapted to local conditions.
Theme 2: Machine Learning for Detecting Solar-Generation Patterns	Machine learning can identify nonlinear, complex and context-dependent relationships between weather conditions and solar-generation output.	Frequently studied approaches include neural networks, ensemble learning, recurrent neural networks, LSTM models, CNNs and hybrid models. LSTM models are useful for temporal relationships, while CNNs can process spatial and visual information. Recent studies also examine hybrid models that combine physical knowledge	More complex models do not automatically produce more useful forecasts. They may require large datasets, greater computational power and specialist knowledge. Additional concerns include limited interpretability and weak transferability between locations.	The main contribution of ML is its ability to represent relationships that are difficult to define using traditional statistical or physical models. However, algorithm selection should be based on both technical performance and practical requirements.	Models should be selected according to the forecasting context, data availability, required accuracy, computational resources and ease of deployment. Hybrid models may provide a balance between flexibility, physical consistency and generalisation.

		with data-driven learning, including Huang et al. (2025).			
Theme 3: Data Quality, Variability and Contextual Dependence	The reliability of ML-based solar forecasting depends strongly on the quality, continuity, representativeness and geographical relevance of the data used for training and testing.	Common data sources include ground-based observations, satellite data, numerical-weather prediction data and reanalysis datasets. Important data issues include missing values, measurement errors, inconsistent sampling rates, sensor limitations, noisy observations and differences between data sources. El-Hendawy et al. (2021) and recent review studies emphasise these concerns.	A model developed in one geographical or climatic region may perform poorly in another because of differences in latitude, cloud regimes, atmospheric conditions, terrain, seasonality and local environmental characteristics. High-resolution, long-term meteorological data may also be limited in developing countries.	ML is not data-neutral. Forecasting reliability is closely connected to whether the training data accurately represents the location, climate and operating conditions where the model will be used.	Data should be cleaned, validated, synchronised and regularly updated before model development. Local calibration, transfer learning, regional validation and uncertainty assessment should be considered before applying a model in a new location.
Theme 4: Explainability, Transparency and Trust	Accurate predictions are not sufficient if energy planners and system operators cannot understand why a model produced a particular forecast.	Explainable Artificial Intelligence techniques include SHAP values, feature importance analysis, attention visualisation and other interpretation methods. Research on interpretable ML, including Yang et al. (2024), highlights the growing importance of transparent solar-PV forecasting.	Explainability methods may be affected by correlated variables, unstable explanations, high computational costs and the possibility that explanations do not accurately represent the model's actual reasoning.	Explainability is not merely a technical addition. It supports meaningful human interaction with forecasting systems and helps users assess whether predictions are physically plausible and consistent with meteorological expectations.	Forecasting systems should provide understandable explanations of important predictors and unusual outputs. Combining ML forecasts with expert knowledge can improve confidence, accountability and decision-making.
Theme 5: Practical Integration for Renewable-Energy Planning and Sustainable Management	Solar forecasting creates value when it supports real energy-system decisions, including grid operation, energy balancing, storage, demand management and photovoltaic-system operation.	Forecast information can support renewable-energy planning, energy-market decisions, battery-storage management, load balancing and the integration of variable renewable energy. Hybrid approaches may improve generalisation and interpretability by combining ML with physical knowledge, as discussed by Santos et al. (2024).	High-performing models may be difficult to deploy where monitoring infrastructure, reliable datasets, computing power, technical expertise and energy-management systems are limited. Technical accuracy alone does not guarantee operational value.	The usefulness of a forecasting model depends on its computational efficiency, local adaptability, transparency and compatibility with existing energy-management practices.	Deployment strategies should be context-sensitive. In developing countries, simpler, interpretable and locally adapted models may be more useful than highly complex systems requiring extensive infrastructure and computing resources.
Overall Synthesis	The findings show that ML-based solar forecasting is a socio-technical process rather than a purely computational task. Successful forecasting depends on the interaction between relevant weather and environmental data, appropriate ML algorithms, high-quality and context-specific datasets, explainable predictions and practical integration with renewable-energy management. Therefore, future progress should focus not only on improving algorithmic accuracy but also on physical consistency, transparency, local				

	adaptability, operational usefulness and sustainable implementation.

CONCLUSION

The dependency of solar energy production and the weather and environment conditions variability is complex and needs a machine learning (ML) approach in understanding and predicting production of solar energy. The current qualitative documentary analysis shows that solar forecast should not be considered as an algorithm selection problem. Rather, in order to make reliable predictions, it is necessary for solar irradiance, temperature, humidity, cloudiness, wind, seasonal variability, environmental variability, data quality, contextual adaptation, and model interpretability to interact. The literature reviewed showed that ML offers promising applications for forecasting renewable energy, which are however dependent significantly on the quality and relevance of the information that models are based on.

Weather and environmental parameters are used as the foundation for prediction of solar generation with one of the main findings. Also, irradiance has direct impact on the availability of sunlight, whereas temperature, humidity, clouds, wind and other atmospheric conditions have impact on PV output variation, and so are more important. Not all such relationships are the same from place to place and time to time. Environmental and climatic context and the sensitivity of relationships to context should thus be taken into account in forecasting systems, and it is important to be aware that the relationships that are derived in one location may not be applicable in another. The study thus adds an interpretive account that solar forecasting is a process that is environment- and technology-dependent.

The analysis also reveals that ML can be particularly beneficial due to the nonlinearity, dynamics and non-predictability of the relationships in the solar generation process. Some ML models can also handle time series and multiple environmental factors and may allow for more flexible forecasting than the models based on prescriptive model assumptions. But this synthesis does not lend itself to the notion that more complex algorithms are better. As well as the sophistication of the model, data availability, the demand for computation, transparency, transferability and usefulness on the field need to be considered. In theory it helps to understand ML not only as a technological approach, but as a part of complex interaction between computational intelligence and knowledge of the environment, and systems using renewable energy.

The results also act as a demonstration of how data quality is a prerequisite for reliable forecasting. Missing observations, inconsistencies in measurements and temporal changes, as well as limitations in the sensors and geographically uneven observations, could limit the use of the ML based predictions. There is also a considerable amount of contextual variation due to climate, terrain, seasonality and the local weather patterns. Hence, it is important to note that models created under one climate regime cannot be expected to be applicable under another. This is particularly important in regions where the long-term and high resolution environmental data and technology are scarce. It is hence, important that it is applied responsibly, adapted to the context and validated.

One of the others that's very relevant is interpretability and trust. The relationships that are discovered by advanced ML models can be helpful for the task, but they can offer limited explanations for its prediction. Explainable AI can be used to identify environmental impacts on models' output, and can make prediction systems more understandable by the technical user and decision maker. Transparency is particularly paramount when disseminating information to support energy supply, storage, grid stability and investments in renewables operations. Explainability is therefore not an extra technical feature, but an integral component of solar forecasting which should be considered as it is part of responsible solar forecasting practice.

The study could impact renewable energy planners, operators, solar system managers, policy makers and researcher. Planning the integration of renewable energy, energy storage, grid balancing and sustainable resource management can be achieved with the help of forecasting systems. However, in practice, conditions in local data and institutional capacity should be considered. However, even the most advanced model might not be the most appropriate model if it cannot be maintained, interpreted and/or localised.

This study also faces some limitations in qualitative approach of documenting the design. It is not a test of the use of the ML algorithms, nor does it create new forecasting data but rather synthesizes and interprets the literature and institutional works. The results are not empirical comparisons of predictive performance, rather they are conceptual and interpretive. Expand this knowledge via explainable AI, hybrid and physics-informed models, multimodal environmental data, transfer learning, uncertainty-aware forecasting and systematic validation in various climatic and geographical regions. Special attention must also be paid to the contexts of developing countries where there is a lack of data.

In conclusion, while ML has great promise in solar energy forecasting and in energy management, its potential is limited to its responsible use in conjunction with environmental knowledge and practical skills. Machine learning is not meant to be a substitute for domain knowledge, professional judgement or responsible decision making in an energy system.

REFERENCES

Abdel-Basset, M., Hawash, H., Chakrabortty, R. K., & Ryan, M. (2021). PV-Net: An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production. *Journal of Cleaner Production, 303,* 127037. https://doi.org/10.1016/j.jclepro.2021.127037

Alcañiz, A., Grzebyk, D., Ziar, H., & Isabella, O. (2023). Trends and gaps in photovoltaic power forecasting with machine learning. *Energy Reports, 9*, 447–471. https://doi.org/10.1016/j.egyr.2022.11.208 (TU Delft Research Portal)

Alkhayat, G., & Mehmood, R. (2021). A review and taxonomy of wind and solar energy forecasting methods based on deep learning. *Energy and AI*, *4*, 100060. https://doi.org/10.1016/j.egyai.2021.100060

Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research, 21*(1), 37–47. https://doi.org/10.1002/capr.12360 (Wiley Online Library)

Feng, C., Zhang, J., Zhang, W., & Hodge, B.-M. (2022). Convolutional neural networks for intra-hour solar forecasting based on sky image sequences. *Applied Energy, 310*, 118438. https://doi.org/10.1016/j.apenergy.2021.118438 (ScienceDirect)

Feng, C., & Zhang, J. (2020). SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. *Solar Energy, 204*, 71–78. https://doi.org/10.1016/j.solener.2020.03.083

Ghodusinejad, M. H., Rashvand, N., Salmanpour, F., Danehkar, S., & Yousefi, H. (2026). A systematic review of solar irradiance forecasting across time horizons using physical, satellite, and AI-based methods. *Solar Compass.* https://doi.org/10.1016/j.solcom.2025.100154 (Mendeley)

Gneiting, T., Lerch, S., & Schulz, B. (2023). Probabilistic solar forecasting: Benchmarks, post-processing, verification. *Solar Energy*, *252*, 72–80. https://doi.org/10.1016/j.solener.2022.12.054 (DOI)

Guijo-Rubio, D., Durán-Rosal, A. M., Gutiérrez, P. A., Gómez-Orellana, A. M., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S., & Hervás-Martínez, C. (2020). Evolutionary artificial neural networks for accurate solar radiation prediction. *Energy, 210*, 118374. https://doi.org/10.1016/j.energy.2020.118374 (IDEAS/RePEc)

Iheanetu, K. J. (2022). Solar photovoltaic power forecasting: A review. *Sustainability, 14*(24), 17005. https://doi.org/10.3390/su142417005 (MDPI)

International Energy Agency. (2025). *Renewables 2025*. IEA. https://www.iea.org/reports/renewables-2025 (IEA)

International Renewable Energy Agency. (2025). *Renewable power generation costs in 2024*. IRENA. (IRENA)

Jadapalli, S., Datta, G., Swati, G. V., Maitra, S. K., Choudhury, U., Biramo, M. L., & Madhavan, R. (2026). Scoping review of artificial intelligence approaches for solar energy forecasting addressing methodological developments, data practices and operational implications. *Discover Applied Sciences*, *8*, 665. https://doi.org/10.1007/s42452-026-08691-1 (DOI)

Kuckartz, U. (2019). Qualitative text analysis: A systematic approach. In G. Kaiser & N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education*. Springer. https://doi.org/10.1007/978-3-030-15636-7_8 (Springer)

Kumari, P., & Toshniwal, D. (2021). Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production, 318*, 128566. https://doi.org/10.1016/j.jclepro.2021.128566 (ScienceDirect)

Lee, J., Wang, W., Harrou, F., & Sun, Y. (2020). Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. *Energy Conversion and Management, 208*, 112582. https://doi.org/10.1016/j.enconman.2020.112582 (ScienceDirect)

Li, G., Xie, S., Wang, B., Xin, J., Li, Y., & Du, S. (2020). Photovoltaic power forecasting with a hybrid deep learning approach. *IEEE Access*, *8*, 175871–175880. https://doi.org/10.1109/ACCESS.2020.3025860 (Directory of Open Access Journals)

Liu, W., Liu, Y., Zhang, T., Han, Y., Zhou, X., Xie, Y., & Yoo, S. (2022). Use of physics to improve solar forecast: Part II, machine learning and model interpretability. *Solar Energy, 244*, 362–378. https://doi.org/10.1016/j.solener.2022.08.040 (ScienceDirect)

Mahmud, K., Azam, S., Karim, A., Zobaed, S. M., Shanmugam, B., & Mathur, D. (2021). Machine learning based PV power generation forecasting in Alice Springs. *IEEE Access, 9*, 46117–46128. https://doi.org/10.1109/ACCESS.2021.3066494 (Charles Darwin University)

Massaoudi, M., Chihi, I., Abu-Rub, H., Refaat, S. S., & Oueslati, F. S. (2021). Convergence of photovoltaic power forecasting and deep learning: State-of-art review. *IEEE Access, 9*, 136593–136615. https://doi.org/10.1109/ACCESS.2021.3117004 (Hertfordshire Research Profiles)

Matera, N., Longo, M., & Leva, S. (2024). Time-dependent photovoltaic performance assessment on a global scale using artificial neural networks. *Sustainable Energy, Grids and Networks, 38*, 101337. https://doi.org/10.1016/j.segan.2024.101337

Mittal, R. K., Gupta, R., Mishra, A. K., & Singla, P. (2026). State-of-the-art deep learning models for solar forecasting: Models, methodologies, and key challenges. *Electrical Engineering, 108*, 327. https://doi.org/10.1007/s00202-026-03672-4 (DOI)

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*, 1–13. https://doi.org/10.1177/1609406917733847 (Sage Journals)

Paletta, Q., Arbod, G., & Lasenby, J. (2021). Benchmarking of deep learning irradiance forecasting models from sky images—An in-depth analysis. *Solar Energy, 224*, 855–867. https://doi.org/10.1016/j.solener.2021.05.056 (ScienceDirect)

Pérez, E., Pérez, J., Segarra-Tamarit, J., & Beltran, H. (2021). A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. *Solar Energy, 218*, 652–660. https://doi.org/10.1016/j.solener.2021.02.033 (ScienceDirect)

Pazikadin, A. R., Rifai, D., Ali, K., Malik, M. Z., Abdalla, A. N., & Faraj, M. A. (2020). Solar irradiance measurement instrumentation and power solar generation forecasting based on artificial neural networks (ANN): A review of five years research trend. *Science of the Total Environment, 715*, 136848. https://doi.org/10.1016/j.scitotenv.2020.136848 (PubMed)

Piantadosi, G., Dutto, S., Galli, A., De Vito, S., Sansone, C., & Di Francia, G. (2024). Photovoltaic power forecasting: A Transformer based framework. *Energy and AI, 18*, 100444. https://doi.org/10.1016/j.egyai.2024.100444 (Directory of Open Access Journals)

Prema, V., Bhaskar, M. S., Almakhles, D., Gowtham, N., & Uma Rao, K. (2022). Critical review of data, models and performance metrics for wind and solar power forecast. *IEEE Access, 10*, 667–688. https://doi.org/10.1109/ACCESS.2021.3137419

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039 (ScienceDirect)

Singh, U., Singh, N., Malik, H., & Hossaini, M. A. (2026). Photovoltaic power forecasting: A review on models and future research directions. *IET Renewable Power Generation, 20*(1), e70186. https://doi.org/10.1049/rpg2.70186 (IET)

Tawn, R., & Browell, J. (2022). A review of very short-term wind and solar power forecasting. *Renewable and Sustainable Energy Reviews, 153*, 111758. https://doi.org/10.1016/j.rser.2021.111758

Theocharides, S., Makrides, G., Livera, A., Theristis, M., Kaimakis, P., & Georghiou, G. E. (2020). Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing. *Applied Energy, 268*, 115023. https://doi.org/10.1016/j.apenergy.2020.115023 (ScienceDirect)

Tracy, S. J. (2010). Qualitative quality: Eight “Big-Tent” criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837–851. https://doi.org/10.1177/1077800410383121 (Sage Journals)

Visser, L., AlSkaif, T., Hu, J., Louwen, A., & van Sark, W. (2023). On the value of expert knowledge in estimation and forecasting of solar photovoltaic power generation. *Solar Energy, 251*, 86–105. https://doi.org/10.1016/j.solener.2023.01.019 (ScienceDirect)

Voyant, C., Notton, G., Duchaud, J.-L., García Gutiérrez, L. A., Bright, J. M., & Yang, D. (2022). Benchmarks for solar radiation time series forecasting. *Renewable Energy, 191*, 747–762. https://doi.org/10.1016/j.renene.2022.04.065 (ScienceDirect)

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. https://doi.org/10.1016/j.renene.2016.12.095 (ScienceDirect)

Wu, Y.-K., Huang, C.-L., Phan, Q.-T., & Li, Y.-Y. (2022). Completed review of various solar power forecasting techniques considering different viewpoints. *Energies, 15*(9), 3320. https://doi.org/10.3390/en15093320 (MDPI)

Yang, D., Li, W., Yagli, G. M., & Srinivasan, D. (2021). Operational solar forecasting for grid integration: Standards, challenges, and outlook. *Solar Energy, 224,* 930–937. https://doi.org/10.1016/j.solener.2021.04.002 (ScienceDirect)