Engineering Applications of Artificial Intelligence 163 (2026) 113115

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Building-integrated photovoltaics forecasting: Machine learning models for irradiance and power, feasibility, and recommended directions

Savisha Mahalingam [a,*], Abreeza Manap [a,**], Md Shahariar Chowdhury [b], Nurfanizan Afandi [c], Faiz Arith [d], Agung Nugroho [e,***]

[a] Institute of Sustainable Energy, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000, Kajang, Selangor, Malaysia
[b] Health and Environmental Research Center, Faculty of Environmental Management, Prince of Songkla University, Hatyai, Songkhla, 90110, Thailand
[c] Department of Mechanical Engineering, College of Engineering, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000, Kajang, Selangor, Malaysia
[d] Faculty of Electronics and Computer Technology and Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Durian Tunggal, 76100, Melaka, Malaysia
[e] Chemical Engineering, Faculty of Industrial Technology, Universitas Pertamina, Jalan Teuku Nyak Arief, Simprug, Kebayoran Lama, Jakarta, 12220, Indonesia

A R T I C L E I N F O

Keywords:
Building-integrated photovoltaics
Energy
Machine learning
Power output prediction
Solar irradiance prediction

A B S T R A C T

Building-integrated photovoltaics (BIPV) are central to energy-efficient buildings, yet forecasting irradiance and power with operational practicality remains challenging. This review analyzes 70 studies on machine learning approaches for BIPV forecasting and organizes the literature by task and deployment needs. For irradiance forecasting, the survey catalogs common contexts (façade/roof, sky/season, urban geometry), input modalities (telemetry, weather, imagery), model families, and frequently reported error metrics. For power forecasting, horizons from intra-hour to seasonal are compared, distinguishing direct power modeling from approaches that incorporate irradiance estimates. Beyond accuracy, the review examines feasibility and robustness, which determine field viability—namely, latency and memory budgets for edge versus cloud execution, sensitivity to outdoor variability, and risks to generalization across sites and seasons. The review concludes with recommended directions focused on implementable next steps, including shared data/metadata conventions for comparability, methods that can be transferred across buildings and seasons, interpretable operations for building-level decision workflows, and pipelines designed for edge execution. The resulting map of models, inputs, and constraints enables readers to match techniques to forecasting tasks and operational contexts.

Abbreviation

artificial intelligence (AI)
artificial neural networks (ANN)
building-integrated photovoltaics (BIPV)
convolutional neural networks (CNN)
Gaussian process regression (GPR)
linear regression (LR)
Local Interpretable Model-agnostic Explanations (LIME)
long short-term memory (LSTM)
ML (ML)
photovoltaic (PV)
random forests (RF)
recurrent neural networks (RNN)
research question (RQ)
root mean square error (RMSE)
SHapley Additive exPlanations (SHAP)
support vector machines (SVM)
support vector regression (SVR)

1. Introduction

The building sector is responsible for approximately 40 % of global energy use and has emerged as a strategic target for integrating renewable energy solutions (Nejat et al., 2015). Solar energy is harnessed using various types of solar cells, including silicon, thin-film, dye-sensitized solar cells, perovskite, and building-integrated photovoltaics (BIPV) (Mahalingam et al., 2025a, 2025b, 2025c). Among these,

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: savisha@uniten.edu.my (S. Mahalingam), abreeza@uniten.edu.my (A. Manap), agung.n@universitaspertamina.ac.id (A. Nugroho).

https://doi.org/10.1016/j.engappai.2025.113115
Received 16 March 2025; Received in revised form 29 October 2025; Accepted 6 November 2025
Available online 11 November 2025

BIPV stands out as a dual-function innovation, replacing traditional building materials in the envelope (such as roofs, facades, and windows) while simultaneously generating electricity (Bonomo et al., 2024). By merging aesthetics, utility, and sustainability, BIPV systems contribute to the development of zero energy buildings and urban decarbonization.

Unlike conventional photovoltaic (PV) systems mounted on open land or rooftops, BIPV installations are subjected to complex and dynamic urban conditions, including partial shading, variable tilt angles, facade orientations, and microclimate variability (Wang et al., 2023). These conditions introduce significant challenges to accurately predicting solar irradiance and the resulting power output, which are critical inputs for system design, performance evaluation, and grid integration (Liu et al., 2021). Traditional prediction methods, including physical modeling (clear-sky models, radiative transfer simulations) and statistical regression approaches, often fall short in capturing the nonlinear, time-dependent behavior of solar irradiance and PV output, particularly in BIPV contexts with irregular and less-than-ideal panel placement (Diagne et al., 2013). In this landscape, ML (ML) offers powerful data-driven alternatives. These models can learn complex mappings from historical meteorological, temporal, and environmental data to accurately estimate target variables such as solar irradiance and power output. Algorithms like artificial neural networks (ANN), support vector machines (SVM), random forests (RF), and long short-term memory (LSTM) networks have demonstrated remarkable performance in time-series prediction tasks, often outperforming classical methods in terms of root mean square error (RMSE), mean absolute percentage error, and other forecasting metrics (Wazirali et al., 2023).

ML techniques not only enhance short-term forecasts but also facilitate better long-term system planning, design optimization, and energy management in smart building environments (Song et al., 2025; Nourani et al., 2025). Although a considerable body of literature explores ML in general PV systems, a focused synthesis on its applications for solar irradiance and power output prediction in BIPV systems remains lacking. This is a crucial gap, as BIPV systems present unique forecasting challenges due to several factors, including highly variable and building-specific solar access, intermittent shading from surrounding structures, and non-uniform irradiance profiles on architectural surfaces, as well as lower data availability compared to large-scale solar farms.

Thus, an in-depth review is necessary to consolidate findings, highlight the algorithm's strengths and weaknesses, and assess model performance, computational feasibility, and data requirements specific to BIPV environments. This review aims to provide a comprehensive and comparative analysis of ML algorithms used for: (a) predicting solar irradiance on BIPV-integrated surfaces, and (b) forecasting BIPV power output under varying real-world urban and building-integrated conditions. This review makes the following key contributions.

- It systematically synthesizes over 70 recent studies on ML methods for solar irradiance and BIPV power output prediction.
- It critically examines the gaps in current research, including challenges related to model overfitting, interpretability, data availability, and the feasibility of real-time deployment.
- It proposes a functional classification of ML approaches based on data characteristics and operational requirements in BIPV systems.

Accordingly, we pose four research questions (RQ): RQ1 (Irradiance forecasting): Which ML approaches and input features yield the most reliable BIPV solar-irradiance predictions across façades/roofs and sky/season conditions? RQ2 (Power-output forecasting): Across relevant time horizons (intra-hour to seasonal), which model families and feature sets most reliably predict BIPV power output, and under what data/operational constraints? RQ3 (Feasibility & robustness): What accuracy–latency–resource trade-offs and robustness factors (outdoor-condition variability, overfitting/generalizability) determine suitability for real-time/edge deployment? RQ4 (Recommended directions): What cross-cutting priorities (datasets/standardization, transferability, interpretability) should guide research and practice? To orient the reader, Sections 2 and 3 introduce BIPV and ML, respectively. Section 4.1 addresses RQ1; Section 4.2, RQ2; Sections 4.3–4.5, RQ3; Section 5, RQ4; Section 6 concludes. Fig. 1 summarizes how this review is organized around four research questions focused on BIPV-enabled, energy-efficient building systems integrated with ML. The architecture is a left-to-right sequence: (1) Irradiance forecasting → (2) Power-output forecasting → (3) Feasibility & robustness → (4) Recommended directions. Each arrow denotes how evidence and insights flow from one question to the next, and each block maps to a dedicated section of the paper.

2. Recent developments of BIPV

A BIPV system is an effective technique for reducing building energy consumption, as it utilizes the building envelope to collect solar energy or radiation and generate both thermal and electrical energy (Yang and Athienitis, 2016). Integrating photovoltaic cells within solar devices is a promising method for generating power and reducing building cooling demands (Pillai et al., 2022; Yu et al., 2021). Besides generating renewable energy, BIPV also addresses the issue of rising CO_2 emissions and global warming. The International Energy Agency report also stated that the construction and building industry contributes the most to total energy usage or consumption. These two industries contributed to a total of 36 % of total energy consumption. Apart from that, these two industries also account for 39 % of energy-related CO_2 emissions (I.E.A. and the U.N.E.P, 2019). This layer's presence prevents the Earth from cooling and causes an increase in global temperatures (Ahmed Ali et al., 2020).

2.1. BIPV research advancements

The BIPV industry has witnessed numerous research advancements, which have ultimately led to the development of new technologies and products. Three generations of advancements in photovoltaic technologies are used to create BIPV modules. The three generations are opaque silicon, followed by transparent thin films, semitransparent, and the developing third generation. The third generation is broad, not fully developed, and still being studied and investigated (Peng et al., 2011). The third generation consists of perovskite, organic PV, and dye-sensitized solar cells (Pillai et al., 2022; El Chaar et al., 2011; Acciari et al., 2019). The technological design options for BIPV development have been growing excellently in improving performance (Kuhn et al., 2021). Generally, high-efficiency BIPVs are expensive due to high energy and labor requirements. Monocrystalline PV is an example of the first generation, accounting for approximately 80 % of the BIPV market. It will remain the industry controller until a PV technology that is more effective and affordable is produced (Reddy et al., 2020). Polycrystalline PV is an effort to lower costs and increase production in the photovoltaic industry. This technology is becoming more appealing due to its lower manufacturing cost, even though its efficiency is slightly less than that of mono-crystalline by about 15 % (Manna and Mahajan, 2007). The benefit of switching from Monocrystalline and polycrystalline in the fabrication of crystalline solar cells is to reduce defects caused by metal contamination and crystal structure issues (Reddy et al., 2020; Manna and Mahajan, 2007). Compared to crystalline silicon cells, thin film technology, part of the second generation of BIPV technology, is a promising choice for reducing PV array costs by lowering material and manufacturing costs without compromising the cells' lifetime or posing any environmental risks. However, the efficiency of second-generation technology is lower (6–10 %) than that of first-generation crystalline silicon (10–19 %) (Roy et al., 2020).

Regarding BIPV applications, there are two primary areas of application: the roof and façades (Pillai et al., 2022; Reddy et al., 2020). BIPV modules installed on building roofs must also provide essential building envelope features, including durability, wind resistance, water

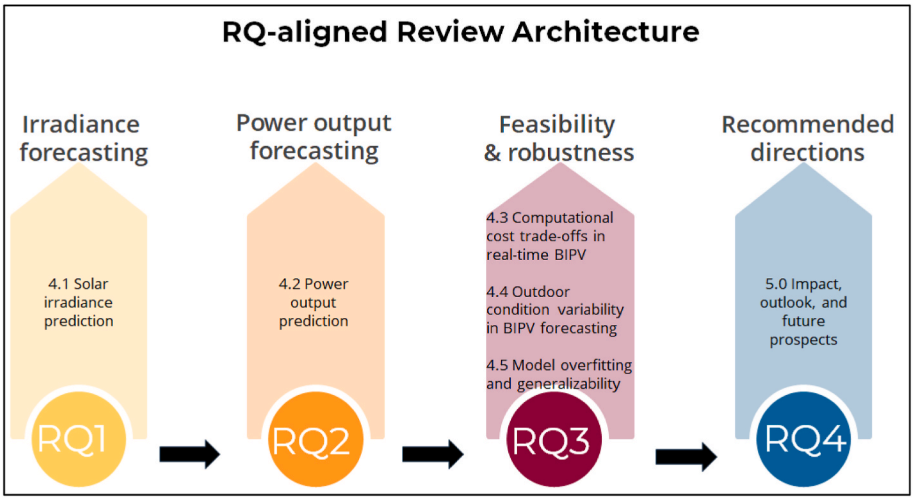

Fig. 1. Review architecture aligned with research questions (RQ).

resistance, and sound acoustic dampening (S.E.A.C.T. Netherland, 2017). The building skin is evolving from a passive barrier to a perceptive, active, and adaptive layer in response to the growing demand for zero energy buildings. In terms of industry, market, and public awareness, BIPV is experiencing significant expansion. The versatility of PV is increasingly acknowledged as a prerequisite that gradually encourages the creation of building and product systems (S.E.A.C.T. Netherland, 2015). Thus, the development of energy-efficient buildings and the forthcoming zero energy building criteria necessitate the use of BIPVs on façade components, in addition to roof-integrated BIPVs (Pillai et al., 2022; S.E.A.C.T. Netherland, 2015). Thin-film technologies are commonly used due to their adaptability and controllable transparency. To accommodate various functional objectives, a range of roof types, including skylights, cold roofs, canopies, and prefabricated multifunctional roofs, as well as façade applications such as curtain-wall façades, rain-screen façades, prefabricated multifunctional façades, accessories, and double-skin façades, have been commercialized (Biyik et al., 2017). The standard roof system has well-established applications and is easily handled and efficient. However, the BIPV system has limited aesthetic value due to its visibility (Shukla et al., 2017). With these recent developments, there are also numerous customization options, such as wafers, transparent BIPV modules, and modules featuring colored glass. As a result, a significant improvement in building aesthetics is expected to accelerate the deployment of BIPV systems in the coming years (S.E.A.C.T. Netherland, 2017; S.E.A.C.T. Netherland, 2015).

2.2. Global BIPV market outlook

As the years passed, the two primary sources of renewable energy generation worldwide have been solar photovoltaic and wind energy. According to the most recent International Energy Agency market forecast, distributed power generation, which includes BIPV technologies, will account for almost 40 % of the increase in global solar energy capacity between 2030 and 2050, when it is expected to be six to eighteen times greater (2840 GWp by 2030 and 8519 GWp by 2050) (I.E. Agency, 2020). BIPV is therefore anticipated to expand rapidly as utility-scale projects develop, driven by improved energy transformation regulations and government support programs (Building-integrated-photovoltaics-BIPV market). Thus, BIPV remains a promising sector due to its contributions to the PV and building and construction industries. European countries, such as Switzerland, Germany, Spain, and many others, have been the primary market stakeholders, as these countries have a vast and expansive share of the global BIPV market, along with Japan. However, the USA and other Asian countries, such as Japan and Korea, are gathering momentum to embrace BIPV, drastically changing the statistics due to recent technological improvements.

Fig. 2 highlights the regional disparity in BIPV market penetration and technological adoption. For instance, thin-film BIPV is more prevalent in regions such as Japan and Germany due to their established supply chains and flexibility in integration. In contrast, crystalline silicon technologies dominate the United States and Chinese markets, where efficiency and cost competitiveness are key drivers. Limited adoption in regions such as Africa and South America can be attributed to a lack of financing mechanisms, weak policy incentives, and limited awareness of the benefits of BIPV. This geographic divergence reflects both technical and economic barriers, including inadequate net metering policies, the absence of building code mandates for PV integration, and high upfront costs. Market trends indicate an increasing hybridization of smart building technologies and the expansion of retrofitting solutions in mature markets.

2.3. BIPV outdoor test

To overcome environmental challenges as well as structural challenges without losing energy production or generation, BIPV integration should ideally start with planning (Pillai et al., 2022). In this respect, the development of reliable, efficient, and safe products is aided by the outdoor testing of BIPV systems (Wan et al., 2015; Test bed for building integrated). Additionally, accurate scale BIPV testbeds help evaluate various mounting system solutions and comprehend the dangers and risks associated with the installation (Pelle et al., 2020). Test bedding is beneficial in optimizing material and production costs, reducing the risk of damage, and confirming the product's reliability and viability before making a significant investment (Wan et al., 2015). Furthermore, testbeds could demonstrate to various parties and stakeholders the novel and practical aesthetic and technical capabilities of BIPV systems, boosting consumer confidence in blending aesthetics, commercial reach, and financial viability (Pelle et al., 2020). Therefore, prior testing, such as test bedding before implementation, is essential and viable. However, overall system performance may not be accurate because specific crucial parameters must be investigated and assessed over a prolonged period. As a result, many BIPV systems are tested outside after installation and used as functional testbeds, which is a more practical and realistic method.

Additionally, a data-collecting system for long-term performance monitoring is an inherent requirement for such testbeds. Fig. 3 shows the chart of installed capacity and installed area of BIPV technologies in test bed locations across various countries. The figure shows that most test-bedding is located in Korea, and BIPV technology is used on the

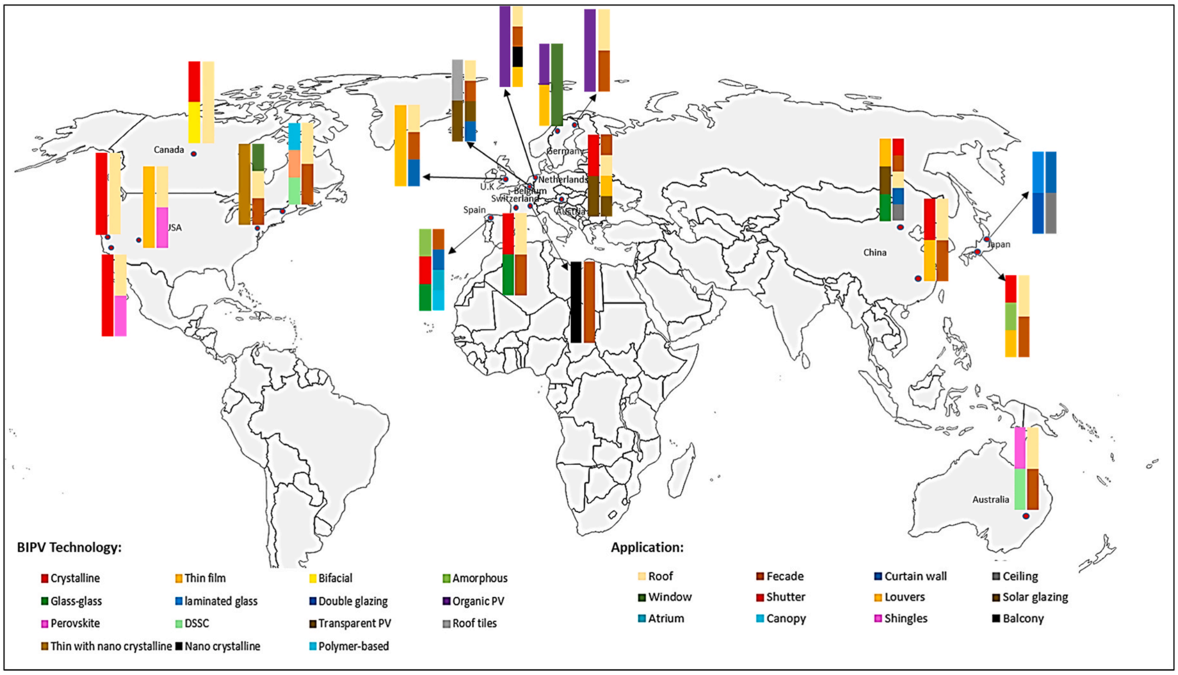

Fig. 2. The major commercial BIPV market stakeholder with their technologies and application around the World.

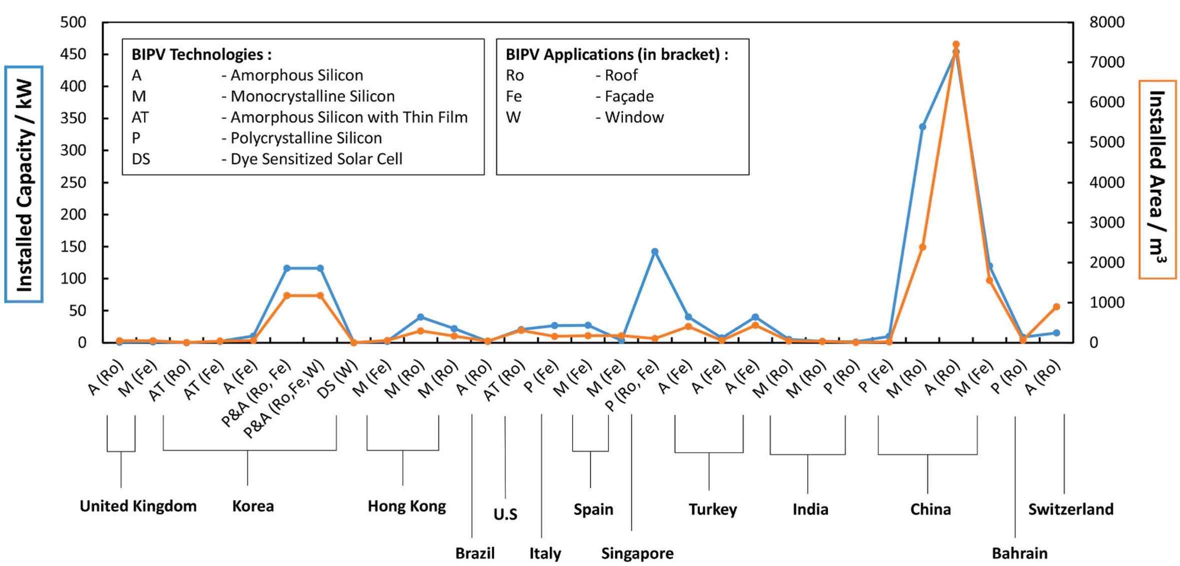

Fig. 3. BIPV technologies-applications with installed capacity and installed area (Pola et al., 2007; Alnaser, 2018; Eke and Demircan, 2015; Wittkopf et al., 2012; Martín-Chivelet et al., 2018; Ban-Weiss et al., 2013; Rüther et al., 2010; Lu and Yang, 2010; Li et al., 2013; Yang et al., 2004; Lee and Yoon, 2018; Lee et al., 2014, 2017; Wang et al., 2016, 2018; Choi et al., 2019; Yoon et al., 2011; Song et al., 2008; Omer et al., 2003; Li and Fang, 2013; Ramanan et al., 2019; Chel and Tiwari, 2011; Aaditya et al., 2013; Eke and Senturk, 2013; Araz et al., 2019).

roof, façade, and window. Then, China followed with a wide range of installed capacity and installed area. Testing of bedding in Spain and Turkey focuses on BIPV applications on the façade, whereas India focuses on the roof. Crystalline silicon systems exhibit higher capacity per unit area, making them preferable for space-limited applications such as rooftops. However, thin-film and emerging technologies (such as dye-sensitized solar cells or organic PVs) offer larger area deployment for the same capacity, suggesting a trade-off in efficiency but improved flexibility, aesthetics, and lightweight integration, particularly in façades and atriums. The figure also reveals application-specific dominance. Rooftop BIPV systems account for the majority of installed capacity due to structural feasibility and solar exposure. Façade-integrated BIPV, although growing, faces limitations due to partial shading, non-optimal angles, and architectural constraints, which explain their lower overall contribution despite higher surface coverage.

Manna et al. have studied a typical characteristic of third-generation PV, which is that it absorbs more sunlight (Manna and Mahajan, 2007). This study has demonstrated that nanotech solar cells can increase conversion efficiency. These technology cells can improve the optical losses due to reflection and carrier losses. However, there are still significant obstacles to understanding and realizing PV devices before they are mass-produced. One limitation of this technology is the fabrication of arrays with micrometer-scale dimensions. One of the third-generation PVs, hot carrier cells, is also required to create a mono-energetic contact between the hot carriers inside the absorber and the cold carriers within the contacts. In 2022, Chuyao and his team investigated the hybrid BIPV technology façade on the plateau (Wang et al., 2022). This study is done to overcome the challenge of the BIPV façade, which must work appropriately to contribute to energy savings effectively. An ANN method can be applied to enhance the indoor thermal environment. As a result, ANN decreased more than 40 % of the overall amount of energy used in the examined plateau areas (Wang et al., 2022).

On the other hand, a study by Alrashidi et al. investigated one potential BIPV application, semitransparent PV glazing. This study evaluated the thermal performance of semitransparent PV glazing made of cadmium telluride with varying transparencies for south and southwest orientations in the United Kingdom. As a result, cadmium telluride PV glazing with the least transparency has a U-value of just 1.52 $W/m^2 \cdot K$ (Alrashidi et al., 2022). Compared to standard clear glass, the least transparent PV windows can reduce solar heat gains by 96 % and cooling energy by 23.2 % when utilized in a southwest orientation (Alrashidi et al., 2022). A 2022 study by Chen and his team investigated the effect of BIPV windows. This study adopted a newly developed BIPV coupled with a single-layer canopy model in Phoenix, USA. A sophisticated Markov chain Monte Carlo approach is used to study the sensitivity and uncertainty of the BIPV window effect in the built environment. The finding of this study shows that window coverage, power generation efficiency, and canyon aspect ratio significantly influence how much power the BIPV window generates (Chen et al., 2022).

3. ML fundamentals for BIPV

3.1. ML overview in renewable energy forecasting

The increasing complexity of energy systems, coupled with the variability of renewable sources such as solar and wind, has driven the adoption of data-driven techniques for accurate forecasting (Unni and Channi, 2025). Among these, ML has emerged as a powerful tool to model the nonlinear, dynamic, and often uncertain behavior of renewable energy resources. Unlike traditional physical or statistical models, ML techniques can automatically learn from historical data without being explicitly programmed with system equations, making them particularly effective in handling complex environmental interactions (Willard et al., 2022).

In the solar energy sector, ML has been widely applied for solar irradiance prediction, PV power output forecasting, and energy yield estimation (Essam et al., 2022; Voyant et al., 2017; Ledmaoui et al., 2023). The strength of ML lies in its ability to capture nonlinear relationships between input variables (e.g., humidity, temperature, and solar angles) and target outputs (e.g., irradiance and energy output) (Hissou et al., 2023). This adaptability makes it suitable for localized renewable energy applications, particularly when using regional or site-specific datasets. Handle noisy, incomplete, or high-dimensional data, which are common in weather-dependent systems (Hassan et al., 2025). Moreover, advancements in deep learning, particularly recurrent neural networks (RNN) and convolutional neural networks (CNN) (e.g., LSTM, CNN-LSTM hybrids), have enhanced the ability to model temporal and spatial dependencies in solar resource forecasting. These capabilities are increasingly relevant for urban-integrated systems, such as BIPV, where irradiance conditions vary significantly due to building geometry and environmental obstructions.

3.2. Role of ML in solar irradiance prediction

Traditional irradiance prediction methods, such as clear-sky models (Yang, 2020) or empirical regression techniques (Citakoglu, 2015), often struggle to capture these nuanced variations, particularly in urban environments where irradiance is highly site-specific and transient. ML provides a data-driven approach that can learn complex spatial and temporal relationships from historical data. Instead of relying on fixed equations, ML models can be trained to predict irradiance based on a combination of inputs such as time of day, solar angles, weather conditions, and location data (Citakoglu, 2015). These models can generalize across different contexts when trained on diverse datasets, making them suitable for highly variable BIPV environments.

Among the commonly used algorithms, ANNs have been widely applied for irradiance forecasting due to their ability to approximate nonlinear functions (Pazikadin et al., 2020). SVMs, particularly in regression mode, have shown effectiveness in handling smaller datasets and performing well in noisy conditions. Deo et al. reported the development of a wavelet-coupled SVM model for forecasting global incident solar radiation using limited meteorological datasets (Deo et al., 2016). The study highlighted the SVM model's capability to handle complex, non-linear patterns in solar irradiation data, particularly when dealing with small datasets. Additionally, Shuvo et al. proposed a multi-hybrid model integrating the Hodrick–Prescott Filter, Discrete Wavelet Transform, and SVM to predict solar radiation (Pal Shuvo et al., 2025). The study emphasized that previous research had shown challenges in predicting solar radiation due to the highly nonlinear and noisy nature of climate data. The proposed model aimed to address these limitations by enhancing the SVM's ability to handle such data, thereby improving prediction accuracy.

RFs are known for their robustness and ability to handle high-dimensional input features without extensive preprocessing. Villegas-Mier et al. highlighted that RF models can process large volumes of data and capture complex relationships between input variables without the need for comprehensive data transformations (Villegas-Mier et al., 2022). This characteristic makes RFs particularly suitable for solar radiation prediction tasks that involve multiple meteorological variables. Tandon et al. evaluated various ML models for solar irradiance prediction and found that the RF model achieved an RMSE of 0.64, outperforming the linear regression (LR) model by approximately 19.47 % (Tandon et al., 2025). The results emphasize the superior predictive capabilities of ML models over traditional methods in solar irradiance forecasting.

3.3. Role of ML in power output forecasting

ML offers a robust alternative by learning complex nonlinear relationships between multiple input variables and the corresponding electrical output of PV modules. ANNs have been extensively used in power output forecasting due to their flexibility and high prediction accuracy across different BIPV geometries and climates. Ghenai et al. evaluated the performance of bifacial solar PV systems installed on flat roofs with varying surface albedo. An ANN-based forecasting model was developed to predict power output, considering the enhanced albedo effect. The study found that the ANN model accurately anticipated power generation, demonstrating its adaptability to different climatic conditions and system configurations (Ghenai et al., 2022).

Support vector regression (SVR) has shown competitive performance in scenarios with limited or noisy training data. Abuella and Chowdhury presented an SVR model to produce solar power forecasts on a rolling basis for 24 hours ahead over a one-year period. The study considered twelve weather variables and found that the SVR model performed competitively compared to ANNs and multiple LR models, particularly in handling noisy and limited datasets (Abuella and Chowdhury, 2017a). RFs with their ensemble learning structure are effective in handling large input spaces and offer interpretable models that can rank the importance of input features such as ambient temperature or irradiance levels (Abuella and Chowdhury, 2017b). LSTM networks have emerged as a powerful tool for forecasting sequential energy output patterns, capturing the temporal dependencies that exist in power generation profiles across different hours or days (Harrou et al., 2020). In BIPV-specific scenarios, these ML models often incorporate additional inputs such as façade orientation, building layout, and thermal conditions, which are less commonly considered in conventional rooftop PV modeling. For instance, time-series models using LSTM have been shown to capture delayed thermal effects and the shadowing impact from adjacent structures more accurately than feedforward models (Gao et al., 2019; Kappler et al., 2023).

3.4. Comparison with traditional methods

Deterministic models, such as clear-sky models and transposition models based on radiative transfer equations, aim to simulate solar irradiance or energy output using physical laws. For instance, the Bird Clear Sky Model, developed by the National Renewable Energy Laboratory, is a widely used deterministic model that calculates clear-sky solar radiation components using radiative transfer equations (Palani et al., 2017). These models require precise input parameters, including solar geometry, panel orientation, albedo, and atmospheric turbidity, and often assume idealized or homogeneous conditions (Palani et al., 2017; Quan and Yang, 2020). However, BIPV installations typically occur in urban environments where such assumptions are not applicable. Facade-mounted or vertical PV modules are subject to partial shading, obstructions, and asymmetric irradiance profiles, making accurate deterministic modeling extremely challenging (Ishaque et al., 2011).

Statistical methods, including multiple LR, autoregressive integrated moving average, and polynomial regressions, rely on predefined relationships between historical variables and forecast targets (Effrosynidis et al., 2023). While simpler to implement, these models struggle to capture nonlinear and time-dependent relationships, especially in multivariate scenarios with interactions between irradiance, temperature, humidity, and building-specific geometries (Almaghrabi et al., 2024; Jacques Molu et al., 2024). They also assume stationarity in data trends, which is often violated in real-time BIPV applications (Diagne et al., 2013; Mohanasundaram and Rangaswamy, 2025). Moreover, both deterministic and traditional statistical approaches lack the adaptability to rapidly changing environmental conditions, such as sudden cloud cover or shadow movement from adjacent structures — scenarios that are common in urban BIPV installations.

ML techniques offer substantial advantages over traditional methods in several key areas, such as nonlinearity and multivariable handling, temporal and spatial pattern learning, robustness to data variability, real-time adaptability, and feature importance and interpretability. ML models such as ANN (Djeldjli et al., 2024), RF (Soleymani and Mohammadzadeh, 2023), and LSTM (Mukhoty et al., 2019) networks are inherently capable of modeling nonlinear relationships between multiple input features and the target variable. For instance, ML models can account for the complex influence of solar position, shading patterns, and weather variables on PV output, without requiring manual feature engineering or equation derivation.

Especially with deep learning methods like LSTM and CNN-LSTM, ML can capture temporal dependencies and spatial correlations in sequential data (Yu et al., 2024; Khan et al., 2024). This is particularly valuable for short-term irradiance and power output forecasting, where data varies minute by minute across building surfaces. ML models have shown greater resilience to noise, missing values, and small datasets (via ensemble techniques or kernel-based models like SVM), as evidenced in recent comparative studies (Deo et al., 2016; Caiafa et al., 2021). With retraining mechanisms or online learning variants, ML models can be continuously updated as new data becomes available, making them suitable for real-time deployment in smart building systems. This dynamic capability is absent mainly in static regression-based or physics-based models.

3.5. Limitations and considerations in ML for BIPV

3.5.1. Data limitations and preprocessing challenges

BIPV systems are typically customized to specific architectural designs, orientations, and urban contexts. This results in high heterogeneity in datasets, limiting the availability of large, standardized datasets needed to train robust ML models. Unlike utility-scale PV systems, which are deployed in open fields with uniform modules and consistent irradiance exposure, BIPV installations vary significantly in terms of module type, shading, tilt, azimuth, and location-specific climate conditions. Moreover, acquiring high-quality, continuous, and labeled datasets from BIPV systems is often hindered by sensor limitations, data logging errors, or building-specific privacy constraints (Singh et al., 2022; Liu et al., 2023). Singh et al. discuss the challenges in BIPV systems, highlighting issues such as the rise in temperature of PV modules, the occurrence of various faults, and the accumulation of dust particles (Singh et al., 2022). These factors can affect the performance and data quality of BIPV systems. The study emphasizes the need to integrate digital technologies, such as the Internet of Things and artificial intelligence (AI), to monitor and address these challenges effectively.

3.5.2. Overfitting and generalization challenges

Another concern in applying ML to BIPV systems is overfitting, especially when models are trained on small or non-representative datasets (Hamad et al., 2025). Overfitting occurs when a model captures noise or spurious patterns in the training data rather than generalizable trends. Hamad et al. highlighted that, due to the nonlinear nature of power generation in PV systems, which is influenced by fluctuating weather conditions, managing this nonlinear data effectively remains a significant challenge (Hamad et al., 2025). The study emphasizes that models trained on limited or non-representative datasets are prone to overfitting, leading to poor generalization on unseen data.

This issue is especially relevant in BIPV applications, where the model trained on data from one building or climate zone may not generalize well to another with different façade orientations, material reflectivity, or shading profiles. Deep learning models, such as LSTM and CNN, although powerful, are particularly prone to overfitting in low-data scenarios (Salman et al., 2024; Kumari and Toshniwal, 2021). This highlights the need for cross-validation, transfer learning, or ensemble methods to enhance model robustness across diverse BIPV settings.

3.5.3. Interpretability and deployment feasibility

While hybrid models such as CNN-LSTM and deep neural networks have demonstrated strong predictive performance in solar irradiance and power forecasting, their “black-box” nature limits interpretability, posing challenges for decision-makers and system operators in the energy sector (Samek et al., 2017). For BIPV applications, transparency is crucial for verifying model behavior under diverse environmental conditions and building geometries. To address this, explainable AI techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) have gained traction.

SHAP assigns an importance score to each input feature for individual predictions, helping stakeholders understand which variables (e.g., solar angle, temperature, irradiance) drive output fluctuations. LIME approximates complex models locally using interpretable surrogates, offering human-readable insights for particular input conditions. Applying SHAP and LIME to CNN-LSTM models in BIPV systems can enhance trust and facilitate better integration with building energy management systems, particularly in settings where safety, compliance, or performance benchmarking are crucial.

Future reviews or experimental work should address the diagnostic capabilities of ML models under real-world BIPV variability, including comparative analysis of their sensitivity, specificity, and resilience to noise. Additionally, integrating explainable AI methods, such as SHAP, LIME, or rule-based surrogates, could help operators better interpret model outputs, especially in edge cases where opaque model decisions hinder fault tracing or maintenance planning (Hassan et al., 2025).

Furthermore, real-time deployment of ML models in smart buildings and BIPV-integrated energy management systems requires lightweight architectures that can function on embedded hardware or edge computing platforms (Singh et al., 2022; Jouini et al., 2024). Many state-of-the-art ML models are computationally intensive, limiting their feasibility for real-time control unless they are simplified or optimized through pruning, quantization, or hardware acceleration (Liang et al., 2021; Roth et al., 2024).

The real-world deployment of ML models in BIPV systems must consider the hardware limitations of embedded systems and edge devices. Many deep learning models, such as CNN or RNN, are computationally intensive and may not be feasible on microcontrollers or low-power processors commonly used in smart buildings. As a result, model compression techniques—including pruning, quantization, and knowledge distillation—are increasingly important to reduce inference time and memory usage without significantly sacrificing accuracy. Hardware accelerators, such as Graphics Processing Units, Tensor Processing Units, and edge AI chips (e.g., NVIDIA Jetson, Google Coral), can also support real-time ML inference in energy-constrained environments. However, these solutions introduce trade-offs between energy consumption, latency, and model complexity. Lightweight algorithms such as decision trees or SVM (with linear kernels) may offer a better balance between predictive performance and deployability for specific BIPV tasks. Future research should explore co-design strategies that optimize ML models in conjunction with hardware constraints, enabling scalable and real-time deployment.

4. ML in BIPV

BIPV's performance depends on many factors, which can be categorized into solar irradiance at the specific location and BIPV energy output (Fig. 4).

4.1. Solar irradiance prediction

This section addresses RQ1 by comparing input features, model families, contexts (façade/roof, sky/season), and evaluation metrics for BIPV irradiance forecasting. As the primary factor affecting BIPV output, solar irradiance will fluctuate significantly due to various factors, including cloud movement, surface orientation, building view factor, and ground view factor. Researchers have developed several ML algorithms, considering several factors to predict solar irradiance. Table 1 lists ML applications in BIPV for predicting solar irradiance. A solar irradiation mapping model can be used to indicate the optimal location for installing a BIPV solar cell, which is critical for the operation of innovative energy systems.

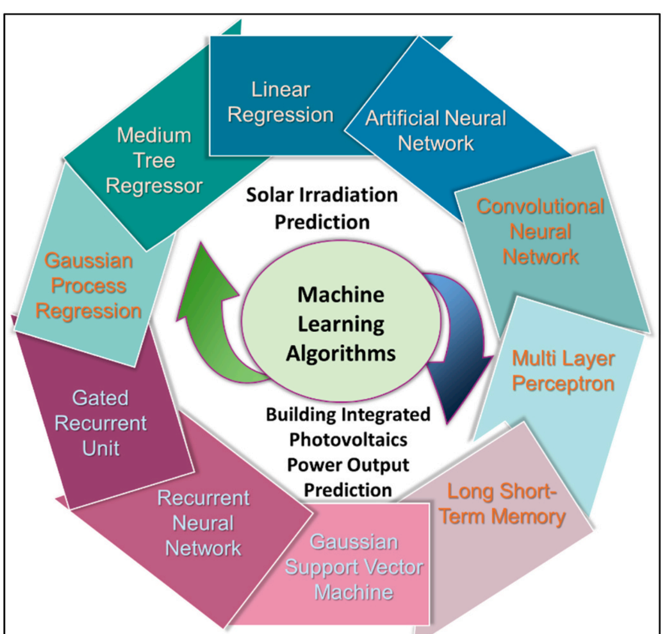

Fig. 4. ML algorithms in predicting solar irradiation and BIPV power output.

Jiang et al. propose ResnetTL, a hybrid deep network designed to estimate hourly global solar radiation using geostationary satellite images of China (Jiang et al., 2019). ResnetTL comprises two blocks (Fig. 5): a CNN for extracting spatial patterns from satellite images and a multilayer perceptron for establishing the complex, underlying nonlinear relationship between inputs and target global solar radiation values. ResnetO utilizes spatial patterns extracted from satellite imagery, while ANN_P leverages pixel spectral information from satellite data, along with auxiliary local time (month, day, and hour) and location (longitude, latitude, and altitude) information. The accuracy of the hourly GSR predictions was assessed by comparing them to ground measurements at 90 training stations and eight validation stations. ResnetTL demonstrated superior performance compared to ANN-P and ResnetO.

Zhen et al. proposed that an accurate mapping model between sky images and surface solar irradiance is critical for ultra-short-term solar PV power forecasts (Zhen et al., 2020). This study addresses the shortcomings of single-model methods such as ANN or CNN. This results in low precision for future irradiance prediction, as the future sky image differs from the historical sky image due to the growth and deformation of clouds. They compare the ANN, CNN, LSTM, and a hybrid of CNN-LSTM to predict solar irradiance based on the input of a sky image. The sky data was preprocessed and classified into three classes before being applied to the ML algorithms. The classified data are input into the hybrid mapping model, and the simulation results are compared with those of the model, which has the highest accuracy among the other three models. The proposed hybrid mapping model exhibits better accuracy and a narrower error range compared to other deep learning models.

Bredemeier et al. used ML to forecast the annual sum of irradiance on rooftop and facade surfaces in an urban area of Hanover, Germany (Bredemeier et al., 2021). As inputs, the ML regressor uses indicators obtained from skyline profiles (surroundings of rooftops and facades). As shown in Fig. 6a and b, the skyline profiles provide a two-dimensional representation of a specific surface's surroundings. The grey-hatched section represents the area of the environment that is cast in shadow by the surface itself. The green area is where the surface meets the earth. Other structures impede the view of the surface in the red section. The surface of the blue section looks up towards the sky. From this image, they extract several factors, including surface tilt, surface orientation, sky view factor, ground view factor, building view factor, and sun coverage factor, which can then be used as input for various algorithms. The applied algorithms are LR, medium tree regressor, medium Gaussian SVM, and Gaussian process regression (GPR).

They discovered that ML algorithms can forecast the total yearly irradiation for all types of building surfaces at all tilts and orientations. The RMSE between predicted values and ground truth values decreases from top to bottom, 58 kWh/(m^2 a) for the LR, 40 kWh/(m^2 a) for the decision tree regressor, 34 kWh/(m^2 a) for the Gaussian SVM regressor and 32 kWh/(m^2 a) for the Gaussian process regressor. They observed that GPR shows a lower RMSE value compared to SVM. However, the computational time was increased significantly. Thus, they used an SVM regressor for further analysis. They found that the increased error in the prediction for façade surfaces is partly due to the larger amount of reflected light impinging on façades compared to roofs.

Zaiani et al. (2024) used the image satellite from the Meteosat Second Generation geostationary satellite, which is divided into several channels VIS06 (0.6 μm), VIS08 (0.8 μm), HRV, and NIR (1.6 μm) (Zaiani et al., 2024). These image satellite data were used as input, and the Daily global solar radiation was predicted using ANN and SVM. Global solar radiation measurements were carried out between 2014 and 2017 in Algeria using the Ghardaïa radiometric station. The station

Table 1
ML in BIPV in solar irradiance prediction.

No.	Input variable	Output variable	ML algorithm	RMSE	Year	Ref.
1	Multifunctional Transport Satellite - ResnetTL: spatial pattern and point (pixel) information - ResnetO: spatial pattern - ANN_P: point information	Surface global solar radiation	- ResnetTL and ResnetO: Convolutional neural networks-Multilayer perceptron - ANN_P: feed-forward artificial neural networks	- ResnetTL: 84.18 W/m^2 - ResnetO: 90.38 W/m^2 - ANN_P: 89.75 W/m^2	2019	Jiang et al., 2019 (Jiang et al., 2019)
2	Sky image data from sky imager. The sky imager and irradiance meter are deployed in a meteorological station located in solar PV plant	Prediction of solar irradiance data	Artificial neural networks Convolutional neural networks Long short-term memory Convolutional neural networks-Long short-term memory	137.49 W/m^2 110.99 W/m^2 127.55 W/m^2 107.66 W/m^2	2020	Zhen et al., 2020 (Zhen et al., 2020)
3	(LoD-2) city geometries of Hanover (Germany): • surface tilt • surface orientation • sky view factor • ground view factor • building view factor • sun coverage factor	Prediction of annual sum of solar irradiation on rooftop and façade surfaces	Linear regression Medium tree regressor Medium Gaussian SVM Gaussian process regression	58 kWh/(m^2 a) 40 kWh/(m^2 a) 34 kWh/(m^2 a) 32 kWh/(m^2 a)	2021	Bredemeier et al., 2021 (Bredemeier et al., 2021)
4	The normalized reflectance obtained from Meteosat Second Generation images taken between 2014 and 2017.	Prediction of Daily Global Solar Radiation from 4 channels: VIS06+VIS08+HRV + IR017	Artificial neural networks Support vector machines	212.21 Wh/m^2 441.95 Wh/m^2	2024	Zaiani et al., 2024 (Zaiani et al., 2024)
5	Solar elevation angle - solar declination - extraterrestrial solar radiation	forecast surface solar irradiance using feature selection method	Convolutional neural networks-Long short-term memory Convolutional neural networks-Gated recurrent unit Artificial neural networks Convolutional neural networks Long short-term memory Gated recurrent unit	0.1966 MJ/m^2 0.1986 MJ/m^2 0.2202 MJ/m^2 0.2179 MJ/m^2 0.2124 MJ/m^2 0.2088 MJ/m^2	2024	Kim, J., 2024 (Kim et al., 2024)

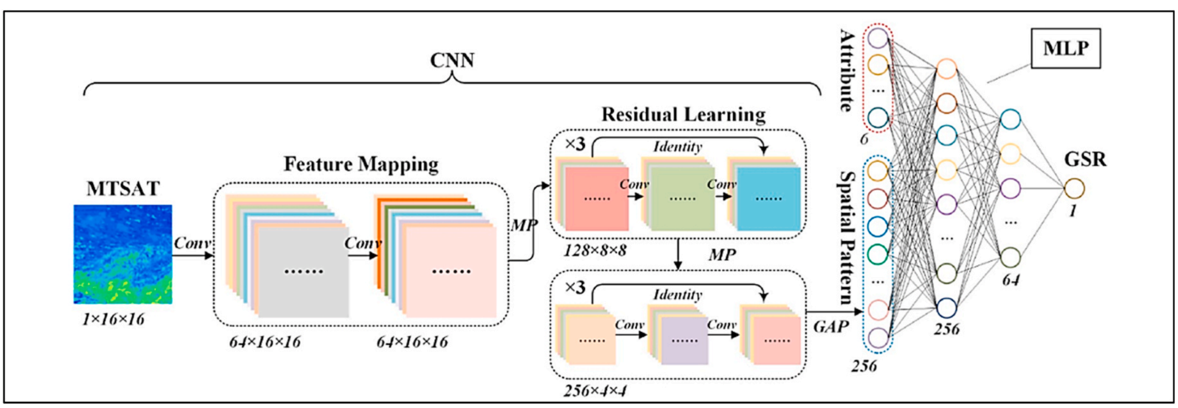

Fig. 5. Structure of proposed deep network consisting of a convolutional neural network (CNN) and a multilayer perceptron (MLP) with Multifunctional Transport Satellite (MTSAT) data as input and global solar radiation (GSR) as output (Jiang et al., 2019).

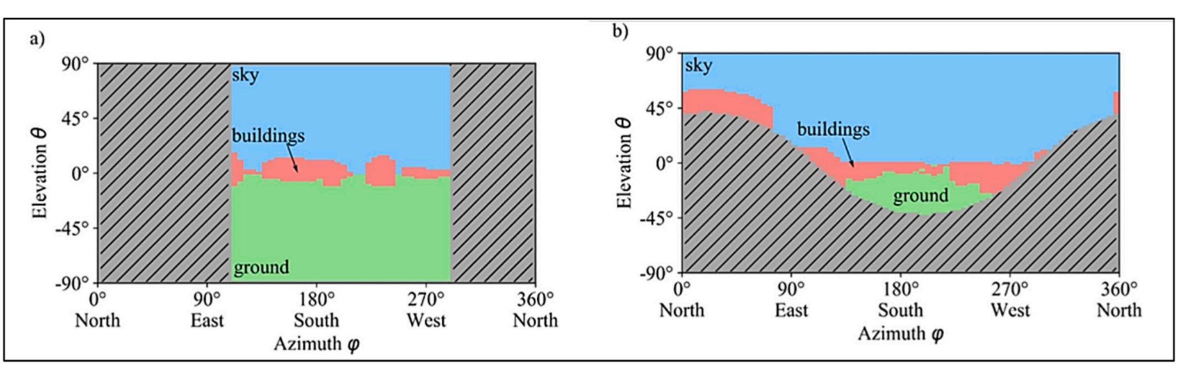

Fig. 6. (a) Surroundings of a façade surface facing towards south and (b) a roof surface with an atilt of approximately 40° and oriented to south (Bredemeier et al., 2021).

is equipped with three instruments to measure the direct, diffuse, and global components of solar radiation every minute, as well as temperature and humidity. The ANN model yielded lower error values than the SVM model. This study presents interesting and promising insights into the contribution of ML to predicting ground observations from satellite images and solar radiation at the ground level in land areas.

Kim et al. used image satellite data from the GEO-KOMPSAT-2A satellite (16 channels) of the Republic of Korea as the input data (Kim et al., 2024). The automated synoptic observation system, run by the Korea Meteorological Administration on the ground, measured the surface solar irradiance. They considered the surface solar irradiance characteristics and utilized a spatio-temporal deep learning model to predict solar irradiance. The procedure has four steps: feature selection, data analysis and preprocessing, forecasting model design and evaluation, and data collection. According to the experimental results, this proposed spatio-temporal deep learning model (CNN-LSTM, CNN-Gated recurrent unit) outperforms ANN, CNN, LSTM, and Gated recurrent unit in terms of overall performance (RMSE ranging from 0.20 to 0.53 MJ/m^2). Furthermore, it was verified that features chosen with the proposed Deep-feature selection perform better in forecasting than those obtained with the conventional approach. Finally, it was verified that surface solar irradiance forecasting performs noticeably better, regardless of the forecasting model, when three solar geometry parameters — solar declination, solar elevation angle, and extraterrestrial solar radiation — are employed.

It is essential to acknowledge that many of the models reviewed in Table 1 rely on localized or short-term datasets, which can restrict the generalizability and robustness of their predictions. The application of ML to estimate solar irradiance becomes particularly challenging when models are used in different climate zones, urban environments, or building typologies. These limitations may lead to overfitting and reduced model performance when the model is used outside the context of the original dataset. To address this issue, future research should focus on developing more diverse and representative datasets, ideally incorporating data from multiple locations and periods. This would help improve the transferability and resilience of ML models used in BIPV

Table 2
ML in BIPV power output prediction.

No.	Input Variables	Output Variable	Prediction Algorithm	RMSE	Year	Ref.
1	· Normative sky index · Wind direction · Relative humidity · Solar altitude · Wind speed · Outdoor air temperature · Precipitable probability	Short-term hourly BIPV power output predictions (kW)	Artificial neural networks Support vector machines Classification and regression trees Chi-square automatic interaction detection Random forests Recurrent neural network	0.76 0.83 0.76 0.98 0.75 0.51	2019	Lee et al., 2019 (Lee et al., 2019)
2	· Temperature · Humidity · Dew point · Wind speed	Forecasted PV power on a sunny day (W) Forecasted PV power on a partially cloudy day (W) Forecasted PV power on a cloudy day (W)	Similarity-based forecasting models based on K-nearest neighbors	2.89 10.81 14.16	2020	Sangrody et al., 2020 (Sangrody et al., 2020)
3	· Solar radiation · Wind speed · Relative humidity · Temperature	PV power prediction on east façade (W) PV power prediction on the south façade (W) PV power prediction on the south façade (W)	Artificial neural networks Quadratic Support vector machines Decision tree	4.42 16.8 8.76	2021	Kabilan et al., 2021 (Kabilan et al., 2021)
4	· Temperature · Irradiance · Voltage for estimation · Current for estimation	Power (kW) P=V*I Voltage (V) Current (I)	Feed-forward neural networks Regression neural network Regression neural network	0.0754 13.27 0.26437	2022	Shin et al., 2022 (Shin et al., 2022)
5	· Spectral irradiance · Incident spectral angle · Solar module temperature	Predicted annual output energy 282.54 (rooftop), 105.07 (east), 174.71 (south), and 90.79 (west) kWh/m^2	Artificial neural networks	1.26	2022	Nguyen & Ishikawa, 2022 (Nguyen and Ishikawa, 2022)
6	· Illuminated fraction of array · Plane on the array irradiance · Ambient temperature · Module temperature · Cosine of the incident angle	Predicted power on south array (kWh/m^2) Predicted power on west array (kWh/m^2) Predicted power on east array (kWh/m^2)	Artificial neural networks	0.19 0.11 0.17	2022	Polo et al., 2022 (Polo et al., 2022)
7	· Diffused normal irradiance · Diffused horizontal · Irradiance · Global horizontal irradiance	BIPV power output - winter 1kWh/m^2 - spring 13 kWh/m^2 - summer 16 kWh/m^2 - autumn 4 kWh/m^2	Random forests Artificial neural networks Linear regression Support vector regression Decision tree regression	0.182 0.231 0.279 1.16 0.257	2023	Dourhmi et al., 2023 (Dourhmi et al., 2023).
8	· Solar irradiance · Incident light's angle · PV module's temperature · Perovskite absorber's thickness · Perovskite absorber's bandgap	Annual output energy (kWh/m^2)	Artificial neural networks	0.15111	2023	Nguyen and Ishikawa, 2023 (Nguyen and Ishikawa, 2023)

applications.

4.2. Power output prediction

This section addresses RQ2, synthesizing evidence on power-output forecasting across horizons (intra-hour to seasonal) and orientations, and contrasting direct power modeling with chained approaches that use predicted irradiance. In a BIPV system, it is essential to forecast short-term load demand to effectively balance supply and demand with power producers and handle the fluctuations in power consumption on the grids (Muneer et al., 2007). Thus, forecasting the energy produced by weather-dependent renewable sources, such as solar energy, is essential. Accurate estimates of BIPV power output depend on weather data such as cloudiness and solar irradiance since solar power generation varies with changes in solar irradiation throughout the day. Researchers have recently employed the ML approach to predict BIPV power output using various models, as summarized in Table 2.

Lee et al. compared six ML models, including the ANN, chi-square automatic interaction detection, classification and regression trees, SVM, RNN, and RF algorithms (Lee et al., 2019). The ML algorithm was used to predict BIPV power output by considering historical weather data from 4 different types of meteorological circumstances (clear, slightly cloudy, cloudy, and overcast days). The findings showed that the RNN achieved the lowest RMSE among other models over 15 days of varying sky conditions. The RNN outperformed the different approaches in terms of hourly predicted accuracy and achieved superior long-term daily prediction performance over a 64-day period.

Sangrody et al. used ML algorithms such as ANN and a similarity-based forecasting model based on K-nearest neighbors to predict the day-ahead solar PV generation (Sangrody et al., 2020). The PV panels were installed on the rooftop of the Engineering and Science Building at the State University of New York at Binghamton. At the same time, weather data were collected from three nearby weather stations. The study results show that the accuracy of categorical similarity-based forecasting is slightly better than that of basic similarity-based forecasting when using sky-cover data. However, the forecasting accuracy of the hierarchical similarity-based forecasting model (RMSE = 1.43 %) is significantly better than that of the basic similarity-based forecasting model (RMSE = 6.67 %) and the categorical similarity-based forecasting model (RMSE = 5.67 %), particularly when temperature, humidity, and irradiance are applied hierarchically in the similarity analysis.

Kabilan et al. utilized a ML algorithm to predict the short-term power efficiency of a BIPV system installed on flat, south-east-oriented, and west-oriented facades of a building (Kabilan et al., 2021). A comparison of three ML models (ANN, decision tree, and quadratic SVM) has been applied to forecast PV power output using four inputs: solar radiation, wind speed, relative humidity, and temperature. The ANN algorithm is the best among all the other models used. This finding suggests that ANN may be a reliable predictor of the performance and reliability of power grid management processes.

Shin et al. (2022) use LR to predict voltage and current as output using irradiance, module temperature, voltage, and current data. Then, the power output can be calculated by multiplying the expected voltage and current. The BIPV module, available in three different colors (red, blue, and grey), along with its electrical and weather characteristics, was used as input data for model training (Shin et al., 2022). When the LR model's prediction value was compared to the feed-forward neural networks model's prediction value for the test data, it was found that the feed-forward neural networks model estimates values more closely than the LR model for power values of colored BIPV systems. The R^2 values for the feedforward neural network model's voltage and current values were 5 % and 2 % higher, respectively. Compared to the LR model's RMSE of 0.1581 kW, the proposed ML model has a lower RMSE of 0.0754 kW. The power generation of a vertically positioned BIPV system with a colored module built with dots on glass can be accurately predicted by the proposed ML model. This approach enables the simple identification of a BIPV system's failure and minimizes energy losses through regular maintenance, allowing for precise power predictions.

Nguyen and Ishikawa performed a study on predicting the output energy of high-efficiency perovskite/silicon tandem solar cells via an ANN based on the specific direction of the PV module and spectral irradiance data for August (Nguyen and Ishikawa, 2022). The ANN model optimization considered a six-neuron input consisting of the perovskite layer's bandgap and thickness, module temperature, incident angle, near-infrared irradiance, and visible irradiance. Thus, it could predict the output power density of the perovskite/silicon tandem solar cell. The expected model indicates that the optimum perovskite layer's bandgap and thickness vary according to the facades of the PV module on the building rooftop, with an RMSE of 1.26 and a correlation coefficient of 0.99979.

Using an ANN algorithm, Polo et al. modeled a BIPV system attached to a building façade, taking into account the shadowing effect brought on by nearby structures, trees, and other urban features (Polo et al., 2022). The input parameters for the ANN model were separated into two groups: the geometry, orientation, and shading of the modules, which included the cosine of the sunlight incident angle, the azimuth of the façade, and the illuminated fraction of the array at each time step; and the PV direct conversion parameters, which included plane of array irradiance, module temperature, and ambient temperature. The suggested ANN model was an easy-to-implement technology that could accurately predict the power of every individual PV array. In 5 minutes, the mean relative error in estimating power varied between 6 % and 15 %. As a result, digital twin approaches for BIPV systems can effectively utilize this type of model, enabling the detection of abnormal behaviors.

Dourhmi et al. utilized the French weather forecast database to evaluate the performance of BIPV predictions. They suggested using several models to maximize PV and meteorological factors. Various ML models, including RF, ANN, LR, SVR, and decision tree regression, are examined, and comparison experiments are carried out to compare the results. The suggested RF and ANN models yield the best results in terms of performance, as well as in making accurate predictions for each season (Dourhmi et al., 2023).

Nguyen and Ishikawa also performed a study on predicting the output energy of BIPV via an ANN (Nguyen and Ishikawa, 2023). In this work, the input is divided into environmental input and structural input variables. The BIPV uses a 4T tandem PV cell, which they compare to their previous work on a 2T tandem PV cell (Nguyen and Ishikawa, 2022) on various directions (east, roof, south, and west). This work examined five ANNs with various hidden layers. They conclude that the ANN-3-hidden-layer model, with an architecture including one input layer (5 neurons), three hidden layers (19, 19, and 16 neurons for the 1st, 2nd, and 3rd layers, respectively), and one output layer (2 neurons), is the optimal model for the predictions. Comparing the 2T and 4T p-Si tandem PV cells revealed that the 4T tandem PV cell outperforms the 2T tandem PV cell in BIPV applications.

4.3. Computational cost trade-offs in real-time BIPV

This section addresses RQ3 by evaluating accuracy–latency–memory trade-offs. In practical BIPV applications, especially those requiring real-time forecasting, optimization, or control, the computational efficiency of ML models becomes a critical concern. High-performing algorithms, such as GPR or hybrid deep learning architectures like CNN-LSTM, offer impressive accuracy but demand considerable computational resources. For example, GPR has a time complexity of $O(n^3)$, which is often impractical for large-scale, continuous real-time data processing (Rasmussen and Williams, 2005). In contrast, lightweight models, such as decision tree regressors, RF, and shallow neural networks, are more suitable for on-device or embedded use due to their lower latency and resource consumption. To address these constraints in real-world BIPV systems, several strategies can be considered.

a). Model simplification: Pruning, quantization, and knowledge distillation can reduce model size and inference time without significantly compromising accuracy (Cheng et al., 2018).
b). Dimensionality reduction: Techniques like principal component analysis or feature selection help minimize input complexity and computation (García et al., 2016). In a BIPV-specific study, Lee et al. (2019) showed that feature reduction significantly improved model simplicity and performance in short-term power forecasting using RNN, making it more feasible for real-time applications.
c). Hardware acceleration: Utilizing edge AI chips, graphics processing units, or tensor processing units can significantly reduce model execution time and power consumption (Zou and Revolutionizing, 2024).

Implementing such optimizations ensures that BIPV systems remain energy-efficient, cost-effective, and responsive under operational constraints. In contexts such as smart buildings, where real-time responsiveness and low power consumption are critical, balancing accuracy with computational feasibility is crucial for successful ML integration.

4.4. Outdoor condition variability in BIPV forecasting

This subsection extends RQ3, examining sensitivity to sky/season/site variability and its impact on the stability of irradiance and power forecasts. Outdoor environmental factors such as soiling, degradation, and cloud variability can significantly impact the real-world performance of BIPV systems. While most ML models focus on clean, ideal datasets, a few recent studies have made strides in incorporating environmental uncertainty into forecasting. For instance, Zhen et al. (2020) used sky image classification to enhance irradiance predictions in facade-integrated systems (Zhen et al., 2020). Their approach included visual sky segmentation and weather scenario classification, such as overcast and scattered clouds, which helped capture the dynamic influence of cloud deformation on irradiance without relying on expensive instrumentation. Although the study focused on short-term irradiance patterns, the underlying method is adaptable to real-time outdoor variability.

Bredemeier et al. (2021) developed a geometry-aware model that incorporated skyline profiles and horizon obstructions in BIPV environments (Bredemeier et al., 2021). By embedding structural shading and local urban geometry into the ML model, the study indirectly accounted for environmental shading and partial irradiance conditions—factors often overlooked in purely satellite- or sensor-based models. Similarly, Shin et al. (2022) performed real-world testing of color-modified BIPV panels under outdoor conditions (Shin et al., 2022). Their model was trained with measured data across different times of day and seasons, capturing variations in spectral response and temperature that can influence module output. While the model did not explicitly quantify degradation or soiling, it represented a meaningful attempt to work with naturally variable outdoor datasets.

Despite these advances, none of the reviewed studies in Section 4 explicitly incorporated long-term module degradation or surface soiling in their modeling pipelines. This gap presents an opportunity for future research to enhance the reliability of ML forecasts for BIPV systems by integrating environmental sensor feedback, cleaning event records, or degradation models.

4.5. Model overfitting and generalizability

This subsection extends RQ3, assessing overfitting risks in small/local datasets and the transferability of models across buildings and seasons. A notable challenge in applying ML to BIPV systems is overfitting, particularly due to the modest size and heterogeneity of BIPV datasets. Unlike centralized PV systems with uniform configurations and abundant sensor data, BIPV setups are customized to individual buildings, resulting in small and often non-representative datasets. The U.S. Department of Energy's Solar Energy Technologies Office has summarized the challenges and opportunities for BIPV, highlighting the lack of standardized data sources necessary for evaluating BIPV systems. Specifically, it notes that data for individual buildings is required, including various levels of building characteristics such as geometry, energy systems, and envelope information. This underscores the customized nature of BIPV installations and the resulting challenges in data standardization and availability (U.S. Department of Energy, 2022). This increases the risk of a model learning noise or building-specific idiosyncrasies, leading to poor generalization (Xu et al., 2023).

Despite this concern, none of the studies reviewed in Section 4 explicitly discussed regularization techniques (L1/L2 penalties, dropout layers) or robust validation strategies (k-fold cross-validation, leave-one-out) as mechanisms to prevent overfitting. The omission suggests a potential research gap, particularly in benchmarking ML performance across diverse building geometries, climates, and sensor setups. To ensure generalizability, best practices from related solar ML literature should be adopted in future BIPV work. These include.

a). Using cross-validation to assess consistency across datasets (Kohavi and Li, 1995).
b). Applying regularization techniques to penalize model complexity (Lever et al., 2016).
c). Implementing early stopping during training to avoid overfitting (Hussein and Shareef, 2024).
d). Leveraging ensemble models to smooth out local biases (Ranglani, 2024).
e). Performing sensitivity analysis to test robustness to missing or noisy inputs (Ankenbrand et al., 2021).

Addressing these methodological aspects will improve the reliability of ML models in BIPV applications, especially for deployment in real-time control, forecasting, or energy management systems.

5. Impact, outlook, and future prospects

This section answers RQ4, translating the impact into cross-cutting priorities for research and practice with recommended directions. Future research in ML for BIPV must move beyond conceptual integration and focus on technically specific and experimentally verifiable pathways. For instance, applying reinforcement learning for dynamic BIPV-envelope control. Integrating phase change materials, thermochromic films, or electrochromic glass into BIPV façades presents a promising avenue for intelligent thermal and daylight regulation (Jaffar et al., 2024; Liu and Wu, 2022). Here, reinforcement learning algorithms, particularly deep Q-networks or proximal policy optimization, can develop dynamic control policies that optimize energy yield and indoor comfort based on real-time environmental inputs (irradiance, ambient temperature, indoor thermal loads) (Guo et al., 2025; Silvestri, 2024). Experimental validation should involve phase change materials or electrochromic-integrated BIPV testbeds equipped with real-time sensors for irradiance, surface temperature, and occupancy, enabling the reinforcement learning agent to learn and adapt policy decisions over time.

Moreover, supervised learning models such as RF or SVM can be trained on long-term monitoring data from inverter status, current-voltage curves, or temperature anomalies to predict system faults before performance is significantly degraded (Thakfan and Bin Salamah, 2024; Chang and Han, 2024). Additionally, a comparative analysis table (Table 3) has been added to summarize the algorithm's strengths, limitations, and performance metrics, including RMSE and R^2, across specific BIPV tasks. This provides readers with a decision-support framework for selecting appropriate ML models based on application needs, dataset sizes, and system constraints.

Table 3
Strengths, limitations, and performance indicators (RMSE or R^2) of various ML algorithms across specific BIPV tasks.

ML algorithms	Task	Strengths	Limitations	Example studies	Typical RMSE/R^2
Artificial neural networks	Power output forecasting	Handles nonlinearities, widely used	Overfitting, data-hungry	Kabilan et al. (2021), Polo et al. (2022) (Kabilan et al., 2021; Polo et al., 2022).	RMSE ≈0.2–4.4 W/m^2
Support vector machines	Irradiance prediction	Good with small data	Limited in multivariate problems	Bredemeier et al. (2021) (Bredemeier et al., 2021)	RMSE ≈ 32–58 kWh/(m^2·a)
Convolutional neural networks-Long short-term memory	Short-term solar forecast	Captures spatial + temporal patterns	Computationally heavy	Zhen et al. (2020), Kim et al. (2024) (Zhen et al., 2020; Kim et al., 2024)	RMSE ≈ 107.6–137.5 W/m^2
Random Forest	Seasonal forecasting	Fast, interpretable	May underperform on deep sequences	Dourhmi et al. (2023) (Dourhmi et al., 2023)	RMSE ≈ 0.18–0.25
Recurrent neural networks	Time-series power output	Effective for sequential data	Training instability	Lee et al. (2019) (Lee et al., 2019)	RMSE ≈ 0.51 kW

Going forward, future research should not only aim at improving predictive accuracy but also prioritize the robustness and interpretability of ML models under noisy, real-world BIPV conditions. There is also a need for standardized benchmarks and open datasets that enable fair comparisons across models. The integration of hardware acceleration, lightweight neural architectures, and online learning techniques will likely be critical in pushing BIPV-ML solutions toward practical, real-time implementation.

Based on the analysis of the reviewed literature, the following core research priorities were identified and ranked.

- Data scarcity and diversity
  - The limited availability of high-resolution, annotated datasets from operational BIPV installations, particularly across diverse climates and building typologies, is the most significant bottleneck. Publicly accessible, standardized BIPV datasets with labeled metadata must be prioritized for model benchmarking and reproducibility.
- Model generalization and transferability
  - ML models trained on narrow, site-specific data often fail to generalize across different buildings. Research should focus on transfer learning, domain adaptation, and meta-learning approaches that enable models to adapt with minimal retraining.
- Interpretability and trust
  - Given the complexity of modern BIPV systems and their interaction with smart building controls, model transparency is essential. Explainable ML techniques, such as SHAP and LIME, or interpretable architectures, should be integrated to support deployment in safety-critical or regulatory settings.
- Computational efficiency for real-time applications
  - Many state-of-the-art models are computationally intensive. Future work must explore model pruning, quantization, or the use of lightweight architectures suitable for edge computing, enabling integration with real-time BIPV control and monitoring platforms.
- Integration with smart building systems
  - There is untapped potential in linking ML-driven BIPV forecasting with heating, ventilation, air conditioning, lighting, and shading systems for holistic energy optimization. Research in this direction will require co-simulation platforms and testbeds that combine building energy models with real-time ML predictions.

By focusing on these ranked priorities, future research can accelerate the practical deployment of ML in BIPV systems, enhancing their energy performance, economic viability, and integration with smart urban infrastructure.

5.1. Proposed perspectives and contributions

While this review has synthesized over 70 studies on the use of ML in BIPV applications, it also offers several original contributions to how ML can be better harnessed to improve the performance and deployability of BIPV systems. These contributions go beyond literature summarization and serve as a foundation for future research directions. This review presents a functional categorization of ML methods based on BIPV data characteristics (e.g., size, resolution, and heterogeneity) and model requirements (e.g., interpretability and real-time deployment). For instance, SVM and RF are highlighted as suitable for small to moderate urban datasets with minimal preprocessing needs. In contrast, LSTM networks are better suited for temporal data streams but may require larger training samples and more computing resources.

An original contribution of this review is the identification of ML's role across the BIPV value chain: from solar irradiance prediction and power output forecasting to control integration with heating, ventilation, and air conditioning systems. This systemic perspective is often lacking in prior studies, which tend to focus narrowly on model accuracy. By framing ML within real-world deployment contexts, such as edge computing and smart building integration, the paper emphasizes the need for scalable and explainable solutions.

This review uniquely emphasizes two underexplored issues: (i) the critical need for interpretable ML models in BIPV settings where energy decisions affect building operations, and (ii) the absence of standardized, publicly available BIPV datasets that enable fair benchmarking. These gaps are synthesized into a ranked research agenda to guide future academic and industrial efforts.

Finally, the review proposes a high-level deployment framework for real-time BIPV applications. It identifies strategies such as model quantization, ensemble pruning, and hybrid control loops (e.g., combining ML with rule-based systems) as promising solutions for integrating ML into resource-constrained building energy management platforms. These perspectives aim to bridge the disconnect between algorithmic development and practical implementation, providing a pathway for ML to support more reliable, efficient, and autonomous BIPV systems.

6. Conclusion

This review aims to clarify how ML approaches can support energy-efficient building systems with BIPV by improving forecasting accuracy and real-time deployability. We organized the evidence around four research questions. RQ1 (irradiance forecasting). Across façades/roofs, urban geometries, and sky/season conditions, deep spatio-temporal models perform best when rich inputs (images or dense telemetry) are available, while support-vector and tree-based methods remain competitive and simpler to configure for small, tabular datasets. Features that encode solar geometry and surface orientation consistently reduce error. RQ2 (power-output forecasting). For short horizons with dense signals, sequence/attention models lead; with modest data or stricter operational simplicity, tree ensembles and SVR deliver robust accuracy. Chaining irradiance predictions into power models helps where on-site irradiance sensing is limited. RQ3 (feasibility & robustness). Edge-suitable solutions favor lightweight ensembles and pruned/quantized neural networks, meeting latency and memory budgets while maintaining accuracy. Robust operation benefits from drift monitoring, retraining triggers, and cross-site/season validation to manage outdoor variability and overfitting risks. RQ4 (recommended directions). The field would benefit most from standardized datasets and metadata, methods that transfer across sites and seasons, interpretable models for operations, and deployment pipelines designed explicitly for edge constraints. In summary, these findings provide actionable guidance for selecting models by task, horizon, data richness, and deployment constraints and outline a clear agenda for advancing BIPV-focused, ML-enabled building energy management.

CRediT authorship contribution statement

Savisha Mahalingam: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Abreeza Manap:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Md Shahariar Chowdhury:** Writing – original draft, Investigation, Formal analysis, Data curation. **Nurfanizan Afandi:** Writing – review & editing, Validation. **Faiz Arith:** Writing – original draft, Investigation, Formal analysis. **Agung Nugroho:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Ministry of Higher Education, Malaysia, through the Fundamental Research Grant Scheme (FRGS), under the project code of FRGS/1/2023/STG04/UNITEN/03/1.

Data availability

Data will be made available on request.

References

Aaditya, G., Pillai, R., Mani, M., 2013. An insight into real-time performance assessment of a building integrated photovoltaic (BIPV) installation in Bangalore (India). Energy Sustain. Dev. 17, 431–437. https://doi.org/10.1016/j.esd.2013.04.007.

Abuella, M., Chowdhury, B., 2017a. Solar power forecasting using support vector regression. https://arxiv.org/abs/1703.09851.

Abuella, M., Chowdhury, B., 2017b. Random forest ensemble of support vector regression models for solar power forecasting. In: 2017 IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf., pp. 1–5. https://doi.org/10.1109/ISGT.2017.8086027.

Acciari, G., Adamo, G., Ala, G., Busacca, A., Caruso, M., Giglia, G., Imburgia, A., Livreri, P., Miceli, R., Parisi, A., Pellitteri, F., Pernice, R., Romano, P., Schettino, G., Viola, F., 2019. Experimental investigation on the performances of innovative PV vertical structures. Photonics 6. https://doi.org/10.3390/photonics6030086.

Ahmed Ali, K., Ahmad, M.I., Yusup, Y., 2020. Issues, impacts, and mitigations of carbon dioxide emissions in the building sector. Sustainability 12. https://doi.org/10.3390/su12187427.

Almaghrabi, S., Rana, M., Hamilton, M., Saiedur Rahaman, M., 2024. Multivariate solar power time series forecasting using multilevel data fusion and deep neural networks. Inf. Fusion 104, 102180. https://doi.org/10.1016/j.inffus.2023.102180.

Alnaser, N.W., 2018. First smart 8.64kW BIPV in a building in awali town at Kingdom of Bahrain. Renew. Sustain. Energy Rev. 82, 205–214. https://doi.org/10.1016/j.rser.2017.09.041.

Alrashidi, H., Issa, W., Sellami, N., Sundaram, S., Mallick, T., 2022. Thermal performance evaluation and energy saving potential of semi-transparent CdTe in façade BIPV. Sol. Energy 232, 84–91. https://doi.org/10.1016/j.solener.2021.12.037.

Ankenbrand, M.J., Shainberg, L., Hock, M., Lohr, D., Schreiber, L.M., 2021. Sensitivity analysis for interpretation of ML based segmentation models in cardiac MRI. BMC Med. Imag. 21, 27. https://doi.org/10.1186/s12880-021-00551-1.

Araz, M., Biyikt, E., Hepbasli, A., 2019. A long-term period performance assessment of a building integrated photovoltaic system. In: E3S Web Conf., 122. https://doi.org/10.1051/e3sconf/201912202007.

Ban-Weiss, G., Wray, C., Delp, W., Ly, P., Akbari, H., Levinson, R., 2013. Electricity production and cooling energy savings from installation of a building-integrated photovoltaic roof on an office building. Energy Build. 56, 210–220. https://doi.org/10.1016/j.enbuild.2012.06.032.

Biyik, E., Araz, M., Hepbasli, A., Shahrestani, M., Yao, R., Shao, L., Essah, E., Oliveira, A. C., del Caño, T., Rico, E., Lechón, J.L., Andrade, L., Mendes, A., Atlı, Y.B., 2017. A key review of building integrated photovoltaic (BIPV) systems. Eng. Sci. Technol. Int. J. 20, 833–858. https://doi.org/10.1016/j.jestch.2017.01.009.

Bonomo, P., Frontini, F., Loonen, R., Reinders, A.H.M.E., 2024. Comprehensive review and state of play in the use of photovoltaics in buildings. Energy Build. 323, 114737. https://doi.org/10.1016/j.enbuild.2024.114737.

Bredemeier, D., Schinke, C., Gewohn, T., Wagner-Mohnsen, H., Niepelt, R., Brendel, R., 2021. Fast evaluation of rooftop and façade PV potentials using backward ray tracing and ML. In: 2021 IEEE 48th Photovolt. Spec. Conf., pp. 294–299. https://doi.org/10.1109/PVSC43889.2021.9518660.

Building-integrated-photovoltaics-BIPV market, (n.d.). https://www.grandviewresearch.com/industry-analysis/building-integrated-photovoltaics-bipv-market (accessed September 19, 2022).

Caiafa, C.F., Sun, Z., Tanaka, T., Marti-Puig, P., Solé-Casals, J., 2021. ML methods with noisy, incomplete or small datasets. Appl. Sci. 11. https://doi.org/10.3390/app11094132.

Chang, Z., Han, T., 2024. Prognostics and health management of photovoltaic systems based on deep learning: a state-of-the-art review and future perspectives. Renew. Sustain. Energy Rev. 205, 114861. https://doi.org/10.1016/j.rser.2024.114861.

Chel, A., Tiwari, G.N., 2011. A case study of a typical 2.32kWP stand-alone photovoltaic (SAPV) in composite climate of new Delhi (India). Appl. Energy 88, 1415–1426. https://doi.org/10.1016/j.apenergy.2010.10.027.

Chen, L., Yang, J., Li, P., 2022. Modelling the effect of BIPV window in the built environment: uncertainty and sensitivity. Build. Environ. 208, 108605. https://doi.org/10.1016/j.buildenv.2021.108605.

Cheng, Y., Wang, D., Zhou, P., Zhang, T., 2018. Model compression and acceleration for deep neural networks: the principles, progress, and challenges. IEEE Signal Process. Mag. 35, 126–136. https://doi.org/10.1109/MSP.2017.2765695.

Choi, W.J., Joo, H.J., Park, J.-W., Kim, S., Lee, J.-B., 2019. Power generation performance of building-integrated photovoltaic systems in a zero energy building. Energies 12. https://doi.org/10.3390/en12132471.

Citakoglu, H., 2015. Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. Comput. Electron. Agric. 118, 28–37. https://doi.org/10.1016/j.compag.2015.08.020.

Deo, R.C., Wen, X., Qi, F., 2016. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl. Energy 168, 568–593. https://doi.org/10.1016/j.apenergy.2016.01.130.

Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N., 2013. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renew. Sustain. Energy Rev. 27, 65–76. https://doi.org/10.1016/j.rser.2013.06.042.

Djeldjli, H., Benatiallah, D., Tanougast, C., Benatiallah, A., 2024. Solar radiation forecasting based on ANN, SVM and a novel hybrid FFA-ANN model: a case study of six cities south of Algeria. AIMS Energy 12, 62–83. https://doi.org/10.3934/energy.2024004.

Dourhmi, M., Benlamine, K., Abouelaziz, I., Zghal, M., Masrour, T., Jouane, Y., 2023. Improved hourly prediction of BIPV photovoltaic power building using artificial learning machine: a case study. In: Ben Ahmed, M., Abdelhakim, B.A., Ane, B.K., Rosiyadi, D. (Eds.), Emerg. Trends Intell. Syst. Netw. Secur. Springer International Publishing, Cham, pp. 270–280. https://doi.org/10.1007/978-3-031-15191-0_26.

Effrosynidis, D., Spiliotis, E., Sylaios, G., Arampatzis, A., 2023. Time series and regression methods for univariate environmental forecasting: an empirical evaluation. Sci. Total Environ. 875, 162580. https://doi.org/10.1016/j.scitotenv.2023.162580.

Eke, R., Demircan, C., 2015. Shading effect on the energy rating of two identical PV systems on a building façade. Sol. Energy 122, 48–57. https://doi.org/10.1016/j.solener.2015.08.022.

Eke, R., Senturk, A., 2013. Monitoring the performance of single and triple junction amorphous silicon modules in two building integrated photovoltaic (BIPV) installations. Appl. Energy 109, 154–162. https://doi.org/10.1016/j.apenergy.2013.03.087.

El Chaar, L., Lamont, L.A., El Zein, N., 2011. Review of photovoltaic technologies. Renew. Sustain. Energy Rev. 15, 2165–2175. https://doi.org/10.1016/j.rser.2011.01.004.

Essam, Y., Ahmed, A.N., Ramli, R., Chau, K.-W., Idris Ibrahim, M.S., Sherif, M., Sefelnasr, A., El-Shafie, A., 2022. Investigating photovoltaic solar power output forecasting using ML algorithms. Eng. Appl. Comput. Fluid Mech. 16, 2002–2034. https://doi.org/10.1080/19942060.2022.2126528.

Gao, M., Li, J., Hong, F., Long, D., 2019. Short-term forecasting of power production in a large-scale photovoltaic plant based on LSTM. Appl. Sci. 9. https://doi.org/10.3390/app9153192.

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J.M., Herrera, F., 2016. Big data preprocessing: methods and prospects. Big Data Anal. 1, 9. https://doi.org/10.1186/s41044-016-0014-0.

Ghenai, C., Ahmad, F.F., Rejeb, O., Bettayeb, M., 2022. Artificial neural networks for power output forecasting from bifacial solar PV system with enhanced building roof surface albedo. J. Build. Eng. 56, 104799. https://doi.org/10.1016/j.jobe.2022.104799.

Guo, F., woo Ham, S., Kim, D., Moon, H.J., 2025. Deep reinforcement learning control for co-optimizing energy consumption, thermal comfort, and indoor air quality in an office building. Appl. Energy 377, 124467. https://doi.org/10.1016/j.apenergy.2024.124467.

Hamad, S.A., Ghalib, M.A., Munshi, A., Alotaibi, M., Ebied, M.A., 2025. Evaluating ML models comprehensively for predicting maximum power from photovoltaic systems. Sci. Rep. 15, 10750. https://doi.org/10.1038/s41598-025-91044-6.
Harrou, F., Kadri, F., Sun, Y., 2020. Forecasting of photovoltaic solar power production using LSTM approach. https://doi.org/10.5772/intechopen.91248.
Hassan, H., Sakr, F., jawazneh, F., Alsmadi, M., Abdulatief, I., Abdallah, S., 2025. Data-DrivenWeather prediction: integrating deep learning and ensemble models for robust weather forecasting. J. Cybersecurity Inf. Manag. 15, 260–284. https://doi.org/10.54216/JCIM.150220.
Hissou, H., Benkirane, S., Guezzaz, A., Azrour, M., Beni-Hssane, A., 2023. A novel ML approach for solar radiation estimation. Sustainability 15. https://doi.org/10.3390/su151310609.
Hussein, B.M., Shareef, S.M., 2024. An empirical study on the correlation between early stopping patience and epochs in deep learning. In: ITM Web Conf.. https://doi.org/10.1051/itmconf/20246401003, 64.
I.E. Agency, 2020. Renewable Capacity Statistics 2019.
I.E.A. and the U.N.E.P, 2019. Global alliance for buildings and construction, 2019 global status report for buildings and construction: towards a zero-emission, efficient and resilient buildings and construction sector. https://www.unep.org/resources/publication/2019-global-status-report-buildings-and-construction-sector.
Ishaque, K., Salam, Z., Taheri, H., Syafaruddin, 2011. Modeling and simulation of photovoltaic (PV) system during partial shading based on a two-diode model. Simulat. Model. Pract. Theor. 19, 1613–1626. https://doi.org/10.1016/j.simpat.2011.04.005.
Jacques Molu, R.J., Tripathi, B., Mbasso, W.F., Dzonde Naoussi, S.R., Bajaj, M., Wira, P., Blazek, V., Prokop, L., Misak, S., 2024. Advancing short-term solar irradiance forecasting accuracy through a hybrid deep learning approach with Bayesian optimization. Results Eng. 23, 102461. https://doi.org/10.1016/j.rineng.2024.102461.
Jaffar, Haleema, Javed, Nasir, Riaz, Ahmad, Ahmad, Hafiz Abrar, Iqbal, Abdul Mueed, Zhou, Chao, Zhang, Jili, 2024. Critical analysis of smart materials in adaptive transparent systems for building façade. Energy Environ., 0958305X241300428 https://doi.org/10.1177/0958305X241300428.
Jiang, H., Lu, N., Qin, J., Tang, W., Yao, L., 2019. A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data. Renew. Sustain. Energy Rev. 114, 109327. https://doi.org/10.1016/j.rser.2019.109327.
Jouini, O., Sethom, K., Namoun, A., Aljohani, N., Alanazi, M.H., Alanazi, M.N., 2024. A survey of ML in edge computing: techniques, frameworks, applications, issues, and research directions. Technologies 12. https://doi.org/10.3390/technologies12060081.
Kabilan, R., Chandran, V., Yogapriya, J., Karthick, A., Gandhi, P.P., Mohanavel, V., Rahim, R., Manoharan, S., 2021. Short-term power prediction of building integrated photovoltaic (BIPV) system based on ML algorithms. Int. J. Photoenergy 2021, 5582418. https://doi.org/10.1155/2021/5582418.
Kappler, T., Starosta, A.S., Munzke, N., Schwarz, B., Hiller, M., 2023. Detection of Shading for Solar Power Forecasting Using ML Techniques.
Khan, S., Mazhar, T., Khan, M.A., Shahzad, T., Ahmad, W., Bibi, A., Saeed, M.M., Hamam, H., 2024. Comparative analysis of deep neural network architectures for renewable energy forecasting: enhancing accuracy with meteorological and time-based features. Discov. Sustain. 5, 533. https://doi.org/10.1007/s43621-024-00783-5.
Kim, J., Kim, E., Jung, S., Kim, M., Kim, B., Kim, S., 2024. Spatio-temporal deep learning-based forecasting of surface solar irradiance: leveraging satellite data and feature selection. Remote Sens. 16. https://doi.org/10.3390/rs16050888.
Kohavi, R., Li, C.-H., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and M o d e l Selection. In: Proc. Fourteenth Int. Jt. Conf. Artif. Intell. IJCAI, International Joint Conferences on Artificial Intelligence, p. 115. In: https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf.
Kuhn, T.E., Erban, C., Heinrich, M., Eisenlohr, J., Ensslen, F., Neuhaus, D.H., 2021. Review of technological design options for building integrated photovoltaics (BIPV). Energy Build. 231, 110381. https://doi.org/10.1016/j.enbuild.2020.110381.
Kumari, P., Toshniwal, D., 2021. Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. Appl. Energy 295, 117061. https://doi.org/10.1016/j.apenergy.2021.117061.
Ledmaoui, Y., El Maghraoui, A., El Aroussi, M., Saadane, R., Chebak, A., Chehri, A., 2023. Forecasting solar energy production: a comparative study of ML algorithms. Energy Rep. 10, 1004–1012. https://doi.org/10.1016/j.egyr.2023.07.042.
Lee, H.M., Yoon, J.H., 2018. Power performance analysis of a transparent DSSC BIPV window based on 2 year measurement data in a full-scale mock-up. Appl. Energy 225, 1013–1021. https://doi.org/10.1016/j.apenergy.2018.04.086.
Lee, J.B., Park, J.W., Yoon, J.H., Baek, N.C., Kim, D.K., Shin, U.C., 2014. An empirical study of performance characteristics of BIPV (building Integrated Photovoltaic) system for the realization of zero energy building. Energy 66, 25–34. https://doi.org/10.1016/j.energy.2013.08.012.
Lee, H.M., Yoon, J.H., Kim, S.C., Shin, U.C., 2017. Operational power performance of south-facing vertical BIPV window system applied in office building. Sol. Energy 145, 66–77. https://doi.org/10.1016/j.solener.2016.07.056.
Lee, D., Jeong, J., Yoon, S.H., Chae, Y.T., 2019. Improvement of short-term BIPV power predictions using feature engineering and a recurrent neural network. Energies 12. https://doi.org/10.3390/en12173247.
Lever, J., Krzywinski, M., Altman, N., 2016. Regularization. Nat. Methods 13, 803–804. https://doi.org/10.1038/nmeth.4014.
Li, D.J., Fang, H., 2013. Study on the power generation data of high-power BIPV grid-connected PV power station. Adv. Mater. Res. 608–609, 132–136. https://doi.org/10.4028/www.scientific.net/AMR.608-609.132.
Li, D.H.W., Chow, S.K.H., Lee, E.W.M., 2013. An analysis of a medium size grid-connected building integrated photovoltaic (BIPV) system using measured data. Energy Build. 60, 383–387. https://doi.org/10.1016/j.enbuild.2013.02.007.
Liang, T., Glossner, J., Wang, L., Shi, S., Zhang, X., 2021. Pruning and quantization for deep neural network acceleration: a survey. Neurocomputing 461, 370–403. https://doi.org/10.1016/j.neucom.2021.07.045.
Liu, X., Wu, Y., 2022. A review of advanced architectural glazing technologies for solar energy conversion and intelligent daylighting control. Archit. Intell 1, 10. https://doi.org/10.1007/s44223-022-00009-6.
Liu, Z., Zhang, Y., Yuan, X., Liu, Y., Xu, J., Zhang, S., He, B., 2021. A comprehensive study of feasibility and applicability of building integrated photovoltaic (BIPV) systems in regions with high solar irradiance. J. Clean. Prod. 307, 127240. https://doi.org/10.1016/j.jclepro.2021.127240.
Liu, Z., Guo, Z., Chen, Q., Song, C., Shang, W., Yuan, M., Zhang, H., 2023. A review of data-driven smart building-integrated photovoltaic systems: challenges and objectives. Energy 263, 126082. https://doi.org/10.1016/j.energy.2022.126082.
Lu, L., Yang, H.X., 2010. Environmental payback time analysis of a roof-mounted building-integrated photovoltaic (BIPV) system in Hong Kong. Appl. Energy 87, 3625–3631. https://doi.org/10.1016/j.apenergy.2010.06.011.
Mahalingam, S., Manap, A., Lau, K.S., Floresyona, D., Chia, C.H., Afandi, N., Chelvanathan, P., Arith, F., Nugroho, A., 2025a. Improving dye-sensitized solar cell longevity: an aerogel encapsulation approach. J. Taiwan Inst. Chem. Eng. 174, 106207. https://doi.org/10.1016/j.jtice.2025.106207.
Mahalingam, S., Manap, A., Floresyona, D., Rabeya, R., Afandi, N., Hasan, Z., Nugroho, A., 2025b. Advancements in flexible perovskite solar cells enabling self-powered systems. Renew. Sustain. Energy Rev. 213, 115488. https://doi.org/10.1016/j.rser.2025.115488.
Mahalingam, S., Manap, A., Lau, K.S., Rabeya, R., Chia, C.H., Xue, K., Chelvanathan, P., 2025c. Recombination suppression in TiO_2/boron-doped reduced graphene oxide-based dye-sensitized solar cells. Renew. Sustain. Energy Rev. 209, 115088. https://doi.org/10.1016/j.rser.2024.115088.
Manna, T.K., Mahajan, S.M., 2007. Nanotechnology in the development of photovoltaic cells. In: 2007 Int. Conf. Clean Electr. Power, pp. 379–386. https://doi.org/10.1109/ICCEP.2007.384240.
Martín-Chivelet, N., Gutiérrez, J.C., Alonso-Abella, M., Chenlo, F., Cuenca, J., 2018. Building retrofit with photovoltaics: construction and performance of a BIPV ventilated façade. Energies 11. https://doi.org/10.3390/en11071719.
Mohanasundaram, V., Rangaswamy, B., 2025. Photovoltaic solar energy prediction using the seasonal-trend decomposition layer and ASOA optimized LSTM neural network model. Sci. Rep. 15, 4032. https://doi.org/10.1038/s41598-025-87625-0.
Mukhoty, B.P., Maurya, V., Shukla, S.K., 2019. Sequence to sequence deep learning models for solar irradiation forecasting. In: 2019 IEEE Milan PowerTech, pp. 1–6. https://doi.org/10.1109/PTC.2019.8810645.
Muneer, T., Younes, S., Munawwar, S., 2007. Discourses on solar radiation modeling. Renew. Sustain. Energy Rev. 11, 551–602. https://doi.org/10.1016/j.rser.2005.05.006.
Nejat, P., Jomehzadeh, F., Taheri, M.M., Gohari, M., Majid, M.Z. Abd, 2015. A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries). Renew. Sustain. Energy Rev. 43, 843–862. https://doi.org/10.1016/j.rser.2014.11.066.
Nguyen, D.C., Ishikawa, Y., 2022. Artificial neural network for predicting annual output energy of building-integrated photovoltaics based on the 2-terminal perovskite/silicon tandem cells under realistic conditions. Energy Rep. 8, 10819–10832. https://doi.org/10.1016/J.EGYR.2022.08.233.
Nguyen, D.C., Ishikawa, Y., 2023. On predicting annual output energy of 4-terminal perovskite/silicon tandem PV cells for building integrated photovoltaic application using ML. Heliyon 9, e18097. https://doi.org/10.1016/j.heliyon.2023.e18097.
Nourani, V., Behfar, N., Booij, M.J., Ng, A., Zhang, C., Mohammadisepasi, S., 2025. Uncertainty quantification in sequential hybrid deep transfer learning for solar irradiation predictions. Eng. Appl. Artif. Intell. 141, 109874. https://doi.org/10.1016/j.engappai.2024.109874.
Omer, S.A., Wilson, R., Riffat, S.B., 2003. Monitoring results of two examples of building integrated PV (BIPV) systems in the UK. Renew. Energy 28, 1387–1399. https://doi.org/10.1016/S0960-1481(02)00257-4.
Pal Shuvo, S., Pal Shibazee, S., Paul, G., Paul Mita, M., Das, C., Malakar, K., 2025. A novel integration of hodrick–prescott filter (Hp-filter) and wavelet transform (WT) with optimize support vector machine (PSO-SVM) in predicting solar radiation. Sci. Rep. 15, 14410. https://doi.org/10.1038/s41598-024-82950-2.
Palani, K., Kota, R., Azad, A.P., Arya, V., 2017. Blue skies: a methodology for data-driven clear sky modelling. In: Proc. Twenty-Sixth Int. Jt. Conf. Artif. Intell. {IJCAI-17}, pp. 3777–3783. https://doi.org/10.24963/ijcai.2017/528.
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. Sci. Total Environ. 715, 136848. https://doi.org/10.1016/j.scitotenv.2020.136848.
Pelle, M., Lucchi, E., Maturi, L., Astigarraga, A., Causone, F., 2020. Coloured BIPV technologies: methodological and experimental assessment for architecturally sensitive areas. Energies 13. https://doi.org/10.3390/en13174506.
Peng, C., Huang, Y., Wu, Z., 2011. Building-integrated photovoltaics (BIPV) in architectural design in China. Energy Build. 43, 3592–3598. https://doi.org/10.1016/j.enbuild.2011.09.032.
Pillai, D.S., Shabunko, V., Krishna, A., 2022. A comprehensive review on building integrated photovoltaic systems: emphasis to technological advancements, outdoor testing, and predictive maintenance. Renew. Sustain. Energy Rev. 156, 111946. https://doi.org/10.1016/j.rser.2021.111946.

Pola, I., Chianese, D., Bernasconi, A., 2007. Flat roof integration of a-Si triple junction modules laminated together with flexible polyolefin membranes. Sol. Energy 81, 1144–1158. https://doi.org/10.1016/j.solener.2007.04.010.

Polo, J., Martín-Chivelet, N., Sanz-Saiz, C., 2022. BIPV modeling with artificial neural networks: towards a BIPV digital twin. Energies 15. https://doi.org/10.3390/en15114173.

Quan, H., Yang, D., 2020. Probabilistic solar irradiance transposition models. Renew. Sustain. Energy Rev. 125, 109814. https://doi.org/10.1016/j.rser.2020.109814.

Ramanan, P., Kalidasa Murugavel, K., Karthick, A., Sudhakar, K., 2019. Performance evaluation of building-integrated photovoltaic systems for residential buildings in southern India. Build. Serv. Eng. Res. Technol. 41, 492–506. https://doi.org/10.1177/0143624419881740.

Ranglani, H., 2024. Empirical analysis of the bias-variance tradeoff across ML models. Mach. Learn. Appl. An Int. J. 11. https://doi.org/10.2139/ssrn.5086450.

Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Processes for ML. The MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001.

Reddy, P., Gupta, M.V.N.S., Nundy, S., Karthick, A., Ghosh, A., 2020. Status of BIPV and BAPV system for less energy-hungry building in india—A review. Appl. Sci. 10. https://doi.org/10.3390/app10072337.

Roth, W., Schindler, G., Klein, B., Peharz, R., Tschiatschek, S., Fröning, H., Pernkopf, F., Ghahramani, Z., 2024. Resource-efficient neural networks for embedded systems. J. Mach. Learn. Res. 25, 1–51. http://jmlr.org/papers/v25/18-566.html.

Roy, A., Ghosh, A., Bhandari, S., Sundaram, S., Mallick, T.K., 2020. Perovskite solar cells for BIPV application: a review. Buildings 10. https://doi.org/10.3390/buildings10070129.

Rüther, R., Nascimento, L., Urbanetz, J., Pfitscher, P., Viana, T., 2010. Long-term performance of the first grid-connected, building-integrated amorphous silicon PV installation in Brazil. In: 2010 35th IEEE Photovolt. Spec. Conf., pp. 2283–2286. https://doi.org/10.1109/PVSC.2010.5617021

Salman, D., Direkoglu, C., Kusaf, M., Fahrioglu, M., 2024. Hybrid deep learning models for time series forecasting of solar power. Neural Comput. Appl. 36, 9095–9112. https://doi.org/10.1007/s00521-024-09558-5.

Samek, W., Wiegand, T., Müller, K.-R., 2017. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models.

Sangrody, H., Zhou, N., Zhang, Z., 2020. Similarity-based models for day-ahead solar PV generation forecasting. IEEE Access 8, 104469–104478. https://doi.org/10.1109/ACCESS.2020.2999903.

S.E.A.C.T. Netherland, 2015. Building Integrated.

S.E.A.C.T. Netherland, 2017. Building Integrated Photovoltaics: Product Overview for Solar Building Skins.

Shin, W.-G., Shin, J.-Y., Hwang, H.-M., Park, C.-H., Ko, S.-W., 2022. Power generation prediction of building-integrated photovoltaic system with colored modules using ML. Energies 15. https://doi.org/10.3390/en15072589.

Shukla, A.K., Sudhakar, K., Baredar, P., 2017. Recent advancement in BIPV product technologies: a review. Energy Build. 140, 188–195. https://doi.org/10.1016/j.enbuild.2017.02.015.

Silvestri, A., 2024. Reinforcement learning control for energy efficient buildings. ETH Zurich A4 - Schlueter, Arno A4 - Zeilinger, Melanie N. A4 - Fiorentini, Massimo. https://doi.org/10.3929/ethz-b-000674480.

Singh, D., Akram, S.V., Singh, R., Gehlot, A., Buddhi, D., Priyadarshi, N., Sharma, G., Bokoro, P.N., 2022. Building integrated photovoltaics 4.0: digitization of the photovoltaic integration in buildings for a resilient infra at large scale. Electronics 11. https://doi.org/10.3390/electronics11172700.

Soleymani, S., Mohammadzadeh, S., 2023. Comparative analysis of ML algorithms for solar irradiance forecasting in smart grids. https://arxiv.org/abs/2310.13791.

Song, J.-H., An, Y.-S., Kim, S.-G., Lee, S.-J., Yoon, J.-H., Choung, Y.-K., 2008. Power output analysis of transparent thin-film module in building integrated photovoltaic system (BIPV). Energy Build. 40, 2067–2075. https://doi.org/10.1016/j.enbuild.2008.05.013.

Song, D., Rehman, M.S.U., Deng, X., Xiao, Z., Noor, J., Yang, J., Dong, M., 2025. Accurate solar power prediction with advanced hybrid deep learning approach. Eng. Appl. Artif. Intell. 148, 110367. https://doi.org/10.1016/j.engappai.2025.110367.

Tandon, A., Awasthi, A., Pattnayak, K.C., Tandon, A., Choudhury, T., Kotecha, K., 2025. ML-driven solar irradiance prediction: advancing renewable energy in Rajasthan. Discov. Appl. Sci. 7, 107. https://doi.org/10.1007/s42452-025-06490-8.

Test bed for building integrated PV (BIPV), (n.d.). https://www.ri.se/en/what-we-do/projects/test-bed-for-building-integrated-pv-bipv (accessed September 19, 2022).

Thakfan, A., Bin Salamah, Y., 2024. Artificial-intelligence-based detection of defects and faults in photovoltaic systems: a survey. Energies 17. https://doi.org/10.3390/en17194807.

Unni, A., Channi, H.K., 2025. Data analytics for performance optimization in renewable energy. In: Optim. Sustain. Energy, pp. 291–327. https://doi.org/10.1002/9781394242139.ch9.

U.S. Department of Energy, 2022. Summary: Challenges and Opportunities for Building-Integrated Photovoltaics RFI. Office of Energy Efficiency and Renewable Energy, pp. 1–7. Washington, DC, Available from: https://www.energy.gov/eere/solar/summary-challenges-and-opportunities-building-integrated-photovoltaics-rfi.

Villegas-Mier, C.G., Rodriguez-Resendiz, J., Álvarez-Alvarado, J.M., Jiménez-Hernández, H., Odry, Á., 2022. Optimized random forest for solar radiation prediction using sunshine hours. Micromachines 13. https://doi.org/10.3390/mi13091406.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., Fouilloy, A., 2017. ML methods for solar radiation forecasting: a review. Renew. Energy 105, 569–582. https://doi.org/10.1016/j.renene.2016.12.095.

Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., Hu, Z., 2015. Photovoltaic and solar power forecasting for smart grid energy management. CSEE J. Power Energy Syst. 1, 38–46. https://doi.org/10.17775/CSEEJPES.2015.00046.

Wang, W., Liu, Y., Wu, X., Xu, Y., Yu, W., Zhao, C., Zhong, Y., 2016. Environmental assessments and economic performance of BAPV and BIPV systems in shanghai. Energy Build. 130, 98–106. https://doi.org/10.1016/j.enbuild.2016.07.066.

Wang, Y., Li, M., Hassanien, R.H.E., Ma, X., Li, G., 2018. Grid-connected semitransparent building-integrated photovoltaic system: the comprehensive case study of the 120 kWp plant in kunming, China. Int. J. Photoenergy 2018, 6510487. https://doi.org/10.1155/2018/6510487.

Wang, C., Ji, J., Yu, B., Xu, L., Wang, Q., Tian, X., 2022. Investigation on the operation strategy of a hybrid BIPV/T façade in Plateau areas: an adaptive regulation method based on artificial neural network. Energy 239, 122055. https://doi.org/10.1016/J.ENERGY.2021.122055.

Wang, W., Yang, H., Xiang, C., 2023. Green roofs and facades with integrated photovoltaic system for zero energy eco-friendly building – a review. Sustain. Energy Technol. Assessments 60, 103426. https://doi.org/10.1016/j.seta.2023.103426.

Wazirali, R., Yaghoubi, E., Abujazar, M.S.S., Ahmad, R., Vakili, A.H., 2023. State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, ML, and deep learning techniques. Elec. Power Syst. Res. 225, 109792. https://doi.org/10.1016/j.epsr.2023.109792.

Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V., 2022. Integrating scientific knowledge with ML for engineering and environmental systems. ACM Comput. Surv. 55. https://doi.org/10.1145/3514228.

Wittkopf, S., Valliappan, S., Liu, L., Ang, K.S., Cheng, S.C.J., 2012. Analytical performance monitoring of a 142.5kWp grid-connected rooftop BIPV system in Singapore. Renew. Energy 47, 9–20. https://doi.org/10.1016/j.renene.2012.03.034.

Xu, P., Ji, X., Li, M., Lu, W., 2023. Small data ML in materials science. npj Comput. Mater. 9, 42. https://doi.org/10.1038/s41524-023-01000-z.

Yang, D., 2020. Choice of clear-sky model in solar forecasting. J. Renew. Sustain. Energy 12, 26101. https://doi.org/10.1063/5.0003495.

Yang, T., Athienitis, A.K., 2016. A review of research and developments of building-integrated photovoltaic/thermal (BIPV/T) systems. Renew. Sustain. Energy Rev. 66, 886–912. https://doi.org/10.1016/j.rser.2016.07.011.

Yang, H., Zheng, G., Lou, C., An, D., Burnett, J., 2004. Grid-connected building-integrated photovoltaics: a Hong Kong case study. Sol. Energy 76, 55–59. https://doi.org/10.1016/j.solener.2003.09.007.

Yoon, J.-H., Song, J., Lee, S.-J., 2011. Practical application of building integrated photovoltaic (BIPV) system using transparent amorphous silicon thin-film PV module. Sol. Energy 85, 723–733. https://doi.org/10.1016/j.solener.2010.12.026.

Yu, G., Yang, H., Luo, D., Cheng, X., Ansah, M.K., 2021. A review on developments and researches of building integrated photovoltaic (BIPV) windows and shading blinds. Renew. Sustain. Energy Rev. 149, 111355. https://doi.org/10.1016/j.rser.2021.111355.

Yu, Y., Cheng, Z., Xiong, B., Li, Q., 2024. Multi-site solar irradiance prediction based on hybrid spatiotemporal graph neural network. J. Renew. Sustain. Energy 16, 43701. https://doi.org/10.1063/5.0207462.

Zaiani, M., Irbah, A., Delanoe, J., Guermoui, M., Boualtit, S.B., Gairaa, K., 2024. Effective prediction of global solar radiation using ML algorithms and satellite images. In: 2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), pp. 530–534. https://doi.org/10.1109/M2GARSS57310.2024.10537514. Oran, Algeria.

Zhen, Z., Liu, J., Zhang, Z., Wang, F., Chai, H., Yu, Y., Lu, X., Wang, T., Lin, Y., 2020. Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image. IEEE Trans. Ind. Appl. 56, 3385–3396. https://doi.org/10.1109/TIA.2020.2984617.

Zou, C., Revolutionizing, M.L., 2024. Harnessing hardware accelerators for enhanced AI efficiency. Appl. Comput. Eng. 47, 141–146. https://doi.org/10.54254/2755-2721/47/20241256.