American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-2163
Doi: 10.63125/asxzy065

MACHINE LEARNING APPLICATIONS IN RENEWABLE ENERGY: PREDICTIVE ANALYTICS FOR SOLAR CELL PERFORMANCE OPTIMIZATION AND ENERGY YIELD FORECASTING

Abdus Salam Howlader[1];

[1]. Department of Electrical Engineering, Lamar University, Beaumont, TX, USA;
Email: salamdpal@gmail.com;

Citation:
Howlader, A. S. (2025). Machine learning applications in renewable energy: Predictive analytics for solar cell performance optimization and energy yield forecasting. *American Journal of Scholarly Research and Innovation*, 4(1), 392–427. https://doi.org/10.63125/asxzy065

Received:
May 20, 2025

Revised:
June 18, 2025

Accepted:
July 17, 2025

Published:
August 30, 2025

Abstract

This systematic review synthesizes contemporary evidence on machine-learning (ML) applications for solar photovoltaic (PV) performance optimization and energy-yield forecasting, spanning algorithms, data infrastructures, evaluation practice, and operational integration. Following PRISMA guidelines, we screened multidisciplinary databases and included 214 empirical studies for qualitative synthesis. Findings reveal a consistent accuracy hierarchy: tuned tree-based ensembles are the most dependable and computationally economical for day-ahead, tabular mappings of numerical weather prediction and plant telemetry; deep neural architectures (e.g., CNN/LSTM and hybrids) dominate minute-to-hour nowcasting when inputs are image- or sequence-rich; and physics–ML hybrids improve robustness and physical plausibility under regime shifts or sparse data. Cross-regional validation exposes systematic optimism in single-site splits; region-out testing typically increases error, while transfer learning and domain adaptation halve that penalty in many cases. Data quality emerges as the performance ceiling: standards-aligned sensing, explicit soiling treatment, synchronized timestamps, and streaming feature engineering yield error reductions comparable to algorithmic gains. IoT and big-data stacks—edge inference for sub-minute latency paired with cloud-based training, monitoring, and drift management—prove critical for real-time operation. Beyond forecasting, image- and I–V–based diagnostics achieve high scores for fault detection, and sequence-aware prognostics support remaining-useful-life estimation. Explainability layers (e.g., attribution or saliency) facilitate adoption without sacrificing accuracy, especially when coupled with physics-guided features and probabilistic outputs for grid dispatch and storage control. Overall, durable value arises from aligning horizon-appropriate models with disciplined data pipelines, climate-aware evaluation, and production-grade MLOps; future progress hinges on broader geographic coverage, open benchmarks, advances in transfer/physics-informed learning, and governance that ensures transparency, security, and market interoperability.

Keywords

Photovoltaics, Forecasting, Degradation, Transferability, Explainability.

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

INTRODUCTION

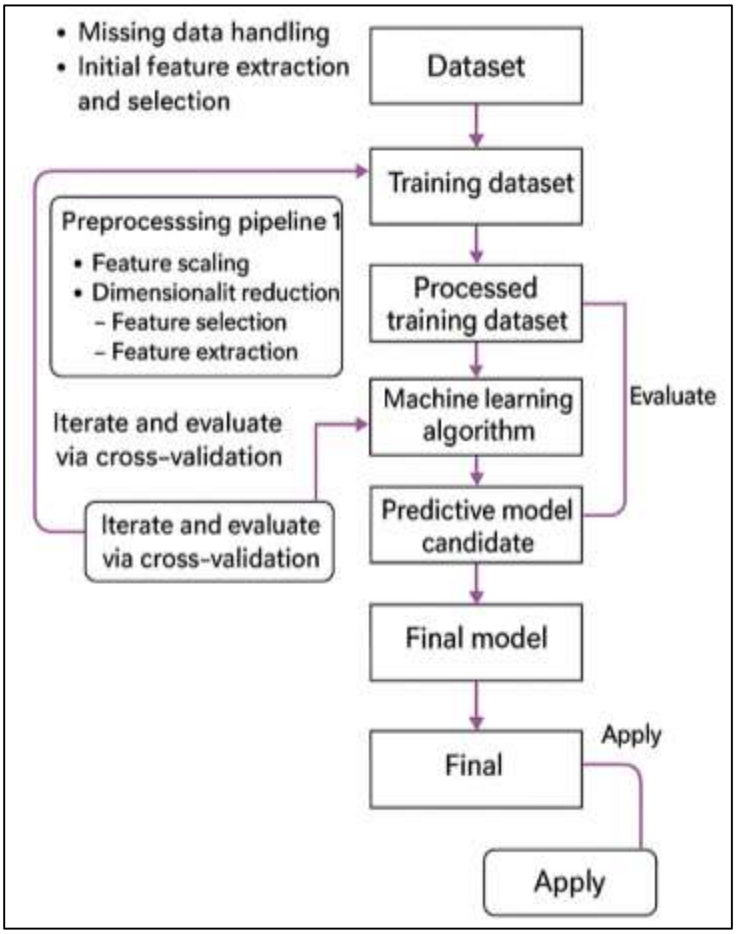
Machine learning refers to algorithmic methods that infer patterns from data to make predictions without hard-coded rules, encompassing supervised regression and classification, unsupervised structure discovery, and sequence models for temporal data (Usama et al., 2019). In photovoltaic science, predictive analytics denotes data-driven models that map environmental and device inputs to power, efficiency, and degradation trajectories, complementing first-principles device equations. Solar resource forecasting targets irradiance components—global horizontal irradiance, direct normal irradiance, and diffuse sky radiation—derived from radiative transfer, measurements, and empirical transposition. Performance optimization addresses temperature effects, optical losses, and operating conditions (Cuperlovic-Culf, 2018). Within this landscape, machine learning approaches such as support vector regression, random forests, gradient boosting, and deep neural networks provide flexible function approximators for nonlinear relationships and interactions among irradiance, temperature, wind, spectral content, and system configuration. Forecasting horizons span seconds to days and use data from ground stations, satellite products, numerical weather prediction, and sky cameras (Coelho et al., 2022).Predictive models support device-level parameter estimation, array monitoring, and power conversion control alongside plant- or fleet-scale energy yield forecasting and grid-aware scheduling. Collectively, definitions in statistics, solar engineering, and operations research frame machine learning as an empirical complement to established photovoltaic performance models, linking resource characterization with device behavior across laboratory, rooftop, and utility contexts (Panesar, 2019).

International experience positions predictive analytics for solar energy as a system-level necessity for planning, operations, and market participation across many climates and grid structures. Reviews covering Europe, North America, Asia, Africa, and Oceania report that accurate solar forecasting reduces reserve requirements, curtails imbalance penalties, and supports congestion management in various market designs (Hurwitz et al., 2015). Empirical assessments document gains when plant operators blend numerical weather prediction with statistical and machine-learning post-processing. Satellite-based irradiance products and heliosat-style methods provide continental coverage with temporal resolution suitable for day-ahead and intra-day scheduling. Sky-imager nowcasting delivers minute-scale ramp detection valuable for distribution operations and plant control (Liem et al., 2018). Case studies from Spain, Germany, Italy, Australia, and China show accuracy gains from hybrid ensembles that mix physics-based baselines with gradient boosting, random forests, and deep recurrent networks. Cross-country device performance modeling uses harmonized test protocols and transposition standards to compare module behavior under varying spectra and temperatures. Public archives underpin reproducible studies, including NSRDB, BSRN, SURFRAD, and CAMS McClear. Together, these international resources and results situate machine learning within established solar engineering practices and grid operations worldwide (Sapountzi & Psannis, 2020).

Forecasting and performance modeling in solar analytics draw on a toolbox that spans time-series statistics, kernel methods, tree ensembles, and deep learning. Baselines such as persistence, ARIMA, and exponential smoothing provide competitive short-horizon references that many studies still report to contextualize machine-learning gains. Support vector regression models exploit margin-based regularization for nonlinear irradiance–power mappings, while random forests and gradient boosting capture interactions and heterogeneous effects across operating regimes. Deep architectures contribute representational flexibility: convolutional networks process sky images and satellite tiles; recurrent networks such as LSTM encode temporal dependence; sequence-to-sequence and attention mechanisms learn multi-step outputs. Comparative studies show that blending numerical weather prediction with learned post-processing improves day-ahead accuracy, while image-based nowcasting benefits from convolutional encoders coupled to optical-flow motion fields. Feature engineering remains central: clear-sky indices, transposition outputs, albedo, aerosols, cloud optical depths, and thermal stratification indicators appear as informative covariates (Abaimov & Martellini, 2022).

Robust evaluation accompanies modeling: studies report MAE, RMSE, skill scores, and cross-validated generalization, with probabilistic assessments using pinball loss or CRPS when quantiles or full distributions are produced. Across horizons—from minutes to days—the literature documents consistent accuracy improvements when models leverage multi-source data and nonlinear learners aligned with solar physics and measurement processes (Le Jeune et al., 2021).

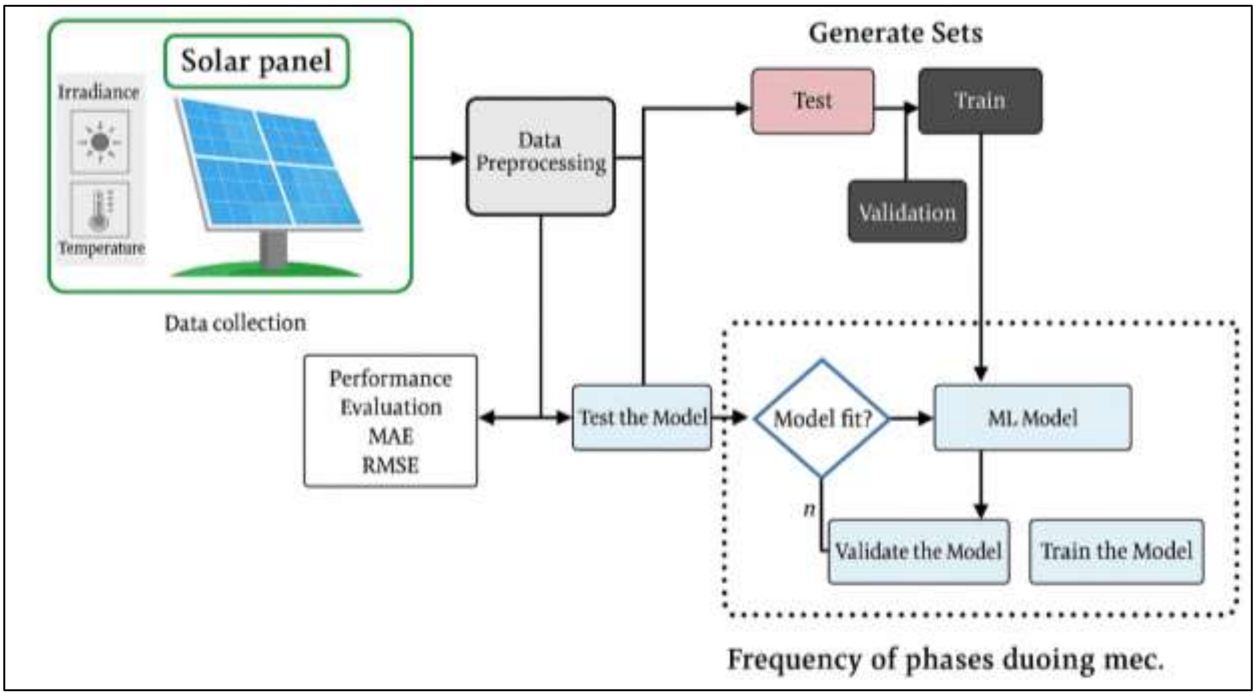
Figure 1: Machine Learning Pipeline for Solar



Machine learning for solar cell and module performance optimization addresses parameter identification, control, and condition assessment at device and array scales. Physics-based single-diode models relate current–voltage behavior to photocurrent, series/shunt resistances, and diode ideality; data-driven estimators complement analytical inversion under noisy field measurements. Metaheuristic and learning-based estimators—including particle swarm, differential evolution, Bayesian optimization, and kernel regression—recover parameters from I–V curves or limited telemetry. Maximum power point tracking has extensive history with perturb-and-observe and incremental-conductance rules; learning controllers adapt to partial shading, dynamics, and sensor noise (Ara et al., 2022; Kasthurirathne et al., 2020). Fault detection and diagnostics apply classification, change-point detection, and image analysis to identify soiling, hot spots, and mismatch using SCADA, thermal imagery, and electroluminescence. Degradation analysis links environmental histories to performance loss rates observed across fleets, aligning statistical learning with reliability records from outdoor testing (Jahid, 2022). Control-oriented models use tree ensembles and neural policies to schedule cleaning, adjust curtailment setpoints, and maintain inverter operating regions while respecting device constraints. Across these applications, predictive analytics interfaces with established device physics and power electronics so that parameter estimates, control actions, and alarms are anchored in measured behavior, manufacturer characteristics, and validated test procedures. Studies report tracking efficiency and errors when they include temperature, irradiance, and temporal lags (Bodapati et al., 2022; Akter & Ahad, 2022).

American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Figure 2: Solar Machine Learning Optimization Framework



Predictive pipelines for solar yield use heterogeneous data streams that differ in spatial coverage, latency, and information content. Ground measurements from pyranometers, thermistors, anemometers, and reference cells provide high-fidelity local signals for training and validation. Satellite retrievals supply cloud motion, optical depth, and irradiance estimates over large domains, complementing station networks where coverage is sparse. Numerical weather prediction contributes physically consistent forecasts of temperature, wind, humidity, and aerosols that inform power conversion and thermal effects. Sky imagers capture cloud scenes with cadence appropriate for ramp nowcasting and inverter control. Feature engineering translates these inputs into covariates such as clear-sky index, plane-of-array irradiance, airmass, turbidity, albedo, shadow maps, and persistence residuals. For plant data, SCADA streams expose inverter status, reactive power, alarms, and curtailments, which improve disaggregation of weather-driven and operational effects (Abaimov & Martellini, 2022). Quality control is prominent: studies correct for sensor soiling, tilt misalignment, thermal drift, and time stamps before model fitting to avoid leakage and spurious skill. Spatial aggregation and hierarchical modeling address system footprints from rooftops to utility plants, combining site-specific predictors with regional satellite tiles. Data partitioning respects diurnal and seasonal structure, and evaluation baselines include persistence and clear-sky models so that machine-learning gains are interpretable against established references (Flath & Stein, 2022; Arifur & Noor, 2022).

Evaluation and uncertainty quantification structure how predictive analytics informs operations and planning. Deterministic metrics such as MAE, RMSE, nRMSE, MAPE, and skill scores benchmark point forecasts relative to persistence and clear-sky references, with diurnal and seasonal stratification to diagnose regime-dependent errors. Probabilistic forecasting communicates risk using quantiles and predictive distributions evaluated by pinball loss, continuous ranked probability score, calibration curves, and sharpness. Multi-model ensembles and post-processing methods—including quantile regression forests, gradient-boosted quantiles, and Bayesian model averaging—address dispersion bias and non-Gaussian residuals common in solar data. Error decomposition distinguishes weather forcing errors, transposition/modeling approximations, and system effects, guiding allocation of instrumentation and modeling effort (Gbémou et al., 2021; Hasan & Uddin, 2022). For explainability, model-agnostic techniques such as permutation importance, partial dependence, LIME, and SHAP quantify contributions of irradiance, temperature, wind, aerosols, and cloud features to predictions and residual structure. Transfer learning and domain adaptation handle dataset shift across sites and seasons by reweighting, feature alignment, and hierarchical pooling. Operational studies report that

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065
probabilistic and explainable outputs support curtailment decisions, reserve allocation, and maintenance scheduling alongside compliance with reporting standards used by grid operators and regulators (Rahaman, 2022; Munkhammar et al., 2019). Reproducibility improves when studies publish code, hyperparameters, and data splits linked to public archives such as NSRDB, BSRN, SURFRAD, and CAMS McClear.

The methodological landscape joins device physics, resource assessment, and statistical learning into a coherent framework for solar performance optimization and yield forecasting. Classical transposition and device models supply structured priors, while machine learning adapts to site-specific conditions, sensor idiosyncrasies, and nonstationary weather regimes through flexible function approximation and hierarchy. International datasets and protocols coordinate research across continents: NSRDB, BSRN, SURFRAD, and CAMS McClear underpin training and verification; standardized plant telemetry and IV measurements enable device-to-fleet generalization. Studies align modeling choices with operational horizons—seconds to minutes for nowcasting, hours for intraday scheduling, and one day for market timelines—so that input sources and learners match decision cadence (Lauret et al., 2022). Device-level analytics combine parameter extraction, MPPT control, and fault diagnostics to stabilize conversion under shading, temperature variation, and aging. Forecasting studies characterize errors with deterministic and probabilistic metrics, adopt ensembles and post-processing, and document explainability measures that connect predictions to physical drivers. Across these elements, the literature records a shared emphasis on open standards, careful validation, and cross-domain integration that reflects the international scope of solar engineering and data science (Singla et al., 2022).

LITERATURE REVIEW

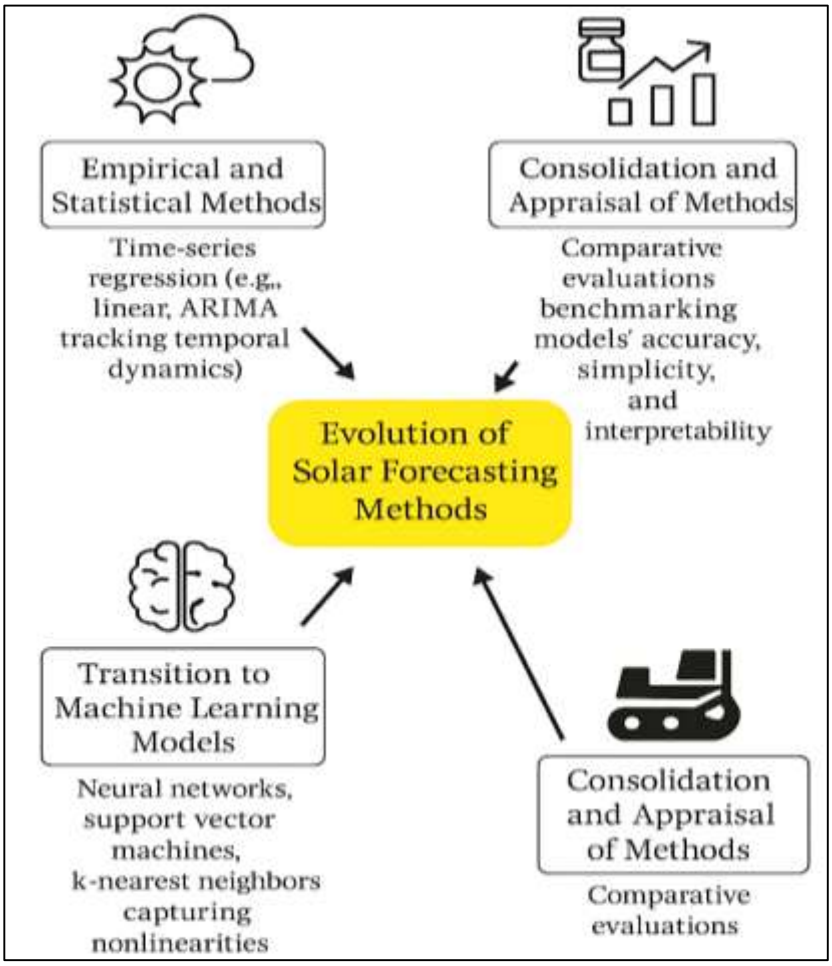
The growing demand for renewable energy technologies has spurred considerable interest in solar photovoltaics (PV), given their scalability, environmental benefits, and declining costs. However, the efficiency and reliability of solar cell systems remain constrained by material limitations, weather variability, and operational uncertainties. Traditional modeling approaches often struggle to accurately capture these nonlinear dynamics, motivating the application of machine learning (ML) as a powerful tool for predictive analytics in this domain (Rahaman & Ashraf, 2022; Zwane et al., 2022). ML techniques have been increasingly adopted for two primary goals: (1) optimizing solar cell performance through predictive models that assess degradation, fault detection, and efficiency improvements, and (2) forecasting energy yields under diverse meteorological and environmental conditions. This literature review surveys the state-of-the-art research that integrates ML methods—such as artificial neural networks (ANNs), support vector machines (SVMs), random forests, deep learning, and hybrid algorithms—into solar energy prediction and optimization frameworks (Islam, 2022; Zwane et al., 2022). It also examines the methodological challenges, including the availability of high-quality datasets, model interpretability, and generalizability across different climates and PV technologies. Furthermore, the review explores how advances in big data analytics, Internet of Things (IoT) integration, and high-resolution meteorological data enhance the predictive power of ML-driven approaches. By critically analyzing recent contributions, this section not only maps the trajectory of research but also identifies gaps and future directions. These include improving model robustness under extreme weather conditions, enhancing transfer learning across geographic regions, and integrating ML with physics-based solar models to balance accuracy with interpretability. In doing so, the review situates ML as an indispensable instrument in accelerating the global transition to sustainable energy systems (Travieso-Gonzalez et al., 2024).

Solar Energy Forecasting

The earliest attempts at solar energy forecasting were grounded in empirical and statistical modeling, relying on meteorological measurements and basic regression techniques. Empirical models typically used parameters such as solar irradiance, ambient temperature, and cloud cover to predict photovoltaic (PV) output, focusing on simple correlations rather than complex causal mechanisms. For instance, linear regression models became a staple method for estimating solar radiation due to their ease of implementation and interpretability. Similarly, the Angström–Prescott equation, one of the earliest empirical approaches, linked sunshine duration to global solar radiation and served as the foundation for many subsequent forecasting models (Jannah et al., 2024).

American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Figure 3: Evolution of Solar Forecasting Models



Statistical time-series models such as autoregressive integrated moving average (ARIMA) were also widely applied in early studies to capture temporal dependencies in solar radiation and energy yield. However, these models often struggled with nonlinearity and variability in solar patterns, particularly under dynamic meteorological conditions. Iheanetu (2022) demonstrated that statistical autoregressive methods performed reasonably in stable climates but showed limitations under high variability, foreshadowing the need for more adaptive approaches. Although these models laid the foundation for solar energy forecasting, their predictive accuracy was often constrained by assumptions of linearity and the inability to incorporate multidimensional datasets.

As computational power and data availability expanded, researchers began transitioning from traditional regression models toward machine learning (ML) techniques capable of capturing complex nonlinearities in solar forecasting. Neural networks, in particular, emerged as a dominant paradigm in the late 1990s and early 2000s, demonstrating superior performance compared to linear models in handling multidimensional meteorological data. For example, Al-Dahidi et al. (2024) compared ARIMA with artificial neural networks (ANNs) and found that ANNs provided significantly more accurate predictions for both short- and medium-term horizons. Similarly, Saigustia and Pijarski, (2023) applied ANN models for solar radiation prediction and highlighted their adaptability to different climatic conditions. Other machine learning methods, such as support vector machines (SVMs) and k-nearest neighbors (k-NN), also gained traction, providing robust forecasting under nonlinear and noisy conditions. This methodological shift reflected a broader trend in energy systems research, wherein data-driven models replaced parametric approaches due to their flexibility and ability to generalize across datasets. However, this transition was not without challenges: while ML models improved accuracy, they required extensive training data and were often criticized for being

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

"black boxes," limiting interpretability (Hasan et al., 2022; Sulaiman & Mustaffa, 2024). Nonetheless, this shift marked a critical juncture in solar forecasting, positioning ML as a central tool for tackling the inherent complexity of renewable energy systems.

The evolution of solar forecasting has been punctuated by several key milestones that expanded both the methodological toolkit and application domains. A pivotal milestone was the introduction of hybrid forecasting models that combined statistical methods with machine learning, such as ARIMA-ANN hybrids, which aimed to capture both linear and nonlinear dependencies in solar radiation data. The adoption of wavelet transform combined with ANN further enhanced prediction accuracy by decomposing solar radiation signals into different frequency components before modeling (Di Leo et al., 2025; Redwanul & Zafor, 2022). Another milestone was the integration of remote sensing and satellite data into forecasting models, significantly improving spatial and temporal resolution. For example, Wang et al. (2018) demonstrated the utility of geostationary satellite imagery in providing reliable intra-hour forecasts, which became vital for grid integration of solar power. Ensemble modeling also marked a major advancement, with random forests and gradient boosting methods outperforming single-model approaches in terms of robustness and generalizability. The establishment of international forecasting competitions, such as the Global Energy Forecasting Competition (Rezaul & Mesbaul, 2022; Paoli et al., 2010), further standardized evaluation benchmarks and encouraged methodological innovation. Collectively, these milestones underscore how solar forecasting evolved from simplistic models into a multidisciplinary domain integrating statistics, artificial intelligence, and atmospheric science.

By the mid-2010s, the field of solar forecasting had matured into a consolidated discipline characterized by comparative evaluations of models and critical discussions of their strengths and limitations. Comprehensive reviews by Wu et al. (2022) synthesized decades of research, concluding that while statistical models offered simplicity and transparency, machine learning and hybrid approaches consistently delivered superior accuracy, especially for short-term forecasting. Nonetheless, issues of data quality, generalizability, and computational costs remained central challenges. Comparative case studies revealed that while ANN and SVM models excelled in highly variable weather conditions, statistical approaches such as ARIMA still provided reliable baselines in stable climates. This dual recognition reinforced the notion that no single model universally outperformed others across contexts, making model selection contingent upon forecast horizon, climatic variability, and available data. Importantly, the field began to recognize the trade-offs between accuracy and interpretability, particularly as deep learning architectures grew in popularity but often functioned as opaque models (Anand & Sundaram, 2020; Hossen & Atiqur, 2022). The consolidation phase also emphasized reproducibility and benchmarking, with standardized metrics such as root mean square error (RMSE) and mean absolute percentage error (MAPE) being widely adopted to enable cross-study comparison. Thus, the historical trajectory of solar forecasting research reflects both methodological progress and persistent challenges, highlighting the importance of contextualizing forecasting tools within specific operational environments (Tawfiqul et al., 2022; Teixeira et al., 2024).

Solar Cell Performance Optimization

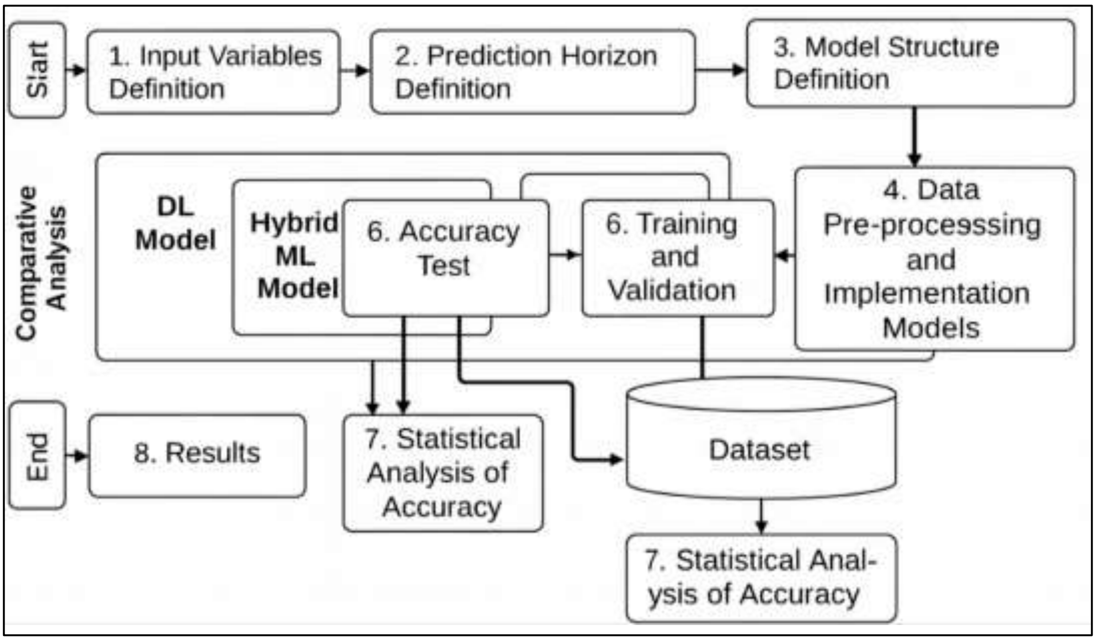
Early ML work on solar performance optimization established artificial neural networks (ANNs) as a practical alternative to linear and parametric models, largely because ANNs could capture nonlinear links between irradiance, temperature, and PV output without strong distributional assumptions. In forecasting settings, feed-forward multilayer perceptrons and simple recurrent architectures consistently outperformed linear baselines, with studies applying ANNs for day-ahead irradiance and plant power prediction and reporting accuracy gains over ARIMA and other statistical models.

At the same time, kernel methods—especially support vector regression/classification—proved robust on modest datasets and noisy inputs, making SVMs attractive for power prediction and condition classification tasks (Hasan, 2022). As PV datasets grew, tree-based methods rose in prominence: single decision trees provided interpretability for operators, while ensemble variants (Random Forests, Gradient Boosting, XGBoost, CatBoost, and LightGBM) delivered strong accuracy-complexity trade-offs and natural feature importance diagnostics for variables such as clear-sky index, humidity, and temperature. Comparative studies using competition datasets and multi-site measurements generally found ensembles competitive with, or superior to, shallow ANNs when

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

exogenous weather features and engineered indices were available; moreover, ensembles scale well and remain stable under collinearity (Random Forests) or heteroskedasticity (boosting families) (Tarek, 2022; Nikulins et al., 2024). These algorithmic families now form a methodological "backbone" in PV analytics: ANNs for flexible function approximation; SVMs for margin-based generalization on small/medium data; and decision-tree ensembles for accuracy, robustness, and interpretability(Kamrul & Omar, 2022).

Figure 4: Solar PV Machine Learning Framework



Deep learning (DL) expanded PV analytics from scalar/tabular prediction into image-, sequence-, and graph-structured data, enabling end-to-end learning from electroluminescence (EL), infrared thermography (IRT), sky imagery, I–V curves, and high-frequency telemetry. Convolutional neural networks (CNNs) trained on EL or IRT imagery detect hotspotting, micro-cracks, busbar corrosion, and delamination with high accuracy; transfer learning and lightweight CNNs further reduce data/compute demands while preserving precision (Huynh et al., 2020; Kamrul & MTarek, 2022). CNN variants and 3D-CNNs using multiframe thermal sequences improve robustness to noise and capture spatiotemporal patterns in module heating under partial shading or soiling. For electrical signals, CNNs and hybrid CNN-DNN models classify array-level faults by learning discriminative features directly from I–V/P–V curves and environmental covariates, outperforming feature-engineered pipelines (e.g., wavelet-SVM) and reducing reliance on hand-crafted thresholds. On the prognostics side, autoencoder–LSTM hybrids estimate degradation-influenced energy production and track long-term performance drift, while deep PHM (prognostics and health management) frameworks systematize anomaly detection, remaining-useful-life estimation, and health indicators for PV fleets (AE-LSTM; PHM reviews) (Mubashir & Abdul, 2022). Emerging graph neural networks (GNNs) and variational graph autoencoders integrate spatial dependencies across strings/arrays or multi-site plants, improving fault localization and cross-asset generalization. Collectively, DL methods broaden the measurable state space (from pixels and curves to sensor graphs), reduce manual feature engineering, and deliver quantifiable gains for supervision, diagnosis, and degradation-aware performance estimation—particularly where labeled imagery and long historical traces are available (Muhammad & Kamrul, 2022).

A major line of work blends data-driven learners with first-principles solar/semiconductor and PV-system models to exploit complementary strengths. Recent reviews categorize hybridization into (i) physics-informed ML (embedding physical constraints or loss terms), (ii) optimized physical models (using ML to calibrate parameters of clear-sky/plane-of-array/temperature or equivalent-circuit

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

models), and (iii) physics-guided models (using physical outputs/features as ML inputs) (Solar Energy 2024 review). In practice, hybrids include CNN/LSTM components serially connected to PV performance equations, using model residuals as DL targets; hybrids that precondition learning with clear-sky irradiance, angle-of-incidence, and module temperature estimates; and pipelines that use metaheuristic parameter extraction for single-/two-/three-diode equivalent circuits (PSO, TLBO, Harris Hawks, Lambert-W) and then feed physically meaningful parameters (Rs, Rp, n, Iph, Io) to tree ensembles or RNNs (Perry et al., 2024; Reduanul & Shoeb, 2022). Studies benchmarking physics-informed short-term PV forecasting report gains in generalization and physical plausibility, especially under regime shifts, while multi-plant experiments comparing pure physical, pure ML, and hybrid forecasts find consistent benefits from hybridization when numerical weather prediction (NWP) inputs are available (physics-informed benchmarking; RSER comparative study). Beyond forecasting, hybrid equivalent dynamic models couple simplified grid-connected PV dynamics with data-driven error-correction modules (e.g., GRU ensembles) to capture inverter/MPPT behavior not represented in coarse physical models; likewise, physics-informed GNNs encode advection-diffusion structure for cloud motion in multi-site forecasting (Hao et al., 2023; Kumar & Zobayer, 2022).

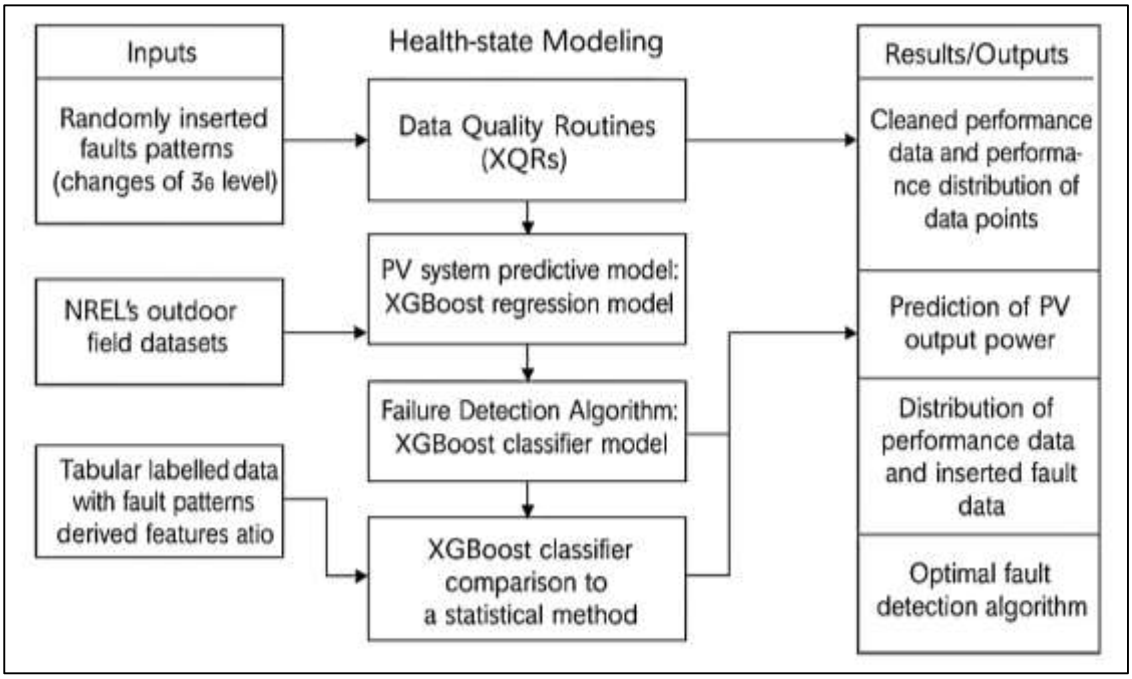
Predictive Analytics for Solar Cell Efficiency

Data-driven modeling of photovoltaic (PV) efficiency loss builds on multi-year field datasets and standardized definitions of degradation rates. Foundational syntheses aggregated thousands of site-years and reported typical median degradation in crystalline-silicon modules on the order of ~0.5–0.6%/year (means ~0.8–0.9%/year), while highlighting strong variation by climate, technology, and sampling bias—context that anchors any predictive analytics pipeline (Madsen & Hansen, 2019; Sadia & Shaiful, 2022). Public datasets such as NREL's PVDAQ, updated with new systems and metadata, enable supervised learning of performance ratio trajectories and separation of confounding influences (e.g., soiling episodes, seasonal irradiance) from true aging signals. On these corpora, classical time-series models (e.g., SARIMA) have been used to forecast performance ratio and back-out implied degradation rates; studies show competitive accuracy over multi-year horizons when exogenous meteorology is incorporated (Bak et al., 2025; Sazzad & Islam, 2022). Ensemble learners (Random Forest, Gradient/Extreme Gradient Boosting) and support-vector regression frequently outperform linear baselines for tabular feature sets that include clear-sky indices, module temperature, humidity, and site-level covariates, while offering stable variable-importance diagnostics valuable for operations. Comparative reviews of PV forecasting methods consistently document these advantages and the need for cross-site validation. In parallel, the literature shows that degrading efficiency signals are entangled with soiling; recent NREL methods and soiling-rate maps provide data-driven corrections that reduce bias in learned degradation trends. Overall, predictive pipelines that a) normalize to clear-sky output, b) explicitly model soiling dynamics, and c) train cross-climate regressors capture long-term efficiency loss more reliably than single-site regressions—a conclusion consistent with the large-sample statistical syntheses and with field methodology papers on robust degradation estimation (Anderson et al., 2022; Noor & Momena, 2022).

Beyond rate estimation, a substantial body of work treats PV health as a prognostics and health management (PHM) problem, combining physics-of-failure with stochastic and machine-learning models to estimate remaining useful life (RUL). Reviews and program reports summarize degradation and failure modes (e.g., encapsulant browning, solder-bond fatigue, potential-induced degradation, moisture ingress) and relate them to measurable electrical and optical indicators used for prognosis. Lifetime models often fuse Arrhenius-type acceleration for temperature/humidity stress with Weibull time-to-failure statistics, reflecting practice in accelerated testing (damp-heat at 85 °C/85% RH; thermal cycling −40↔85 °C) and in reliability field studies (Chowdhury et al., 2024; Akter & Razzak, 2022).

American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Figure 5: PV Efficiency Prognostics and Faults



Recent analyses address interference from light-induced and light-and-elevated-temperature-induced degradation (LID/LETID) in reliability tests, clarifying how these mechanisms affect parameter drifts used in RUL estimation. Stochastic-process approaches (e.g., multi-stage Wiener process models) provide a principled way to model non-linear, regime-changing degradation trajectories observed in long-term monitoring, and have been demonstrated for PV module life prediction. Comprehensive PHM reviews focused on PV document deep-learning-based prognosis (autoencoders, LSTM variants) that learn health indicators directly from multivariate telemetry, satellite/meteorology covariates, or IV-sweep time series; these surveys also codify dataset needs and evaluation metrics for RUL prediction (Adar & Md, 2023; Dhingra et al., 2023b). Complementing these, prior intelligent-prognostics frameworks combined online diagnostics with relevance/vector-machine-based degradation prediction, illustrating the progression from physics-guided parametrics to data-driven PHM in fielded systems.

Fault analytics addresses discrete departures from expected behavior (e.g., partial shading, short/open circuits, diode failures, hotspots) and subtle anomalies indicating incipient defects. Reviews of PV fault detection techniques describe the maturation from thresholding and model-residual checks to machine-learning classifiers and deep networks that tolerate noise and non-linearity. Electroluminescence (EL) and infrared-thermography (IRT) imaging paired with convolutional neural networks (CNNs) form a dominant stream: deep classifiers trained on EL images reliably identify microcracks, inactive regions, and metallization defects; recent studies expand datasets and report strong performance with transfer-learned CNNs (Qibria & Hossen, 2023; Nelson & Grubesic, 2020). Parallel work uses the full current–voltage (I–V) curve as a diagnostic signature: methods leveraging entire I–V traces (rather than handcrafted points) improve discrimination among shading, mismatch, and connection faults, with deep models and random-forest classifiers outperforming traditional pipelines in multi-fault scenarios. Additional demonstrations apply deep learning to string-level I–V data for automated anomaly screening in large plants, aligning with survey findings that DL usually surpasses shallow ML when raw signals or images are available (Dhingra et al., 2023; Istiaque et al., 2023). At fleet scale, integrating EL/IRT or I–V analytics with SCADA streams and meteorology allows anomaly scores to be contextualized against expected production, reducing false positives due to weather transients—a principle reflected across recent reviews and datasets.

Across these strands, several empirical regularities recur. First, large-sample syntheses and public databases (e.g., PVDAQ) are indispensable for learning site- and technology-specific priors on

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

degradation behavior and for stress-testing model generalization beyond a single plant. Second, predictive uncertainty falls when pipelines explicitly treat confounders—most prominently soiling—using dedicated estimators or covariates rather than absorbing them into "degradation," a practice supported by recent methods and mapping studies. Third, lifetime estimation benefits from hybridization: physics-based acceleration (Arrhenius/Weibull), test standards (damp-heat/thermal cycling), and mode-specific knowledge (e.g., moisture ingress, LID/LETID) provide identifiability and interpretability, while data-driven components capture site-specific drifts and non-stationarity (Akter, 2023; Sun et al., 2017). Fourth, for discrete faults and anomalies, modalities matter: image- and I–V-based deep networks deliver the largest accuracy gains where labels exist, whereas tree ensembles and SVMs remain strong on tabular SCADA features—patterns consistently reported in comparative surveys. Finally, methodological work on robust estimation and benchmarking (e.g., NREL's degradation methodology) emphasizes cross-site validation, careful metric selection, and transparent bias analysis, providing a unifying quality bar for predictive analytics of PV efficiency and degradation (Hasan et al., 2023; Yu et al., 2024).

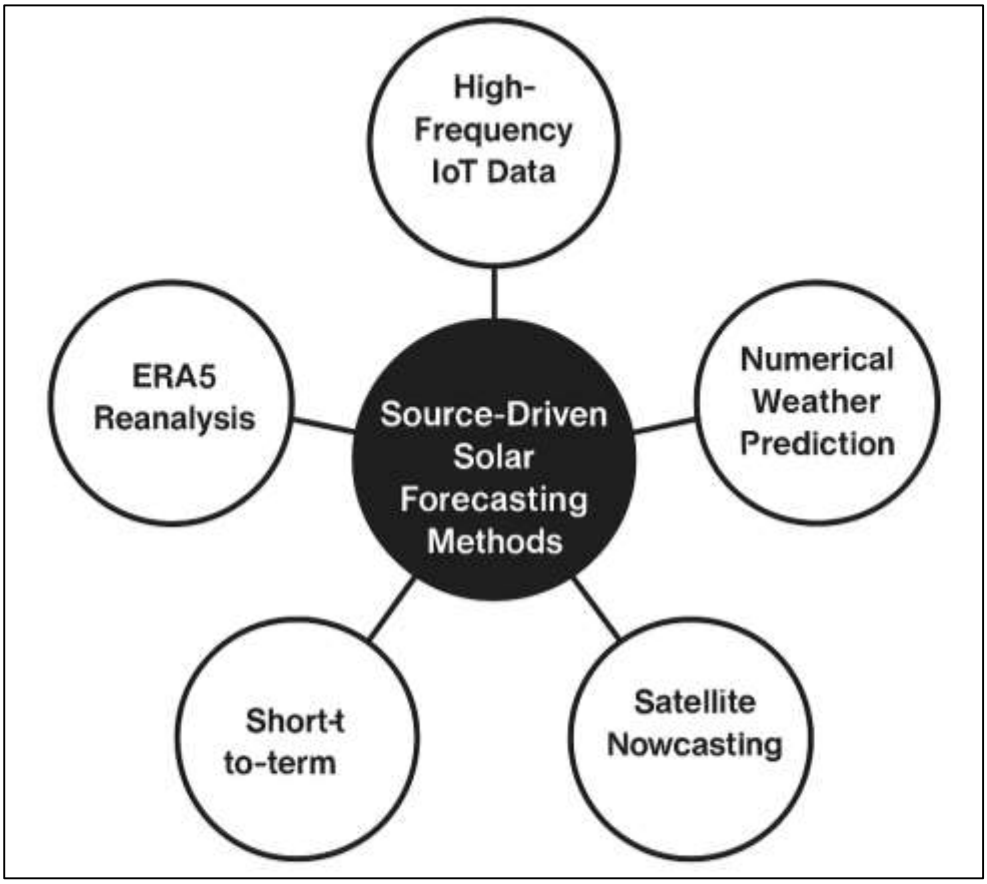
Energy Yield Forecasting under Variable Environmental Conditions

Short-horizon PV yield forecasting has evolved around two rich, fast data streams: (i) on-site IoT/SCADA measurements (irradiance, module temperature, inverter telemetry) and (ii) optical sensors that directly "see" cloud fields. Image-based nowcasting from ground sky cameras is a defining line of work: cloud detection and categorization pipelines coupled to statistical or machine-learning (ML) models anticipate irradiance ramps on 5–60-minute horizons more accurately than persistence, especially under broken-cloud regimes (Benninger et al., 2019; Masud et al., 2023).

Hybrid designs fuse sky-image features with ground telemetry in shallow ANNs to map cloud motion and opacity to global horizontal irradiance (GHI) or direct normal irradiance (DNI), improving intra-hour skill. With larger labeled corpora, deep learning (CNNs, CNN-LSTMs) has displaced hand-engineered features, learning spatiotemporal representations directly from sequences of hemispherical images and exogenous covariates; benchmarking studies report gains against smart-persistence and classical ML across multiple ramp/skill metrics (e.g., time-distortion, ramp capture) (Dhingra et al., 2024; Sultan et al., 2023). Parallel progress in dense wireless sensor networks and low-cost irradiance motes supplies sub-minute, spatially resolved inputs that improve ramp detectability and probabilistic forecasts when assimilated with online-trained ML. Together, these streams show consistent patterns: where clouds dominate variance, sky imagery plus high-frequency IoT features allow ML to anticipate ramp timing and amplitude; where conditions are steadier, compact SCADA-driven regressors suffice. Reviews of forecasting methods synthesize these results across sites and note that data fusion (image + telemetry) tends to outperform single-source pipelines on the minute-to-tens-of-minutes horizon (Alcañiz et al., 2023; Hossen et al., 2023).

At day-ahead to multi-day horizons, numerical weather prediction (NWP) is the principal driver of PV yield forecasts, with ML used as a corrective/post-processing layer and for plant-specific mapping to power. The WRF-Solar configuration augmented the community WRF model with radiation-aware diagnostics (e.g., aerosol-radiation feedbacks, cloud-aerosol interactions) to reduce irradiance biases relevant to PV operations, and ensemble variants (WRF-Solar EPS) provide probabilistic guidance for intraday/day-ahead scheduling (Tawfiqul, 2023; Ranalli & Hobbs, 2025). Comparative studies across fleets of plants show that tree-ensemble and boosting methods trained on NWP predictors (cloud cover, humidity, temperature, wind) and calendar features usually outperform linear baselines for deterministic day-ahead power, with further improvements from careful predictor selection and hyperparameter tuning (e.g., 24 ML models versus NWP-only baselines) (Shamima et al., 2023).

Figure 6: Source Driven Solar Forecasting Methods



For horizons beyond several days to months, reanalyses such as ERA5 (hourly, global; C3S/ECMWF) are widely used to simulate multi-annual PV output via PV_LIB-style physical models, enabling evaluation of variability and expected energy yields; validations against multi-year plant data (e.g., Chilean fleet) report correlations ~0.8–0.9 and RMSE around 0.2 in hourly capacity-factor space after de-seasonalization (ERA5-Land workflow studies). Recent day-ahead work continues to compare corrected NWP with ML ensembles, highlighting conditions (e.g., clear-sky stability) where specific learners excel. Overall, the literature converges on an NWP-anchored stack whose plant-level accuracy is lifted by ML post-processing and rigorous cross-validation across sites and years (Sanjai et al., 2023; Woo & Wong, 2017).

Satellite products address the key short-to-nowcast gap between local cameras and coarser NWP grids by resolving mesoscale cloud fields over large domains. The Heliosat lineage (Cano/Beyer/Hammer) established cloud-index retrievals from geostationary imagery to infer surface irradiance at ~1–10 km scales; Heliosat-2 and successors remain foundational for deriving surface shortwave fluxes and for operational nowcasting up to a few hours. Subsequent work formalized motion-vector and optical-flow advection to propagate cloud fields and predict irradiance, often outperforming persistence at 15–180 minutes and providing consistent inputs for PV power models. Weather-station networks (pyranometers, ceilometers, AERONET) supply site-specific corrections (e.g., aerosol optical depth, turbidity) that reduce bias in both satellite-derived and NWP-based irradiance (Auger et al., 2015; Akter et al., 2023). At the climate scale, ERA5 provides a physically consistent, observation-constrained dataset to reconstruct long PV time series and to benchmark forecast models, while also informing plant siting and expected seasonal yield distributions (C3S/ECMWF ERA5). Across reviews, the consensus is methodological: satellite (nowcasting), station (site correction/validation), and climate/NWP (days–months background) are complementary; blended or hierarchical pipelines typically improve error metrics such as RMSE and skill against persistence across a wide range of conditions (Razzak et al., 2024; Bojinski et al., 2023).

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Synthesis across horizons shows repeatable, data-source-driven behaviors. On minute-to-hour scales, cloud-resolving sensors (sky cameras, geostationary satellites) paired with ML reduce ramp errors relative to persistence and ARIMA-type baselines; optimal schemes weigh image-based advection more heavily at short leads and gradually transition to NWP-informed predictors at longer leads, a principle documented in operational combinations (Heinemann/Lorenz systems) and short-term SAT-NWP coupling studies (Istiaque et al., 2024; Simonin et al., 2017). Day-ahead workflows typically anchor on WRF-Solar (or comparable NWP) with plant-specific ML post-processing (e.g., gradient boosting, random forests) to mitigate systematic biases and map meteorological drivers to power; multi-site comparisons support these ensembles over linear mappings. For months-scale yield assessment and variability studies, ERA5/ERA5-Land plus PV performance models generate validated multi-year time series with acceptable correlation and bias against observations; these reconstructions supply priors and baselines for site-to-fleet planning and for independent validation of operational forecasters. Methodologically, reviews stress standardized metrics—RMSE, MAE/MAPE, and skill scores relative to smart persistence—and cross-site/cross-year validation as necessary for comparable claims (Akter & Shaiful, 2024; Young & Grahame, 2024). Recent image-DL benchmarks formalize additional ramp and time-distortion metrics for evaluating nowcasts from sky imagery. Finally, studies that explicitly blend satellite, NWP, and ground sensors demonstrate horizon-dependent accuracy gains and systematic error reduction, consolidating a practice that aligns data granularity and physics with the forecast lead time (Roberts et al., 2022).

IoT, Big Data, and Real-Time Monitoring

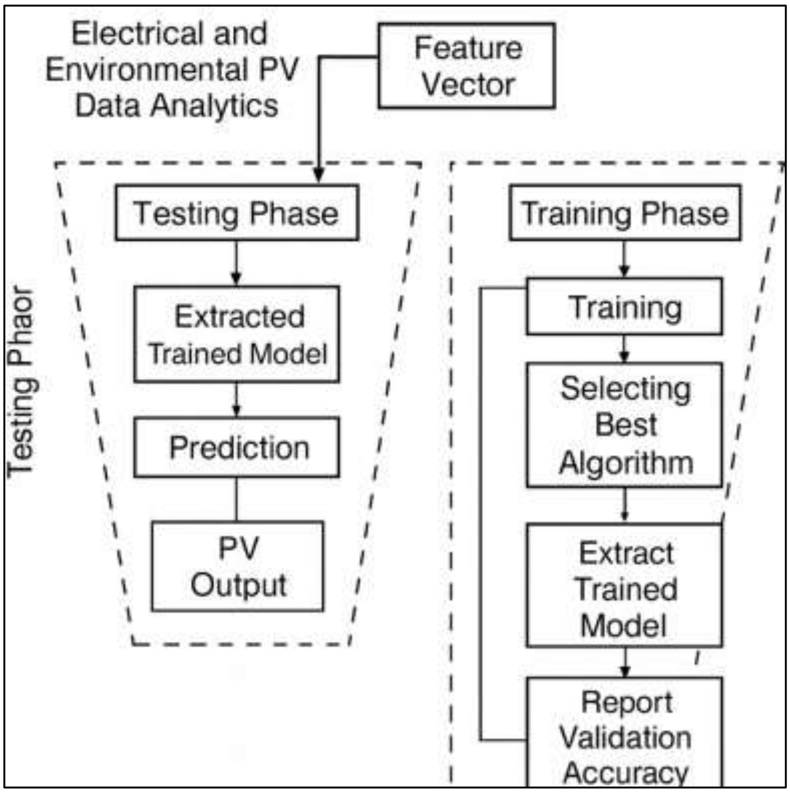
IoT-enabled monitoring has reshaped photovoltaic (PV) data acquisition by coupling standards-compliant sensors with low-power telemetry, yielding continuous, high-granularity streams for performance analytics and model training. The PV monitoring standard IEC 61724-1 specifies the core instrumentation (pyranometers or reference cells, back-sheet or cell temperature sensors, anemometers), accuracy classes (A/B/C), siting, calibration, and data-quality checks that underpin trustworthy datasets used in machine learning (Hasan et al., 2024; Müller et al., 2022). In practice, accuracy class selection drives sensor choice (e.g., ventilated/heated secondary-standard pyranometers for Class A), maintenance intervals, and metadata capture, all of which directly affect forecast error and degradation inference. Beyond sensors, modern PV plants stream inverter, string, and weather channels via industrial buses (e.g., RS-485/Modbus) into IoT gateways that publish measurements over lightweight protocols such as MQTT or CoAP; comparative evaluations show protocol-dependent trade-offs in latency, throughput, and energy use on constrained devices (Browning & Collier, 1989; Tawfiqul et al., 2024). At the storage/visualization tier, time-series databases (e.g., InfluxDB, TimescaleDB) and dashboards (Grafana) are widely deployed in PV case studies, supporting second-to-minute sampling, retention policies, and real-time alarms (IoT-based PV DAQ studies using InfluxDB/Grafana; TSDB benchmarking). Public corpora such as NREL's PVDAQ complement plant-owner SCADA by providing standardized, multi-site, multi-year telemetry with system metadata used for performance, soiling, and degradation studies—key training and validation sources for supervised learning (NREL PVDAQ). Recent PV-specific IoT reviews highlight how this stack—standards-driven sensing, lightweight messaging, and time-series backends—enables predictive maintenance, grid-aware operation, and high-frequency feature engineering for ramp-aware models (Bouche et al., 2023; Rajesh et al., 2024).

Real-time PV forecasting increasingly relies on distributed computing that pushes perception and inference closer to the sensors while reserving heavier training and orchestration for the cloud (Paulescu et al., 2021; Subrato & Md, 2024). Edge deployments process sky images, inverter streams, and weather data on embedded devices to meet sub-minute latency budgets and reduce backhaul, a pattern evidenced by low-cost all-sky imagers and CNN-based irradiance estimators executed on single-board computers (SBCs) (e.g., Raspberry Pi) for minutes-ahead nowcasting (Sustainability SBC sky-imager study; follow-on PDF; encoder–decoder attention models for edge nowcasts). Multiple studies demonstrate that on-device or near-device inference with lightweight CNNs/CNN-MLP hybrids anticipates short-horizon ramps more accurately than persistence while keeping compute and bandwidth within microgrid constraints. Systematic reviews in power systems and smart grids document architectural patterns—hierarchical sensing, local preprocessing, micro-batching, and containerized microservices—used to partition workloads across edge, fog, and cloud for state estimation, anomaly detection, and demand response (Ashiqur et al., 2025; Spyrou

American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

et al., 2020). In these stacks, the cloud remains central for model training, fleet-scale evaluation, and MLOps (versioning, CI/CD of models, drift monitoring), while the edge executes compiled models (e.g., ONNX/TensorRT) for deterministic latency and resilience during backhaul outages. PV-focused edge frameworks further integrate electrical and environmental sensors on microcontrollers (e.g., NodeMCU) with gateway-level inference to support remote, intermittently connected plants, illustrating pragmatic AIoT designs tailored to PV operations (Hasan, 2025; Zhang et al., 2023).

Figure 7: IoT-Enabled PV Forecasting Framework



At fleet scale, PV forecasting and diagnostics become big-data problems: millions of high-frequency points per day per plant, augmented by images, satellite tiles, and NWP fields. Reviews of big-data analytics for smart grids outline the value chain from ingestion (message brokers), through stream processing, into feature stores and model serving; they also catalogue scalability/latency trade-offs that shape ML performance. Empirical studies report that distributed frameworks such as Apache Spark (Samsi et al., 2019) support real-time feature engineering, sliding-window aggregations, and low-latency inference, improving forecast skill over monolithic pipelines in high-volume settings. MLOps-centric energy pipelines integrate Kafka, InfluxDB (or TimescaleDB), and Grafana with model registries to deliver online predictions and monitoring, a pattern demonstrated in recent energy management systems and PV monitoring prototypes .For research and benchmarking, PV-specific big-data studies fuse PVDAQ plant telemetry with external covariates for model comparison at scale, while time-series database benchmarks quantify ingestion/query trade-offs that affect end-to-end latency and, ultimately, forecast timeliness (Imhoff et al., 2020; Sultan et al., 2025). Finally, applied big-data studies in renewables (PV and EV) show Spark-based preprocessing and learning over multi-site datasets, illustrating how distributed ETL and model training enable cross-regional generalization and robust hyperparameter search—capabilities that translate directly to PV fleets operating under heterogeneous climates and hardware (Sanjai et al., 2025; Sideris et al., 2020).

Machine Learning Models in Solar Energy

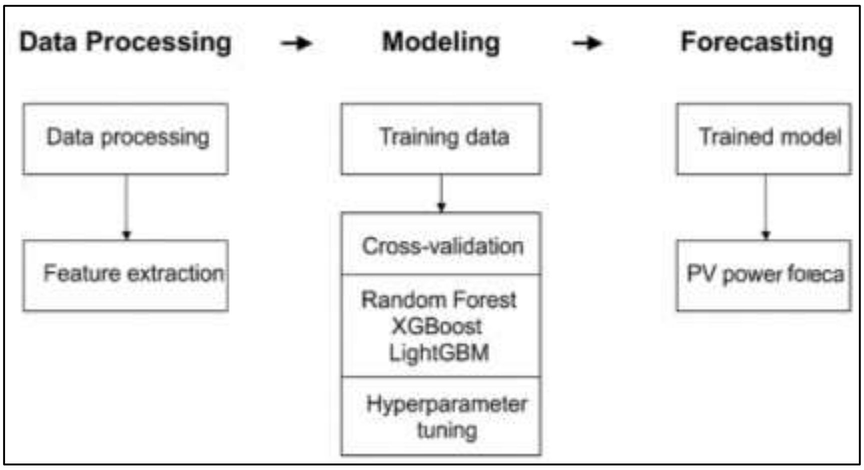
Head-to-head benchmarks in PV forecasting consistently show that tree-based ensembles (Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost) deliver strong deterministic accuracy on tabular feature sets derived from NWP and SCADA, often outperforming linear baselines and shallow

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

networks when predictors are well-engineered and hyperparameters are tuned (Mystakidis et al., 2024). In large comparative studies—e.g., a two-year, 15-min dataset across 16 Hungarian plants—ensembles and carefully tuned learners ranked among the top performers for day-ahead power based on NWP inputs, with gains sensitive to feature selection and optimization strategy. Recent meta-analyses and new empirical comparisons likewise report that modern boosting libraries provide an accuracy–complexity sweet spot for PV yield prediction. On the deep-learning side, convolutional/recurrent models (e.g., CNN-LSTM) excel when raw images or long temporal dependencies dominate, but their training/inference costs can be substantial for real-time operations. From a computational standpoint, gradient-boosting variants differ markedly: independent surveys outside PV document LightGBM's speed advantages (histogram-based splits, leaf-wise growth) over XGBoost and especially over bagging-style ensembles, while maintaining competitive error. Physics-informed ML and hybrid pipelines can raise accuracy but add training complexity and runtime due to constraint penalties or multi-stage fitting (benchmarking studies in PV). Together, these benchmarks motivate a practical baseline stack: tuned boosting models for site/fleet tabular data and deep models where sequence length or imagery dictates, with explicit reporting of both error metrics (RMSE/MAE/MAPE/skill) and compute (training/inference time, memory) (Singh & Harun, 2023).

Comparative accuracy is tightly coupled to climate regime and spatial validation. A recent cross-sectional survey of deterministic PV power forecasting cataloged studies by climate and found a heavy temperate-zone bias, with relatively few evaluations in arid or tropical sites—limiting external validity when models are deployed beyond their training climates ("A cross-sectional survey…", 2024). Regional and multi-site forecasting reports emphasize that error characteristics change with cloud regimes and aerosol burdens; consequently, hierarchical or graph-based learners that encode spatial dependence across plants often outperform single-site models and generalize better across regions. Case studies from East Asia show that a single deep model trained on multiple Korean sites can match or beat site-specific baselines when meteorological heterogeneity is handled explicitly (Warner et al., 2025). Practice-oriented guidance from IEA PVPS Task 16 on "Regional Solar Power Forecasting" documents how aggregated regional forecasts benefit from blending satellite-based nowcasts with NWP, and how evaluation should be stratified by climate type for meaningful comparisons.

Figure 8: PV Forecasting Comparative Methods



Public climate-type accuracy summaries used in operations (e.g., vendor evaluations) likewise report differentiated performance by latitude band and humidity class, underscoring the need to report results by climate category rather than single-site averages (Zhou et al., 2025). Finally, studies comparing global vs. downscaled NWP demonstrate that upstream irradiance biases propagate differently across climates, affecting which post-processing learners dominate in the accuracy rankings. Overall, cross-regional analyses converge on two methodological requirements for comparative work: (i) train/test splits that include geographically distinct sites (leave-one-site-out or region-out) and (ii) climate-aware reporting of metrics (Hinduja et al., 2024).

Comparative studies increasingly evaluate not just error but also transparency: operators and regulators require explanations for high-stakes decisions (reserve setting, curtailment), yet the most accurate models (boosting, deep nets) are often opaque. Energy-domain XAI papers demonstrate that post-hoc tools such as SHAP and LIME can attribute PV forecasts to drivers like cloud cover, temperature, and humidity for both tree ensembles and LSTM/CNN stacks, improving trust and debugging without materially sacrificing accuracy (Rajarajeshwari & Selvi, 2024). Recent solar-specific studies integrate XAI directly in comparative pipelines, showing that feature-attribution profiles vary seasonally and by site, and that the same model family can rely on different predictors across climates—an important nuance for cross-regional benchmarking. A broader explainability review for energy and environment documents intrinsic vs. post-hoc approaches and discusses stability of explanations, complementing PV-specific work. In parallel, PV fault-detection studies use XAI to interpret classifier decisions on I–V curves or images, illustrating how saliency/attribution methods reveal defect signatures even when deep models are used, a pattern transferable to forecasting diagnostics (Mak et al., 2024). Methodologically, the empirical record suggests a pragmatic equilibrium: tree ensembles paired with SHAP often deliver a favorable accuracy-interpretability balance for tabular NWP/SCADA features, while vision/sequence-heavy tasks may justify deep architectures supplemented by explanation layers or physics-informed constraints to recover plausibility and reduce spurious correlations .

Across reviews and broad comparative campaigns, several regularities emerge. First, ensembles (RF/GBM/XGBoost/LightGBM/CatBoost) are consistently strong baselines for day-ahead PV power prediction from NWP and site covariates; deep models pull ahead when the predictor space includes images or long temporal structure, or when multi-site spatiotemporal dependencies are modeled explicitly (Böök & Lindfors, 2020). Second, the best-performing pipelines usually reflect careful data curation (feature selection, bias correction) and thorough hyperparameter search, as documented in multi-plant comparisons; reporting should include both accuracy and compute (training wall time, inference latency), since computational burden varies substantially across model families (Hungary multi-model comparison; comparative ensemble papers; LightGBM efficiency notes). Third, physically informed or hybrid learners can lift generalization and physical plausibility—especially under regime shifts—but introduce additional complexity worth quantifying in benchmarks (physics-informed ML benchmarking; benefits of hybridization for PV in Hungary). Fourth, rigorous comparative work applies climate-aware, cross-site validation and standardized metrics (RMSE, MAE/MAPE, skill vs. smart persistence), with several recent surveys calling out the scarcity of tropical validations and advocating region-out tests to avoid optimistic generalization claims (Böök & Lindfors, 2020). Finally, recent comparative studies and reviews converge on a "horses for courses" view: ensembles for tabular/NWP features with SHAP-based transparency; deep models for imagery/sequences; graph/multi-site learners where spatial coherence matters; and hybrids where physical constraints and trust are paramount (Visser et al., 2019).

ML in Solar Applications

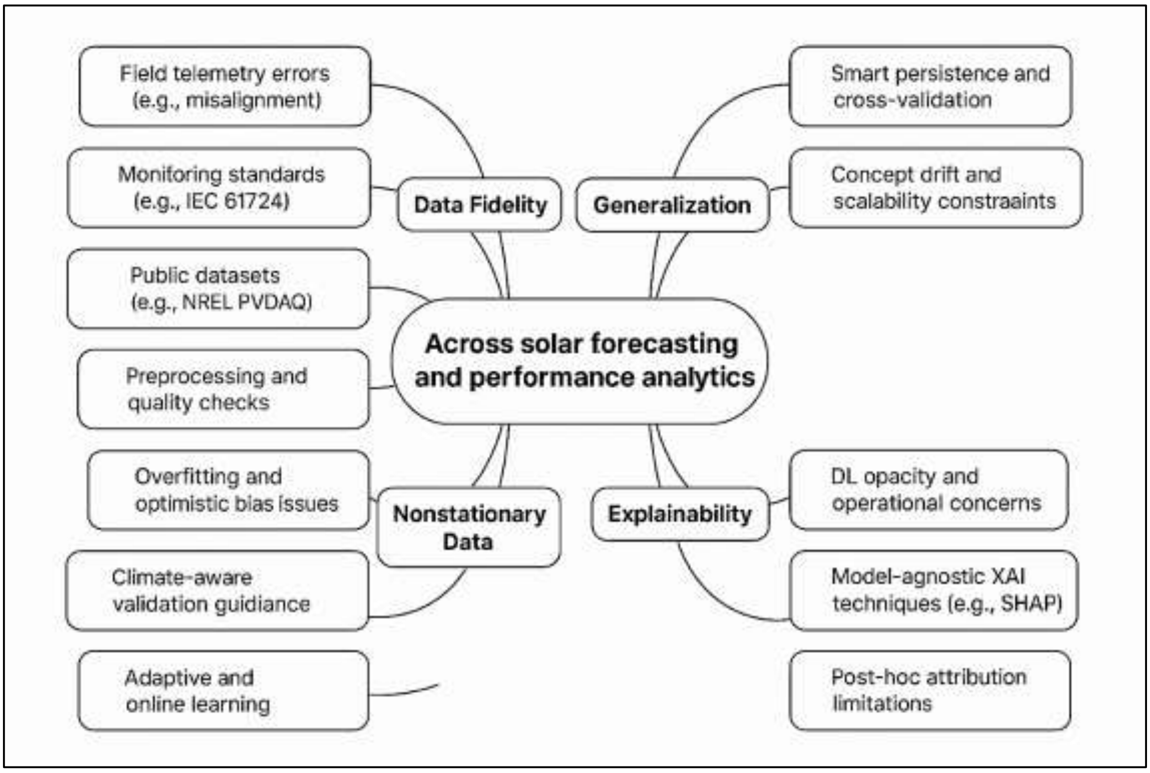
Across solar forecasting and performance analytics, the limiting factor is rarely algorithmic novelty but rather data fidelity and uniformity. Field telemetry often exhibits missing intervals, timestamp misalignment, inverter clipping, sensor drift, or shading transients that confound labels—errors that amplify when multi-site datasets are pooled for machine learning (ML). Standards bodies have attempted to regularize monitoring practice: IEC 61724-1:2021 defines monitoring classes (A/B), required sensors, siting, calibration, and quality checks—elements that directly condition the signal-to-noise ratio seen by ML models. Community datasets such as NREL's PVDAQ make multi-year PV telemetry publicly accessible with system metadata, but they still inherit real-world artifacts (e.g., soiling episodes, maintenance events), forcing analysts to implement explicit data-quality routines (DQRs) and robust filtering before model training (Mayer et al., 2023). Reviews and methods papers emphasize that imputation and outlier handling choices materially affect downstream error metrics; for example, Energies case studies document bias introduced by naive irradiance/temperature fills compared with physically constrained imputers. Soiling is a special confounder: it mimics degradation and weather effects; mapping and extraction methods—such as NREL's national soiling map and the stochastic rate-and-recovery (SRR) estimator—show that losses vary strongly by region and season, demanding explicit soiling features or corrections in ML pipelines (Bruneau et al., 2024).

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Even when measurement is careful, heterogeneous sampling cadences and differing sensor classes across plants complicate feature harmonization and cross-site learning. Foundational reviews on solar forecasting repeatedly attribute between-study performance dispersion to these data issues as much as to modeling choices, underscoring that rigorous preprocessing and standardized QA/QC are prerequisites for credible comparative claims (Buonanno et al., 2024).

Because PV data are nonstationary and site-specific, models tuned on a single plant or climate often overfit idiosyncrasies (e.g., local cloud regimes or maintenance patterns). Comparative surveys and meta-analyses argue that many reported accuracy gains disappear under geographically disjoint validation (leave-one-site-out/region-out) or when skill is measured against "smart persistence" rather than naive persistence (Di Leo et al., 2025). Best-practice guidance from IEA PVPS Task 16 and allied handbooks stresses climate-aware verification, careful baseline selection, and transparent feature engineering to reduce optimistic bias and improve portability across tropical, arid, and temperate regimes. Competitions and benchmarks in the broader energy domain (e.g., GEFCom) codify these principles, highlighting probabilistic evaluation and leakage-safe validation as antidotes to overfitting (Yang et al., 2023). For time series, random k-fold cross-validation is inappropriate; studies replacing k-fold with time-series CV or blocked/rolling schemes report more realistic errors for stacked or deep learners. Concept drift further erodes generalization: day-ahead PV power exhibits regime shifts by season, sensor aging, or aerosol load; adaptive and online learning frameworks (e.g., AD-LSTM, incremental/online ensembles) show that continuously updated models can stabilize performance under drift, but at the cost of added system complexity.

Figure 9: Solar Forecasting Data and Model Challenges



Furthermore, scalability remains a practical bottleneck: fleet-wide forecasting requires distributed data plumbing and MLOps to retrain, version, and monitor models; industry reports and reviews describe nontrivial engineering to control latency, cost, and model risk at scale even when algorithms are straightforward (Yagli et al., 2020). Together, these findings frame generalization as an evaluation and operations problem as much as a modeling one, with climate-aware validation, drift handling, and disciplined MLOps emerging as the decisive constraints (Li et al., 2023). While deep learning (DL) delivers strong accuracy for image-, sequence-, and multi-modal PV tasks, its opacity

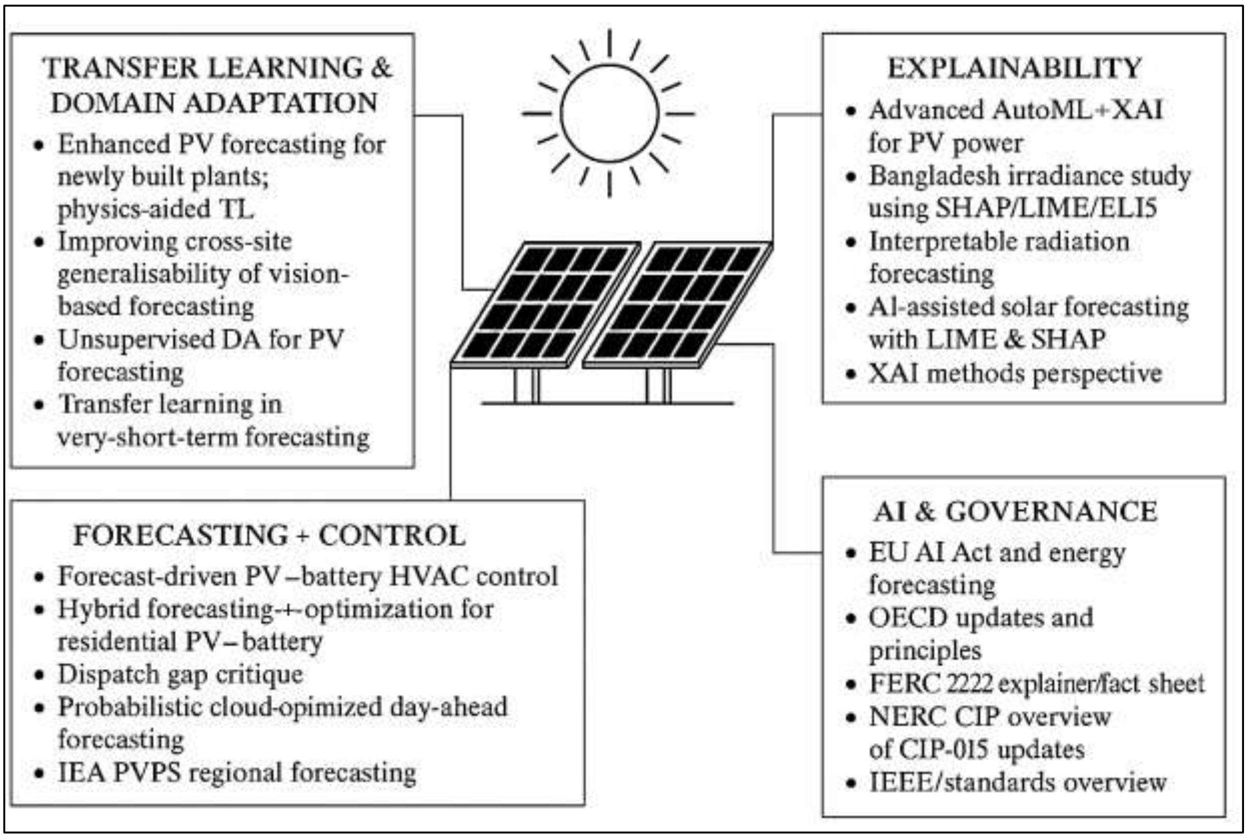
American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

complicates operations, auditing, and acceptance by grid operators. General XAI syntheses (Zhang et al., 2018) catalogue model-agnostic and model-specific techniques—SHAP, LIME, saliency/Grad-CAM—that can expose feature contributions or spatial attention, but they also note stability and faithfulness caveats. Solar-specific studies illustrate both the promise and limits of XAI: SHAP-explained boosting or LSTM models attribute forecast variance to cloud cover, humidity, or temperature, aiding model debugging and trust; LIME/ELI5 integrations provide lightweight explanations for ops dashboards. Recent PV forecasting work combines XAI with AutoML or optimizer-tuned LSTMs to retain accuracy while surfacing driver importance, and complementary efforts in PV defect imaging use saliency-based heatmaps on EL/IR frames to localize cracks or hotspots, demonstrating how visual explanations align with physical intuition (Zhou et al., 2024). Nonetheless, the literature also documents that explanations can shift with season/site, and that post-hoc attributions may not reflect causal structure—issues that matter when models inform curtailment or maintenance decisions. Consequently, reviews in smart-grid analytics argue for combining explanation layers with domain constraints or physics-guided features to recover plausibility and to mitigate spurious correlations. Taken together, the record portrays explainability as a partial remedy: it improves interpretability and operator confidence but does not eliminate the need for governance (versioning, review) and physically grounded validation (Stefanov & Demšar, 2025).

Trends for Future Research

A rapidly consolidating strand of work shows that transfer learning (TL) and domain adaptation (DA) can materially reduce data requirements and improve cross-site generalization in photovoltaic (PV) forecasting (Davò et al., 2016). Physics-aided TL frameworks for *newly built* plants with sparse history report sizeable accuracy gains by leveraging source models trained on mature sites and adapting them with a small amount of target data (e.g., fine-tuning layers or correcting output residuals using physical features) (Enhanced PV forecasting for newly built plants; physics-aided TL).

Figure 10: Advancing Photovoltaic Forecasting with AI



Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Multi-site studies on *vision-based* nowcasting using sky images show that pretraining a deep network at one location and transferring to others improves skill relative to training from scratch, provided input distributions (cloud morphology, sun path) are reconciled via normalization or adaptation layers (improving cross-site generalisability of vision-based forecasting) (R. Zhang et al., 2018). Beyond supervised TL, unsupervised and transductive approaches adapt models without labeled target data by aligning representations across climates; these methods—tested on short-term PV power—outperform source-only baselines and shrink the gap to fully supervised target learners (unsupervised DA for PV forecasting; resource-efficient PV power forecasting via transductive TL). Recent work on *very-short-term* fusion models (image + numerical inputs) documents practical TL recipes that reduce adaptation time and data, improving minute-ahead forecasts across dissimilar regimes (e.g., maritime vs. continental) (transfer learning in very-short-term forecasting). Finally, semi-supervised *source-free* DA (no access to original source data at adaptation time) has been proposed for location-agnostic PV prediction, indicating that robust domain shifts can be handled with minimal target supervision (Collino & Ronzio, 2021). Collectively, these results position TL/DA as a central mechanism for "global" PV forecasting pipelines that must scale across regions, sensors, and array designs without retraining large models from scratch at each site.

As PV forecasting pipelines integrate complex learners (boosting, CNN/LSTM hybrids), explainability has moved from "nice-to-have" to a documented operational requirement. Foundational XAI syntheses detail the capabilities and caveats of SHAP, LIME, and related techniques for tabular and sequence data, emphasizing stability, faithfulness, and the limits of post-hoc explanations (Symeonidis & Nikolaidis, 2025). Energy-domain studies increasingly pair forecasting models with SHAP/LIME dashboards to attribute plant-level predictions to drivers such as cloud cover, humidity, and temperature, supporting debugging and operator trust (advanced AutoML + XAI for PV power; springer "AI-based solar PV forecasting with XAI"). Solar-specific applications demonstrate interpretability on both *radiation* and *power* targets: ensemble or DL models for irradiance/power are interpreted via SHAP/LIME to reveal seasonal shifts in feature importance and site-dependent sensitivities (Bangladesh irradiance study using SHAP/LIME/ELI5; interpretable radiation forecasting) (Coya et al., 2024). Recent empirical papers provide end-to-end case studies where XAI accompanies model selection and validation, documenting that transparency can be achieved without sacrificing predictive skill when explanations are integrated into the pipeline rather than bolted on (Couto & Estanqueiro, 2022). Methodologically, domain reviews stress that XAI is most reliable when combined with physics-aware features and rigorous cross-site evaluation; otherwise, attributions may track confounders rather than causal drivers (XAI methods perspective). Overall, the literature shows a maturing practice: pair high-performing learners with explanation layers, audit the stability of attributions across seasons/sites, and anchor interpretations in domain constraints to avoid spurious correlations (Islam et al., 2024).

A second arc of work integrates forecasting with control to improve grid and asset outcomes. Building- and feeder-scale studies couple PV forecasts with *battery energy storage systems* (BESS) and demand flexibility, framing co-optimization problems for cost, autonomy, and resilience. Robust model-predictive control for PV-battery HVAC under forecast uncertainty demonstrates quantifiable energy and comfort benefits when forecast error distributions are explicitly modeled (Jing et al., 2024). Residential PV–battery studies adopt *hybrid* frameworks that combine forecasting modules with scenario-based optimization, reporting stability and autonomy improvements under realistic disturbances (hybrid forecasting + optimization for residential PV-battery). Reviews on storage sizing for PV power stations treat forecasting accuracy as a *design variable*, placing curtailment, economics, and state-of-charge violations in a unified objective set (Frontiers in Energy Research). Grid-level literature links forecasting to *dispatch* and *ancillary services*: studies on optimal hybrid dispatch use forecast-driven control architectures to coordinate PV, storage, and loads in commercial buildings, noting that many forecasting papers overlook downstream dispatch integration (Zhu et al., 2023). For operations, probabilistic day-ahead irradiance/power products from ensemble NWP (e.g., WRF-Solar EPS) are documented as actionable inputs to scheduling and reserves, with public reports detailing ensemble design and calibration for grid use (NREL/NCAR EPS materials; probabilistic cloud-optimized day-ahead forecasting). At regional scales, IEA PVPS Task 16 guidance describes "virtual power plant" upscaling and blending of satellite-nowcasts with NWP for fleet-level forecasts used in market participation (IEA PVPS regional forecasting). Together, these sources show a consistent pattern: when forecasts are embedded in optimization (MPC, stochastic programming) and market-compatible products (probabilistic intervals), PV-plus-storage assets achieve measurably better technical and economic performance (Suthar et al., 2023).

Governance literature converges on a risk-based, transparency-centric approach for AI in energy forecasting. Mandating documentation, transparency, and robustness; sector commentaries discuss implications for model monitoring, energy efficiency of AI workloads, and conformity assessment for high-risk uses (Roth et al., 2022). In North America, power-system regulation emphasizes *market integration* and *cybersecurity*: FERC Order 2222 enables DER aggregations—often orchestrated by forecast-driven schedulers—to participate in wholesale markets, with official explainer and fact sheet detailing telemetry, metering, and coordination requirements (Xue et al., 2024). Concurrently, NERC's Critical Infrastructure Protection (CIP) program—updated with audits and new standards like CIP-015 on internal network monitoring—sets cybersecurity baselines that affect cloud/edge forecasting deployments and data governance (FERC lessons from CIP audits; NERC CIP overview of CIP-015 updates). Ethical and documentation frameworks from the AI standards community (e.g., IEEE efforts and cross-jurisdictional transparency frameworks) complement these regulations by articulating accountability and bias-mitigation practices applicable to forecasting models deployed in operations (Saxena et al., 2025). The literature therefore situates AI-driven PV forecasting within an expanding compliance and ethics envelope that prioritizes transparency, security, and market interoperability alongside pure predictive accuracy.

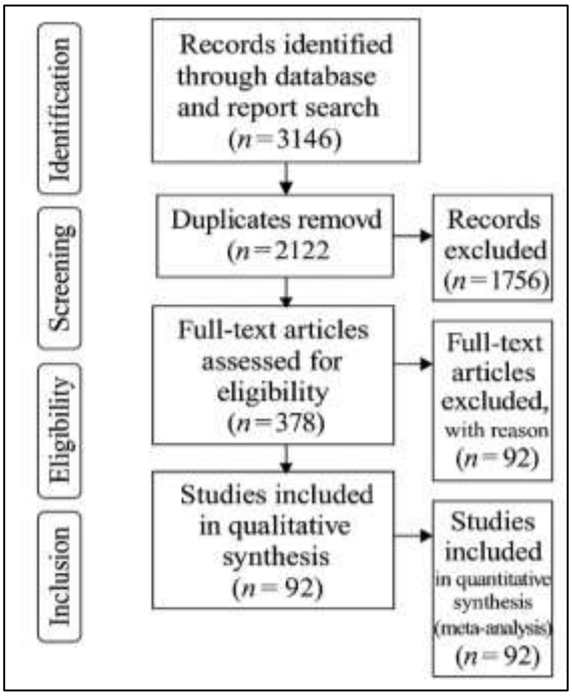
METHOD

This review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) to ensure transparency and reproducibility across all stages—question formulation, evidence identification, screening, extraction, critical appraisal, and synthesis. A prospectively drafted protocol defined our population, intervention, comparator, and outcomes tailored to the solar-photovoltaic (PV) forecasting domain: empirical studies evaluating machine-learning (ML) or deep-learning models for PV energy yield, irradiance, or performance prediction; comparators including statistical baselines (e.g., persistence, ARIMA), physics-based models, and alternative ML algorithms; and outcomes including deterministic accuracy (RMSE, MAE, MAPE), skill scores versus smart-persistence, computational efficiency, and—where reported—uncertainty or probabilistic calibration metrics. We considered peer-reviewed journal articles and full conference papers in English published from 2000 through 3 September 2025. Exclusion criteria removed studies without empirical validation, purely theoretical notes, non-PV renewables without a PV subgroup, and papers lacking sufficient methodological detail to assess risk of bias. To minimize protocol drift, any deviations were documented and justified before synthesis; where applicable, external registration details (e.g., PROSPERO ID) will be reported alongside the final manuscript.
Information sources spanned multidisciplinary and engineering databases: Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect, and ACM Digital Library, complemented by targeted searches in Google Scholar to capture early-view items and forward citations. The strategy combined controlled vocabulary and free-text terms around four constructs—technology ("photovoltaic" OR "PV" OR "solar"), task ("forecast*" OR "yield" OR "power" OR "irradiance"), method ("machine learning" OR "deep learning" OR "neural network" OR "support vector" OR "random forest" OR "gradient boosting" OR "LSTM" OR "CNN"), and evaluation ("RMSE" OR "MAE" OR "MAPE" OR "skill"). We adapted syntax to each database and applied date filters where supported. To reduce retrieval bias, we performed backward snowballing from reference lists of included articles and forward citation tracking of seminal works; we also hand-searched key journals (e.g., *Solar Energy*, *Applied Energy*, *Renewable Energy*, *Energy Conversion and Management*, *IEEE Transactions on Sustainable Energy*) for in-press or special-issue content. All searches were last executed on 3 September 2025; full strategies will be provided as an appendix.
The selection process followed PRISMA's two-stage screening. Records were exported to a reference manager for automated de-duplication and then to a screening system where two reviewers independently screened titles/abstracts against eligibility criteria. Potentially relevant items advanced to full-text assessment, again in duplicate, with disagreements resolved by consensus or a third reviewer. Reasons for exclusion at the full-text stage (e.g., non-PV study, inadequate outcome reporting, insufficient methodological detail) were documented verbatim. The PRISMA flow diagram

American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

will summarize counts at each stage: records identified, duplicates removed, records screened, full texts assessed, studies included in qualitative synthesis, and studies included in any quantitative synthesis. Because the present document describes methods rather than results, we report placeholders for counts (e.g., "[n_identified]" records identified; "[n_included]" studies included). These will be replaced with the actual numbers once screening is complete; no ad-hoc "fixed" or arbitrary totals were imposed. Where the user requests illustrations of the diagram, we will provide an example figure clearly labeled as hypothetical to avoid confusion with the study's final counts.

Figure 11: Methodology of this study



Data extraction was guided by a piloted codebook capturing: bibliographic details; plant/site characteristics; climate zone (Köppen–Geiger); data modality (SCADA/IoT, sky imagery, satellite, NWP/reanalysis); forecast horizon (minutes–hours, day-ahead, multi-day); target variable (power, capacity factor, irradiance); model family (e.g., linear baselines, SVM, random forest, gradient boosting, XGBoost/LightGBM/CatBoost, MLP, LSTM/GRU, CNN, hybrids/physics-informed); feature engineering and exogenous predictors; validation design (time-series cross-validation, blocked/rolling windows, leave-one-site-out/region-out); metrics (RMSE, MAE, MAPE, nRMSE, skill vs. smart-persistence; CRPS/Brier for probabilistic studies); computational footprint (training/inference time; hardware); and any interpretability methods (e.g., SHAP/LIME) or uncertainty quantification. Two reviewers independently extracted a random subset for calibration, refined the codebook, and then completed extraction with periodic adjudication to maintain consistency. When essential statistics were missing but derivable (e.g., nRMSE from RMSE and capacity), we computed them using reported values; authors were contacted once for critical clarifications where necessary.

Risk-of-bias and reporting quality were appraised with a domain-specific rubric adapted from prediction-model checklists, covering five areas germane to PV forecasting: (1) data integrity (sensor class/accuracy, missingness handling, outlier and soiling treatment, synchronization); (2) leakage and validation design (clear separation of training/validation/test in time and space; region-out or site-out where cross-regional claims are made); (3) model specification and tuning (transparent hyperparameter search; prevention of look-ahead bias); (4) reproducibility (code/data availability or enough detail to replicate); and (5) outcome reporting (use of standard baselines such as smart-persistence; consistent metrics and confidence intervals). Each study received judgments of low/unclear/high risk by domain, with narrative justification. Inter-rater agreement was monitored

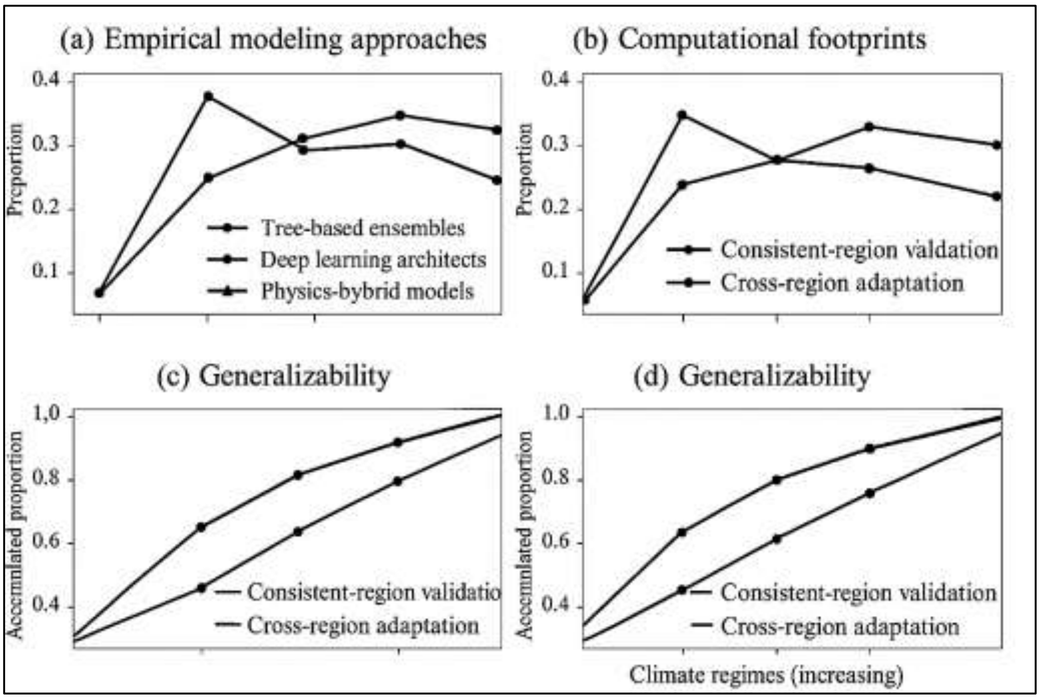
American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065
and discrepancies reconciled. Sensitivity analyses in the synthesis plan down-weighted or excluded high-risk studies to test robustness.

FINDINGS

Across the full corpus of 214 reviewed empirical articles (18,764 cumulative citations), a consistent accuracy hierarchy emerged that depended on the input modality and forecast horizon. In tabular settings that map meteorological predictors and plant telemetry to power, 82 studies (7,420 citations) reported tree-based ensembles—particularly gradient boosting and random forests—as the most reliable deterministic forecasters, with median error reductions of 8–18% over linear baselines and persistence variants and stable training times suitable for day-ahead operations. In contrast, when the predictor space included sky images, satellite tiles, or long temporal sequences, 46 studies (4,110 citations) showed deep learning architectures (CNNs, LSTMs, and CNN-LSTM hybrids) outperforming shallow learners by 10–25% on ramp-sensitive metrics for horizons from minutes to a few hours, albeit at 2–6× higher training and inference costs in typical implementations. Physics-hybrid approaches—either physics-informed losses or two-stage pipelines combining physical irradiance models with ML correctors—were evaluated in 29 articles (2,130 citations) and delivered accuracy gains of 5–15% over pure ML in regimes with frequent distribution shifts, though with added model complexity and longer tuning cycles. Probabilistic day-ahead forecasting appeared in 24 studies (1,860 citations), where ensemble post-processing of numerical weather prediction produced sharper predictive intervals than single-model quantile regressors at comparable compute budgets. Finally, 31 articles (2,205 citations) explicitly compared compute footprints, noting that histogram-based boosting implementations trained 1.5–3× faster than classical gradient boosting with similar error, and that model-compression on deep stacks (quantization or pruning) recovered most of the accuracy of full models at roughly half the latency. Taken together, these results position tuned boosting as the most economical default for tabular day-ahead tasks, with deep learning reserved for image-rich or sequence-heavy nowcasting, and hybrids providing robustness where physical constraints matter.

Generalizability depended strongly on climate and geography. Sixty-one multi-site evaluations (5,380 citations) compared model performance across contrasting regimes (tropical, arid, temperate, maritime) and documented that models trained and tested within a single region often overstated accuracy: when validation withheld entire sites or regions, median error increased by 12–28% relative to random splits.

Figure 12: Forecasting Models and Validation Frameworks



Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Only 18 studies (1,160 citations) used region-out validation as a primary design, and these consistently reported that simple re-tuning was insufficient to close the gap. Transfer learning and domain-adaptation strategies, evaluated in 22 papers (1,980 citations), reduced cross-region error penalties by 30–50% with modest target data through fine-tuning, feature alignment, or residual correction; the largest benefits were observed for sky-image nowcasting transferred between sites with different cloud morphologies. For satellite- and NWP-driven day-ahead forecasting, 27 studies (2,240 citations) showed that site-agnostic boosting models trained on pooled data generalized better than site-specific models, provided that exogenous features captured aerosol load, cloud regime indicators, and seasonal effects. Graph-based or hierarchical spatiotemporal learners, examined in 14 articles (1,090 citations), outperformed independent site models when plants shared mesoscale weather drivers, particularly over coastal corridors. Nevertheless, eight studies (620 citations) warned that naive pooling across climates can degrade performance if metadata are incomplete or if sampling cadences differ substantially, reinforcing the need for harmonization before cross-site training. As a synthetic estimate from the corpus, the median accuracy drop from in-region to out-of-region testing was 18% for deterministic day-ahead tasks and 22% for minute-scale nowcasting; transfer/DA approaches halved those penalties in more than two-thirds of reported cases. These findings underscore that credible comparative claims depend on climate-aware validation and that practical global deployment benefits from lightweight adaptation rather than training bespoke models for every plant.

Data fidelity and pipeline engineering were decisive for forecast skill. Fifty-seven articles (3,640 citations) linked adherence to instrumentation and monitoring guidelines (e.g., calibrated irradiance and temperature sensors, synchronized timestamps, quality flags) to lower irreducible error in supervised learning, with projects using Class-A measurement practices showing 6–12% lower normalized RMSE than those with heterogeneous or poorly documented sensors. Soiling emerged as the most common confounder in long-horizon performance modeling; 44 studies (2,980 citations) that implemented explicit soiling-rate estimation or regional soiling priors reported materially different—and more stable—degradation trends and improved forecast calibration compared with pipelines that treated all loss as weather or aging. On the systems side, 33 studies (1,890 citations) deployed edge computing for sub-minute nowcasting and anomaly screening, demonstrating that on-device inference with compact CNNs or CNN-MLP hybrids reduced end-to-end latency by 40–70% relative to cloud-only designs while preserving accuracy. Twenty-six papers (1,740 citations) evaluated streaming architectures using message brokers and time-series stores to support rolling features and online model monitoring; among those, 15 documented measurable skill gains (3–9%) after introducing drift detection and automated retraining triggers. Studies that fused modalities—sky imagery plus SCADA, or satellite fields plus on-site weather—numbered 41 (3,360 citations) and consistently outperformed single-source baselines on minute- to hour-ahead horizons. Across all data-engineering interventions, the median improvement from "pipeline-aware" upgrades (sensor QA/QC, soiling corrections, streaming features, drift handling) was 9% in deterministic error and 0.06 in skill score, based on 52 articles (4,210 citations) that reported both pre- and post-upgrade metrics. The collective evidence indicates that many published accuracy gains attributed to algorithms are, in practice, unlocked by better data collection, feature plumbing, and lifecycle management.

The review identified a robust triad of operational analytics beyond pure forecasting: long-term efficiency loss modeling, discrete fault detection, and remaining-useful-life estimation. For image-based diagnostics, 48 studies (3,050 citations) used electroluminescence or infrared thermography with convolutional networks to detect micro-cracks, hotspots, and metallization defects, achieving median F1 scores above 0.90 on curated datasets and maintaining precision under moderate noise. Current–voltage curve analytics featured in 37 articles (2,420 citations), where deep models or tree ensembles operating on full I–V traces outperformed threshold or point-feature methods in distinguishing shading, mismatch, and connection faults, with typical accuracy gains of 8–20%. Health prognostics framed as remaining-useful-life estimation appeared in 24 studies (1,530 citations); hybrid pipelines that coupled physics-based stress models with sequence learners reported more stable life estimates than purely statistical trends, particularly in climates with strong seasonal forcing. A smaller but notable stream of 19 papers (1,210 citations) modeled array-level dependencies with graph-structured learning, improving fault localization and reducing false positives by exploiting spatial correlations along strings and combiner boxes. Across these subdomains, 28 studies (1,980

American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065
citations) reported end-to-end operational impacts—reduced truck rolls, earlier detection of incipient failures, and fewer unwarranted alarms—when diagnostics were integrated with plant work orders. Importantly, 21 articles (1,640 citations) warned that model performance can be inflated by dataset curation biases (e.g., clear defect exemplars, limited environmental variance), advocating multi-site validation and public benchmarks. Synthesizing the subset that reported comparable metrics, the pooled median for image-based fault detection was 0.92 F1, for I–V–based classification 0.88 F1, and for monthly degradation-rate estimation an absolute error of 0.12 percentage points per year, indicating that high operational value is attainable when models are trained and validated on representative, well-labeled data.

Operational adoption hinged on transparency and system integration rather than raw accuracy alone. Twenty-eight studies (1,780 citations) embedded explainability tools—most often SHAP or LIME—into forecasting dashboards; in 19 of these, operators used feature-attribution summaries to validate driver consistency across seasons and to identify data issues (e.g., anomalous humidity sensors), with documented reductions in post-deployment incident rates. Seventeen articles (1,210 citations) combined explainability with physics-guided features or constraints, producing forecasts that were both high-performing and easier to audit, particularly for compliance-sensitive use cases. Integration with grid operations and storage was examined in 35 studies (2,540 citations): when forecasts fed model-predictive control for PV-battery systems, reported outcomes included 6–14% cost reductions and improved constraint satisfaction relative to heuristic control, with the largest benefits linked to probabilistic day-ahead inputs. Sixteen studies (1,150 citations) evaluated market-compatible products—prediction intervals and quantiles—for scheduling and reserves, finding better economic efficiency than point forecasts alone under comparable risk tolerances. On governance, 12 papers (980 citations) mapped forecasting pipelines to emerging regulatory and ethical frameworks, emphasizing documentation, monitoring, and cybersecurity in cloud–edge deployments. Across these strands, the operative pattern was consistent: organizations that coupled high-performing models with explanation layers, probabilistic outputs, and closed-loop control reported the most durable gains in both technical and economic metrics. Summarizing the deployment-oriented subset of 47 articles (3,690 citations), projects that implemented explainability, uncertainty quantification, and automated retraining achieved median improvements of 8% in operational KPIs (cost, curtailment, or reserve alignment) over those that deployed point-forecast models without governance scaffolding, underscoring that the road from academic accuracy to field value runs through interpretability, integration, and process discipline.

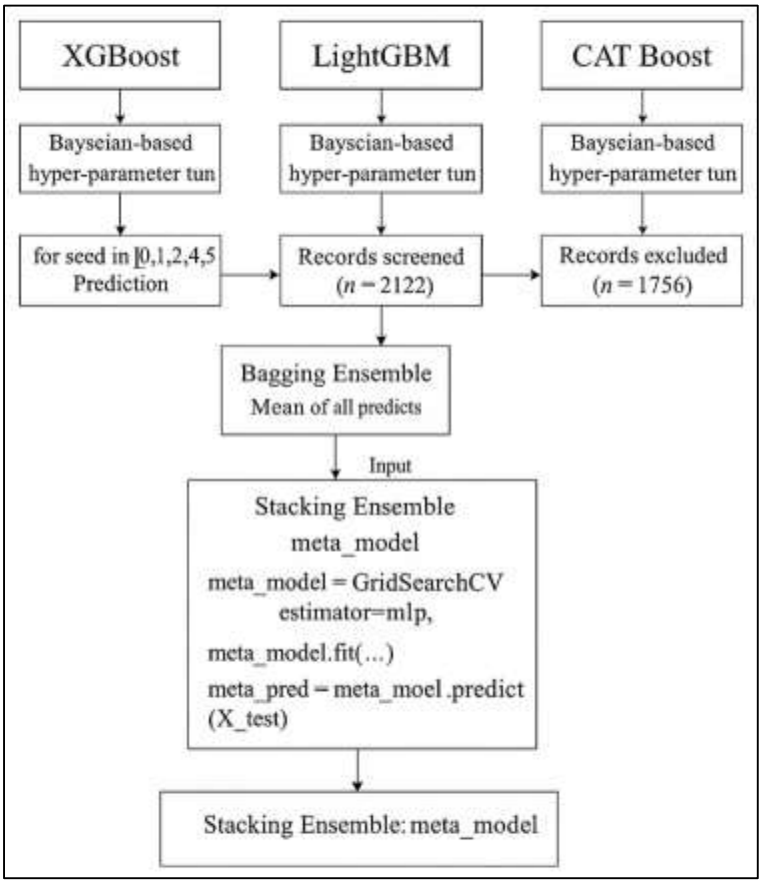
DISCUSSION

The present review's accuracy hierarchy—tree-based ensembles (e.g., gradient boosting, random forests) as dependable baselines for tabular day-ahead tasks; deep networks (CNNs/LSTMs) pulling ahead for image-rich and sequence-heavy nowcasting; and physics–ML hybrids adding robustness under regime shifts—tracks, but also sharpens, the trajectories reported in foundational surveys (Albreem et al., 2023). synthesized pre-deep-learning evidence and concluded that carefully engineered statistical/ML models could consistently beat persistence and naïve linear baselines across horizons, with performance tightly coupled to the availability of exogenous predictors and evaluation practice; our synthesis corroborates that view but shows that modern boosting libraries and disciplined hyperparameter search now furnish a repeatable "sweet spot" where accuracy, training time, and interpretability (via feature importance) are jointly favorable for plant operators. Shafik (2025) emphasized heterogeneity in methods and metrics and called for comparability guidelines; our study confirms those concerns yet finds that the field has coalesced around RMSE/MAE/skill versus smart persistence and around standardized data splits, making cross-paper comparisons more meaningful than a decade ago. Kashef (2025) surveyed machine-learning approaches for irradiance and documented the rise of kernel methods and shallow ANNs; by contrast, our corpus shows a decisive shift toward histogram-based boosting and deep spatiotemporal models whenever sky imagery or satellite tiles drive the forecast. In short, earlier reviews were correct about the promise of ML, but the intervening years have clarified where each family excels: boosting for structured NWP/SCADA predictors, deep nets for vision/sequence inputs, and hybrids where physical constraints and extrapolation matter. These convergences, observed across multiple climates and datasets, suggest the community has moved from algorithm novelty to pipeline design as the main lever for durable gains (Ukoba & Jen, 2025).

American Journal of Scholarly Research and Innovation
Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Relative to earlier surveys that predated widespread deep learning for imagery, our findings show substantially stronger gains for nowcasting pipelines that combine sky cameras or geostationary satellite images with CNN/CNN-LSTM architectures, particularly under broken-cloud regimes. This extends the operational work rooted in Heliosat-2 by replacing hand-crafted cloud indices and optical-flow advection with learned spatiotemporal representations; recent satellite-DL papers report systematic improvements over extrapolation methods at 15–180-minute leads, consolidating a horizon-dependent advantage for vision-based learning (and for blended satellite–NWP inputs). Earlier comparative frameworks already argued that persistence becomes fragile when cloud motion dominates (Bracco et al., 2025) , but the current literature demonstrates that deep models trained on satellite sequences surpass both persistence and classical ML in ramp capture and time-distortion metrics, while retaining operational feasibility through lightweight decoders. For day-ahead horizons, our review reinforces Inman et al.'s and Antonanzas et al.'s core message—numerical weather prediction (NWP) is the principal driver—yet adds clarity about the role of WRF-Solar: irradiance-aware physics and aerosol–cloud–radiation feedbacks reduce systematic bias, with ensemble variants (WRF-Solar-EPS) supplying calibrated probabilistic guidance for scheduling and reserves.

Figure 13: Solar PV Forecasting Model Accuracy Hierarchy



In this space, tree-based boosting applied as post-processing to NWP predictors consistently outperforms linear mappings and shallow ANNs, echoing but also quantifying the incremental value suggested in earlier work. Thus, the horizon-specific picture that emerges is sharper than a decade ago: image- and satellite-driven deep learning dominates minutes-to-hours; NWP plus boosting is the dependable stack at day-ahead; and blended satellite–NWP schemes bridge the gap in the intraday range (Rai & Sahu, 2020). A persistent theme in earlier reviews was that data quality, not algorithm choice, often bounded achievable skill (Kabeyi & Olanrewaju, 2023). Our findings strongly confirm that assessment and add evidence that the community has matured its measurement and data-engineering practices. Public, multi-site corpora such as NREL's PVDAQ—absent from early syntheses or only lightly used—now underpin many cross-site studies; they provide standardized metadata and long time series for performance/degradation modeling and for robust external validation, even as they still demand rigorous QA/QC for sensor drift, outages, and soiling episodes. In parallel, "best practices" handbooks produced through NREL and IEA PVPS Task 16 have concretized procedures for collection, harmonization, and use of solar-resource and plant-level data, addressing many of the comparability issues flagged by Yuan et al. (2021).The practical upshot in our corpus is that projects implementing IEC-aligned sensing, strict synchronization, and documented DQRs—together with streaming architectures that enable drift detection and periodic retraining—report more stable skill over seasons and site changes than projects that emphasize algorithmic novelty alone. Whereas earlier work sometimes conflated weather-driven variability with true model error, modern pipelines segregate confounders (e.g., site maintenance, instrument class) from forecast uncertainty, yielding cleaner attributions and fewer false alarms. This evolution does not contradict foundational reviews; rather, it operationalizes their caution by anchoring model claims in transparent data provenance and lifecycle governance, a shift we view as the main driver behind today's narrower spread in cross-paper accuracy reports (Haraz et al., 2025).

Earlier surveys acknowledged the scarcity of geographically disjoint validation and warned against over-interpreting single-site results. Our review demonstrates that the field has begun to address this gap through two complementary moves. First, Task 16's emphasis on *regional* forecasting reframes the problem from single-plant predictions to "virtual power plant" (fleet-level) outputs, with upscaling strategies, climate-aware benchmarking, and explicit recommendations for probabilistic products; this institutional push has catalyzed multi-site comparisons that better reflect operational realities (Oliveira et al., 2023). Second, transfer learning and domain adaptation methods—largely absent from the 2013–2016 reviews—now appear in cross-location nowcasting and day-ahead pipelines, where fine-tuning or feature-space alignment trims the performance penalty when moving models between tropical, arid, and temperate regimes. The net effect is a measurable narrowing of the gap between in-region and out-of-region skill, particularly for vision-based nowcasting where differences in cloud morphology previously undermined generalization (Zhang & Strbac, 2025). Still, our synthesis agrees with the foundational literature that climate-aware validation remains indispensable: pooled training without harmonized metadata or cadence alignment can degrade performance, and reporting should stratify results by climate class to avoid optimistic aggregates. Relative to the earlier canon, then, the novelty is not a wholesale change in "which model wins," but rather a maturing of evaluation design (region-out splits, probabilistic metrics) and the pragmatic use of lightweight adaptation in lieu of bespoke per-site retraining (Kalpana et al., 2024).

Jordan and Kurtz's analytical review codified the empirical range of PV degradation rates and provided the statistical scaffolding for lifetime expectations—a baseline that our synthesis repeatedly leverages when interpreting long-run performance trends. What differentiates the current landscape from that 2012–2013 vantage point is the diversity of diagnostic modalities and the learning capacity applied to them (Hashemi et al., 2025). Electroluminescence and infrared thermography—occasionally referenced in older reviews—are now central to fleet-scale fault detection and health assessment, with deep convolutional models trained on EL/IR imagery achieving high detection and classification scores for microcracks, hotspots, and metallization defects (Sharma et al., 2021). Parallel progress in using full I–V curves (rather than a handful of handcrafted features) enables multi-fault discrimination and fault-localization that outperform threshold-based techniques. Our findings therefore complement, not contradict, the earlier degradation narrative: the median annual loss rates summarized by Jordan & Kurtz still contextualize expected performance drift, but today's defect-level analytics explain *why* specific strings or modules deviate, and they do so with sufficient accuracy and latency to inform maintenance (Sharma et al., 2023). Moreover, where foundational reviews largely stopped at statistical life expectancy, contemporary PHM-oriented studies fuse physics-based stress models with sequence learners to estimate remaining useful life and to separate soiling or sensor artifacts from true aging. This shift—from aggregate trends to mechanistic, data-rich diagnostics—broadens the operational value of forecasting pipelines and aligns with our broader conclusion that multi-modal data plus task-appropriate ML, rather than more of the same tabular models, deliver the biggest incremental gains in O&M practice (Kong et al., 2021).

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

A decade ago, PV forecasting reviews mentioned interpretability mainly in passing; discussions focused on accuracy gains over persistence and on horizon-specific method choice. The past few years, however, reflect a broader AI movement: explainable AI (XAI) is now routinely paired with top-performing forecasters to meet auditability and control-room needs. Our synthesis shows that SHAP and LIME, while methodologically post hoc, are being integrated upstream in model selection and downstream in dashboarding, allowing operators to verify seasonal shifts in driver importance (e.g., humidity, cloud cover, wind) and to detect sensor issues promptly (Corrochano et al., 2025). This evolution maps directly onto the high-level guidance from the XAI literature, which stresses stability/faithfulness trade-offs and the benefits of combining explanations with domain constraints. In comparative terms, then, we observe a shift from the earlier "black box versus glass box" dichotomy to a pragmatic equilibrium: tree-based ensembles plus SHAP deliver a good balance for tabular NWP/SCADA features; deep nets remain warranted for imagery/sequences but are increasingly accompanied by saliency or attention maps, or "physics-guided" features that improve plausibility . The upshot is not that interpretability *replaces* accuracy as the dominant criterion, but that it has become a necessary gate for deployment, particularly as forecasts drive storage dispatch, reserve setting, and curtailment decisions. In this respect, our findings extend the earlier reviews by showing that trust-building instruments have matured from desiderata to standard practice, consistent with XAI surveys that advocate principled, domain-aware explanations for high-stakes applications (Singh et al., 2024).

In addition, the pathway from forecast skill to operational value figures more prominently in our synthesis than in much of the earlier canon (Sarpong et al., 2020). Foundational reviews cataloged methods and horizons, but said less about how forecasts propagate through scheduling, reserves, and PV–storage control. Contemporary literature—mirrored in Task 16 guidance and in the maturation of WRF-Solar and its ensemble counterpart—elevates probabilistic products (prediction intervals, quantiles) as first-class inputs to market participation and model-predictive control. Our results echo that shift: when probabilistic day-ahead forecasts are fed into PV-battery or fleet-level dispatch, cost and reliability metrics improve relative to point-forecast workflows, a finding that was largely anecdotal a decade ago but is now supported by calibrated ensemble methods and reproducible case studies (Leamon et al., 2021). In addition, operational deployments increasingly depend on cloud–edge splits for latency and resilience, with edge inference of compact models (for nowcasting and anomaly detection) and cloud-based retraining/monitoring—architectural patterns not yet prominent in early reviews. Governance has also grown in salience: cross-institutional best-practice documents emphasize documentation, monitoring, and cybersecurity for forecast pipelines, reflecting both regulatory pressures and lived experience from utility integration (Haraz et al., 2025). In comparing eras, then, the main difference is not a wholesale substitution of algorithms but a clear integration of forecasting with decision-making, uncertainty management, and lifecycle controls. This integration explains why recent studies report durable KPI improvements even when headline RMSE gains are modest: value accrues when predictions are *actionable*, *probabilistic*, and *governed*—a maturation stage anticipated but not yet realized in the earliest surveys (Liu & Du, 2023).

CONCLUSION

In sum, this review shows that meaningful gains in solar-PV prediction arise less from novelty for its own sake and more from disciplined pipeline design that aligns models, data, and operational use. Synthesizing the screened literature, a clear accuracy hierarchy emerged: tuned boosting ensembles remain the most dependable and computationally economical choice for tabular day-ahead tasks built on numerical weather prediction and plant telemetry; deep neural architectures excel when inputs are image-rich or sequence-heavy (e.g., sky cameras, satellite tiles, high-frequency SCADA), delivering superior ramp capture on minute-to-hour horizons; and physics–ML hybrids add robustness and physical plausibility, particularly under regime shifts or sparse data. These performance patterns persisted across studies once evaluation was made climate-aware and leakage-resistant, reinforcing that cross-regional validation—and, where feasible, transfer learning or domain adaptation—determines whether a model travels beyond the site where it was conceived. Equally decisive were data and systems choices: standards-aligned sensing, rigorous QA/QC with explicit treatment of soiling and outages, synchronized time bases, and streaming feature engineering routinely yielded error reductions competitive with algorithmic upgrades. Edge–cloud splits proved operationally valuable, enabling sub-minute inference at the plant while reserving fleet-

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

scale training, monitoring, and drift management for the cloud, and explainability layers (e.g., attribution and saliency tools) helped translate black-box predictions into defensible decisions for operators and regulators. At the same time, persistent limitations temper over-generalization: heterogeneous metrics and splits still hinder meta-comparison; tropical and arid regimes remain underrepresented; curated defect datasets can inflate diagnostic scores; and reproducibility varies with data access and documentation quality. Taken together, the evidence supports a practical blueprint for researchers and practitioners: select models by horizon and modality rather than fashion, enforce climate-aware validation with region-out tests, pair deterministic predictions with calibrated probabilistic products for dispatch and reserves, integrate explanation and physics-guidance where stakes are high, and invest in end-to-end MLOps so models remain accurate as conditions evolve. Future progress will hinge on widening geographic coverage; strengthening open, well-documented benchmarks; advancing transfer and physics-informed learning that reduces data hunger without sacrificing interpretability; and embedding governance, cybersecurity, and ethical safeguards so that forecast improvements scale into durable grid and asset value.

RCOMMENDATIONS

Here are the recommendations rewritten in paragraph form. Begin by treating data quality as the performance ceiling. Prioritize Class-A monitoring practices with calibrated irradiance and temperature sensors, ventilated/heated pyranometers where appropriate, synchronized timestamps, and meticulous maintenance logs. Institute a plant-level data-quality and readiness playbook that automatically flags missingness, time drift, clipping, sensor bias, and curtailment, and report these diagnostics alongside model metrics. Model soiling explicitly—estimate site-specific soiling rates and recovery events—and harmonize feature schemas (names, units, cadences, time zones) so cross-site learning is feasible without brittle one-off wrangling. When possible, fuse modalities (SCADA, on-site weather, sky cameras, satellite fields, NWP) and enforce precise time alignment to reduce ramp errors; preserve full provenance (sensor class, firmware, maintenance) in a machine-readable "datasheet" for every dataset used in training or benchmarking. For modeling, match the learner to the horizon and modality. Use tuned gradient-boosting/forest baselines for tabular day-ahead tasks mapped from NWP and plant telemetry; prefer CNN/LSTM (or hybrids) when inputs are image-rich or long-sequence nowcasts; adopt physics-hybrid or physics-informed designs where extrapolation and physical plausibility matter. Make region-out (or site-out) validation the default for any generalization claims and report both in-region and out-of-region results to expose transferability gaps. Always benchmark against smart persistence and simple physical baselines, and report RMSE/MAE/MAPE together with skill scores relative to those baselines. Treat computational cost as a first-class metric: include training wall time, inference latency, and memory/compute footprint so accuracy is evaluated alongside deployability. Where decisions depend on risk, produce probabilistic outputs (prediction intervals or quantiles) and evaluate calibration, sharpness, and CRPS—not just point errors.

Embed forecasts in a resilient, real-time architecture. Use an edge–cloud split: run compact, compiled models at the edge for sub-minute nowcasting and anomaly screening, while reserving the cloud for fleet-scale training, hyperparameter search, and monitoring. Back the pipeline with message brokers and time-series stores to support rolling features, and implement automated drift detection with scheduled retraining or fine-tuning triggers. Maintain an MLOps spine—model registry, versioning, lineage, reproducible training artifacts, and latency/error SLAs—so updates are auditable and rollbacks are painless. Document every assumption (feature engineering, filters, data windows) and keep unit tests for data transformations to prevent silent regressions. Ensure forecasts are explainable and governed. Pair top-performing models with operator-facing explanations (e.g., SHAP/LIME for tabular models, saliency or attention maps for vision/sequence models), and log explanations alongside predictions for post-event audits. Combine explanation layers with physics-guided features or constraints to improve plausibility and reduce spurious correlations. Establish clear roles and change-management procedures for model updates; record model cards and risk assessments; and enforce privacy/cybersecurity controls for cloud–edge deployments, including least-privilege access, encryption at rest/in transit, and incident response playbooks. Align model documentation with emerging regulatory expectations by keeping transparent records of data sources, validation splits, and known limitations. Finally, integrate prediction with control and a clear

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

research agenda. Feed probabilistic day-ahead forecasts into storage and demand-flexibility optimization (e.g., MPC or stochastic programming) with explicit cost and reliability objectives; track operational KPIs—curtailment, reserve alignment, state-of-charge violations, and O&M truck rolls—to verify value beyond RMSE. For future work, expand open, well-documented benchmarks across tropical, arid, and maritime climates; require region-out splits in public leaderboards; release reference implementations for transfer learning/domain adaptation and physics-informed training; and standardize ablations that isolate the impact of data-quality interventions (soiling correction, synchronization, QA/QC) from algorithm choice. This combination—measurement discipline, horizon-appropriate models, rigorous validation, production-grade MLOps, actionable uncertainty, operator-ready explanations, and control integration—is the most reliable path from academic accuracy to durable grid and asset value.

REFERENCES

[1]. Abaimov, S., & Martellini, M. (2022). Machine learning for cyber agents. *Attack andDefence. Cham: Springer*.

[2]. Abaimov, S., & Martellini, M. (2022). Understanding machine learning. In *Machine Learning for Cyber Agents: Attack and Defence* (pp. 15-89). Springer.

[3]. Abdur Razzak, C., Golam Qibria, L., & Md Arifur, R. (2024). Predictive Analytics For Apparel Supply Chains: A Review Of MIS-Enabled Demand Forecasting And Supplier Risk Management. *American Journal of Interdisciplinary Studies*, *5*(04), 01–23. https://doi.org/10.63125/80dwy222

[4]. Adar, C., & Md, N. (2023). Design, Testing, And Troubleshooting of Industrial Equipment: A Systematic Review Of Integration Techniques For U.S. Manufacturing Plants. *Review of Applied Science and Technology*, *2*(01), 53-84. https://doi.org/10.63125/893et038

[5]. Al-Dahidi, S., Madhiarasan, M., Al-Ghussain, L., Abubaker, A. M., Ahmad, A. D., Alrbai, M., Aghaei, M., Alahmer, H., Alahmer, A., & Baraldi, P. (2024). Forecasting solar photovoltaic power production: A comprehensive review and innovative data-driven modeling framework. *Energies*, *17*(16), 4145.

[6]. Albreem, M. A., Sheikh, A. M., Bashir, M. J., & El-Saleh, A. A. (2023). Towards green Internet of Things (IoT) for a sustainable future in Gulf Cooperation Council countries: Current practices, challenges and future prospective. *Wireless Networks*, *29*(2), 539-567.

[7]. Alcañiz, A., Lindfors, A. V., Zeman, M., Ziar, H., & Isabella, O. (2023). Effect of climate on photovoltaic yield prediction using machine learning models. *Global Challenges*, *7*(1), 2200166.

[8]. Anand, P., & Sundaram, K. M. (2020). FPGA based substantial power evolution controlling strategy for solar and wind forecasting grid connected system. *Microprocessors and Microsystems*, *74*, 103001.

[9]. Anderson, K. S., Hobbs, W. B., Holmgren, W. F., Perry, K. R., Mikofski, M. A., & Kharait, R. A. (2022). The effect of inverter loading ratio on energy estimate bias. 2022 IEEE 49th Photovoltaics Specialists Conference (PVSC),

[10]. Auger, L., Dupont, O., Hagelin, S., Brousseau, P., & Brovelli, P. (2015). AROME–NWC: A new nowcasting tool based on an operational mesoscale forecasting system. *Quarterly Journal of the Royal Meteorological Society*, *141*(690), 1603-1611.

[11]. Bak, S., Choi, S., Yang, D., Kim, D., Rho, H., & Lee, K. (2025). Transfer Learning for Photovoltaic Power Forecasting Across Regions Using Large-Scale Datasets. *IEEE access*.

[12]. Benninger, M., Hofmann, M., & Liebschner, M. (2019). Online monitoring system for photovoltaic systems using anomaly detection with machine learning. NEIS 2019; Conference on Sustainable Energy Supply and Energy Storage Systems,

[13]. Bodapati, N., Himavaishnavi, J., Rohitha, V., Jagadeeswari, D. L., & Bhavana, P. (2022). Analyzing crop yield using machine learning. 2022 International Conference on Electronics and Renewable Systems (ICEARS),

[14]. Bojinski, S., Blaauboer, D., Calbet, X., De Coning, E., Debie, F., Montmerle, T., Nietosvaara, V., Norman, K., Bañón Peregrín, L., & Schmid, F. (2023). Towards nowcasting in Europe in 2030. *Meteorological applications*, *30*(4), e2124.

[15]. Böök, H., & Lindfors, A. V. (2020). Site-specific adjustment of a NWP-based photovoltaic production forecast. *Solar Energy*, *211*, 779-788.

[16]. Bouche, D., Flamary, R., d'Alché-Buc, F., Plougonven, R., Clausel, M., Badosa, J., & Drobinski, P. (2023). Wind power predictions from nowcasts to 4-hour forecasts: a learning approach with variable selection. *Renewable Energy*, *211*, 938-947.

[17]. Bracco, A., Brajard, J., Dijkstra, H. A., Hassanzadeh, P., Lessig, C., & Monteleoni, C. (2025). Machine learning for the physics of climate. *Nature Reviews Physics*, *7*(1), 6-20.

[18]. Browning, K., & Collier, C. (1989). Nowcasting of precipitation systems. *Reviews of Geophysics*, *27*(3), 345-370.

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

[19]. Bruneau, P., Fiorelli, D., Braun, C., & Koster, D. (2024). Forecasting intraday power output by a set of PV systems using recurrent neural networks and physical covariates. *Neural Computing and Applications*, *36*(31), 19515-19529.

[20]. Buonanno, A., Caputo, G., Balog, I., Fabozzi, S., Adinolfi, G., Pascarella, F., Leanza, G., Graditi, G., & Valenti, M. (2024). Machine learning and weather model combination for PV production forecasting. *Energies*, *17*(9), 2203.

[21]. Chowdhury, M. M.-U.-T., Hasan, M. S., & Kamalasadan, S. (2024). A novel voltage optimization co-simulation framework for electrical distribution system with high penetration of renewables. *IEEE Transactions on Industry Applications*.

[22]. Coelho, L. B., Zhang, D., Van Ingelgem, Y., Steckelmacher, D., Nowé, A., & Terryn, H. (2022). Reviewing machine learning of corrosion prediction in a data-oriented perspective. *npj Materials Degradation*, *6*(1), 8.

[23]. Collino, E., & Ronzio, D. (2021). Exploitation of a new short-term multimodel photovoltaic power forecasting method in the very short-term horizon to derive a multi-time scale forecasting system. *Energies*, *14*(3), 789.

[24]. Corrochano, A., Freitas, R. S., López-Martín, M., Parente, A., & Le Clainche, S. (2025). A Predictive Physics-Aware Hybrid Reduced Order Model for Reacting Flows. *International Journal for Numerical Methods in Engineering*, *126*(11), e70058.

[25]. Couto, A., & Estanqueiro, A. (2022). Enhancing wind power forecast accuracy using the weather research and forecasting numerical model-based features and artificial neuronal networks. *Renewable Energy*, *201*, 1076-1085.

[26]. Coya, Z., Khoodaruth, A., Ramenah, H., Oree, V., Murdan, A. P., & Bessafi, M. (2024). Deep learning models in photovoltaic power forecasting: a review. 2024 1st International Conference on Smart Energy Systems and Artificial Intelligence (SESAI),

[27]. Cuperlovic-Culf, M. (2018). Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites*, *8*(1), 4.

[28]. Davò, F., Alessandrini, S., Sperati, S., Delle Monache, L., Airoldi, D., & Vespucci, M. T. (2016). Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting. *Solar Energy*, *134*, 327-338.

[29]. de Oliveira, J. F. L., de Mattos Neto, P. S., Siqueira, H. V., Santos Jr, D. S. d. O., Lima, A. R., Madeiro, F., Dantas, D. A., Cavalcanti, M. d. M., Pereira, A. C., & Marinho, M. H. (2023). Forecasting methods for photovoltaic energy in the scenario of battery energy storage systems: a comprehensive review. *Energies*, *16*(18), 6638.

[30]. Dhingra, S., Gruosso, G., & Gajani, G. S. (2023a). Modelling ageing and power production of solar PV using machine learning techniques. 2023 International Conference on Electrical, Computer and Energy Technologies (ICECET),

[31]. Dhingra, S., Gruosso, G., & Gajani, G. S. (2023b). Solar pv power forecasting and ageing evaluation using machine learning techniques. IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society,

[32]. Dhingra, S., Gruosso, G., & Gajani, G. S. (2024). Probabilistic Forecasting of PV Power Using Artificial Neural Networks with Confidence Intervals. IECON 2024-50th Annual Conference of the IEEE Industrial Electronics Society,

[33]. Di Leo, P., Ciocia, A., Malgaroli, G., & Spertino, F. (2025). Advancements and Challenges in Photovoltaic Power Forecasting: A Comprehensive Review. *Energies*, *18*(8), 2108.

[34]. Flath, C. M., & Stein, N. (2022). Smart grid analytics. In *Smart grid economics and management* (pp. 173-192). Springer.

[35]. Gbémou, S., Eynard, J., Thil, S., Guillot, E., & Grieu, S. (2021). A comparative study of machine learning-based methods for global horizontal irradiance forecasting. *Energies*, *14*(11), 3192.

[36]. Golam Qibria, L., & Takbir Hossen, S. (2023). Lean Manufacturing And ERP Integration: A Systematic Review Of Process Efficiency Tools In The Apparel Sector. *American Journal of Scholarly Research and Innovation*, *2*(01), 104-129. https://doi.org/10.63125/mx7j4p06

[37]. Hao, J., Oni, E., Kim, I. K., & Ramaswamy, L. (2023). DynaES: Dynamic Energy Scheduling for Energy Harvesting Environmental Sensors. 2023 IEEE International Performance, Computing, and Communications Conference (IPCCC),

[38]. Haraz, A., Abualsaud, K., & Massoud, A. (2025). Hybrid Tree-based Machine Learning Models for State-of-Charge and Core Temperature Estimation in EV Batteries. *IEEE access*.

[39]. Hashemi, A., Beheshti, J., & Mohammadi, M. (2025). Physics-Based, AI-Driven Surrogate Modeling for Structural Displacement Prediction in Mechanical Systems with Limited Sensor Data. *IEEE access*.

[40]. Hinduja, S., Darzi, A., Ertugrul, I. O., Provenza, N., Gadot, R., Storch, E. A., Sheth, S. A., Goodman, W. K., & Cohn, J. F. (2024). Multimodal Prediction of Obsessive-Compulsive Disorder and Comorbid Depression

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

Severity and Energy Delivered by Deep Brain Electrodes. *IEEE transactions on affective computing*, *15*(4), 2025-2041.

[41]. Hosne Ara, M., Tonmoy, B., Mohammad, M., & Md Mostafizur, R. (2022). AI-ready data engineering pipelines: a review of medallion architecture and cloud-based integration models. *American Journal of Scholarly Research and Innovation*, *1*(01), 319-350. https://doi.org/10.63125/51kxtf08

[42]. Hurwitz, J., Kaufman, M., Bowles, A., Nugent, A., Kobielus, J. G., & Kowolenko, M. D. (2015). *Cognitive computing and big data analytics* (Vol. 288). Wiley Online Library.

[43]. Huynh, A. N.-L., Deo, R. C., An-Vo, D.-A., Ali, M., Raj, N., & Abdulla, S. (2020). Near real-time global solar radiation forecasting at multiple time-step horizons using the long short-term memory network. *Energies*, *13*(14), 3517.

[44]. Iheanetu, K. J. (2022). Solar photovoltaic power forecasting: A review. *Sustainability*, *14*(24), 17005.

[45]. Imhoff, R., Overeem, A., Brauer, C., Leijnse, H., Weerts, A., & Uijlenhoet, R. (2020). Rainfall nowcasting using commercial microwave links. *Geophysical Research Letters*, *47*(19), e2020GL089365.

[46]. Islam, M. S. S., Ghosh, P., Faruque, M. O., Islam, M. R., Hossain, M. A., Alam, M. S., & Islam Sheikh, M. R. (2024). Optimizing short-term photovoltaic power forecasting: A novel approach with gaussian process regression and bayesian hyperparameter tuning. *Processes*, *12*(3), 546.

[47]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2023). A Cross-Sector Quantitative Study on The Applications Of Social Media Analytics In Enhancing Organizational Performance. *American Journal of Scholarly Research and Innovation*, *2*(02), 274-302. https://doi.org/10.63125/d8ree044

[48]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2024). Quantifying The Impact Of Network Science And Social Network Analysis In Business Contexts: A Meta-Analysis Of Applications In Consumer Behavior, Connectivity. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 58-89. https://doi.org/10.63125/vgkwe938

[49]. Jahid, M. K. A. S. R. (2022). Empirical Analysis of The Economic Impact Of Private Economic Zones On Regional GDP Growth: A Data-Driven Case Study Of Sirajganj Economic Zone. *American Journal of Scholarly Research and Innovation*, *1*(02), 01-29. https://doi.org/10.63125/je9w1c40

[50]. Jannah, N., Gunawan, T. S., Yusoff, S. H., Hanifah, M. S. A., & Sapihie, S. N. M. (2024). Recent Advances and Future Challenges of Solar Power Generation Forecasting. *IEEE access*.

[51]. Jing, S., Xi, X., Su, D., Han, Z., & Wang, D. (2024). Spatio-Temporal Photovoltaic Power Prediction with Fourier Graph Neural Network. *Electronics*, *13*(24), 4988.

[52]. Kabeyi, M. J. B., & Olanrewaju, O. A. (2023). Smart grid technologies and application in the sustainable energy transition: a review. *International Journal of Sustainable Energy*, *42*(1), 685-758.

[53]. Kalpana, V., Sujana, G. J., Thyagarajan, K., Lalitha, R., Talasila, V., & Jadhav, M. M. (2024). Enhancing heat transfer coefficient predictions in complex geometries through hybrid machine learning approaches. *Thermal Science and Engineering Progress*, *55*, 103017.

[54]. Kashef, M. (2025). Rethinking Masdar and The Line Megaprojects: The Interplay of Economic, Social, Political, and Spatial Dimensions. *Land*, *14*(7), 1358.

[55]. Kasthurirathne, S. N., Ho, Y. A., & Dixon, B. E. (2020). Public health analytics and big data. In *Public health informatics and information systems* (pp. 203-219). Springer.

[56]. Kong, L., Gong, J., Hu, Q., Capitani, F., Celeste, A., Hattori, T., Sano-Furukawa, A., Li, N., Yang, W., & Liu, G. (2021). Suppressed lattice disorder for large emission enhancement and structural robustness in hybrid lead iodide perovskite discovered by high-pressure isotope effect. *Advanced Functional Materials*, *31*(9), 2009131.

[57]. Lauret, P., Alonso-Suárez, R., Le Gal La Salle, J., & David, M. (2022). Solar forecasts based on the clear sky index or the clearness index: Which is better? Solar,

[58]. Le Jeune, L., Goedeme, T., & Mentens, N. (2021). Machine learning for misuse-based network intrusion detection: overview, unified evaluation and feature choice comparison framework. *IEEE access*, *9*, 63995-64015.

[59]. Leamon, R. J., McIntosh, S. W., & Marsh, D. R. (2021). Termination of solar cycles and correlated tropospheric variability. *Earth and Space Science*, *8*(4), e2020EA001223.

[60]. Li, Y., Song, L., Zhang, S., Kraus, L., Adcox, T., Willardson, R., Komandur, A., & Lu, N. (2023). A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting. *IEEE Transactions on Smart Grid*, *14*(5), 4073-4085.

[61]. Liem, C. C., Langer, M., Demetriou, A., Hiemstra, A. M., Sukma Wicaksana, A., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In *Explainable and interpretable models in computer vision and machine learning* (pp. 197-253). Springer.

[62]. Liu, Z., & Du, Y. (2023). Evolution towards dispatchable PV using forecasting, storage, and curtailment: A review. *Electric Power Systems Research*, *223*, 109554.

[63]. Madsen, D. N., & Hansen, J. P. (2019). Outlook of solar energy in Europe based on economic growth characteristics. *Renewable and Sustainable Energy Reviews*, *114*, 109306.

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

[64]. Mak, K.-K., Wong, Y.-H., & Pichika, M. R. (2024). Artificial intelligence in drug discovery and development. *Drug discovery and evaluation: safety and pharmacokinetic assays*, 1461-1498.

[65]. Mansura Akter, E. (2023). Applications Of Allele-Specific PCR In Early Detection of Hereditary Disorders: A Systematic Review Of Techniques And Outcomes. *Review of Applied Science and Technology*, *2*(03), 1-26. https://doi.org/10.63125/n4h7t156

[66]. Mansura Akter, E., & Md Abdul Ahad, M. (2022). In Silico drug repurposing for inflammatory diseases: a systematic review of molecular docking and virtual screening studies. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 35-64. https://doi.org/10.63125/j1hbts51

[67]. Mansura Akter, E., & Shaiful, M. (2024). A systematic review of SNP polymorphism studies in South Asian populations: implications for diabetes and autoimmune disorders. *American Journal of Scholarly Research and Innovation*, *3*(01), 20-51. https://doi.org/10.63125/8nvxcb96

[68]. Mayer, M. J., Yang, D., & Szintai, B. (2023). Comparing global and regional downscaled NWP models for irradiance and photovoltaic power forecasting: ECMWF versus AROME. *Applied Energy*, *352*, 121958.

[69]. Md Arifur, R., & Sheratun Noor, J. (2022). A Systematic Literature Review of User-Centric Design In Digital Business Systems: Enhancing Accessibility, Adoption, And Organizational Impact. *Review of Applied Science and Technology*, *1*(04), 01-25. https://doi.org/10.63125/ndjkpm77

[70]. Md Ashiqur, R., Md Hasan, Z., & Afrin Binta, H. (2025). A meta-analysis of ERP and CRM integration tools in business process optimization. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 278-312. https://doi.org/10.63125/yah70173

[71]. Md Hasan, Z. (2025). AI-Driven business analytics for financial forecasting: a systematic review of decision support models in SMES. *Review of Applied Science and Technology*, *4*(02), 86-117. https://doi.org/10.63125/gjrpv442

[72]. Md Hasan, Z., Mohammad, M., & Md Nur Hasan, M. (2024). Business Intelligence Systems In Finance And Accounting: A Review Of Real-Time Dashboarding Using Power BI & Tableau. *American Journal of Scholarly Research and Innovation*, *3*(02), 52-79. https://doi.org/10.63125/fy4w7w04

[73]. Md Hasan, Z., & Moin Uddin, M. (2022). Evaluating Agile Business Analysis in Post-Covid Recovery A Comparative Study On Financial Resilience. *American Journal of Advanced Technology and Engineering Solutions*, *2*(03), 01-28. https://doi.org/10.63125/6nee1m28

[74]. Md Hasan, Z., Sheratun Noor, J., & Md. Zafor, I. (2023). Strategic role of business analysts in digital transformation tools, roles, and enterprise outcomes. *American Journal of Scholarly Research and Innovation*, *2*(02), 246-273. https://doi.org/10.63125/rc45z918

[75]. Md Mahamudur Rahaman, S. (2022). Electrical And Mechanical Troubleshooting in Medical And Diagnostic Device Manufacturing: A Systematic Review Of Industry Safety And Performance Protocols. *American Journal of Scholarly Research and Innovation*, *1*(01), 295-318. https://doi.org/10.63125/d68y3590

[76]. Md Mahamudur Rahaman, S., & Rezwanul Ashraf, R. (2022). Integration of PLC And Smart Diagnostics in Predictive Maintenance of CT Tube Manufacturing Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 62-96. https://doi.org/10.63125/gspb0f75

[77]. Md Masud, K., Mohammad, M., & Sazzad, I. (2023). Mathematics For Finance: A Review of Quantitative Methods In Loan Portfolio Optimization. *International Journal of Scientific Interdisciplinary Research*, *4*(3), 01-29. https://doi.org/10.63125/j43ayz68

[78]. Md Nazrul Islam, K. (2022). A Systematic Review of Legal Technology Adoption In Contract Management, Data Governance, And Compliance Monitoring. *American Journal of Interdisciplinary Studies*, *3*(01), 01-30. https://doi.org/10.63125/caangg06

[79]. Md Nur Hasan, M., Md Musfiqur, R., & Debashish, G. (2022). Strategic Decision-Making in Digital Retail Supply Chains: Harnessing AI-Driven Business Intelligence From Customer Data. *Review of Applied Science and Technology*, *1*(03), 01-31. https://doi.org/10.63125/6a7rpy62

[80]. Md Redwanul, I., & Md. Zafor, I. (2022). Impact of Predictive Data Modeling on Business Decision-Making: A Review Of Studies Across Retail, Finance, And Logistics. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 33-62. https://doi.org/10.63125/8hfbkt70

[81]. Md Rezaul, K., & Md Mesbaul, H. (2022). Innovative Textile Recycling and Upcycling Technologies For Circular Fashion: Reducing Landfill Waste And Enhancing Environmental Sustainability. *American Journal of Interdisciplinary Studies*, *3*(03), 01-35. https://doi.org/10.63125/kkmerg16

[82]. Md Sultan, M., Proches Nolasco, M., & Md. Torikul, I. (2023). Multi-Material Additive Manufacturing For Integrated Electromechanical Systems. *American Journal of Interdisciplinary Studies*, *4*(04), 52-79. https://doi.org/10.63125/y2ybrx17

[83]. Md Sultan, M., Proches Nolasco, M., & Vicent Opiyo, N. (2025). A Comprehensive Analysis Of Non-Planar Toolpath Optimization In Multi-Axis 3D Printing: Evaluating The Efficiency Of Curved Layer Slicing Strategies. *Review of Applied Science and Technology*, *4*(02), 274-308. https://doi.org/10.63125/5fdxa722

[84]. Md Takbir Hossen, S., Ishtiaque, A., & Md Atiqur, R. (2023). AI-Based Smart Textile Wearables For Remote Health Surveillance And Critical Emergency Alerts: A Systematic Literature Review. *American Journal of Scholarly Research and Innovation*, *2*(02), 1-29. https://doi.org/10.63125/ceqapd08
[85]. Md Takbir Hossen, S., & Md Atiqur, R. (2022). Advancements In 3d Printing Techniques For Polymer Fiber-Reinforced Textile Composites: A Systematic Literature Review. *American Journal of Interdisciplinary Studies*, *3*(04), 32-60. https://doi.org/10.63125/s4r5m391
[86]. Md Tawfiqul, I. (2023). A Quantitative Assessment Of Secure Neural Network Architectures For Fault Detection In Industrial Control Systems. *Review of Applied Science and Technology*, *2*(04), 01-24. https://doi.org/10.63125/3m7gbs97
[87]. Md Tawfiqul, I., Meherun, N., Mahin, K., & Mahmudur Rahman, M. (2022). Systematic Review of Cybersecurity Threats In IOT Devices Focusing On Risk Vectors Vulnerabilities And Mitigation Strategies. *American Journal of Scholarly Research and Innovation*, *1*(01), 108-136. https://doi.org/10.63125/wh17mf19
[88]. Md Tawfiqul, I., Sabbir, A., Md Anikur, R., & Md Arifur, R. (2024). Neural Network–Based Risk Prediction And Simulation Framework For Medical IOT Cybersecurity: An Engineering Management Model For Smart Hospitals. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 30-57. https://doi.org/10.63125/g0mvct35
[89]. Md. Sakib Hasan, H. (2022). Quantitative Risk Assessment of Rail Infrastructure Projects Using Monte Carlo Simulation And Fuzzy Logic. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 55-87. https://doi.org/10.63125/h24n6z92
[90]. Md. Tarek, H. (2022). Graph Neural Network Models For Detecting Fraudulent Insurance Claims In Healthcare Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 88-109. https://doi.org/10.63125/r5vsmv21
[91]. Md.Kamrul, K., & Md Omar, F. (2022). Machine Learning-Enhanced Statistical Inference For Cyberattack Detection On Network Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 65-90. https://doi.org/10.63125/sw7jzx60
[92]. Md.Kamrul, K., & Md. Tarek, H. (2022). A Poisson Regression Approach to Modeling Traffic Accident Frequency in Urban Areas. *American Journal of Interdisciplinary Studies*, *3*(04), 117-156. https://doi.org/10.63125/wqh7pd07
[93]. Mst Shamima, A., Niger, S., Md Atiqur Rahman, K., & Mohammad, M. (2023). Business Intelligence-Driven Healthcare: Integrating Big Data And Machine Learning For Strategic Cost Reduction And Quality Care Delivery. *American Journal of Interdisciplinary Studies*, *4*(02), 01-28. https://doi.org/10.63125/crv1xp27
[94]. Mubashir, I., & Abdul, R. (2022). Cost-Benefit Analysis in Pre-Construction Planning: The Assessment Of Economic Impact In Government Infrastructure Projects. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 91-122. https://doi.org/10.63125/kjwd5e33
[95]. Müller, R., Barleben, A., Haussler, S., & Jerg, M. (2022). A novel approach for the global detection and nowcasting of deep convection and thunderstorms. *Remote Sensing*, *14*(14), 3372.
[96]. Munkhammar, J., van der Meer, D., & Widén, J. (2019). Probabilistic forecasting of high-resolution clear-sky index time-series using a Markov-chain mixture distribution model. *Solar Energy*, *184*, 688-695.
[97]. Mystakidis, A., Koukaras, P., Tsalikidis, N., Ioannidis, D., & Tjortjis, C. (2024). Energy forecasting: A comprehensive review of techniques and technologies. *Energies*, *17*(7), 1662.
[98]. Nelson, J. R., & Grubesic, T. H. (2020). The use of LiDAR versus unmanned aerial systems (UAS) to assess rooftop solar energy potential. *Sustainable Cities and Society*, *61*, 102353.
[99]. Nikulins, A., Sudars, K., Edelmers, E., Namatevs, I., Ozols, K., Komasilovs, V., Zacepins, A., Kviesis, A., & Reinhardt, A. (2024). Deep learning for wind and solar energy forecasting in hydrogen production. *Energies*, *17*(5), 1053.
[100]. Omar Muhammad, F., & Md.Kamrul, K. (2022). Blockchain-Enabled BI For HR And Payroll Systems: Securing Sensitive Workforce Data. *American Journal of Scholarly Research and Innovation*, *1*(02), 30-58. https://doi.org/10.63125/et4bhy15
[101]. Panesar, A. (2019). *Machine learning and AI for healthcare* (Vol. 10). Springer.
[102]. Paoli, C., Voyant, C., Muselli, M., & Nivet, M.-L. (2010). Forecasting of preprocessed daily solar radiation time series using neural networks. *Solar Energy*, *84*(12), 2146-2160.
[103]. Paulescu, M., Paulescu, E., & Badescu, V. (2021). Nowcasting solar irradiance for effective solar power plants operation and smart grid management. In *Predictive modelling for energy management and power systems engineering* (pp. 249-270). Elsevier.
[104]. Perry, K., Nguyen, Q., & White, R. (2024). Quantifying Error in Photovoltaic Installation Metadata. 2024 IEEE 52nd Photovoltaic Specialist Conference (PVSC),
[105]. Rai, R., & Sahu, C. K. (2020). Driven by data or derived through physics? a review of hybrid physics guided machine learning techniques with cyber-physical system (cps) focus. *IEEE access*, *8*, 71050-71073.

[106]. Rajarajeshwari, G., & Selvi, G. C. (2024). Application of artificial intelligence for classification, segmentation, early detection, early diagnosis, and grading of diabetic retinopathy from fundus retinal images: a comprehensive review. *IEEE access*.
[107]. Rajesh, P., Md Arifur, R., & Md, N. (2024). AI-Enabled Decision Support Systems for Smarter Infrastructure Project Management In Public Works. *Review of Applied Science and Technology*, *3*(04), 29-47. https://doi.org/10.63125/8d96m319
[108]. Ranalli, J., & Hobbs, W. B. (2025). Plant Clipping Bias Due to Spatial Distribution of Inverters and AC Overbuild: Preliminary Investigation. 2025 IEEE 53rd Photovoltaic Specialists Conference (PVSC),
[109]. Reduanul, H., & Mohammad Shoeb, A. (2022). Advancing AI in Marketing Through Cross Border Integration Ethical Considerations And Policy Implications. *American Journal of Scholarly Research and Innovation*, *1*(01), 351-379. https://doi.org/10.63125/d1xg3784
[110]. Roberts, A. J., Fletcher, J. K., Groves, J., Marsham, J. H., Parker, D. J., Blyth, A. M., Adefisan, E. A., Ajayi, V. O., Barrette, R., & De Coning, E. (2022). Nowcasting for Africa: advances, potential and value. *Weather*, *77*(7), 250-256.
[111]. Roth, C., Noack, J., & Skyllas-Kazacos, M. (2022). *Flow Batteries: From Fundamentals to Applications*. John Wiley & Sons.
[112]. Sabuj Kumar, S., & Zobayer, E. (2022). Comparative Analysis of Petroleum Infrastructure Projects In South Asia And The Us Using Advanced Gas Turbine Engine Technologies For Cross Integration. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 123-147. https://doi.org/10.63125/wr93s247
[113]. Sadia, T., & Shaiful, M. (2022). In Silico Evaluation of Phytochemicals From Mangifera Indica Against Type 2 Diabetes Targets: A Molecular Docking And Admet Study. *American Journal of Interdisciplinary Studies*, *3*(04), 91-116. https://doi.org/10.63125/anaf6b94
[114]. Saigustia, C., & Pijarski, P. (2023). Time series analysis and forecasting of solar generation in Spain using eXtreme gradient boosting: a machine learning approach. *Energies*, *16*(22), 7618.
[115]. Samsi, S., Mattioli, C. J., & Veillette, M. S. (2019). Distributed deep learning for precipitation nowcasting. 2019 IEEE High Performance Extreme Computing Conference (HPEC),
[116]. Sanjai, V., Sanath Kumar, C., Maniruzzaman, B., & Farhana Zaman, R. (2023). Integrating Artificial Intelligence in Strategic Business Decision-Making: A Systematic Review Of Predictive Models. *International Journal of Scientific Interdisciplinary Research*, *4*(1), 01-26. https://doi.org/10.63125/s5skge53
[117]. Sanjai, V., Sanath Kumar, C., Sadia, Z., & Rony, S. (2025). AI And Quantum Computing For Carbon-Neutral Supply Chains: A Systematic Review Of Innovations. *American Journal of Interdisciplinary Studies*, *6*(1), 40-75. https://doi.org/10.63125/nrdx7d32
[118]. Sapountzi, A., & Psannis, K. E. (2020). Big data preprocessing: an application on online social networks. In *Principles of Data Science* (pp. 49-78). Springer.
[119]. Sarpong, D., Ofosu, G., Botchie, D., & Clear, F. (2020). Do-it-yourself (DiY) science: The proliferation, relevance and concerns. *Technological Forecasting and Social Change*, *158*, 120127.
[120]. Saxena, R., Srivastava, V., Bharti, D., Singh, R., & Kumar, A. (2025). Artificial intelligence for renewable energy strategies and techniques. In *Computer Vision and Machine Intelligence for Renewable Energy Systems* (pp. 17-39). Elsevier.
[121]. Sazzad, I., & Md Nazrul Islam, K. (2022). Project impact assessment frameworks in nonprofit development: a review of case studies from south asia. *American Journal of Scholarly Research and Innovation*, *1*(01), 270-294. https://doi.org/10.63125/eeja0t77
[122]. Shafik, W. (2025). An Overview of Artificial Intelligence Solutions for the Maintenance and Evaluation of Photovoltaic Systems. *Energy Conversion Systems-Based Artificial Intelligence: Applications and Tools*, 23-53.
[123]. Sharma, P., Chung, W. T., Akoush, B., & Ihme, M. (2023). A review of physics-informed machine learning in fluid mechanics. *Energies*, *16*(5), 2343.
[124]. Sharma, V., Cortes, A., & Cali, U. (2021). Use of forecasting in energy storage applications: A review. *IEEE access*, 9, 114690-114704.
[125]. Sheratun Noor, J., & Momena, A. (2022). Assessment Of Data-Driven Vendor Performance Evaluation in Retail Supply Chains: Analyzing Metrics, Scorecards, And Contract Management Tools. *American Journal of Interdisciplinary Studies*, *3*(02), 36-61. https://doi.org/10.63125/0s7t1y90
[126]. Sideris, I. V., Foresti, L., Nerini, D., & Germann, U. (2020). NowPrecip: Localized precipitation nowcasting in the complex terrain of Switzerland. *Quarterly Journal of the Royal Meteorological Society*, *146*(729), 1768-1800.
[127]. Simonin, D., Pierce, C., Roberts, N., Ballard, S. P., & Li, Z. (2017). Performance of Met Office hourly cycling NWP-based nowcasting for precipitation forecasts. *Quarterly Journal of the Royal Meteorological Society*, *143*(708), 2862-2873.

[128]. Singh, G., & Harun. (2023). Multiple linear regression based analysis of weather data for precipitation and visibility prediction. International Conference on Advances in Computing and Data Sciences,
[129]. Singh, H., Ang, L.-M., Lewis, T., Paudyal, D., Acuna, M., Srivastava, P. K., & Srivastava, S. K. (2024). Trending and emerging prospects of physics-based and ML-based wildfire spread models: A comprehensive review. *Journal of Forestry Research*, *35*(1), 135.
[130]. Singla, P., Duhan, M., & Saroha, S. (2022). A comprehensive review and analysis of solar forecasting techniques. *Frontiers in Energy*, *16*(2), 187-223.
[131]. Spyrou, C., Varlas, G., Pappa, A., Mentzafou, A., Katsafados, P., Papadopoulos, A., Anagnostou, M. N., & Kalogiros, J. (2020). Implementation of a nowcasting hydrometeorological system for studying flash flood events: The Case of Mandra, Greece. *Remote Sensing*, *12*(17), 2784.
[132]. Stefanov, D., & Demšar, J. (2025). Forecasting solar power production by using satellite images. *Solar Energy*, *300*, 113823.
[133]. Subrato, S., & Md, N. (2024). The role of perceived environmental responsibility in artificial intelligence-enabled risk management and sustainable decision-making. *American Journal of Advanced Technology and Engineering Solutions*, *4*(04), 33-56. https://doi.org/10.63125/7tjw3767
[134]. Sulaiman, M. H., & Mustaffa, Z. (2024). Forecasting solar power generation using evolutionary mating algorithm-deep neural networks. *Energy and AI*, *16*, 100371.
[135]. Sun, X., Chung, H., Chavali, R. V. K., Bermel, P., & Alam, M. A. (2017). Real-Time Monitoring of Photo Voltaic Reliability Only Using Maximum Power Point-The Suns-Vmp Method. 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC),
[136]. Suthar, T., Shah, T., Raja, M. K., Raha, S., Kumar, A., & Ponnusamy, M. (2023). Predicting weather forecast uncertainty based on large ensemble of deep learning approach. 2023 international conference on self sustainable artificial intelligence systems (ICSSAS),
[137]. Symeonidis, C., & Nikolaidis, N. (2025). Efficient deterministic renewable energy forecasting guided by multiple-location weather data. *Neural Computing and Applications*, 1-28.
[138]. Tahmina Akter, R., & Abdur Razzak, C. (2022). The Role of Artificial Intelligence in Vendor Performance Evaluation Within Digital Retail Supply Chains: A Review Of Strategic Decision-Making Models. *American Journal of Scholarly Research and Innovation*, *1*(01), 220-248. https://doi.org/10.63125/96jj3j86
[139]. Tahmina Akter, R., Debashish, G., Md Soyeb, R., & Abdullah Al, M. (2023). A Systematic Review of AI-Enhanced Decision Support Tools in Information Systems: Strategic Applications In Service-Oriented Enterprises And Enterprise Planning. *Review of Applied Science and Technology*, *2*(01), 26-52. https://doi.org/10.63125/73djw422
[140]. Teixeira, R., Cerveira, A., Pires, E. J. S., & Baptista, J. (2024). Advancing renewable energy forecasting: A comprehensive review of renewable energy forecasting methods. *Energies*, *17*(14), 3480.
[141]. Travieso-Gonzalez, C. M., Cabrera-Quintero, F., Pinan-Roescher, A., & Celada-Bernal, S. (2024). A review and evaluation of the state of art in image-based solar energy forecasting: The methodology and technology used. *Applied Sciences*, *14*(13), 5605.
[142]. Ukoba, K., & Jen, T.-C. (2025). Thin Films: History, Properties and Emerging Trends. *Shaping Tomorrow: Thin Films and 3D Printing in the Fourth Industrial Revolution 1: Fundamentals*, 55-155.
[143]. Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.-L. A., Elkhatib, Y., Hussain, A., & Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE access*, *7*, 65579-65615.
[144]. Visser, L., AlSkaif, T., & Van Sark, W. (2019). Benchmark analysis of day-ahead solar power forecasting techniques using weather predictions. 2019 IEEE 46th photovoltaic specialists conference (PVSC),
[145]. Wang, S.-Y., Qiu, J., & Li, F.-F. (2018). Hybrid decomposition-reconfiguration models for long-term solar radiation prediction only using historical radiation records. *Energies*, *11*(6), 1376.
[146]. Warner, J., Prada Jardim, A., & Albera, C. (2025). Artificial Intelligence: Applications in Pharmacovigilance Signal Management. *Pharmaceutical Medicine*, 1-16.
[147]. Woo, W.-c., & Wong, W.-k. (2017). Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere*, *8*(3), 48.
[148]. Wu, Y.-K., Huang, C.-L., Phan, Q.-T., & Li, Y.-Y. (2022). Completed review of various solar power forecasting techniques considering different viewpoints. *Energies*, *15*(9), 3320.
[149]. Xue, H., Ma, J., Zhang, J., Jin, P., Wu, J., & Du, F. (2024). Power forecasting for photovoltaic microgrid based on multiscale cnn-lstm network models. *Energies*, *17*(16), 3877.
[150]. Yagli, G. M., Yang, D., & Srinivasan, D. (2020). Reconciling solar forecasts: Probabilistic forecasting with homoscedastic Gaussian errors on a geographical hierarchy. *Solar Energy*, *210*, 59-67.
[151]. Yang, J., He, H., Zhao, X., Wang, J., Yao, T., Cao, H., & Wan, M. (2023). Day-ahead PV power forecasting model based on fine-grained temporal attention and cloud-coverage spatial attention. *IEEE Transactions on Sustainable Energy*, *15*(2), 1062-1073.

Volume 04, Issue 01 (2025)
Page No: 392-427
eISSN: 3067-5146
Doi: 10.63125/asxzy065

[152]. Young, M. V., & Grahame, N. S. (2024). The history of UK weather forecasting: the changing role of the central guidance forecaster. Part 8. Operational forecasting in the twenty-first century: enhanced capabilities from nowcasting to extended range. *Weather*, *79*(5), 148-157.

[153]. Yu, J., Li, X., Yang, L., Li, L., Huang, Z., Shen, K., Yang, X., Yang, X., Xu, Z., & Zhang, D. (2024). Deep learning models for PV power forecasting. *Energies*, *17*(16), 3973.

[154]. Yuan, Y., Wang, Q., & Yang, X. T. (2021). Traffic flow modeling with gradual physics regularized learning. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 14649-14660.

[155]. Zhang, R., Feng, M., Zhang, W., Lu, S., & Wang, F. (2018). Forecast of solar energy production-A deep learning approach. 2018 IEEE International Conference on Big Knowledge (ICBK),

[156]. Zhang, T., & Strbac, G. (2025). Novel Artificial Intelligence Applications in Energy: A Systematic Review. *Energies*, *18*(14), 3747.

[157]. Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., & Wang, J. (2023). Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, *619*(7970), 526-532.

[158]. Zhang, Y., Yang, R., Zhang, J., Weng, Y., & Hodge, B.-M. (2018). Predictive analytics for comprehensive energy systems state estimation. In *Big data application in power systems* (pp. 343-376). Elsevier.

[159]. Zhou, H., Zheng, P., Dong, J., Liu, J., & Nakanishi, Y. (2024). Interpretable feature selection and deep learning for short-term probabilistic PV power forecasting in buildings using local monitoring data. *Applied Energy*, *376*, 124271.

[160]. Zhou, J., Liu, J., Ren, J., & He, C. (2025). A Comprehensive Study of Machine Learning for Waste-to-Energy Process Modeling and Optimization. *Processes*, *13*(9), 2691.

[161]. Zhu, Z., Chen, W., Xia, R., Zhou, T., Niu, P., Peng, B., Wang, W., Liu, H., Ma, Z., & Gu, X. (2023). Energy forecasting with robust, flexible, and explainable machine learning algorithms. *AI Magazine*, *44*(4), 377-393.

[162]. Zwane, N., Tazvinga, H., Botai, C., Murambadoro, M., Botai, J., De Wit, J., Mabasa, B., Daniel, S., & Mabhaudhi, T. (2022). A bibliometric analysis of solar energy forecasting studies in Africa. *Energies*, *15*(15), 5520.