Energy Reports 13 (2025) 4299–4317

Contents lists available at ScienceDirect

Energy Reports

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

Deep learning augmented medium-term photovoltaic energy forecasting: A coupled approach using PVGIS and numerical weather model data

Muhammad Ehtsham [a], Marianna Rotilio [a,*], Federica Cucchiella [b]

[a] Department of Civil, Construction-Architectural and Environmental Engineering, University of L'Aquila, L'Aquila 67100, Italy
[b] Department of Industrial and Information Engineering and Economics, University of L'Aquila, L'Aquila 67100, Italy

ARTICLE INFO

Keywords:
PV Energy
Deep Learning
Machine Learning
Hybrid Models
PV management

ABSTRACT

Integrating PV energy resources into energy grids is crucial for PV energy organizations, making medium- and short-term PV forecasts important. PV organizations look forward to modern tolls for efficient systems for the most beneficial operations for PV systems. This research proposes, applies, and assesses a modern machine learning and deep learning based short-term PV energy forecasting system. Numerical weather model-based data is utilized for real-time forecasting at an hourly scale for the next four days, an additional analysis is performed by leveraging the PVGIS data in addition to NWM data. The proposed methodology is developed and applied to more than 200 PV installations, both BIPVs and BAPVs. The system was able to produce effective PV energy forecasts with high accuracy and efficiency analysis of 3 different PV installations ranging 17 kWp, 91 kWp, and 386kWp are reported in this paper. The research concluded with the feasibility of the proposed systems and findings further support the efficacy of the proposed framework, which can be adopted by organizations seeking to optimize PV system performance and reliability.

1. Introduction

PV energy being an affordable and green energy source is being recognized globally and a huge shift in PV sources is being observed (Choudhary and Srivastava, 2019). PV technology is continuously improving, not only through advancements in the cells of PV modules but the introduction of PV cells and modules in different components of the buildings in the form of BIPVs is also emerging as modern sophisticated advancements for the architects and energy planners (Basher et al., 2023). The buildings sector which consumes almost 40 % of the total energy production in the EU (Peppas et al., 2021), is at the forefront of benefiting from advancements in the field of PV energy resources. Modern sustainable buildings are being integrated with BIPVS as the BIPVs are way more attractive than the BAPVs as they embed themselves in the building envelopes, without occupying more area or space, are more lightweight, resilient to wind forces, and aesthetically appealing (Ghosh, 2020). However, BIPVsare more prone to heat absorption as compared to BAPVs, this is due to their embedding in the building enveloped and urban inland, making the ventilation sparse. Studies have revealed that the ventilation of PVs plays a vital part in their efficient performance (Touati et al., 2013; Ehtsham and Rotilio, 2024), leading to the difference in PV production of BAPVs and BIPVs in the same atmospheric conditions. There are plenty of online web sources/platforms available that provide technical, economical, system optimization, sensitivity analysis, and GHG analysis for PV plants, these include PVGIS, SolarGIS, PVWatts, RETScreen, BlueSol, PVsyst, HelioScope, PV*SOL premium, Solarius PV, Solar Pro, PV F-Chart, PolySun, SAM, and HOMER (Milosavljević et al., 2022). These estimates, although provide a baseline for the analysis often lack certain accuracy and the actual results may deviate from these estimates. Mainly because production by specific PV modules is often different in real environments from that of the proposed in laboratory settings or controlled environments (Hernández-Callejo et al., 2019). With all the benefits of PV-derived energy, there is a main shortcoming that makes them challenging, it is the unpredictability of the PV energy production (Antonanzas et al., 2016a; Sinsel et al., 2020). It is essential to accurately predict how much energy will be produced by a specific PV plant at a specific time or hour.

With advancements in computational technologies, modern machine learning, and deep learning models have revolutionized the IT industry. Modern AI is revolutionizing the way we live by making computations not only easier but also significantly faster. AI is being leveraged across a wide range of organizations, professions, and industries, driving the

* Corresponding author.
E-mail address: marianna.rotilio@univaq.it (M. Rotilio).

https://doi.org/10.1016/j.egyr.2025.03.058
Received 18 December 2024; Received in revised form 28 February 2025; Accepted 31 March 2025

Nomenclature

Abbreviations	Building Applied Photovoltaics
BAPV	Building Applied Photovoltaics
BIPV	Building Integrated Photovoltaics
CM SAF	Climate Monitoring Satellite Application Facility
CSV	Comma-Separated Values
FNN	Feedforward Neural Network
GBM	Gradient Boosting Machines
Gb(i)	Beam (direct) irradiance on the inclined plane (plane of the array) (W/m2)
Gd(i)	Diffuse irradiance on the inclined plane (plane of the array) (W/m2)
GHG	Greenhouse Gas
GRU	Gated Recurrent Unit
Gr(i)	Reflected irradiance on the inclined plane (plane of the array) (W/m2)
H_sun	Sun height (degree)
HOMER	Hybrid Optimization of Multiple Energy Resources
KNN	K-Nearest Neighbors
KNN	K-Nearest Neighbors
kWp	Kilowatt Peak
LSTM	Long Short-Term Memory
MAE	(Mean Absolute Error)
MLP	Multilayer Perceptron
nRMSE	Normalized RMSE
NWP	Numerical Weather Prediction
PV	Photovoltaic
PVGIS	Photovoltaic Geographical Information System
RES	Renewable Energy Sources
RMSE	Root Mean Squared Error
R^2	Coefficient of Determination
RF	Random Forest
SAM	System Advisor Model
SARAH	Surface Solar Radiation Data Set – Heliosat
T2m	2-m air temperature (degree Celsius)
TF	Transformer
WS10m	10-m total wind speed (m/s)
XGBoost	Extreme Gradient Boosting

digital transformation of Industry 4.0. AI is playing a pivotal role in shaping Industry 5.0, where the focus shifts towards human-centric, sustainable, and resilient industrial practices, enhancing collaboration between humans and intelligent systems (Leng et al., 2022). AI finds applications in various fields including the medical field (Zhang and Kamel Boulos, 2023; Ziller et al., 2024), educational sector (Wang et al., 2023; Baidoo-Anu and Ansah, 2023), economics and finance sector (Rahmani et al., 2023; Jareño and Yousaf, 2023), engineering sciences (Yüksel et al., 2023; Ahmed et al., 2023; Tan et al., 2024), and agriculture field (Wakchaure et al., 2023; Javaid et al., 2023; Akintuyi, 2024).

Rahmani et al. (2023) concluded that AI was able to detect and identify fraudulent accounting transactions by employing advanced models, findings highlighted AI's suitability for stock market predictions, oil price forecasting, and economic impacts (Wakchaure et al., 2023). provided an overview of applications of AI techniques in agriculture and findings of research underscore the effectivity of AI models in a real-time manner, improving decision-making in a time-effective manner and avoiding human errors. AI-based approaches outperformed traditional practices. Similar to other fields, AI models, more specifically machine learning (Scott et al., 2023a; Mahmud et al., 2021), deep learning (Du Plessis et al., 2021; Cantillo-Luna et al., 2023), and hybrid (Li et al., 2020a, 2020b; Khelifi et al., 2023) models have been exploited for the forecasting and efficiency analysis of the PV systems at various scales. To enhance the effectiveness of PV energy resources the forecast horizon may be selected based upon specific uses of the data (Antonanzas et al., 2016b). The accuracy of a PV forecasting model hugely depends upon the forecast horizon (Das et al., 2018). Various researchers focused on different forecast horizons, ranging from very small-term forecasts for less than 30 minutes, small-term forecasts for 30–360 minutes, medium-term forecasts for 6–24 hours, and long-term forecasts for more than 24 hours (Ren et al., 2015). Energy management and maintenance planning employ medium-term forecasts. Daily operations, unit commitment, scheduling functions, net interchange evaluation, and system security analysis all depend heavily on short-term forecasting (Srinivasan and Lee, 1995). Power system control and scheduling depend on the short-term load forecasting of renewable energy, which also has an impact on system dependability.

Researchers have exploited modern techniques including ML and DL algorithms for PV energy forecasting at local and micro scales (Zhen et al., 2021; Sharadga et al., 2020) (Aslam et al., 2021). conducted a comprehensive survey on the deep learning methods for power load and renewable energy forecasting in smart microgrids. Findings revealed that different DL based models have their own strong and weak areas and it was difficult to recommend one single model as optimal for solar energy or load forecasting. The literature revealed that to train the models the researchers took leverage of irradiance data either from numerical-based weather models or locally generated forecasts (Gupta and Singh, 2021; Konstantinou et al., 2021; Mohamad Radzi et al., 2023). The irradiance data is crucial for PV forecasting as it is the most important metrological variable affecting PV generation (Ziane et al., 2021). Solar irradiance data is highly intermittent and complex in nature, which introduces noise to the forecasting target function and renders it non-smooth (Kumari and Toshniwal, 2021). Moreover, access to in situ real-time and forecasted irradiance data is complex and requires sophisticated and economically unviable equipment to record the irradiance data (Nou et al., 2016). This problem is more prominent in case of the small-scale PV plants, which have less flexible budgets, and installation and operation of irradiance meters are rare.

The literature reveals that some recent studies have utilized LSTM, Xgboost, GRU, and CNN models for PV forecasting at isolated or micro scales utilizing PVGIS (Fadoul et al., 2023; Sabareesh et al., 2023; Didavi et al., 2024; Asghar et al., 2024). In addition, (Brester et al., 2023) utilized ANN model for performance assessment for day-ahead solar PV forecasting using different training datasets from ground-based historical weather observations and historical numerical weather prediction datasets, the methodology was applied to three PV installations in Finland (Markovics and Mayer, 2022). evaluated various ML models, including but not limited to SVM, RF, XGboost, KNN, and MLP, for deterministic day-ahead power forecasting based on numerical weather predictions for 16 PV plants in Hungary. Literature reveals that most of the research produced focuses on isolated PV plants with a noticeable lack of a standardized, universal technique for organizational-level PV energy forecasting, that is capable of energy forecasting both BAPVs and BIPVs, utilizing the least available data. Furthermore, no comprehensive methodological framework has been proposed that combines PVGIS and NWM model data for the application and evaluation of ML, DL, and hybrid models at the meso scale.

This research work was conducted to formulate and validate a PV energy forecasting system at the mesoscale utilizing both PVGIS and NWM datasets. This research work proves the proposed system's applicability on organization scales by the creation of a high-accuracy machine and deep learning-based forecasting system. Contrary to forecasting systems reported and exploited by many researchers, the

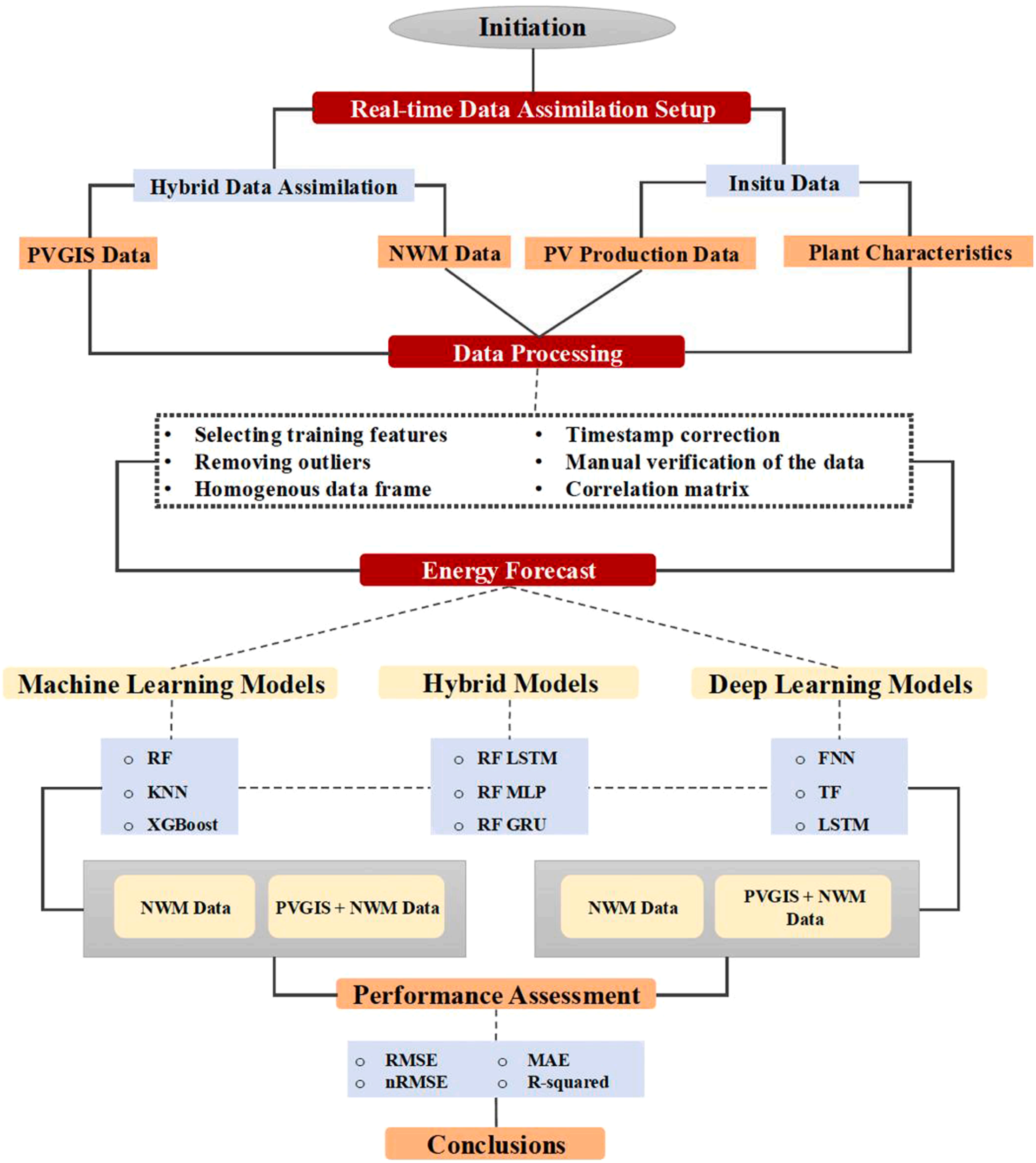


Fig. 1. Flowchart of the proposed methodology.

proposed system can forecast the PV production of a few hundred PV plants situated at different locations. The main advantage of the proposed system is that it leverages minimal data about the plant's design and characteristics. It does not need detailed module and array information, nor does it rely on economically unviable datasets. These highlights make the proposed system an easy-to-implement and economically viable source of datasets for both small and large-scale PV organizations.

2. Methodology

A methodological framework was developed before the initialization of the research work. Considering the implications of most organizations and companies involved in monitoring and management purposes, the authors devised a methodology to formulate a framework to forecast PV energy production at an hourly scale in the absence of real-time irradiance data. The data used in the research work was accessed from OpenWeather (OpenWeatherMap, 2025) and PVGIS (Photovoltaic Geographical Information System, 2025), OpenWeather provided the historical and forecasted real-time data of various meteorological parameters. Some other researchers also exploited the data from the same model and recommended the suitability at their designed study levels (Musah et al., 2022; Uribe et al., 2024; Jurczyński; Retnowati et al., 2021). However, this research work exploits the datasets at the meso scale, involving hundreds of PV plants and inverters. Additionally, the effectiveness of various ML, DL, and Hybrid models has been exploited. Fig. 1 illustrates the various steps undertaken during this research.

2.1. Data acquisition and data assimilation setup

In the initial phase a data acquisition system was formulated, this system was crucial for the successive research stages, as the continuous data accessing and storing was necessary for the real-time application of the proposed system. A Python script was written in Google colab to connect the MySQL database for accessing the historic energy production data of inverters. It is noteworthy here to mention that the data

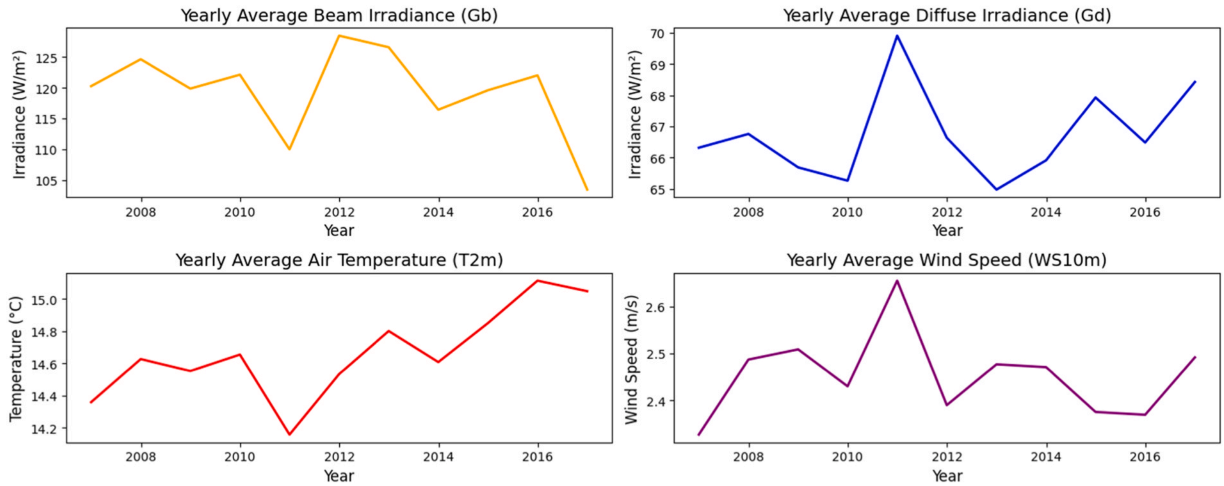


Fig. 2. Yearly averages of Gb, Gd, T2m, and WS10m data accessed from PVGIS for one of the PV plant locations, Latitude:41.35, Longitude:13.43, Elevation 12 m.

Table 1
Different performance metrics employed to assess the performance of the models.

Metric	Formula	Range	Description
RMSE (Root Mean Squared Error)	$RMSE = \sqrt{\left[\frac{\sum_{i=1}^{N}\left(RHE_i^a - PHE_i^f\right)^2}{N}\right]}$	[0, ∞)	Measures the average size of the errors between predicted and actual values. A lower value indicates better prediction.
nRMSE (Normalized RMSE)	$nRMSE = \frac{RMSE}{RHE(m)}$	[0, ∞)	RMSE adjusted by the range of actual values, making it easier to compare across different datasets. A lower value means the model was able to forecast with high accuracy.
MAE (Mean Absolute Error)	$MAE = \frac{\sum_{i=1}^{N}\left\|RHE_i^a - PHE_i^f\right\|}{N}$	[0, ∞)	The average of the absolute differences between predicted and actual values. Smaller values mean better quality forecasts.
R Squared (Coefficient of Determination)	$R2 = 1 - \frac{\sum_{i=1}^{N}\left(RHE_i^a - PHE_i^f\right)^2}{\sum_{i=1}^{N}\left(RHE_i^a - RHE(m)\right)^2}$	(-∞, 1]	Indicates how well the model explains the variability in the actual data. An R^2 value closer to 1 indicates better forecasts.

acquired through the developed script is in real-time and converts the 5-minute records to hourly records. Hence, providing the hourly energy production records till the last completed hour. The data acquisition system in real time makes the system effective for the short-term forecasts making the decision-making process more efficient. A continuous and robust data assimilation setup combines the data transmitted by self-standing dataloggers connected with all the inverters associated with the respective PV plant.

2.2. Numerical weather model data

Weather-based metrological features data was collected from the OpenWeather, which has its numerical weather prediction model, and the data is provided through APIs, the numerical weather prediction (NWP) model of OpenWeather uses several data sources for the modeling. These sources include 1; Global NWP models including NOAA

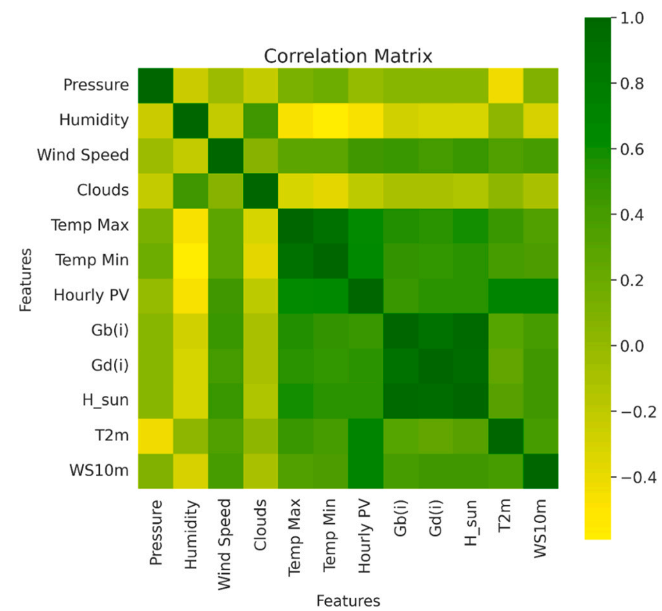


Fig. 3. Correlation matrix between features and target variable for the plant BDS.

Table 2
Plant characteristics of 3 reported plants.

Plant Code	Inverters	Latitude	Longitude	Nominal Power	Activation Date	Location
BDS	4	42.3	14.4	17.07 kWp	2009–08–04	Chietino (CH)
ADI	4	41.3	13.4	91 kWp	2013–05–14	Fondi (LT)
CRA	14	42.6	13.9	386 kWp	2019–10–02	Notaresco (TE)

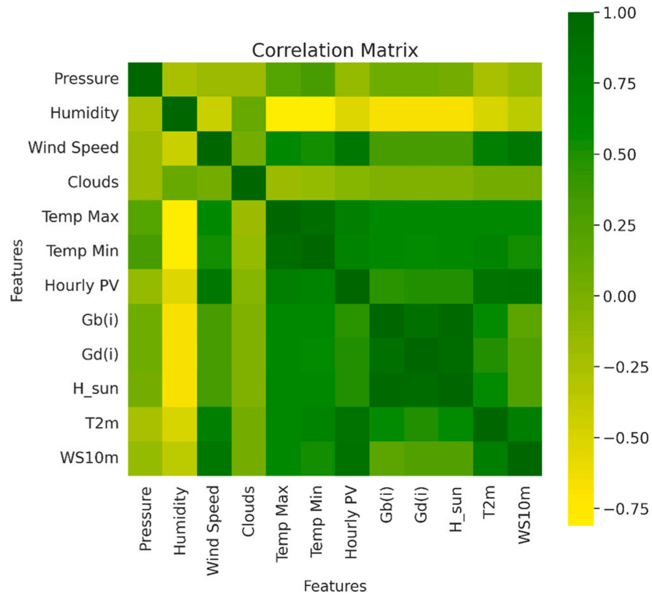


Fig. 4. Correlation matrix between features and target variable for the plant ADI.

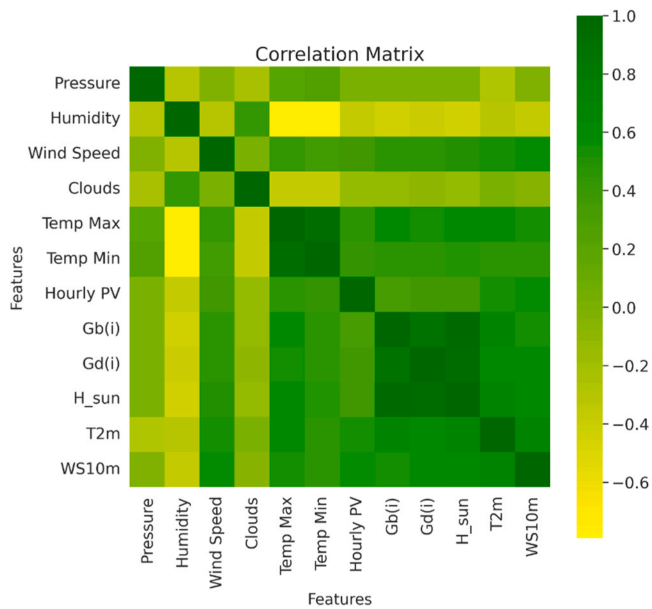


Fig. 5. Correlation matrix between features and target variable for the plant CRA.

GFS 0.25 and 0.5 grid sizes, NOAA CFS, and ECMWF ERA, 2; Weather stations including METAR stations, Users' stations, and Companies' stations, 3; Weather radar data, and 4; Satellite data. OpenWeather access and download the metrological estimates from the sources. Afterward, this data is processed with the algorithms of OpenWeather to enhance the accuracy of the metrological estimates. The data is processed in real time to make sure the availability of real-time data for the users.

OpenWeather data is available globally and for any location on the globe the historic, current, and forecasted metrological data can be assessed using APIs. The data is being utilized for agricultural, industrial, weather applications, and PV-related applications. OpenWeather provides a variety of data services including Current Weather and Forecasts: Access to current conditions, minute-by-minute forecasts for the next hour, hourly forecasts for 48 hours, daily forecasts for up to 8 days, and government weather alerts. Historical Weather Data: Weather data for any location from the past 45 years, including historical hourly data and daily summaries. Weather Maps: Current, forecast, and historical weather maps with various layers such as precipitation, temperature, and wind, available at different resolutions and update intervals. Some of the datasets are made available by using specialized APIs like Solar irradiance, PV energy prediction, air pollution data, and fire weather index. The utilization of specific datasets depends upon one's requirements and needed temporal resolution. For the sake of this research project, the developer plan was accessed which provides current weather and forecasts. The limits are Hourly forecasts for up to 4 days, daily forecasts for up to 16 days, calls per minute 3000, and 3-hour forecasts for 5 days. For the historic data medium plan was accessed which limits the users to 50000 calls per day and provides historical data up to one year back.

2.3. PVGIS data

The data was accessed for the selected location for the period starting from 01st January 2006–31st December 2016. The features collected for the mentioned period were

- Gb: Beam (direct) irradiance on the inclined plane (plane of the array) (W/m2)
- Gd: Diffuse irradiance on the inclined plane (plane of the array) (W/m2)
- Gr: Reflected irradiance on the inclined plane (plane of the array) (W/m2)
- H_sun: Sun height (degree)
- T2m: 2-m air temperature (degree Celsius)
- WS10m: 10-m total wind speed (m/s)

The hourly records were collected and to make homogenous datasets all the records were averaged for the selected period and a total number of 366 days records with 24 values of each parameter were created. The inclusion of leap year data was made with the data available for the leap years only. The PVGIS used data from the CM SAF "SARAH-Edition 2" solar radiation data product. These data were incorporated in PVGIS version 5.2. The period used to calculate the averages is 2005–2020.

All the data collected from the PVGIS was accessed in.csv format and then exported into Microsoft Excel files for cross-validation and verification by the authors. The script was written and executed in Google colab and the results and the output of the code were saved corresponding to the specific plant code of the PV plant under consideration. The script is universal and has the potential to access the data for any latitude and longitude, the data of altitude and longitude for the plant under consideration is always fetched from the MySQL database of the PV organization, making sure the applicability and continuity of the analysis process. For an overview of the data insights regarding the parameters accessed from PVGIS, the yearly averages of four meteorological parameters are depicted in Fig. 2.

2.4. Data processing

A consistent and reliable data processing system was necessary for the continuous application and meso scale scope of the proposed systems, hence the scripts were written in Google colab to specifically tailor the accessed data frames and make them homogeneous for the subsequent phases of the analysis.

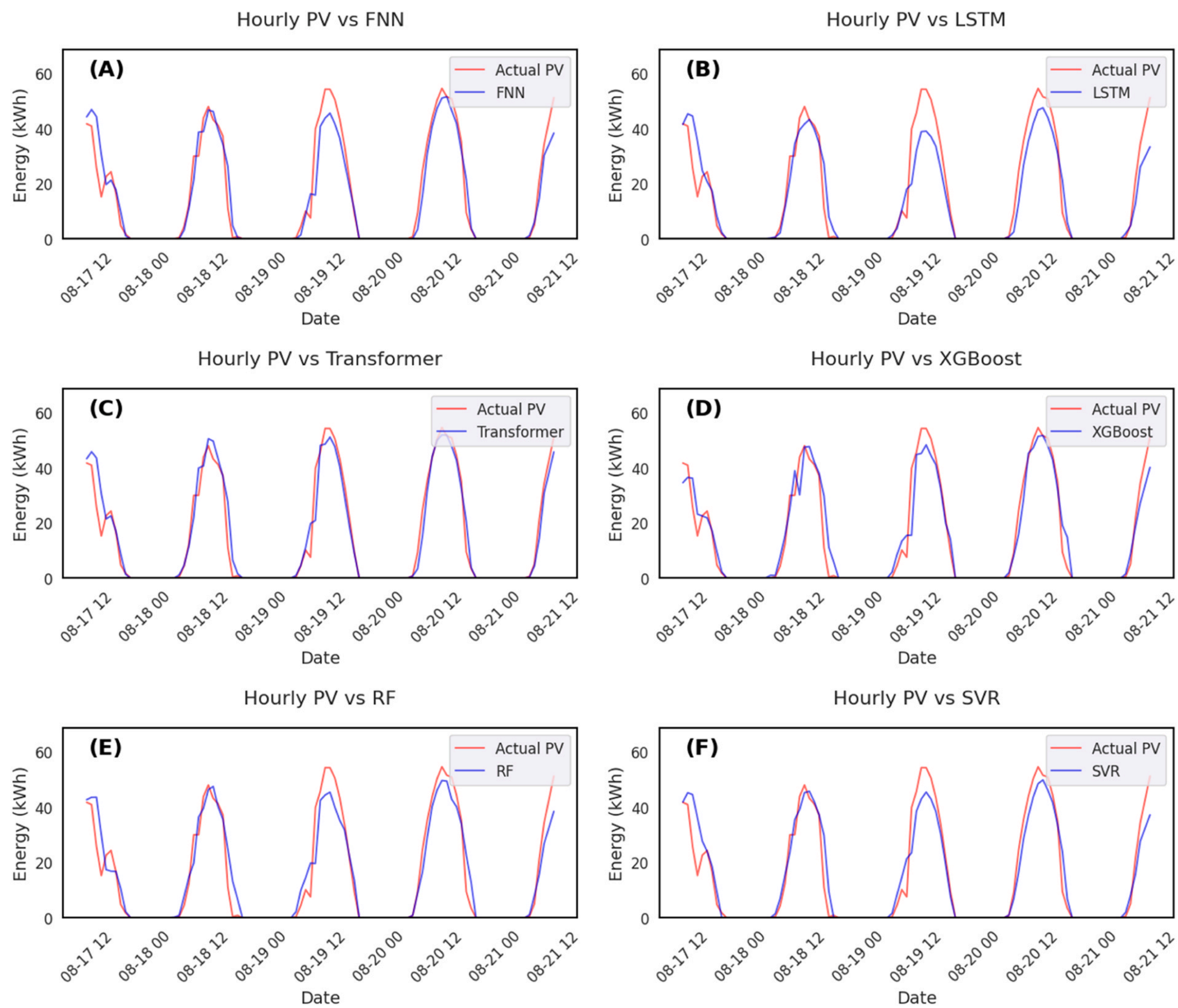


Fig. 6. Comparison between actual energy recorded and forecasted energy generated by ML and DL models for 91 kWp PV plant commissioned on 2013–05–14 (ADI).

2.5. *PV energy forecasts*

2.5.1. *Machine learning models*

Machine learning algorithms have been exploited in different fields of science to make the decision-making processes more efficient and identical to all other fields the PV solar forecasting models have benefitted from the modern machine learning models and interesting results have been achieved and reported by various researchers. Most of the researchers report that the use of ML models is producing efficient forecasts of the PV energy as compared to traditional regression models. Many researchers have utilized ML models like Random Forest (RF) (Niu et al., 2020), Decision trees (DT) (Wang et al., 2018), Support vector regression (SVR) (Wolff et al., 2016), and Gradient Boosting Machines (GBM) (Persson et al., 2017) i.e. models like XGBoost, LightGBM, and CatBoost for PV forecasting at various horizons and for different parts of the world. However, the literature lacks the introduction of an approach feasible and validated for mesoscale application of ML models leveraging PVGIS and NWM data. For this research project, the authors decided to utilize the models that have been reported to produce best quality forecasts in different fields and authors selected RF, SVR, and XGBoost models for forecasting. The working of the ML models for PV forecasting has been already discussed in the literature in detail (Kumar et al., 2022; Scott et al., 2023b) and we will not dive into the details of the working and functioning procedure of each ML model. Instead, we provide the data regarding the tuned hyperparameters of each ML model which depicted optimal results in succeeding sections of the article.

2.5.2. *Deep learning models*

In addition to ML models, in the recent past many modern deep learning models have gained a reputation amongst researchers for the time-series forecasting of different variables (Zeroual et al., 2020; Torres et al., 2021). Similarly, many researchers have utilized DL models for effective short- and long-term PV forecasting at different scales and various time horizons (Abdel-Basset et al., 2021; Li et al., 2020c; Mellit et al., 2021). Neural Networks specifically are crucial, which are sometimes referred to as artificial neural networks, these are a subset of ML and are pivotal to all DL models. The word "neural" comes from the main characteristics of the model which is like the signaling activity of neurons in a human brain. Notable DL models that have been depicted include but are not limited to a multilayer perceptron (MLP), recurrent neural networks (RNN), Feedforward Neural Networks, convolutional neural networks (CNN), and graph neural networks (GNN), LSTM, Transformers. The authors decided to utilize the Multilayer Perceptron (MLP) of FNN, Recurrent Neural Network (RNN) which are also known as an LSTM (Long Short-Term Memory) network, and Transformer-based model for time-series forecasting. Transformers, originally designed for natural language processing, are becoming increasingly popular for time-series due to their ability to capture

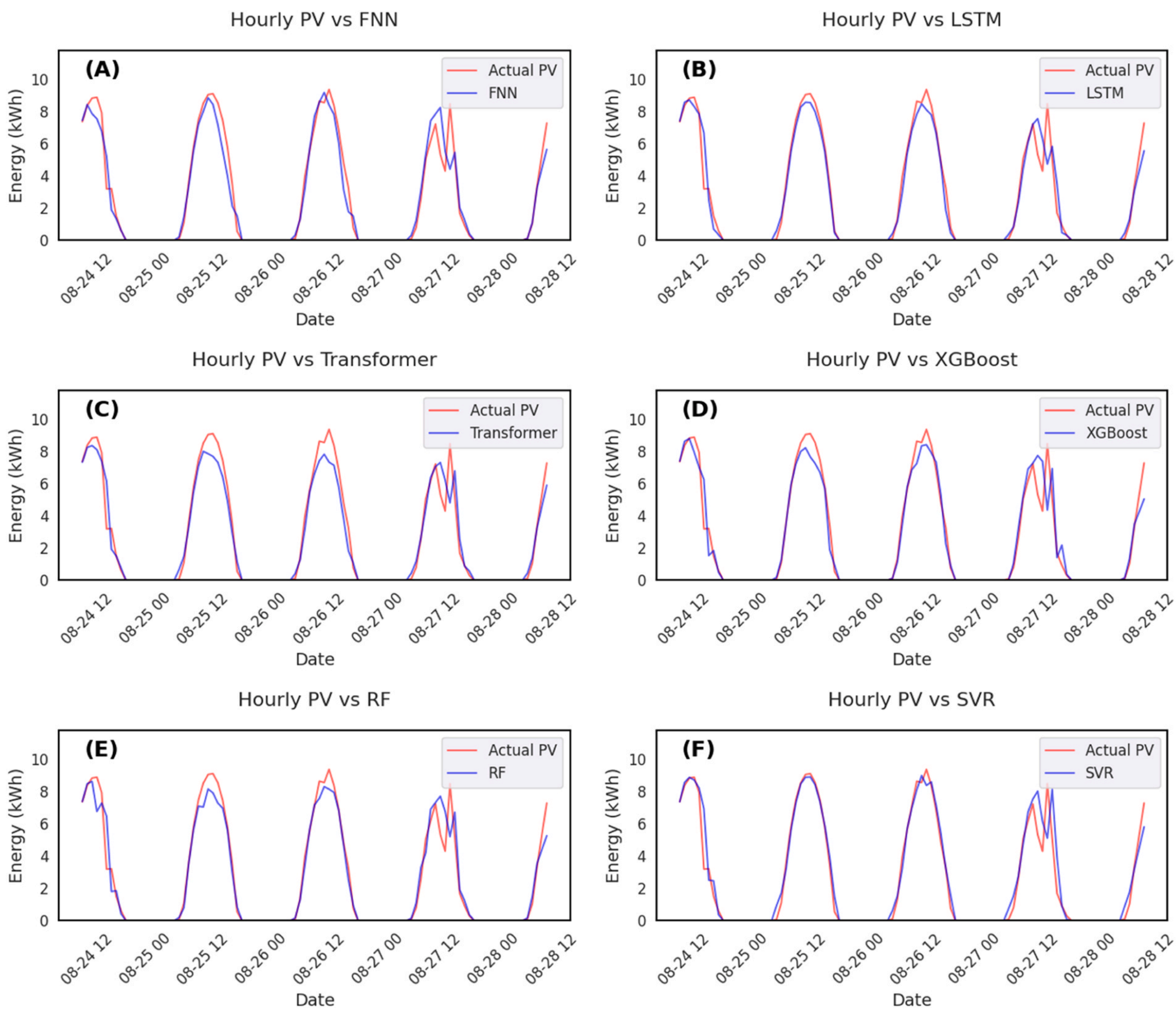


Fig. 7. Comparison between actual energy recorded and forecasted energy generated by ML and DL models for 17 kWp PV plant commissioned on 2009–08–04 (BDS).

long-range dependencies and relationships in sequences (Chen et al., 2023). DL models adopted for energy forecasting include:

- Multilayer Perceptron (MLP) Model
- LSTM (Long Short-Term Memory) Model
- Transformer Model

The MLP model is a feedforward neural network composed of dense layers used for time-series forecasting (Oancea and Ciucu, 2014). The second model is a recurrent neural network with LSTM layers, which is designed to capture long-term temporal dependencies in the time-series data (Hewamalage et al., 2021). The third model is based on the Transformer architecture, using multi-head attention layers to handle time-series forecasting by focusing on different parts of the input sequence (Zeng et al., 2022).

2.5.3. Hybrid models

Hybrid models, combine the traditional ML techniques with DL architecture, and have gained significant spotlight due to their capability of leveraging the strengths of both approaches. Various studies have evaluated the hybrid models for different target variables (Akhter et al., 2022; Qaid et al., 2021; Celik and Inik, 2024). Authors decided to include the hybrid models by combining the random forest model with different notable and effective DL architectures namely LSTM, MLP, and GRU. RF LSTM model was the first among the developed hybrid models, it integrates the strengths of the RF model with LSTM networks, which has proved its effectivity for capturing temporal dependencies and for handling sequential data. This combination allows RF LSTM to leverage both the feature selection capabilities of RF and the sequence learning ability of LSTM, making it particularly effective for time series forecasting. The second developed model is RF MLP combines RF with a Multi-Layer Perceptron, which is a type of feedforward artificial neural network, famous for learning complex nonlinear relationships. The integration of RF with MLP enables the model to harness both the feature importance evaluation of RF and the deep learning power of MLP, making it robust in scenarios requiring both interpretability and deep learning capabilities. The third model was RF GRU, pairing RF with a Gated Recurrent Unit (GRU). GRUs are well known for their learning effectiveness utilizing fewer parameters in time series forecasting as compared to LSTMs. These hybrid models provide an ample balance between complexity and performance. All the aforementioned hybrid models with varying architecture share the common ambition of improving prediction accuracy by leveraging the strengths of both ML and DL techniques.

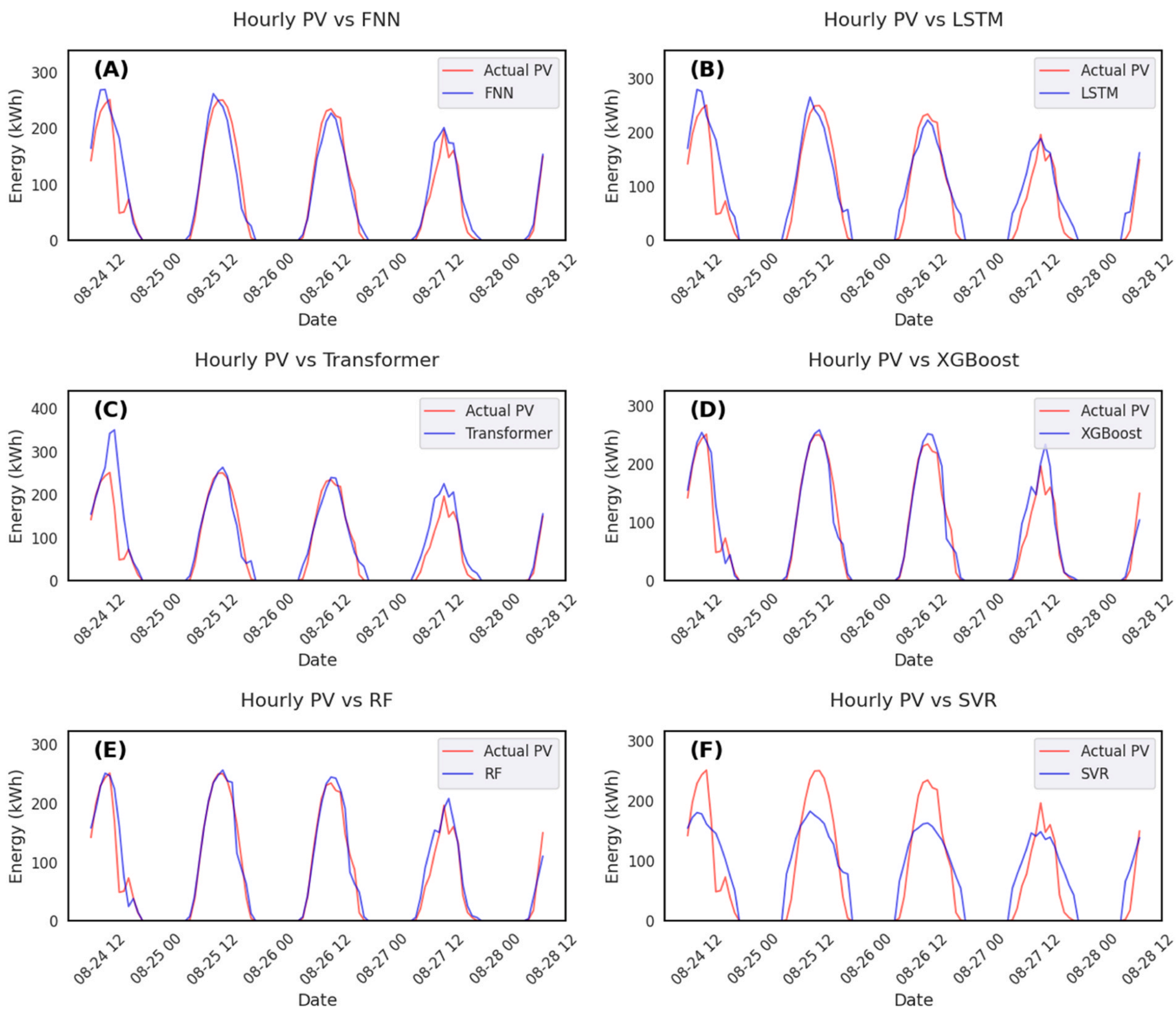


Fig. 8. Comparison between actual energy recorded and forecasted energy generated by ML and DL models for 400 kWp PV plant commissioned on 2021–09–26 (CRA).

2.6. Performance assessment

Forecasts generated by the models need to be assessed for their effectiveness in terms of capturing the energy forecasts correctly and effectively. In this regard, researchers have employed different performance metrics. However, the most common and acceptable performance metrics employed in recent research include RMSE, nRMSE, MAE, and R^2. The authors decided to use all these performance metrics to assess and compare the efficiency of all the employed models. Table 1 contains the performance metrics (RMSE, nRMSE, MAE, and R-squared), formulas, corresponding ranges, and descriptions. In the subsequent sections of the paper, the values of performance metrics are reported for forecasts generated for different PV plants utilizing diverse training features.

RHE indicates the recorded hourly PV energy in kWh, these values are actual values recorded and transmitted by the dataloggers attached to the inverters. PHE indicates the ML and DL model predicted hourly PV energy; units are the same as RHE. These performance metrics help evaluate the accuracy and effectiveness of employed ML and DL models in predicting outcomes. Lower values of RMSE, nRMSE, and MAE indicate better performance, while higher values of R^2 i.e. closer to 1 indicate a better fit of the model.

3. Case study and results

3.1. Scope of case study

The developed methodological framework was verified by deploying to real-life application, which not only proved the suitability of the proposed system but also paves the way for PV organizations associated with the management and installation of PV systems. A notable solar PV organization in the Abruzzo region of southern Italy provided the data of more than 800 inverters which relate to 208 different PV plants. The historic energy production data was accessed from the MySQL database of the company, as the login credentials were provided by the management, hence ensuring the continuous data availability of almost all the plants and inverters separately. The developed framework can forecast the data for any of the plants and the user must add a plant code in the first stage of running the system after that all the successive stages run automatically and combine different databases and fetch data from web-based APIs as previously discussed in detail. Automation of the system not only reduces the workload on human resources but also ensures the resolution of any errors or man-made mistakes.

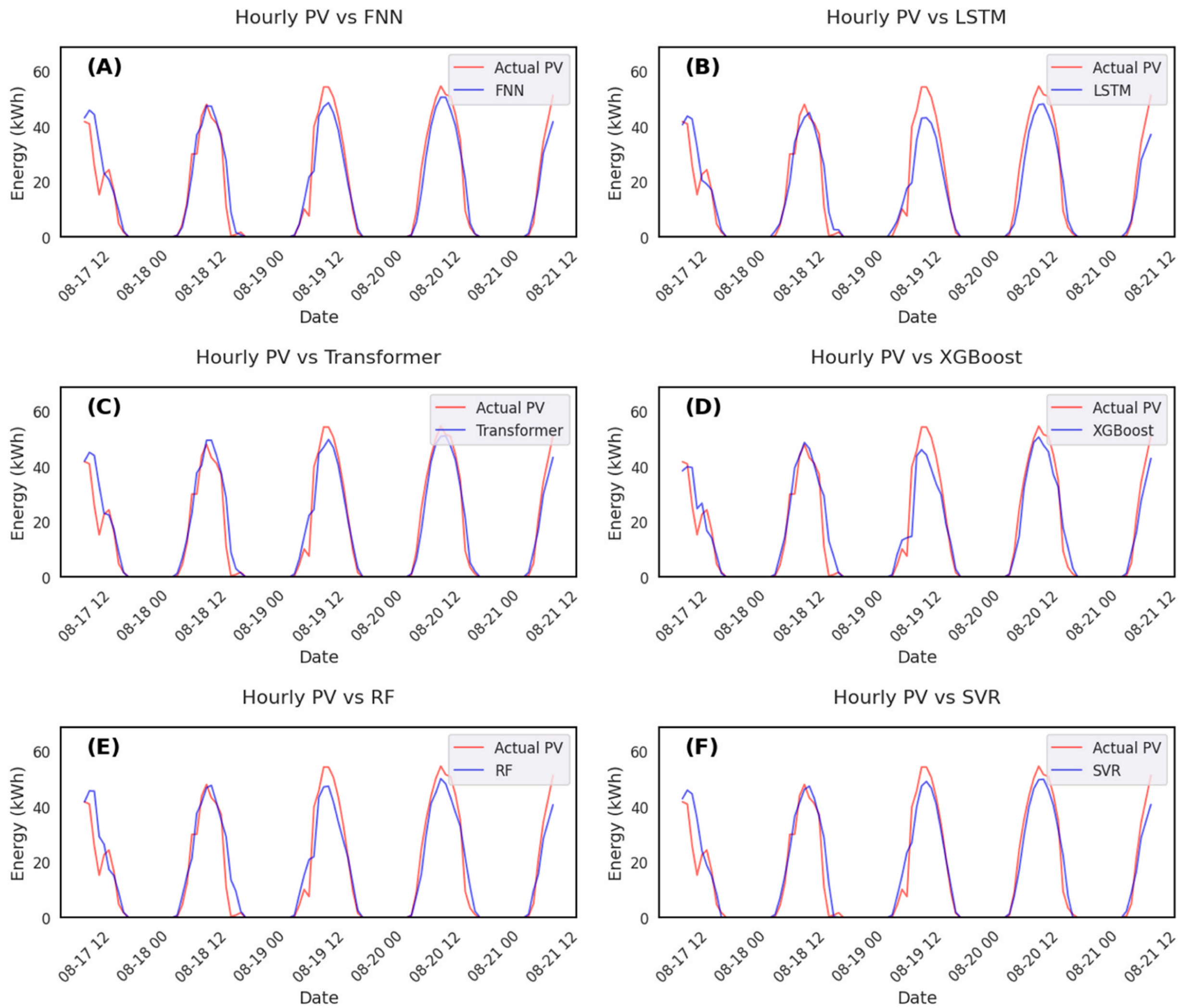


Fig. 9. Comparison between actual energy recorded and forecasted energy generated by ML and DL models for 91 kWp PV plant commissioned on 2013–05–14 (ADI).

3.2. Practical application

The proposed methodological framework was adopted for real-world practical application and a completely functional system was deployed. The monitoring team of the PV organization was taken on board and the forecasts were verified. This not only verified the effectiveness of the framework proposed in this research work but also was helpful for the PV organization for effective hourly forecasts. Although the system can forecast the hourly PV energy for the next 4 days i.e. 96 hourly values in kWh, this is impossible to report the forecast of all the plants, the system was deployed to forecast the energy for the selected PV plant and the automated system ensures the fetching and storing of the data from the predefines sources with certainty. After deploying the system for all the plants, the 4-day forecasts of 3 different PV plants are reported in this paper. The reported PV plants are small-scale, medium-scale, and a large-scale PV plant. Hence, ensuring the effectiveness of the system for all types of PV plants. Plant characteristics of reported PV plants can be seen in Table 2.

3.3. Performance assessment

The Performance of the proposed systems was evaluated using the globally accepted and widely used performance metrics. In the initial phase, the correlation matrix between the features is reported for all three chosen plants. The correlation matrix helps to understand how much a variable independently shares in the forecast of the dependent variable i.e. Hourly forecasts of PV energy. Performance assessment is reported in three phases in the following sections, firstly we have reported the correlation matrix. The second phase highlights the performance of the ML and DL models employing only NWM data and historic plant data. In the third and last phase, the performance of all the models using all the available data i.e. PVGIS, NWM, and historic energy production data is reported. The performance comparison between NWM only and PVGIS + NWM creates an opportunity for the readers to visualize the importance and the effect of both datasets on the efficiency of the models.

3.3.1. Mutual information between features

The mutual correlation matrix indicates the information shared by training features used for forecasting the target variable (Jebli et al., 2021). In Figs. 3, 4, and 5 the mutual information matrix of all three selected PV plants is reported. These aid in the visualization of the features that contribute most to the prediction of the target variable. In all three selected PV plants, it was observed that the feature “Pressure”, which was originally acquired from numerical weather model data is the least contributing feature and has the lowest correlation among all the training features. It was observed that the training features extracted from the PVGIS data contributed more and had a better correlation with

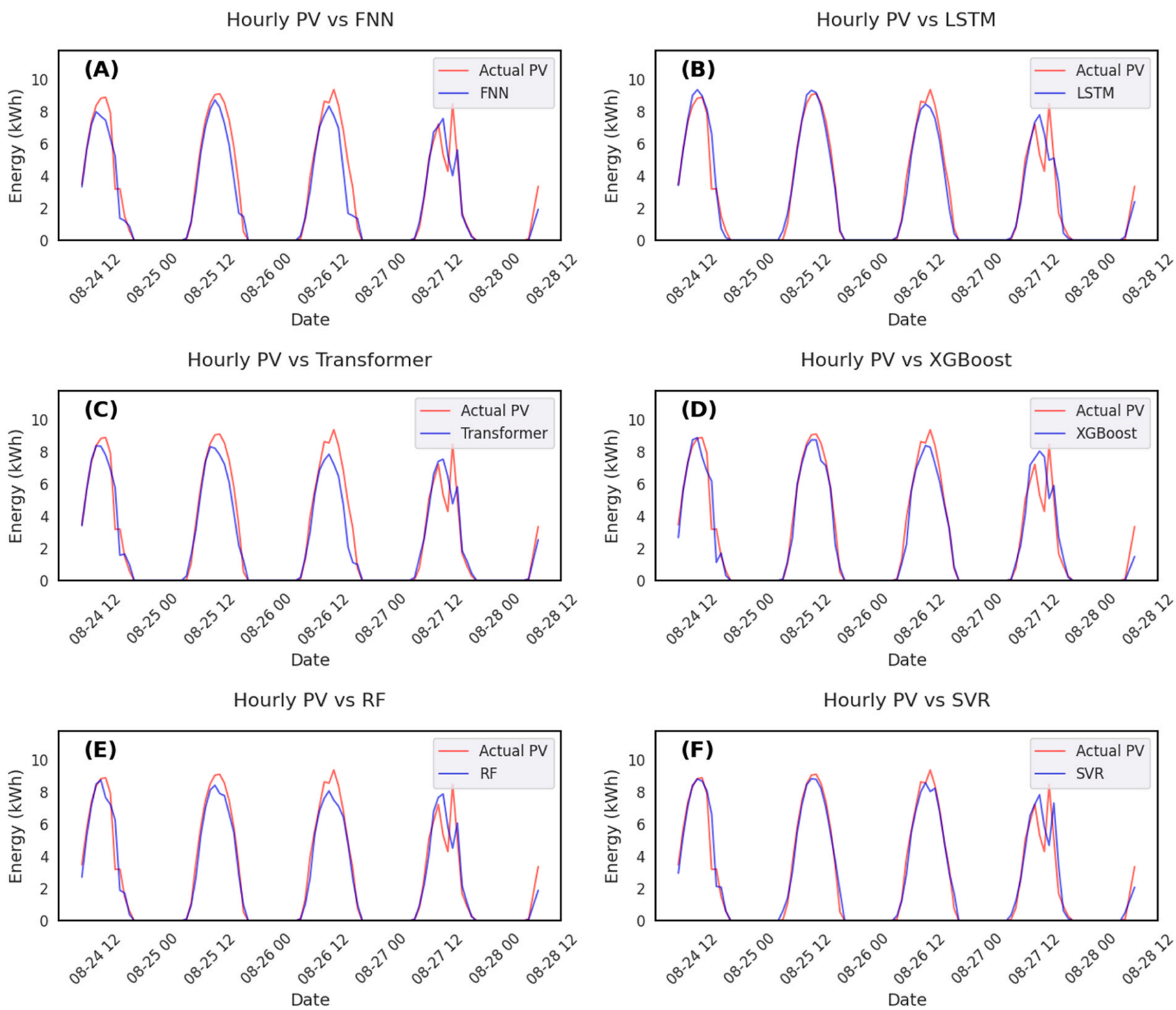


Fig. 10. Comparison between actual energy recorded and forecasted energy generated by ML and DL models for 17.07 kWp PV plant commissioned on 2009–08–04 (BDS).

the target variable as compared to that of the NWM data. Similarly, the correlation indicated that the H_sun (Sun height, degree) has the highest value of correlation in all three reported cases, followed by Gb(i) i.e. Beam (direct) irradiance on the inclined plane (plane of the array) (W/m^2) and Gd(i) i.e. Diffuse irradiance on the inclined plane (plane of the array) (W/m^2).

3.3.2. NWM and plant data

Selected ML and DL models need to be trained with a certain set of training features, for the first phase of the analysis the ML and DL models were trained with only data from numerical weather models and hourly PV energy production as target variable. During the first phase of the forecasts, the training features used were 'Pressure', 'Humidity', 'Wind Speed', 'Clouds', 'Temp Max', and 'Temp Min'. All these training features were extracted from the NWM data, and the target variable was 'Hourly PV Production' in kWh. In Figs. 6, 7, and 8 the comparison between the forecasted PV energy production revealed by the ML and DL models along with the actual hourly PV production is reported. In the case of 386 kWp CRA plant RF model outperformed all the ML and DL models and depicted a R^2 of 0.90. A similar trend was observed in the case of BDS plant where RF model depicted the second-best performance as compared to all other models with a R^2 of 0.90, LSTM showed a R^2 of 0.91. For the 91 kWp, ADI PV plant, it can be interpreted that the transformer model was able to capture the energy prediction with the highest accuracy as compared to all the other models employed. The performance of SVR model was somehow less accurate as compared to other models in two cases.

3.3.3. PVGIS NWM and plant data

In the previous phase ML and DL models were deployed to forecast the hourly PV production using only NWM data, in this phase along with the NWM data and plant data, the data from PVGIS was also incorporated into the training process of the models. This step aids in the in-depth analysis to evaluate the efficiency of these models. The meteorological parameters acquired from the PVGIS API were stored and utilized for each PV plant location. The features selected for the training of the models consisted of 'Pressure', 'Humidity', 'Wind Speed', 'Clouds', 'Gb(i)', 'Gd(i)', 'H_sun', 'T2m', 'WS10m', 'Temp Max', and 'Temp Min', with the target variable 'Hourly PV Production' in kWh. The comparison between 4-day hourly PV energy forecast and the actual PV energy recorded for each PV plant can be seen in Figs. 9, 10, and 11, also the values of the performance metrics can be seen in Table 3. The results revealed that among all the models the SVR model benefited the most from the additional data and the accuracy of the said model was significantly improved in all three reported cases, this can be referred to as the SVR model's inherent capabilities efficiently handling the high dimensional and complex datasets. The SVR model is reported to find optimal hyperplanes in high dimensional spaces, which paves the way

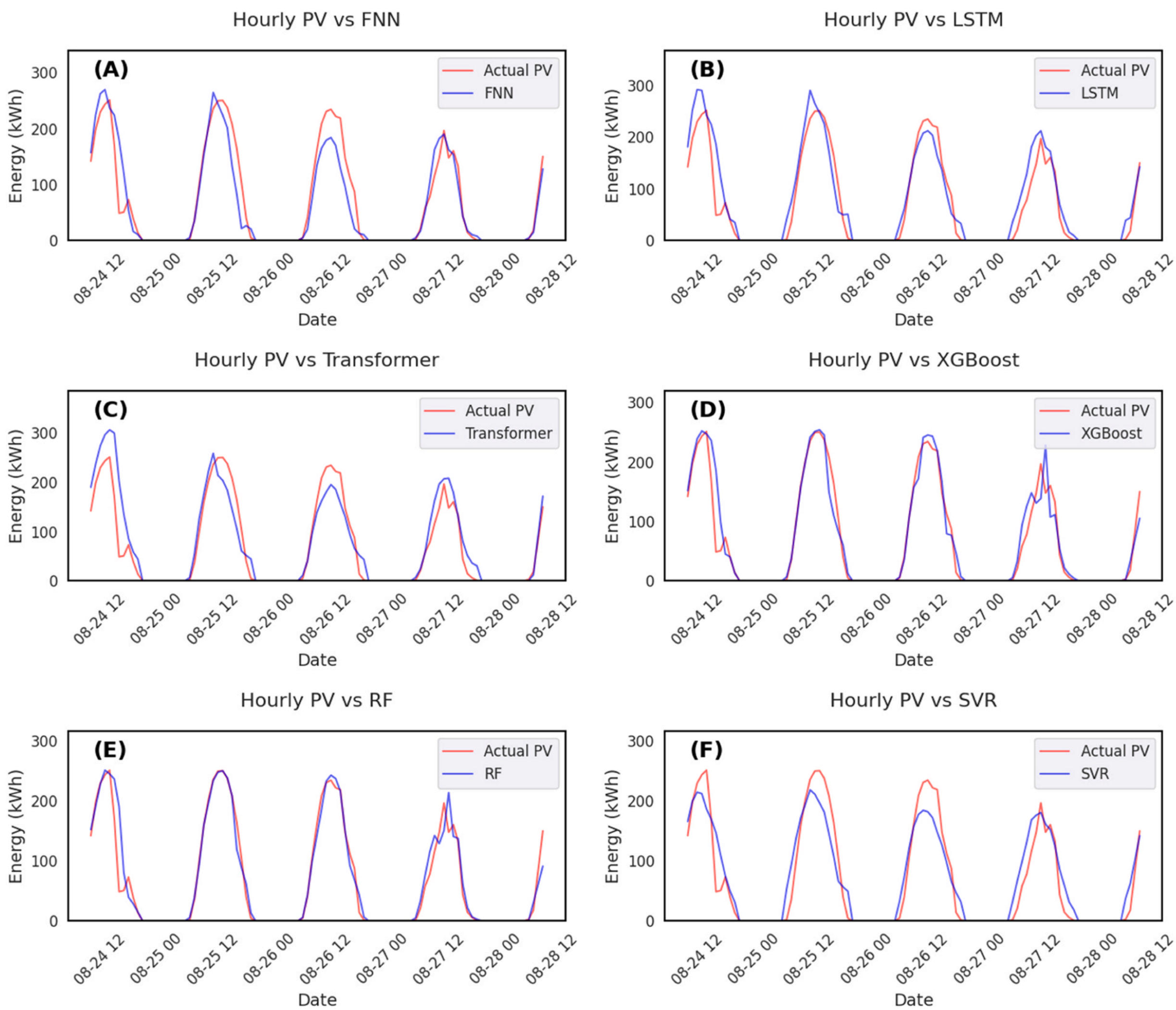


Fig. 11. Comparison between actual energy recorded and forecasted energy generated by ML and DL models for 400 kWp PV plant commissioned on 2021–09–26 (CRA).

for better generalization when additional datasets are provided. The PVGIS data enhanced the model's ability to capture the trends and patterns in the datasets, resulting in increased performance of forecasting capabilities.

All the other models showed almost the same performance trends as we observed in the case of the NWM data only, however, in some cases the accuracy of the model was reduced as compared to the NWM data only case. However, there was no significant trend reported that could be highlighted, for instance, the FNN showed decreased performance for the 2 plants, on the other hand in the third case the performance improved with the addition of new datasets from PVGIS. Overview of the comparative accuracy analysis of all the models concerning perfect predictions or recorded values are depicted in Fig. 12.

The scatter plots presented in Fig. 12 provide an opportunity for readers to compare the accuracy of all ML and DL models in relation to perfect predictions or recorded values.

3.3.4. Hybrid models

In the last phase of the analysis, hybrid models were utilized for hourly PV energy forecasting, three hybrid models depicted suitable performance and the comparison of the forecasts generated by different models for different plants is depicted in Figs. 13, 14, and 15.

In case of BDS plant (17.07 kWp), when using NWM data, the RF LSTM, RF MLP, and RF GRU models showed similar performance trends, the value of RMSE was found between 1.06–1.09 kWh in all three hybrid models. R^2 values of 0.89 for all three models, indicates a high level of accuracy. Same models were utilized for forecasts using training features from both NWM and PVGIS, performance revealed that all models experience an increase in RMSE. The RF LSTM model showed the best performance with an RMSE of 1.26 kWh and an R^2 of 0.85.

The second case reported is of the ADI plant (91 kWp), all the models showed a varying range of performance. When models were trained only using NWM data, the RF LSTM model has the lowest RMSE of 6.64 kWh and the highest R^2 value of 0.88. This trend indicates a strong predictive accuracy of the hybrid model. In contrast, when both datasets are combined for training the models, there is a slight increase in RMSE for all the models. However, the RF LSTM still showcased the best performance with RMSE of 7.24 kWh and an R^2 of 0.85.

For the CRA plant (386 kWp), which is the largest among the three plants reported in this researchwork, the value of RMSE was higher as anticipated in case of large scale plants. When only NWM data was utilized, RF LSTM model performed superior as compared to other models, achieving an RMSE of 26.13 kWh and an R^2 value of 0.91. However, with the addition of PVGIS data, all models depicted a decrease in forecasting performance, specifically the RF MLP was most prone to inaccuracies.

Overall, the hybrid models generally perform well, with RF LSTM consistently showing strong results, especially when using NWM data

Table 3
Overview of the forecast performance for all models across all three plants.

Model	RMSE kWh	nRMSE	MAE kWh	R-squared (R^2)
	BDS NWM (17.07 kWp)			
FNN	1.08	0.12	0.75	0.89
LSTM	0.98	0.11	0.61	0.91
Transformer	1.05	0.11	0.76	0.89
XGBoost	1.13	0.12	0.71	0.88
RF	1.05	0.11	0.66	0.90
SVR	1.07	0.11	0.65	0.89
	BDS PVGIS NWM (17.07 kWp)			
FNN	1.22	0.13	0.83	0.85
LSTM	0.96	0.10	0.57	0.91
Transformer	1.18	0.13	0.83	0.86
XGBoost	1.10	0.12	0.69	0.88
RF	1.03	0.11	0.69	0.90
SVR	1.00	0.11	0.60	0.90
	ADI NWM (91 kWp)			
FNN	6.98	0.13	4.93	0.86
LSTM	8.46	0.16	6.30	0.80
Transformer	6.16	0.11	3.97	0.89
XGBoost	6.80	0.13	4.94	0.87
RF	7.35	0.14	5.61	0.85
SVR	9.14	0.17	7.20	0.76
	ADI PVGIS NWM (91 kWp)			
FNN	6.51	0.12	4.55	0.88
LSTM	7.41	0.14	5.63	0.85
Transformer	6.35	0.12	4.38	0.89
XGBoost	6.97	0.13	5.13	0.86
RF	7.30	0.13	5.52	0.85
SVR	6.78	0.13	4.88	0.87
	CRA NWM (386 kWp)			
FNN	30.71	0.12	21.36	0.87
LSTM	38.35	0.15	31.56	0.80
Transformer	46.91	0.19	28.67	0.70
XGBoost	27.91	0.11	18.88	0.89
RF	26.63	0.11	17.02	0.90
SVR	52.11	0.21	45.88	0.63
	CRA PVGIS NWM (386 kWp)			
FNN	39.53	0.16	28.10	0.79
LSTM	38.62	0.15	31.19	0.80
Transformer	44.57	0.18	34.56	0.73
XGBoost	32.59	0.13	20.57	0.86
RF	28.83	0.12	16.77	0.89
SVR	38.82	0.16	32.98	0.80

alone. The addition of PVGIS data with NWM data did not enhance the efficiency of the forecasts and in most cases the performance of the models was adversely affected. The details of performance metrics of hybrid models can be seen in Table 4.

The scatter plots presented in Fig. 16 offer readers a chance to compare the effectiveness of each hybrid model's forecast against the recorded hourly PV production, utilizing both types of datsets.

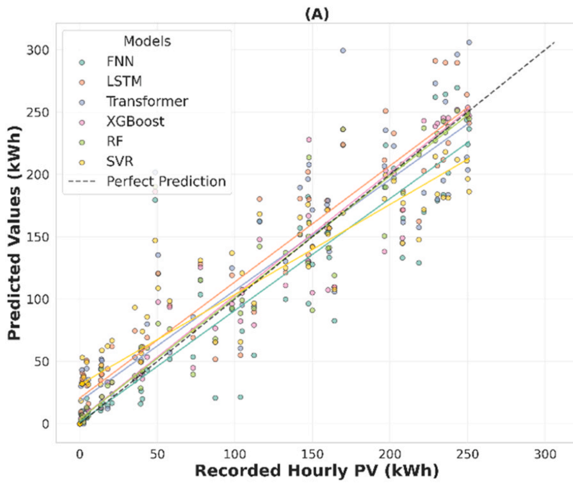
4. Discussion and conclusions

This research work was designed to exploit the numerical weather model data and the PVGIS data for the forecasting of solar energy resources using state-of-the-art machine learning and deep learning techniques. The research work started with the streamlining of the data acquisition and collection setup. The setup was designed such that the applicability of the system can be assessed for real-world conditions and the efficiency could be checked for the forecasting in real-world conditions, contradictory to most of the forecasting techniques exploited in the literature this research work mainly focused on the latest data and the acceptability of the system in the PV organizations. The developed methodology was applied to forecast the PV energy of more than 800 inverters, connected with more than 200 PV plants at various locations. In this research work, the performance of two datasets is assessed in terms of hourly PV forecasting and reported alongside various ML, DL, and hybrid models.

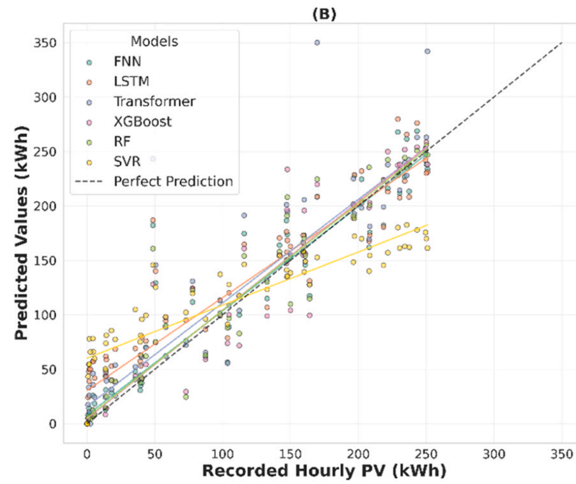
Analysis revealed that most of the models were able to capture effective forecasts when only data from NWM was used for training the models, in some cases the 0.91 R^2 was achieved using the NWM data only. ML models specifically RF models depicted optimal performance as compared to DL models when adopting the NWM data only. In all three reported cases the RF depicted the R^2 of 0.90, 0.90. and 0.85 respectively and nRMSE of 0.11, 0.11, and 0.14. on the other hand, the FNN model showed the best performance among the exploited DL models achieving an R^2 of 0.87, 0.89, and 0.86 and nRMSE of 0.12, 0.12, and 0.13, respectively.

In addition to NWM data, an additional analysis was conducted, in which the forecasts were generated using the NWM and historic PVGIS data. It was reported that there was no significant improvement in the performance of the models. For instance, in the case of the 385 kWp plant, the RF model depicted optimal performance in both datasets, where for NWM data only the metrics were RMSE 26.63 kWh, MAE 17.02 kWh, and R^2 0.90, similarly for NWM + PVGIS data the metrics were RMSE 28.83 kWh, MAE 16.77 kWh, and R^2 0.89.

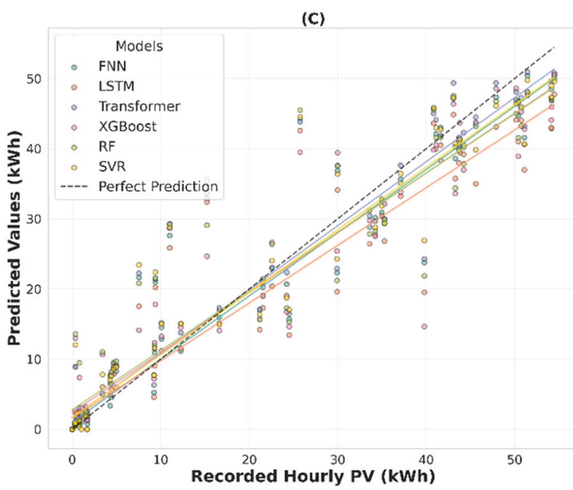
A notable trend was observed in the case of the SVR model, as the performance was significantly improved with the inclusion of PVGIS data the SVR model performed poorest in all three reported cases. SVR depicted MAE and R^2 values of 45.88 kWh and 0.63 for 386 kWp plant, 0.65 kWh and 0.89 for 17.07 kWp plant, and 7.20 kWh and 0.76 for 91 kWp plant respectively. All the metrics improved with the inclusion of PVGIS data for training the models and the model depicted values of MAE and R^2 of 32.98 kWh and 0.80 for 386 kWp plant, 0.60 kWh and 0.90 for 17.07 kWp plant, and 4.88 kWh and 0.87 for 91 kWp plant



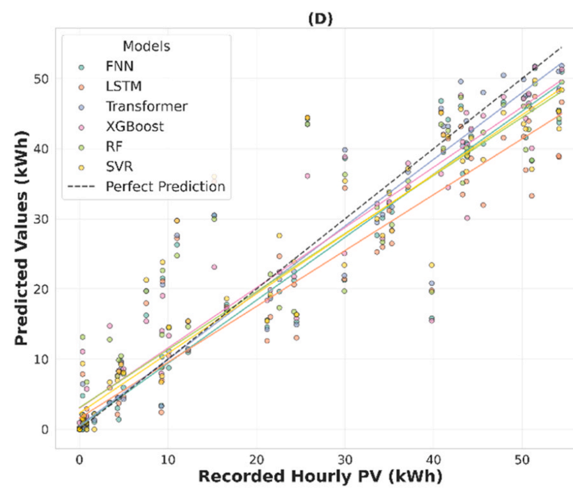
CRA PVGIS and NWM Data



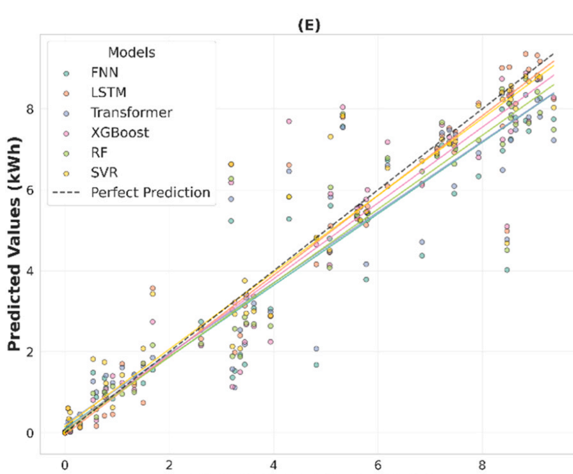
CRA NWM Data



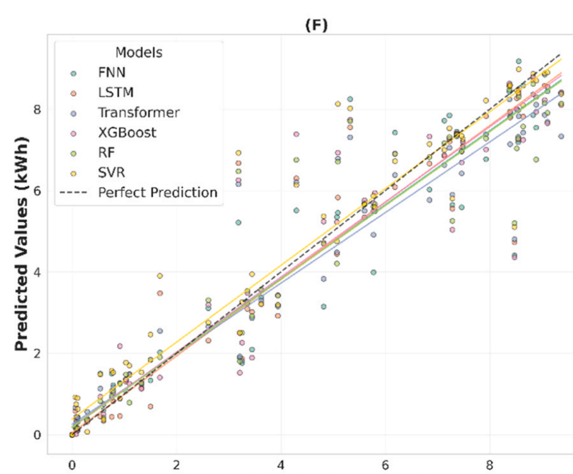
ADI PVGIS and NWM Data



ADI NWM Data



BDS PVGIS and NWM Data



BDS NWM Data

Fig. 12. Overview of the comparative accuracy analysis of all the models concerning perfect predictions or recorded values.

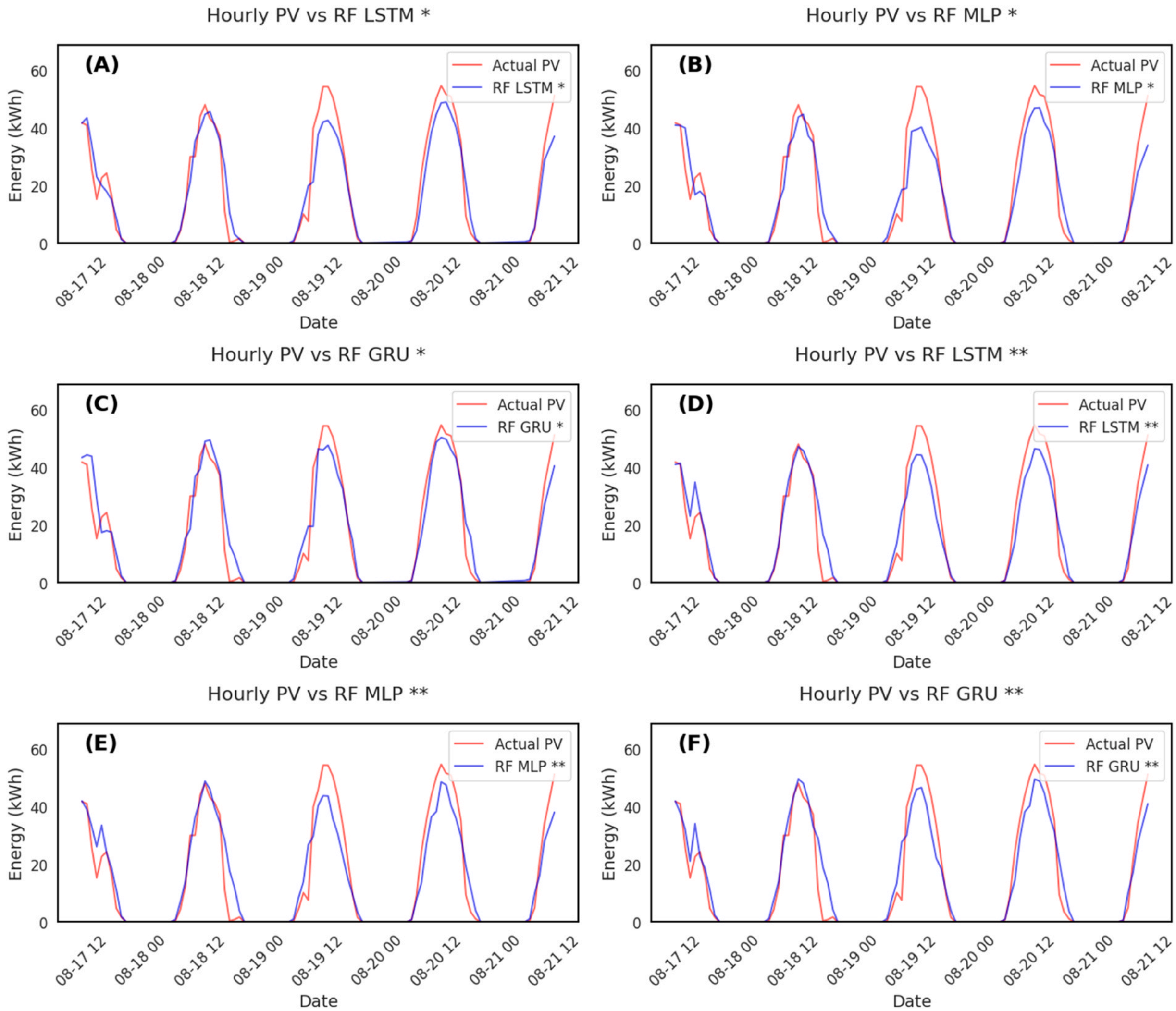


Fig. 13. Comparison between forecasts generated by hybrid models and actual energy recorded for the hybrid models in the case of plant ADI (91 kWp). * Utilizing only NWM data, * * Utilizing both NWM and PVGIS datasets.

respectively.

To further investigate the effectivity, 3 hybrid models were utilized, and a combination of ML and DL model was created. This analysis revealed that all the hybrid models were able to effectively forecasts the energy utilizing only NWM data and addition of PVGIS data had adverse effects on the forecasting capabilities of the models. The RF LSTM model showcased best performance and, in most cases, depicted a R^2 value higher than 0.88.

In conclusion, the ML, DL and hybrid models showed great potential for the effective hourly PV forecasting of solar energy resources, specifically the RF model, which was able to produce effective hourly forecasts even in the presence of minimal data availability, an R^2 of 0.91 was achieved for the hourly forecasts of next 4 days. Among the DL models the FNN and LSTM models showed great promise and good performance metrics were recorded in almost all the cases, and both models were able to capture the hourly PV energy with a high accuracy, depiction of overall R^2 with more than 0.85 in most cases. Among the hybrid models RF LSTM was the top-performing in all the reported cases, particularly when using NWM data alone, but even with the inclusion of PVGIS data, RF LSTM maintained its strong performance.

The main contributions of this research work are

- The study provides a comprehensive analysis of the performance of various ML, DL, and hybrid models for hourly PV energy forecasting, considering both NWM-only and combined NWM + PVGIS data. The developed methodology is applied to over 800 inverters across 200 PV plants in different geographical locations, which reflects real-world conditions and operational setups.
- The study provides valuable insights into the impact of including various datasets on the forecasting models, creating an opportunity for the future researchers and PV managers to select the most effective datasets. It was revealed during the analysis that addition of PVGIS dataset did not always result in enhancing the forecasting capacity of the models.
- The research revealed that RF models outperformed other ML models like SVR and XGBoost and DL model FNN. When using NWM data alone, RF achieved an R^2 value of 0.91. The FNN and LSTM models both showed promising results, but the LSTM outperformed all the other models in overall performance. The performance metrics of this study closely relate to as reported by (Nastić et al., 2024), where highest R^2 scores for three different PV plants were recorded 0.944, 0.974, and 0.973 for the 24 hours hourly forecasts. However, in our proposed and reported cases the hourly forecasts are generated for 4 days.
- This study also highlights the practical applicability of the developed models for real-world PV organizations, showcasing how the models can be deployed for accurate short-term forecasting, even with

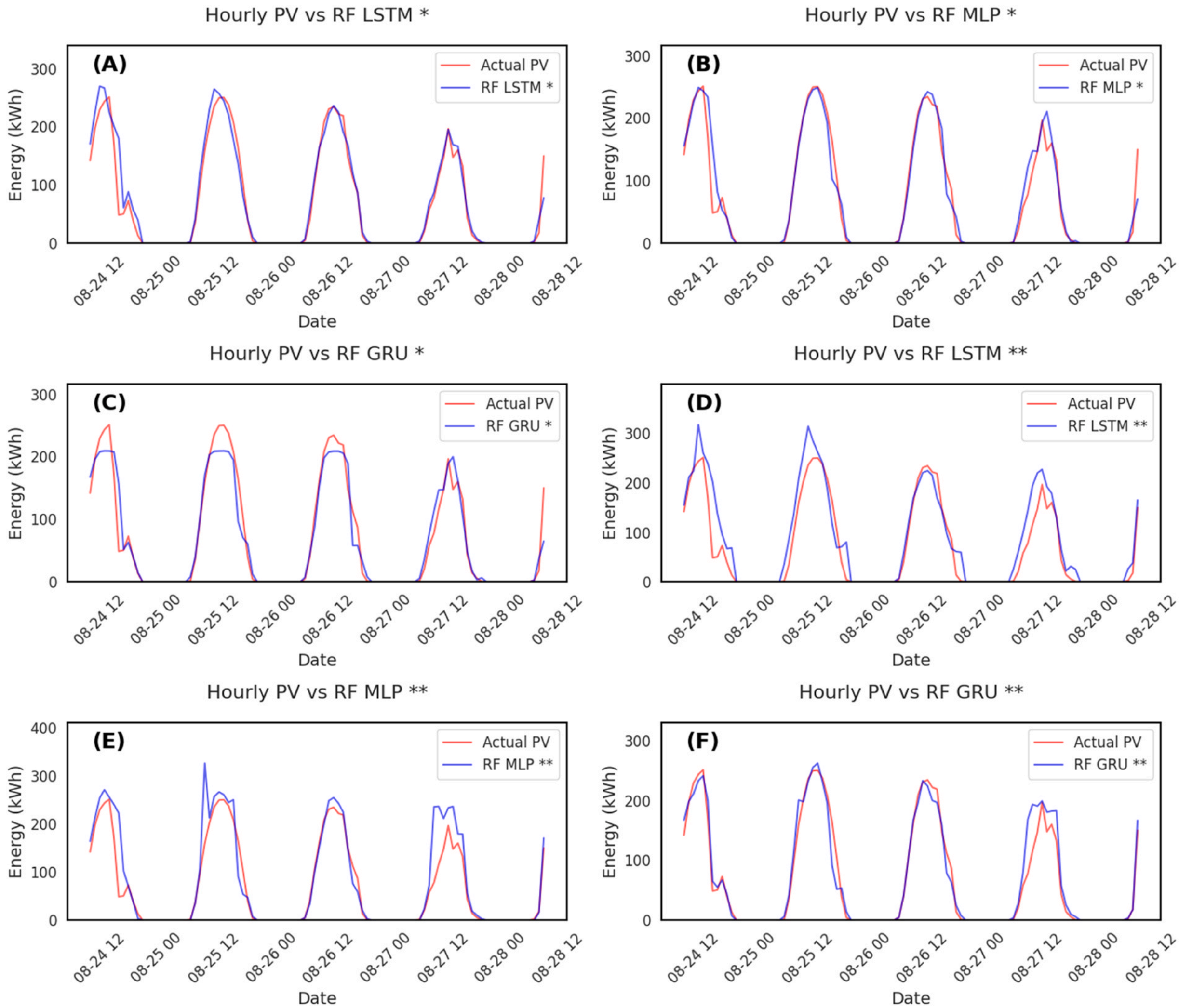


Fig. 14. Comparison between forecasts generated by hybrid models and actual energy recorded for the hybrid models in the case of plant CRA (386 kWp). * Utilizing only NWM data, * * Utilizing both NWM and PVGIS datasets.

minimal data availability using economically viable real time datasets.

- The development and evaluation of a hybrid approach combining machine learning (ML) models with deep learning models, such as LSTM, GRU, and MLP, demonstrated effective forecasting of PV energy production using both NWM and PVGIS data. Notably, when using only NWM data, the hybrid models, particularly the RF-LSTM model, exhibited robust performance across all cases. Similar trend was reported by (Karijadi and Chou, 2022), when utilized RF LSTM hybrid model they concluded that hybrid RF-LSTM based on CEEMDAN for building energy prediction outperformed other prediction methods such as Linear Regression, SVR, RF, LSTM and ANN. This highlights the significant advantage that hybrid approaches can offer in enhancing the forecasting accuracy of PV energy generation.

Still, the possibility of polishing and further defining the forecasts lies in exploiting the different databases and state-of-the-art DL techniques, as the AI is flourishing and widening its wings days by day, researchers need to be proactive in exploring the applicability of these modern techniques for their respective fields.

Further research opportunities lie in

- Exploiting the modern ML and DL models using the same datasets used in this research work. Authors have only adopted the 6 models in total; however, many ML and DL models have showed great promise in the recent past for different scientific forecast, hence the applicability needs to be assessed specifically for PV energy forecasting.
- Applicability evaluation of the proposed framework in different parts of the world, hence creating a opportunity for the global effective assessment of our proposed methodological framework.
- Incorporation of other NWM datasets, satellite based meteorological training features, and ground based meteorological training features for further enhancing the forecasting accuracy.
- Application of the proposed methodological framework at a different temporal resolution and forecast horizon, this may create the opportunity for applicability of the forecast in the different industries, for instance per minute forecast for next couple of hours may produce more potential for industries and make a foreground for effective decision making.

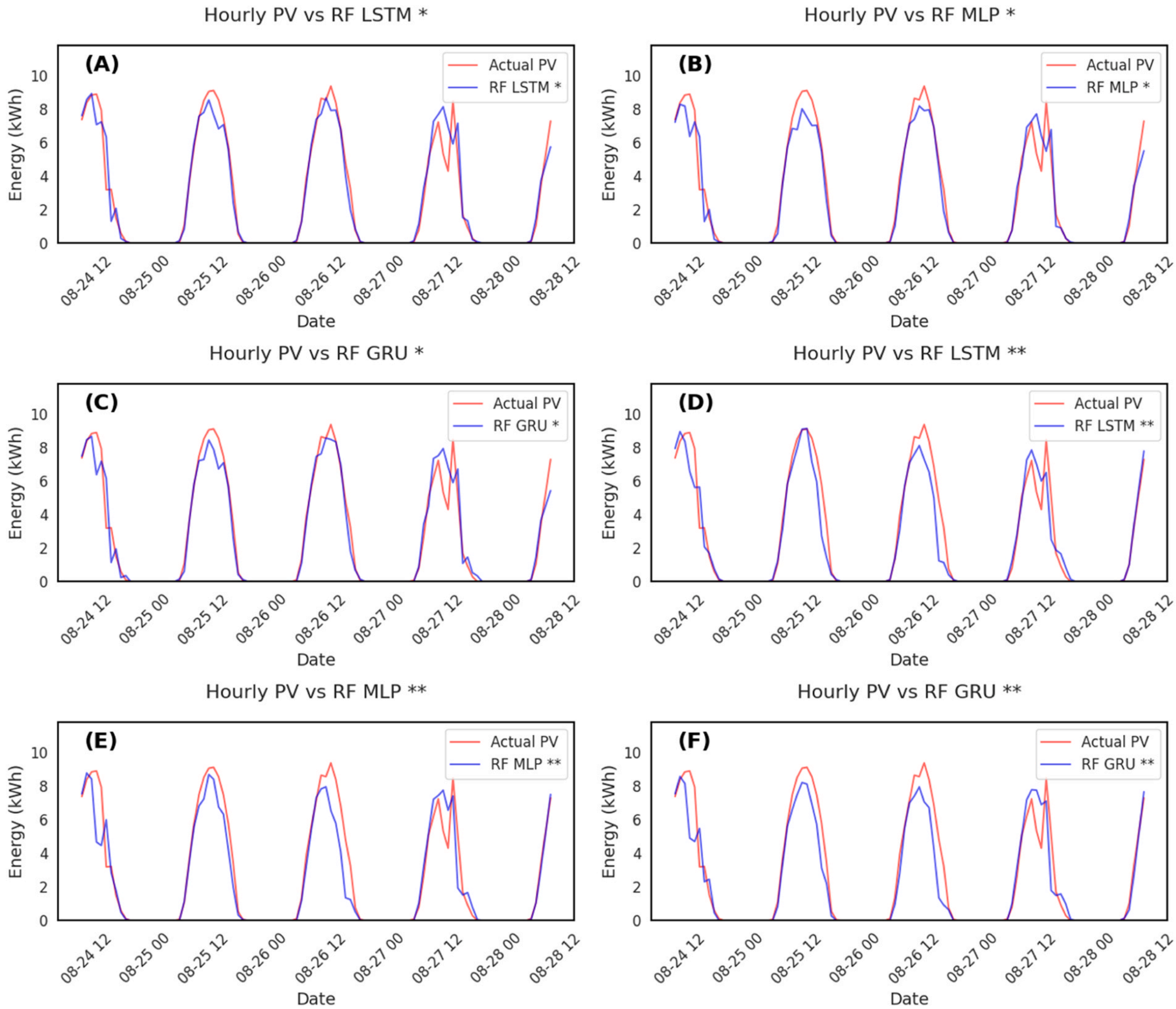


Fig. 15. Comparison between forecasts generated by hybrid models and actual energy recorded for the hybrid models in the case of plant BDS (17.07 kWp). * Utilizing only NWM data, * * Utilizing both NWM and PVGIS datasets.

Table 4
Overview of the forecast performance for all hybrid models across all three plants.

Model	RMSE kWh	nRMSE	MAE kWh	R-squared (R^2)
	BDS NWM (17.07 kWp)			
RF LSTM	1.06	0.11	0.69	0.89
RF MLP	1.09	0.12	0.73	0.89
RF GRU	1.06	0.11	0.71	0.89
	BDS PVGIS + NWM (17.07 kWp)			
RF LSTM	1.26	0.14	0.86	0.85
RF MLP	1.44	0.15	0.94	0.80
RF GRU	1.46	0.16	1.02	0.80
	ADI NWM (91 kWp)			
RF LSTM	6.64	0.12	4.97	0.88
RF MLP	7.87	0.15	6.14	0.83
RF GRU	7.06	0.13	5.26	0.86
	ADI PVGIS + NWM (91 kWp)			
RF LSTM	7.24	0.13	5.58	0.85
RF MLP	7.95	0.15	6.24	0.82
RF GRU	7.42	0.14	5.74	0.85
	CRA NWM (386 kWp)			
RF LSTM	26.13	0.10	16.73	0.91
RF MLP	27.23	0.11	16.71	0.90
RF GRU	30.02	0.12	20.86	0.88
	CRA PVGIS + NWM (386 kWp)			
RF LSTM	44.45	0.18	34.41	0.73
RF MLP	49.62	0.20	28.30	0.67
RF GRU	25.93	0.10	17.15	0.91

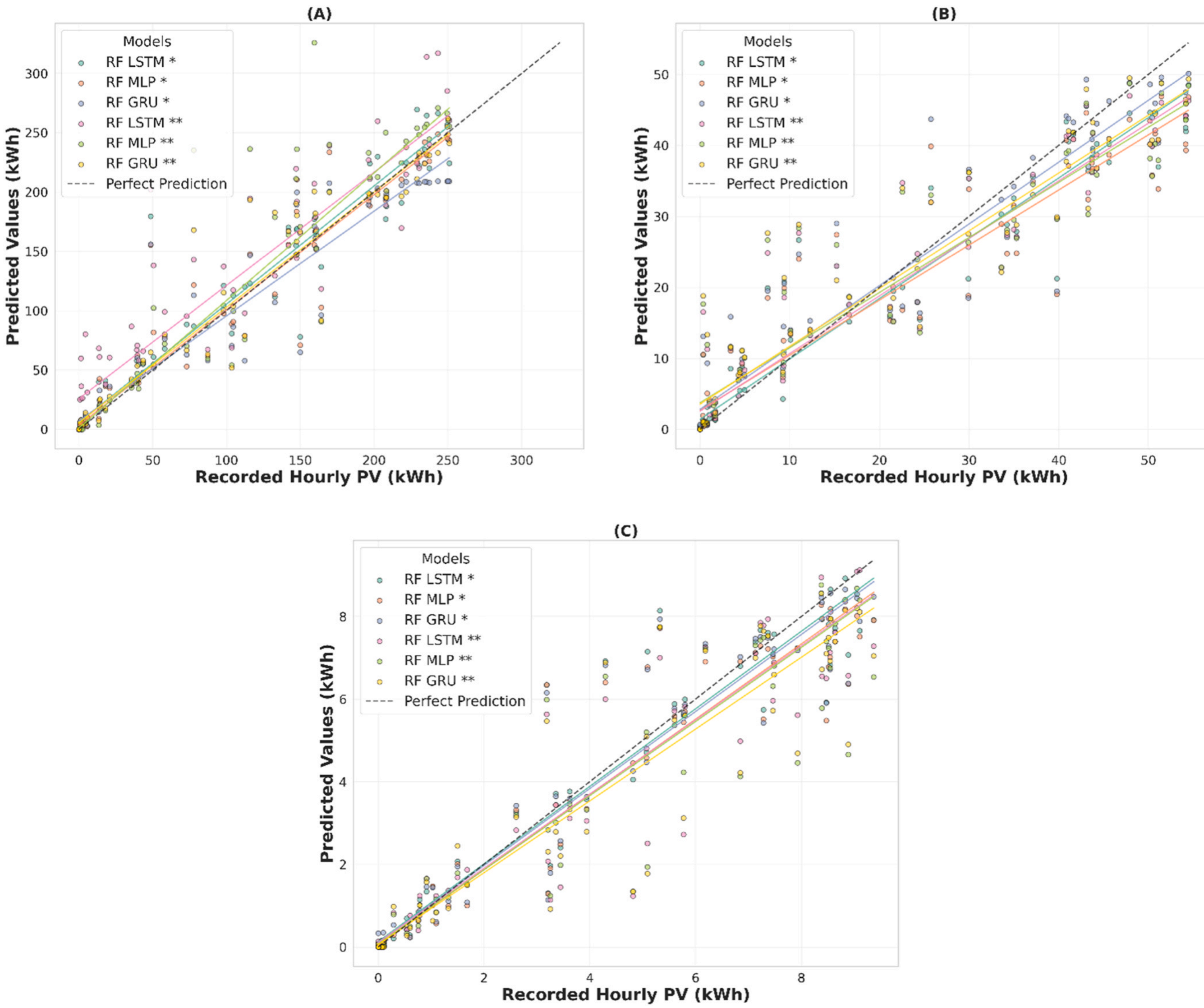


Fig. 16. Overview of the comparative accuracy analysis of all models with respect to perfect predictions or recorded values: (A) Plant CRA 386 kWp, (B) Plant ADI 91 kWp, and (C) Plant BDS 17.07 kWp. * Utilizing only NWM data, * * Utilizing both NWM and PVGIS datasets.

Acknowledgment

The authors express their gratitude to Solis SpA for their invaluable cooperation in this study. This encompasses providing technical and production data for BIPVs and BAPVs, supplying the necessary computer processing system and required software, as well as the contributions of the monitoring team and the company’s technical head in validating the research findings. The authors would also like to express their gratitude to the developers of OpenWeather for providing the meteorological data. Finally, authors would like to express our sincere appreciation to PVGIS (Photovoltaic Geographical Information System) and European Commission’s Joint Research Centre for providing data resources.

CRediT authorship contribution statement

Cucchiella Federica: Validation, Project administration, Conceptualization. **Ehtsham Muhammad:** Writing - Review & Editing, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rotilio Marianna:** Validation, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

References

Abdel-Basset, M., Hawash, H., Chakrabortty, R.K., Ryan, M., 2021. PV-Net: an innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production. J. Clean. Prod. 303, 127037.
Ahmed, F., Xiang, X., Jiang, C., Xiang, G., Yang, S., 2023. Survey on traditional and AI based estimation techniques for hydrodynamic coefficients of autonomous underwater vehicle. Ocean Eng. 268, 113300.
Akhter, M.N., Mekhilef, S., Mokhlis, H., Ali, R., Usama, M., Muhammad, M.A., Khairuddin, A.S.M., 2022. A hybrid deep learning method for an hour ahead power output forecasting of three different photovoltaic systems. Appl. Energy 307, 118185.
Akintuyi, O.B., 2024. AI in agriculture: a comparative review of developments in the USA and Africa. Res. J. Sci. Eng. 10, 60–70.

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J., Antonanzas-Torres, F., 2016b. Review of photovoltaic power forecasting. Sol. Energy 136, 78–111.
Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J., Antonanzas-Torres, F., 2016a. Review of photovoltaic power forecasting. Sol. Energy 136, 78–111.
Asghar, R., Fulginei, F.R., Quercio, M., Ahmad, M., Abusara, M.: Application of Deep Learning Algorithms for BIPV Power Forecasting in Italy. In: 2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe). pp. 1–6. IEEE (2024).
Aslam, S., Herodotou, H., Mohsin, S.M., Javaid, N., Ashraf, N., Aslam, S., 2021. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. Renew. Sustain. Energy Rev. 144, 110992.
Baidoo-Anu, D., Ansah, L.O., 2023. Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. J. AI 7, 52–62.
Basher, M.K., Nur-E-Alam, M., Rahman, M.M., Alameh, K., Hinckley, S., 2023. Aesthetically appealing building integrated photovoltaic systems for net-zero energy buildings. Curr. Status, Chall., Future Dev. —A Rev. Build. 13, 863.
Brester, C., Kallio-Myers, V., Lindfors, A.V., Kolehmainen, M., Niska, H., 2023. Evaluating neural network models in site-specific solar PV forecasting using numerical weather prediction data and weather observations. Renew. Energy 207, 266–274.
Cantillo-Luna, S., Moreno-Chuquen, R., Celeita, D., Anders, G., 2023. Deep and machine learning models to forecast photovoltaic power generation. Energ. (Basel) 16, 4097.
Celik, M., Inik, O., 2024. Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor Multi-Classification. Expert Syst. Appl. 238, 122159.
Chen, Z., Ma, M., Li, T., Wang, H., Li, C., 2023. Long sequence time-series forecasting with deep learning: a survey. Inf. Fusion 97, 101819.
Choudhary, P., Srivastava, R.K., 2019. Sustainability perspectives-a review for solar photovoltaic trends and growth opportunities. J. Clean. Prod. 227, 589–612.
Das, U.K., Tey, K.S., Seyedmahmoudian, M., Mekhilef, S., Idris, M.Y.I., Van Deventer, W., Horan, B., Stojcevski, A., 2018. Forecasting of photovoltaic power generation and model optimization: a review. Renew. Sustain. Energy Rev. 81, 912–928.
Didavi, K.B.A., Agbokpanzo, R.G., Agbomahena, B.M., 2024. LSTM and XGBoost models for 24-hour ahead forecast of PV power from direct irradiation. Renew. Energy Res. Appl. 5, 229–241.
Du Plessis, A.A., Strauss, J.M., Rix, A.J., 2021. Short-term solar power forecasting: investigating the ability of deep learning models to capture low-level utility-scale Photovoltaic system behaviour. Appl. Energy 285, 116395.
Ehtsham, M., Rotilio, M., 2024. Effect of 2023 European Heatwave on Photovoltaic Energy Generation: ACase Study of Central and Southern Italy. In: Renewable Energy Resources and Conservation. Springer, pp. 33–40.
Fadoul, F.F., Hassan, A.A., Çağlar, R., 2023. Assessing the feasibility of integrating renewable energy: decision tree analysis for parameter evaluation and LSTM forecasting for solar and wind power generation in a campus microgrid. IEEE Access 11, 124690–124708.
Ghosh, A., 2020. Potential of building integrated and attached/applied photovoltaic (BIPV/BAPV) for adaptive less energy-hungry building's skin: a comprehensive review. J. Clean. Prod. 276, 123343.
Gupta, P., Singh, R., 2021. PV power forecasting based on data-driven models: a review. Int. J. Sustain. Eng. 14, 1733–1755.
Hernández-Callejo, L., Gallardo-Saavedra, S., Alonso-Gómez, V., 2019. A review of photovoltaic systems: design, operation and maintenance. Sol. Energy 188, 426–440.
Hewamalage, H., Bergmeir, C., Bandara, K., 2021. Recurrent neural networks for time series forecasting: current status and future directions. Int J. Forecast 37, 388–427.
Jareño, F., Yousaf, I., 2023. Artificial intelligence-based tokens: Fresh evidence of connectedness with artificial intelligence-based equities. International Review of Financial. Analysis 89, 102826.
Javaid, M., Haleem, A., Khan, I.H., Suman, R., 2023. Understanding the potential applications of Artificial Intelligence in Agriculture Sector. Adv. Agrochem. 2, 15–30.
Jebli, I., Belouadha, F.-Z., Kabbaj, M.I., Tilioua, A., 2021. Prediction of solar energy guided by pearson correlation using machine learning. Energy 224, 120109.
Jurczyński, D.: Application of satellite technology in air pollution monitoring analysis of carbon monoxide measurements, pm10, pm2. 5 levels and weather index.
Karijadi, I., Chou, S.-Y., 2022. A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction. Energy Build. 259, 111908.
Khelifi, R., Guermoui, M., Rabehi, A., Taallah, A., Zoukel, A., Ghoneim, S.S.M., Bajaj, M., AboRas, K.M., Zaitsev, I., 2023. Short-term PV power forecasting using a hybrid TVF-EMD-ELM strategy. Int. Trans. Electr. Energy Syst. 2023, 1–14.
Konstantinou, M., Peratikou, S., Charalambides, A.G., 2021. Solar photovoltaic forecasting of power output using LSTM networks. Atmosphere 12, 124.
Kumar, A., Dubey, A.K., Ramirez, I.S., Muñoz del Río, A., Márquez, F.P.G., 2022. A review and analysis of forecasting of photovoltaic power generation using machine learning. In: International Conference on Management Science and Engineering Management. Springer, pp. 492–505.
Kumari, P., Toshniwal, D., 2021. Deep learning models for solar irradiance forecasting: a comprehensive review. J. Clean. Prod. 318, 128566.
Leng, J., Sha, W., Wang, B., Zheng, P., Zhuang, C., Liu, Q., Wuest, T., Mourtzis, D., Wang, L., 2022. Industry 5.0: prospect and retrospect. J. Manuf. Syst. 65, 279–295.
Li, G., Xie, S., Wang, B., Xin, J., Li, Y., Du, S., 2020a. Photovoltaic power forecasting with a hybrid deep learning approach. IEEE Access 8, 175871–175880.
Li, P., Zhou, K., Lu, X., Yang, S., 2020b. A hybrid deep learning model for short-term PV power forecasting. Appl. Energy 259, 114216.
Li, P., Zhou, K., Lu, X., Yang, S., 2020c. A hybrid deep learning model for short-term PV power forecasting. Appl. Energy 259, 114216.
Mahmud, K., Azam, S., Karim, A., Zobaed, S., Shanmugam, B., Mathur, D., 2021. Machine learning based PV power generation forecasting in alice springs. IEEE Access 9, 46117–46128.
Markovics, D., Mayer, M.J., 2022. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew. Sustain. Energy Rev. 161, 112364.
Mellit, A., Pavan, A.M., Lughi, V., 2021. Deep learning neural networks for short-term photovoltaic power forecasting. Renew. Energy 172, 276–288.
Milosavljević, D.D., Kevkić, T.S., Jovanović, S.J., 2022. Review and validation of photovoltaic solar simulation tools/software based on case study. Open Phys. 20, 431–451.
Mohamad Radzi, P.N.L., Akhter, M.N., Mekhilef, S., Mohamed Shah, N., 2023. Review on the application of photovoltaic forecasting using machine learning for very short-to long-term forecasting. Sustainability 15, 2942.
Musah, A., Dutra, L.M.M., Aldosery, A., Browning, E., Ambrizzi, T., Borges, I.V.G., Tunali, M., Basibüyük, S., Yenigün, O., Moreno, G.M.M., 2022. An Evaluation of the OpenWeatherMap API versus INMET using weather data from two Brazilian cities: Recife and Campina Grande. Data 2022 7, 106.
Nastić, F., Jurišević, N., Nikolić, D., Končalović, D., 2024. Harnessing open data for hourly power generation forecasting in newly commissioned photovoltaic power plants. Energy Sustain. Dev. 81, 101512.
Niu, D., Wang, K., Sun, L., Wu, J., Xu, X., 2020. Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: a case study. Appl. Soft Comput. 93, 106389.
Nou, J., Chauvin, R., Thil, S., Grieu, S., 2016. A new approach to the real-time assessment of the clear-sky direct normal irradiance. Appl. Math. Model 40, 7245–7264.
Oancea, B., Ciucu, Ş.C.: Time series forecasting using neural networks. arXiv preprint arXiv:1401.1333. (2014).
OpenWeatherMap. Available at: https://openweathermap.org/ Last accessed on January 16, 2025.
Peppas, A., Kollias, K., Politis, A., Karalis, L., Taxiarchou, M., Paspaliaris, I., 2021. Performance evaluation and life cycle analysis of RES-hydrogen hybrid energy system for office building. Int J. Hydrog. Energy 46, 6286–6298.
Persson, C., Bacher, P., Shiga, T., Madsen, H., 2017. Multi-site solar power forecasting using gradient boosted regression trees. Sol. Energy 150, 423–436.
European Commission, The Joint Research Centre: EU Science Hub, Photovoltaic Geographical Information System (P.V.G.I.S.) Available at: ⟨https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis_en⟩, Last accessed on: Apr. 04 2025.
Qaid, T.S., Mazaar, H., Al-Shamri, M.Y.H., Alqahtani, M.S., Raweh, A.A., Alakwaa, W., 2021. Hybrid deep-learning and machine-learning models for predicting COVID-19. Comput. Intell. Neurosci. 2021, 9996737.
Rahmani, A.M., Rezazadeh, B., Haghparast, M., Chang, W.-C., Ting, S.G., 2023. Applications of artificial intelligence in the economy, including applications in stock trading, market analysis, and risk management. IEEE Access.
Ren, Y., Suganthan, P.N., Srikanth, N., 2015. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. Renew. Sustain. Energy Rev. 50, 82–91.
Retnowati, Y., Prayoga, G.A., Tamba, J.D., Fauzandi, F.I., Husni, E.M., Yusuf, R., Yahya, B.N., 2021. Statistical Analysis of Urban Heat Island: A Case Study in Bandung, Indonesia. In: 2021 IEEE 11th International Conference on System Engineering and Technology (ICSET). IEEE, pp. 13–18.
Sabareesh, S.U., Aravind, K.S.N., Chowdary, K.B., Syama, S., VS, K.D., 2023. LSTM based 24 h ahead forecasting of solar PV system for standalone household system. Procedia Comput. Sci. 218, 1304–1313.
Scott, C., Ahsan, M., Albarbar, A., 2023a. Machine learning for forecasting a photovoltaic (PV) generation system. Energy 278, 127807.
Scott, C., Ahsan, M., Albarbar, A., 2023b. Machine learning for forecasting a photovoltaic (PV) generation system. Energy 278, 127807.
Sharadga, H., Hajimirza, S., Balog, R.S., 2020. Time series forecasting of solar power generation for large-scale photovoltaic plants. Renew. Energy 150, 797–807.
Sinsel, S.R., Riemke, R.L., Hoffmann, V.H., 2020. Challenges and solution technologies for the integration of variable renewable energy sources—a review. Renew. Energy 145, 2271–2285.
Srinivasan, D., Lee, M.A., 1995. Survey of hybrid fuzzy neural approaches to electric load forecasting. In: 1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century. IEEE, pp. 4004–4008.
Tan, K., Wu, J., Zhou, H., Wang, Y., Chen, J., 2024. Integrating advanced computer vision and AI algorithms for autonomous driving systems. J. Theory Pract. Eng. Sci. 4, 41–48.
Torres, J.F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., Troncoso, A., 2021. Deep learning for time series forecasting: a survey. Big Data 9, 3–21.
Touati, F.A., Al-Hitmi, M.A., Bouchech, H.J., 2013. Study of the effects of dust, relative humidity, and temperature on solar PV performance in Doha: comparison between monocrystalline and amorphous PVS. Int J. Green. Energy 10, 680–689.
Uribe, R.A., Santamaria, P.A., Castañeda, P.S., 2024. System to Prevent the Development of Dermatological Diseases Generated by Overexposure to UV Radiation during Childhood using Open Weather Map API and IoT Wearables. In: 2024 9th International Conference on Computer and Communication Systems (ICCCS). IEEE, pp. 1002–1007.
Wakchaure, M., Patle, B.K., Mahindrakar, A.K., 2023. Application of AI techniques and robotics in agriculture: a review. Artif. Intell. Life Sci. 3, 100057.
Wang, J., Li, P., Ran, R., Che, Y., Zhou, Y., 2018. A short-term photovoltaic power prediction model based on the gradient boost decision tree. Appl. Sci. 8, 689.

Wang, X., Li, L., Tan, S.C., Yang, L., Lei, J., 2023. Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers' AI readiness. Comput. Hum. Behav. 146, 107798.

Wolff, B., Kühnert, J., Lorenz, E., Kramer, O., Heinemann, D., 2016. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. Sol. Energy 135, 197–208.

Yüksel, N., Börklü, H.R., Sezer, H.K., Canyurt, O.E., 2023. Review of artificial intelligence applications in engineering design perspective. Eng. Appl. Artif. Intell. 118, 105697.

Zeng, P., Hu, G., Zhou, X., Li, S., Liu, P., Liu, S., 2022. Muformer: a long sequence time-series forecasting model based on modified multi-head attention. Knowl. Based Syst. 254, 109584.

Zeroual, A., Harrou, F., Dairi, A., Sun, Y., 2020. Deep learning methods for forecasting COVID-19 time-Series data: a comparative study. Chaos Solitons Fractals 140, 110121.

Zhang, P., Kamel Boulos, M.N., 2023. Generative AI in medicine and healthcare: promises, opportunities and challenges. Future Internet 15, 286.

Zhen, H., Niu, D., Wang, K., Shi, Y., Ji, Z., Xu, X., 2021. Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information. Energy 231, 120908.

Ziane, A., Necaibia, A., Sahouane, N., Dabou, R., Mostefaoui, M., Bouraiou, A., Khelifi, S., Rouabhia, A., Blal, M., 2021. Photovoltaic output power performance assessment and forecasting: impact of meteorological variables. Sol. Energy 220, 745–757.

Ziller, A., Mueller, T.T., Stieger, S., Feiner, L.F., Brandt, J., Braren, R., Rueckert, D., Kaissis, G., 2024. Reconciling privacy and accuracy in AI for medical imaging. Nat. Mach. Intell. 1–11.