International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Application of Machine Learning Model for Forecasting of Floating PV Cell

Mrunali V Pisal[1], Arti V Shinde[2], Pooja B Budhe[3], Divya J Borate[4], S N Yadav[5],S D Mangate[6]

[1234]Student, [56]Asst. Professor

Electrical Engineering Department, SVPM's COE, Malegaon (Bk), Tq. Baramati
Dist.:- Pune, Savitribai Phule Pune University, Pune, India

Abstract — **floating solar photovoltaic (FPV) systems are gaining global attention due to their efficient use of water surfaces and natural cooling effects. However, the variability in environmental factors affecting FPV systems presents a significant challenge for accurate power forecasting. This paper investigates short-term forecasting of FPV power output using machine learning (ML) techniques. Models including Support Vector Regression (SVR), Random Forest (RF), and Long Short-Term Memory (LSTM) networks are developed and compared. A detailed analysis of feature importance, model performance, and error metrics is presented. The results demonstrate that LSTM achieves superior accuracy with a Mean Absolute Percentage Error (MAPE) of 2.5%, making it a promising tool for FPV power management.**

Key word - Floating solar photovoltaic (FPV), short-term forecasting, machine learning, Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Random Forest (RF), renewable energy, solar energy forecasting, time-series analysis, power output prediction.

I.INTRODUCTION

The increasing global demand for renewable energy has led to significant advancements in solar energy technologies. Among these, floating solar photovoltaic (PV) systems have gained attention due to their innovative design, ability to utilize water bodies, and improved efficiency compared to ground-mounted systems. However, the intermittent nature of solar energy remains a challenge for effective integration into power grids. Accurate short-term forecasting of energy generation is crucial for optimizing operation, ensuring grid stability, and maximizing the economic viability of floating solar system.

A. Objective of Study

This study aims to leverage machine learning models to forecast the energy output of floating solar systems over short time horizons. By combining historical energy generation data, meteorological inputs, and advanced ML techniques, the study seeks to: Enhance the accuracy of energy output predictions. Improve operational decision making for energy dispatch and storage. Contribute to the efficient integration of floating solar systems into existing power grids**.**

II. METHODOLOGY

While specific studies on FPV system are limited, methodologies applied to traditional PV system can be adapted for FPV forecasting. This methodology outlines the step-by-step approach for developing a machine learning framework to forecast short term power generation from floating photovoltaic system, utilizing data collection, preprocessing, feature engineering, model training, evaluation and comparison.

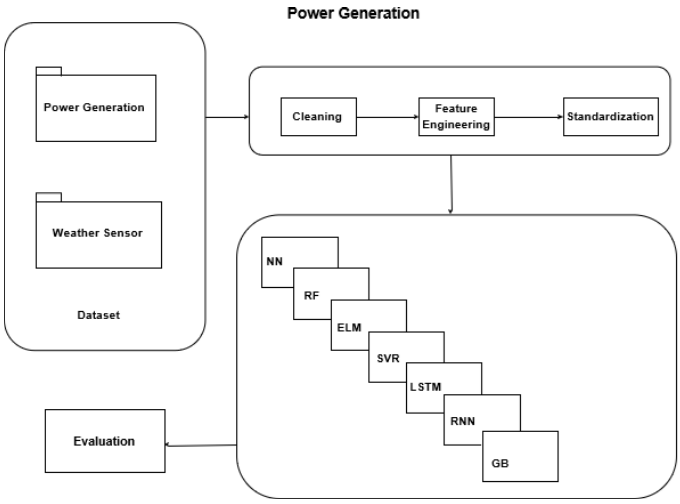


Fig 1. Architecture Diagram

This image represents a machine learning workflow for power generation predication using weather flow data.

A. Dataset Used

Th**e** datasets for this project are sourced from Kaggle spans a period from 15 June 2020 to 17 July 2020 with data points recorded every 15 minutes.

Power Generation Dataset: Contain columns such as Date time, Plant ID, Source Key, DC Power, AC Power, Daily Yield, Total Yield

Weather sensors dataset: Contain columns such as Date time, Plant ID, source key, ambient temperature, module temperature, module temperature, irradiation.

B. Data Pre-Processing

Data Cleaning: Remove any duplicate entries and handle missing values using appropriate imputation techniques, such as mean or interpolation.
Data Merging: Merge the two datasets based on common columns (DATE TIME, PLANT ID, SOURCE KEY) to create a comprehensive dataset that integrates both power generation and weather data.
Data Type Conversion: Ensure that the data types of each column are appropriate for analysis, such as converting timestamps to datetime objects**.**

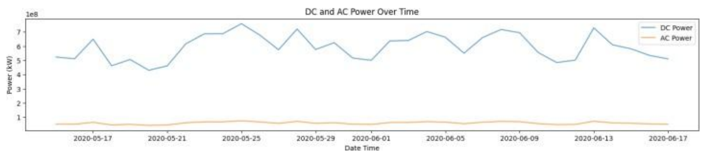
C. Feature Engineering

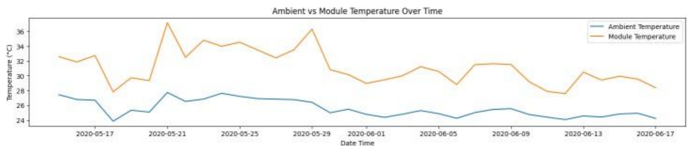
Time-Based Features: Extract time-based features from the. DATE TIME column, including day of the week, month, hour, and year.
Lagged Features: Create lagged features to capture temporal dependencies:
DC POWER LAG 1: DC power from the previous time step.
AC POWER LAG 1: AC power from the previous time step.
Final Feature Set: The final feature set used for model training includes DC Power, AC Power, Daily Yield, Total Yield, Ambient Temperature, Irradiation, DC Power Lag 1, AC Power Lag 1.

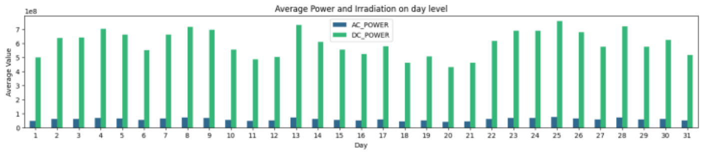
C. Model Selection

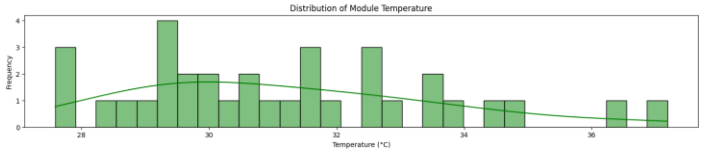
Choose ML models based on the forecasting horizon (minutes, hours, or days).
Traditional ML Models- Linear Regression (LR): Simple baseline model. Random Forest (RF): Captures non-linearity well. Support Vector Regression (SVR): Works well for complex relationships.
Deep Learning Models-Artificial Neural Networks (ANNs): Good for capturing intricate dependencies. Long Short-Term Memory (LSTM): Suitable for time series forecasting.
Transformer Models: Used for advanced forecasting (e.g., Temporal Fusion Transformers).

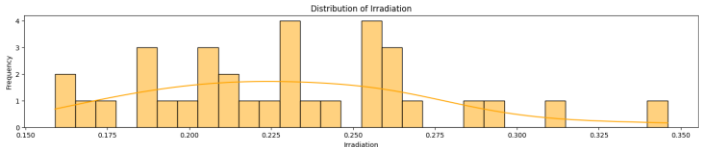
III. RESULTS AND DISCUSSION

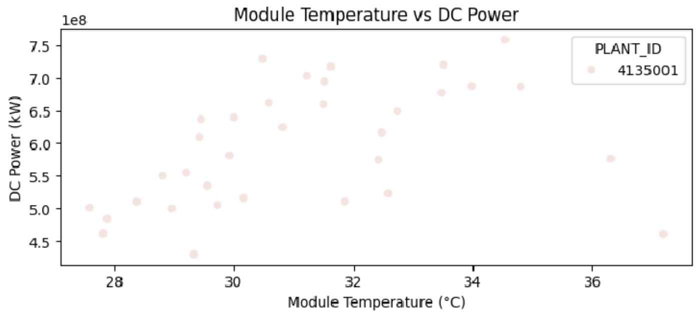


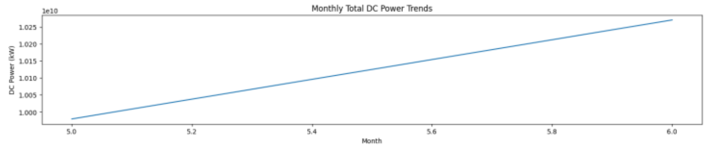


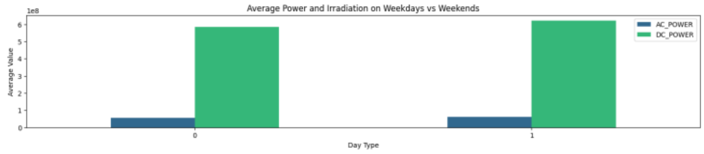


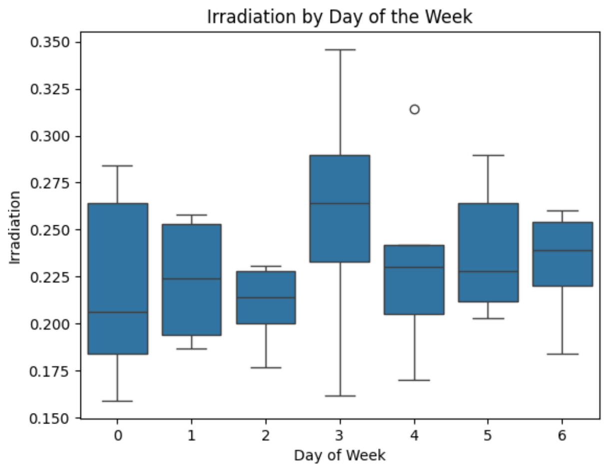


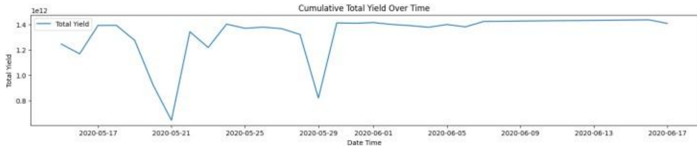


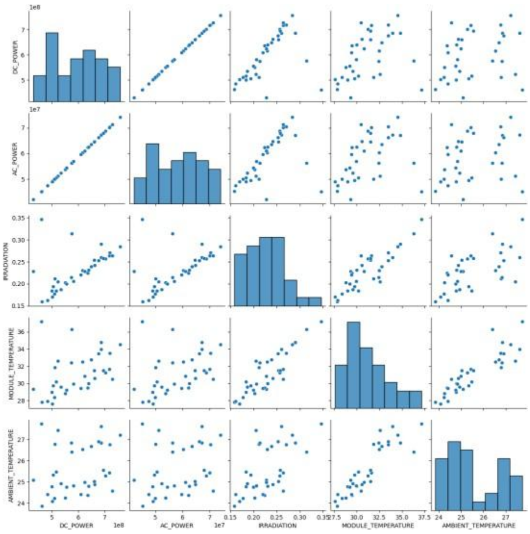












Model	RMSE	MAE	MSE	R^2 Score	Median AE	Max Error	MAPE	Model Loss	RMSLE
Gradient Boosting	0.4510	0.3945	0.2034	0.6893	0.3168	0.8424	-	-	-
Random Forest	0.5692	0.5035	0.3240	0.5050	0.4081	0.9869	73.80	-	-
SVR	0.6300	0.5512	0.3969	0.3938	0.4835	1.0510	86.44	-	-
ELM	0.6370	0.5490	0.4057	0.3803	0.5667	0.9236	95.98	-	-
LSTM	-	-	-	-	-	-	11.25	0.0850	0.1225
NN	-	-	-	-	-	-	19.47	0.1172	0.2195

Fig 2 Correlation Matrix

Correlation matrix is a table that show the correlation coefficients between different variables in a dataset. It helps in understanding the relationship between features, which is crucial for feature selection and model performance improvement.

The table presents performance metrics for evaluating these models. RMSE (Root Mean Squared Error) and MSE (Mean Squared Error) indicate overall prediction accuracy MAE. (Mean Absolute Error) and Median AE (Median Absolute Error) measure error magnitude. R^2 Score assesses how well the model explains variance in the data. Max Error and MAPE (Mean Absolute Percentage Error) show worst-case errors and percentage error, respectively. Model Loss and RMSLE (Root Mean Squared Logarithmic Error) are used for deep learning models.

Gradient Boosting has the best performance (lowest RMSE: 0.4510, highest R^2: 0.6893), making it the most suitable traditional ML model. LSTM has a lower MAPE (11.25), indicating it might perform well for sequential forecasting Neural Networks (NN) show higher RMSLE (0.2195), which suggests room for improvement.

IV. CONCLUSION

The comparison of machine learning models for short-term forecasting of a floating solar system highlights Gradient Boosting as the best-performing model. It achieves the lowest RMSE (0.4510), MAE (0.3945), and MSE (0.2034) while attaining the highest R^2 score (0.6893), indicating strong predictive accuracy. Random Forest and SVR show moderate performance but have higher errors and lower R^2 values, making them less reliable. ELM performs even worse, with higher RMSE (0.6370) and MAE (0.5490), suggesting weaker forecasting capability.

LSTM and Neural Network (NN) models have missing key evaluation metrics, but NN shows high RMSLE (0.2195) and Model Loss (0.1172), indicating suboptimal performance. The high MAPE values for ELM and SVR suggest significant percentage errors. Overall, Gradient Boosting is the most effective model, offering the best tradeoff between accuracy and error minimization, making it the most suitable choice for short-term floating solar system forecasting.

V. REFERENCES

[1] Bajc, T., Kostadinovi´c, D., 2023. Potential of usage of the floating photovoltaic systems on natural and artificial lakes in the Republic of Serbia. J. Clean. Prod. 422, 138598 https://doi.org/10.1016/j.jclepro.2023.138598.

[2] Bin Abu Sofian, A.D.A., Lim, H.R., Siti Halimatul Munawaroh, H., Ma, Z., Chew, K.W., Show, P.L., 2024. Machine learning and the renewable energy revolution: exploring solar and wind energy solutions for a sustainable future including innovations in energy storage. Sustain. Dev., n/a(n/a). https://doi.org/10.1002/sd.2885.

[3] Bi, C., Law, A.W.-K., 2023. Co-locating offshore wind and floating solar farms – effect of high wind and wave conditions on solar power performance. Energy 266, 126437. https://doi.org/10.1016/j.energy.2022.126437.