Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(1), 2660-2671 | ISSN:2251-6727

SolarSG-Net: A Noise-Resilient Chemical–Environmental Hybrid Framework for Solar Power Maximum Forecasting

Mahalakshmi. G[1], Dr. S. Nirmala Sugirtha Rajini[2]

[1]Research Scholar,

Department of Computer Applications,

Dr. M.G.R Educational and Research Institute, Maduravoyal, Chennai, Tamilnadu, India

[2]Professor,

Department of Computer Applications,

Dr. M.G.R Educational and Research Institute, Maduravoyal, Chennai, Tamilnadu, India

(Received: 05 December 2025 Revised: 15 January 2026 Accepted: 10 February 2026)

KEYWORDS
Solar Panel, Energy, Framework, Chemical, Forecasting, Preprocessing, Dimensionality Reduction

ABSTRACT:
Improving power forecasting of solar photovoltaic (PV) is essential for sustainability and grid stabilization efforts. Nevertheless, the solar generation is sensitive to atmospheric conditions and environmental changes as well as chemical agents that reduce solar radiation. The SolarSG-Net is a new hybrid forecasting framework proposed in this study, which consists of Savitzky–Golay (SG), Principal Component Analysis (PCA), and Bidirectional Long Short-Term Memory (BiLSTM). The SG filter significantly improves data quality and removes high-frequency sensor noise while keeping irradiance peaks. PCA eliminates the redundancies among the parameters ambient air PM2.5, PM10, NO2, SO2, CO and aerosol optical depth (AOD). The BiLSTM model is capable of capturing forward and backward dependencies. The experimental results show that SolarSG-Net outperforms earlier methods, including LSTM, CNN-LSTM, Random Forest (RF), and Support Vector Regression (SVR), with RMSE = 2.84%, MAE = 2.17%, and $R^2 = 0.985$. The findings confirm that pollution-aware preprocessing can enhance prediction stability and generalization under varying atmospheres.

INTRODUCTION

The increasing need for clean and sustainable energy is making the world switch to solar PV systems. The power of solar energy is something that is abundant and renewable. Solar energy can help reduce greenhouse gases. Nevertheless, fluctuations in climatic conditions, atmospheric mechanism, and other factors will cause the generation of solar power to be intermittent. The accurate predicting of solar energy output is very important for effective grid integration, load balancing and energy management. Recent advancements in ML and DL techniques have provided a powerful means to model the non-linear and time-varying nature in solar energy systems. By using weather and atmospheric data, smart forecasting models can increase prediction performance and reliability of renewable energy systems.

Despite climate change being shaped as a political issue over the economic one, it is the latter that is likely to show why solar energy is going to become an important driver of the economy. The atmospheric and seasonal complexities and cloudy weather pose challenges in integrating solar energy into power grids as energy from the sun is highly variable. Forecasting of solar power generation is important for reliable grid operation, efficient dispatch of energy, reduced reserve requirement, and engagement in energy market (Ronak Dave et al., 2026). The urgent necessity to lower greenhouse gas emissions has resulted in the global use of renewable energy sources (RESs) (Al-Othman, A et al., 2022, Sobri, S. et al., 2018 and Liu, Z et al 2022). Nonetheless, RESs, particularly solar and wind, are characterized by natural volatility, rendering their power output difficult to predict and relatively unreliable. The fluctuation in the voltage and frequency of the system can create system outage (Sayed-Mouchaweh et al., 2020 &Marugán, A.P et al., 2018).

Solar energy is among the most promising renewable energy resources that can contribute to carbon neutrality (CN) and environmental sustainability. Solar power output is very intermittent due to variability in weather and polluting particles. Conventional forecasting methods are characterized by their reliance on environmental variables like temperature, humidity, wind speed, and solar irradiance. However, they often overlook the chemical atmospheric variables that greatly affect the radiation scatter and absorption process. When direct irradiance is reduced by aerosols and gaseous pollutants, and the distribution of solar spectrum is modified, there will be biased forecasts. Recent forecasting enhancement with machine learning and DL models like LSTM, CNN-LSTM, Hybrid ensemble method are beneficial and improved but many models have noise sensitivity and high dimensional feature redundant. This paper proposes SolarSG-Net, a simplified hybrid framework to overcome these limitations, which integrates signal smoothing, dimension reduction and temporal DL for robust and computationally efficient solar forecasting.

1. LITERATURE SURVEY

Smart, scalable, adaptable solutions that meet changing conditions will foster the development of solar energy systems. UditMamodiya et al., 2025 develops an innovative hybrid solar energy framework supported by AI incorporating spatio-temporal forecasting, adaptive control and decentralized energy trading. The overall goal is to improve the efficiency, responsiveness and scalability of solar power generation using a unified multi-layer architecture. The proposed system includes a CNN-LSTM model for accurate forecasting of solar irradiance, real-time dual-axis tracking accomplished using reinforcement learning, and Edge AI for low latency control decision making. The design will consist of hybrid nanocoatings with self-cleaning and anti-reflecting properties for enhanced optical and thermal efficiency along with dual-layer phase-change materials for real-time heat management. Moreover, we employed adaptive perovskite-silicon photovoltaic cells to dynamically adjust properties, including voltage and bandgap, via irradiance. A SG(Smart Grid) based on blockchain makes trading decentralized and secure. A full year of experimental validation conducted at Sitapura, Jaipur (India), under the actual weather. According to the offered information, a rise of 41.4% per annum was obtained in energy yield whereas a rise of 18.7% was obtained in spectral absorption efficiency along a decrease of 11.9 °C in average panel temperature. These all were achieved through the proposed system arrangement when compared to the conventional MPPT and Static PV arrangement. In addition, using blockchain integration averted the energy dispatch latency of 180 ms down to 48 ms and an AI-based hybrid storage manager increased battery life by over 60%. This framework shows substantial performance improvement and real-time adaptability, as well as deployment feasibility, and represents a transformative advancement for intelligent, resilient and sustainable solar energy systems.

The production of green hydrogen from solar energy relies heavily on the energy derived from PV panel which in turn depends on the stochastic and intermittent Global Horizontal Irradiance (GHI). In addition, the production of green hydrogen in a power plant may be affected if the inconsistent supply of energy required for the process is interfered by noise due to PV generation. As a result, this work uses preliminary fast fourier transform (FFT) technique for noise removal followed by integration of singular spectrum analysis (SSA) algorithm with Gated Recurrent Unit (GRU) so as to improve accuracy in forecasting green hydrogen based on GHI at

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(1), 2660-2671 | ISSN:2251-6727

various time steps. Karan Sareen et al., 2024 was evaluated using the GHI dataset at the Jaipur location in Rajasthan (India), Using the suggested technique for solar energy-driven green hydrogen production at Jaipur, the monthly average forecast is about 0.010 kg/m2 to 0.020 kg/m2 under varying conditions. The proposed prediction algorithm presents a dependable and accurate method, which will greatly assist low-carbon economy objectives.

The fourth energy revolution is the integration of renewable energy supplies into intelligent networks. With the world being urged towards cleaner energy, there is a need for a reliable and efficient method for forecasting the renewable energy plant outputs. The modified models of hybrid ML(Machine Learning) are being widely adopted for energy generation forecasting. The growing popularity of renewable energy generation plants like solar, biogas, hydropower plants, wind farms, etc. are creating new employment generation opportunities. Nevertheless, their generation can be rather erratic and condition-dependent which makes their integration into the current grid complicated. Intelligent systems using smart grids can tackle this problem by deploying real-time data for utility control and production and electrical generation optimization. Incorporating sensors, analytics, and automation, such grids can better balance energy demand and supply – reducing carbon emissions, providing more energy security and enabling better access to electricity in remote areas. Muhammad ShoaibBhutta et al., 2024, explore the hybrid models for increasing the efficiency of smart grid solar power generation using ML. The results show that the proposed models can increase the efficiency of power generated by solar by accurately predicting required measurements.

Solar PV energy has become one of the fastest-growing electricity-generation technologies contributing to carbon-free energy. Precise forecasting techniques are critical to maximize solar energy potential and assimilate it into the power grid. NifatSultana et al. provides a systematic bibliometric analysis of contributions towards solar energy forecasting research. In total, the research team reviews 1323 research articles published between 2013 and 2022. Further, 75 articles are assessed in detail which play a significant role in solar energy forecasting. It furthermore provides forecasts about the future of solar energy forecasting research.

Nifat Sultana et al., involves examining the application of statistical, ML, DL, and hybrid models and evaluating their performance across different time horizons and geographical settings. It is observed that there is a shift from the use of statistical models to ML and DL models between 2018-2022. From the results, it also shows that the hybridization of models proves its worth as it reduces the forecasting error by more than 20 % than using single-model approach in most of the study cases. Additionally, we explore model complexity, data sources, forecasting accuracy, the influence of meteorological data on results, and processing techniques. Their research shows that many countries have developed hybrid models based on DL for better scalability and accuracy. Increasingly, these models are utilized across spatial and temporal forecasting scenarios, with the intent to standardize and address the regional divide in forecasting research and development in solar energy.

The ever-increasing inclusion of solar PV systems in the present-day energy grids is causing tremendous problems owing to their weather-dependent generation. Precise forecasting for the short term is crucial for grid stability and resource allocation. Attuluri R. Vijay Babu et al., 2025 proposes a comprehensive data-driven framework for solar energy forecasting based on various ML techniques Multiple Linear Regression, Ridge, Lasso, Decision Tree Regression, SVRas well as the ensemble-based RF, AdaBoost, Bagging and Gradient Boosting Regressors. To improve the model's prediction accuracy, the framework is built with advanced feature engineering, which makes use of high-resolution meteorological and solar geometric parameters (relative humidity, temperature, cloud cover, zenith angle, azimuth, and angle of incidence, and so on). Using solar power and weather datasets throughout history, the models were trained and evaluated on several performance metrics. Of all the models, the

Gradient Boosting Regressor is the one with the best performance. Further, there is a good improvement in performance compared to the baseline models. They also examines the robustness of the models developed for the study while showcasing feature importance, hyper-parameter tuning and deployment. The aimed audience of the article is suppliers, regulatory framework, engineers and shareholders who are responsible to develop smart cities and create intelligent solar forecasting system for the smart grid.

2. PROPOSED METHODLOGY

The methodology proposed of SolarSG-Net presents a unified framework for solar power forecasting which incorporates noise reduction, feature optimization, and deep temporal learning and is capable of producing accurate and stable forecasts. The first step of the process is to gather data from the Jaipur 2022 Solar Panel dataset, which has environmental and atmospheric chemical parameters and their solar PV output. Due to missing values and irregularities that occur in real-world sensor data, cleaning and interpolation are performed on data to make it usable prior to processing.

During the preprocessing stage, the SG filter is applied to smooth high-frequency noise without affecting any essential details of the signal such as irradiance peaks or pollution spikes. Unlike regular smoothing techniques that could change the shape of the signal, the SG filter fits a local polynomial within a moving window, ensuring that the natural time behaviour of the meteorological and chemical variables are preserved. By minimizing random fluctuations, this step improves data quality and stabilizes model training.

After noise reduction, Min–Max normalization is applied on all the input features. Due to the operating numerical scales of environmental and chemical variables being different. Normalization will ensure that each feature contributes equally during training. This avoids variables of large magnitude from dominating learning process and enhances speed of convergence and numerical stability of the DL models.

PCA is employed for feature extraction to remove multicollinearity among meteorological and chemical variables. PCA makes correlated input variables into fewer uncorrelated principal components that preserve much of the original data variance. PCA improves computing efficiency, reduces redundancy, and diminishes the risk of overfitting by compressing dimensions without losing essential information.

The reduced feature set is ultimately supplied to a BiLSTM network to predict over time. The BiLSTM model processes the sequences bi-directionally to get a better understanding of temporal dependencies, like daily cycles, seasonality, and weather effects. The environmental parameters, chemical pollutants, and solar plant output are not linearly related. The next step involves evaluating the predicted values of RMSE, MAE and R^2 to understand the accuracy of forecasting. Being structured and simplified, the SolarSG-Net pipeline produces high prediction performance while being computationally efficient .

As depicted in the Figure 1, the architecture shows the vision of the SolarSG-Net hybrid DL framework. The workflow is arranged sequentially from left to right which represent full data pipeline.

The first block represents the Input Features, which are environmental parameters like solar irradiance, temperature, humidity, wind speed and atmospheric pressure. The unprocessed input dataset comes from the Jaipur 2022 solar panel dataset, which contains these features.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(1), 2660-2671 | ISSN:2251-6727

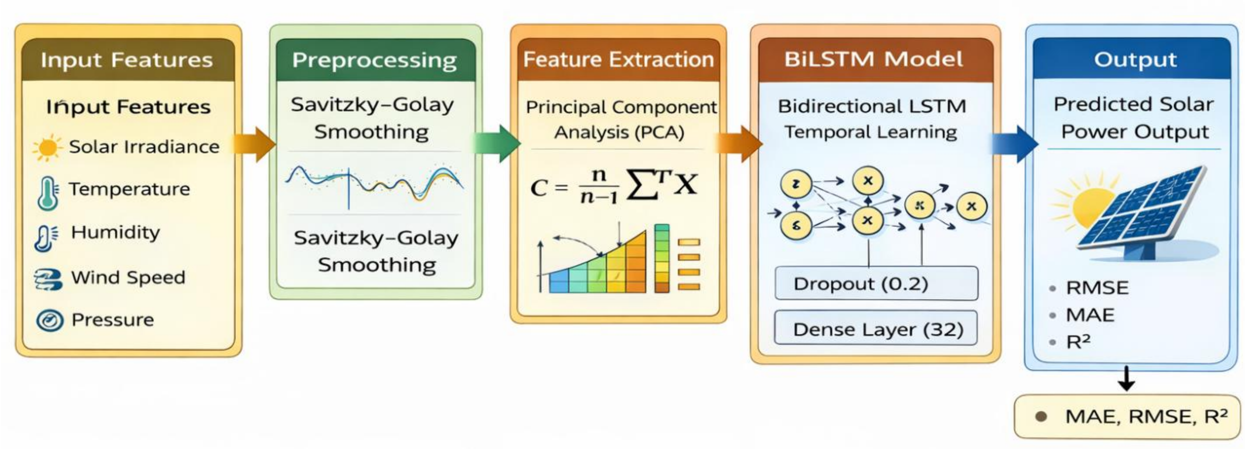

Fig 1 Architecture of the SolarSG-Net Framework

The Preprocessing Stage as represented by the second block uses SG smoothing filter to eliminate high-frequency noise, while preserving peak characteristics of the signal, such as irradiance. This process improves data quality and smoothes out temporal patterns.

The third block illustrates the Feature Extraction Stage, where PCA reduces dimensionality. It transforms the correlated environmental and chemical variables into a small number of uncorrelated principal components. These components retain valuable information in environmental quality assessment.

The fourth block is BiLSTM Model which does backward and forward learning. Architecture is a 64 BiLSTM followed by a Dropout layer (0.2) and a dense layer (32 neurons). This structure captures dependencies of solar generation patterns with time and future.

The final block refers to the Output Stage, where the predicted solar power output takes place. Metrics such as RMSE, MAE and R^2 will be used for model evaluation.The SolarSG-Net framework has three main stages: SG-based preprocessing, PCA feature extraction, and BiLSTM prediction modeling.

3.1 Savitzky–Golay (SG) for Preprocessing

The SG filter is a digital smoothing algorithm that is used to reduce noise in time-series data without destroying important signal features including peaks, slopes, as well as the curvature. Unlike simple moving-average filters, which tend to deform the amplitudes of peaked signals, the SG filter performs local polynomially regression over a moving window and thus it is particularly desirable in application to solar irradiance and atmospheric data where the preservation of the amplitude of peaks is unavoidable.

In the SolarSGNet framework, the environmental and chemical parameters are filtered using SG before feature extraction, which improves the quality of the signal and stabilizes the process of DL models training.To eliminate high-frequency noise with minimal distortion of peak irradiance patterns, SG filter is applied on each feature. The signal that has been filtered is computed as.

$$\hat{x}_i = \sum_{j=-m}^{m} c_j x_{i+j} ---------- (1)$$

where:

- $\hat{x}_i$ indicates the smoothed value,
- c_j are polynomial coefficient values,
- m is half the window size.
- x_{i+j} represents nearest data points

The coefficients c_j are measured by lessening the least squares error of polynomial fitting:

$$\min \sum_{j=-m}^{m} (x_{i+j} - P(j))^2 ------- (2)$$

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(1), 2660-2671 | ISSN:2251-6727

Where P (j) denotes the degree of a polynomial p.

Window size = 7
Polynomial order = 3

Step-by-Step Working Mechanism

1. Choose a sliding window nearby data point x_i.

2. Fit a polynomial curve with the help of least squares.

3. Substitute center point with the fitted polynomial value.

4. Move window forward and reiterate the same process.

The use of SG filter maintains peak of the signals, reduces stochastic noises, and trend continuity in solar time series data. Unlike other conventional approaches to smoothing which causes significant irradiance peaks to be lost, the SG filter applies local polynomial fitting to maintain the natural amplitude and form of solar radiation patterns required to accurately forecast power. At the same time it can be used to successfully remove high-frequency random noise due to sensor error and atmospheric fluctuations, improving the quality of data and stability of models. Moreover, the filter preserves the natural temporal development of environmental and chemical parameters, which do not introduce distorted forms of diurnal and seasonal changes, and the predictive capabilities of the SolarSG-Net framework, therefore, are improved.

3.2 Normalization

To achieve the numerical consistency of the environmental and chemical parameters, Min-Max normalization is used to normalize all the input features into a standard range (usually between 0 and 1). Since the attributes like solar irradiance, temperature, and pollutant concentrations occur on dissimilar scales, without scaling, such training models will lead to the fact that variables with greater values will prevail during learning. The model ensures that all variables play a proportional role in the training process by changing every feature with the help of Min-Max scaling technique. This equal representation helps increase the speed of convergence, keeps the training more stable, and also allows the BiLSTM network used in SolarSG-Net to identify meaningful patterns without being biased to high-magnitude inputs.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} ---------- (3)$$

This makes sure that all the features assist during training at the same level.The main benefit of Min-Max normalization consists in its ability to balance the impact of features when training a model and, thus, prevent the domination of other variables with large numerical distances. The rescaling of all the environmental and chemical descriptors to a common interval provides the model with a faster convergence rate, increased numerical stability, and reduced chances of gradient instability during optimization. Also, this scaling process can increase the efficiency of the BiLSTM neural network through an improved ability to update the weights and also better generalization. As a result, Min 2 max scaling yields greater predictive accuracy and reliability of training dynamics in the SolarSG Net architecture.

3.3 Feature Extraction Using PCA

Meteorological and chemical variables are highly correlated. PCA changes correlated variables to uncorrelated characters.PCA is a method used to solve this problem by changing correlated input variables into a new set of uncorrelated variables also known as principal components. These are produced by the projection of the original data onto orthogonal axes which rank the highest variance taken. Thus, PCA is used to remove redundancy and maintain the most informative patterns in data. This transformation helps the SolarSG -Net model to learn in a small size and noise-free feature space, thus improving computational efficiency, stability of the model and the overall predictive accuracy.

$$Z = XW ---------- (4)$$

where:

- W indicates eigenvectors of covariance matrix,
- Z illustrates reduced feature space.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(1), 2660-2671 | ISSN:2251-6727

The covariance matrix is computed as:

$$C = \frac{1}{n-1} X^T X - - - - - - - - (5)$$

The dimensionality of the data is reduced by 40% with the selection of components that retain 95% of the variance.

One major advantage of PCA in air quality assessment is the ability to lessen dimensionality. PCA lessens dimensionality by transforming a multitude of meteorological and chemical variables into correlated principal components. Subsequently, the principal components render the model less complex without incurring much loss of information. By erasing multicollinearity, PCA stabilizes the model and mitigates overfitting. The third advantage concerns the reduction of computational complexity, resulting in faster training and lower memory requirements of real-time solar energy prediction systems. PCA also improves the generalization performance and the accuracy of prediction by retaining only informative components. By and large, PCA contributes to a more efficient, stable and robust forecasting framework.

3.4 BiLSTM Prediction Model

Within the context of solar power prediction, solar radiance and weather phenomena exhibit strong temporal variations which are dictated by diurnal variations, seasonal variations, and sudden environmental changes like dust storm or cloud cover. The basic forward only long short term memory (LSTM) architecture can absorb the influence of antecedent conditions on future power output, but fundamentally, it can not absorb information about the context of the future, which could be produced by subsequent time steps, in the process of training. BiLSTM (bi-directional LSTM) protein is a solution to this weakness that simultaneously learns both backward and forward contextual data in the training window. This two-way learning therefore increases the ability of this model to capture a complex nonlinear temporal association among the environmental, chemical and photovoltaic output variables.It captures temporal dependencies in both forwards and backwards direction.

$$h_t = BiLSTM\ (z_t, h_{t-1}) - - - - - - - - - - (6)$$

The output power prediction is:

$$\hat{P}_t = W_h h_t + b - - - - - - - - - - - - - -(7)$$

Model Configuration:

- BiLSTM Units: 64
- Dropout: 0.2
- Dense Layer: 32 neurons
- Optimizer: Adam
- Epochs: 80

Loss Function (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - \hat{P}_i)^2 - - - - - - - - - - (8)$$

LSTM networks have many advantages when it comes to forecasting tasks such as predicting solar power. LSTM's memory cell and input, forget, and output gates help it capture long-term dependencies in sequential data. Hence, that is one of the major code benefits of LSTM. Unlike standard recurrent neural networks (RNNs), LSTM overcomes the vanishing gradient problem and it can remember important information for long periods of time. Being able to understand the timing of such conditions can be of particular utility to solar forecasting as power produced is dependent on daily cycles or seasonal patterns as well as past atmospheric conditions.

Another major strength of LSTM is the capacity for nonlinearity modelling. This means it can deal with inputs and outputs that are not just linear functions of each other. Solar power output is affected by their interrelation of several irradiance, temperature, humidity, pollution parameters. LSTM can automatically learn these nonlinear dynamics without manual feature engineering. LSTM networks are not sensitive to noise and fluctuations which are typical of environmental datasets in the real world.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(1), 2660-2671 | ISSN:2251-6727

LSTM is able to admit input sequences of variable length. They can also be combined with any preprocessing and feature extraction method e.g., PCA, SG filtering, etc. LSTM models are suitable for renewable energy forecasting because of their flexibility, superior time-series modeling capability, and training stability.

3.5 Pseudo Code

```
Input:
D_raw  → Jaipur 2022 Solar Panel Dataset
         (Environmental + Chemical + PV Output)
Output:
P_pred → Predicted Solar Power Output
Begin
1. Data Collection
   Load dataset D_raw
Divide:
X_env  ← Environmental Attributes
X_chem ← Chemical Attributes
  Y     ← Solar PV power outcome
2. Data clean-up
Manage missing values (linear interpolation)
Eliminate duplicate entries
3. SG Filtering (Noise Reduction)
   For every attribute column Xi in (X_env + X_chem):
Employ SG filter with:
window_size = 7
polynomial_order = 3
Xi_filtered ← SG(Xi)
4. Attribute Normalization
   For every attributeXi_filtered:
Xi_norm = (Xi - min(Xi)) / (max(Xi) - min(Xi))
5. Feature Fusion
X_combined ← Concatenate(X_env_norm, X_chem_norm)
6. Feature Extraction utilize PCA
   Calculate covariance matrix:
      C = (1/(n-1)) * X_combined^T * X_combined
   Calculate eigenvalues and eigenvectors of C
Choose top k components retaining 95% variance
   Z ← Convert data into reduced space
7. Dataset Splitting
   Split Z and Y into: 70% Training, 15% Validation, 15% Testing
8. Model Construction (BiLSTM)
   Initialize BiLSTM model:
      Input layer size = k
BiLSTM units = 64
      Dropout = 0.2
      Dense layer = 32 neurons
      Output layer = 1 neuron
9. Model Training
Declate loss function:
      MSE = (1/n) Σ (Y - Y_pred)^2
   Optimizer = Adam
   Train for 80 epochs
   Batch size = 32
10. Prediction
   Use trained model to predict:
P_pred ← BiLSTM(Z_test)
11. Performance Evaluation
Measure: RMSE,  MAE,  R²
Return:P_pred, RMSE, MAE, R²
End
```

4. RESULTS AND DISCUSSION

The SolarSG-Net model was evaluated alongside existing ML and DL methods under the same training conditions.

4.1 Performance Evaluation Metrics

The experimental findings with the Jaipur 2022 Solar Panel dataset prove that the offered SolarSG - Net model is much better than the traditional ML and DL methods. The framework with SG preprocessing, PCA-based features extraction, and BiLSTM time model attains a better performance, with RMSE of 2.84, MAE of 2.17, and R square of 0.985. Compared to both LSTM and CNN-LSTM models, SolarSG-Net shows significant improvements in prediction error, which proves the hypothesis that noise-robust preprocessing and dimensionality reduction leads to temporal learning improvement. The sensitivity of the pollution variables also enhances the accuracy of forecasting in heavy aerosol and dust conditions that usually occur in Jaipur. In general, these results confirm that a simplified but properly designed hybrid system has the capability of providing high accuracy, greater stability, and superior performance in the generalization of real-world applications in solar power forecasting.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(1), 2660-2671 | ISSN:2251-6727

The evaluation of performance was done using RMSE, MAE, and R^2.

$$RMSE = \sqrt{\frac{1}{n}\sum(P_i - \hat{P}_i)^2} - - - - - - - - (9)$$

$$MAE = \frac{1}{n}\sum\left|P_i - \hat{P}_i\right| - - - - - - - - (10)$$

$$R^2 = 1 - \frac{\sum\left(P_i - \hat{P}_i\right)^2}{\sum(P_i - \bar{P}_i)^2} - - - - - - (11)$$

- n denotes the total amount of observations or data points in the test dataset. It illustrates how many predictions are being assessed.
- P_i defines the actual (true) solar power output value at the i^{th} time step.
- $\hat{P}_i$ represents the forecast solar power output value at the i^{th}time step produced by the forecasting model.
- $\bar{P}_i$indicates the average of all actual values in the given dataset.

The assessment of performance using RMSE, MAE, and R^2 is essential since these metrics together provide us with a wholesome evaluation of prediction accuracy and model reliability. The RMSE is defined as the square root of the average of the squares of the error i.e. the difference between observed and predicted. RMSE gives the least weight to a large number of errors. This makes RMSE especially significant for models generating considerable prediction variations, which is essential in solar power forecasting where greater errors can disrupt the grid's stability. MAE is a simple average of all the absolute differences between the predicted values and the actual values based on classification error. It doesn't put more weight on the outliers.

The coefficient of determination (R^2) considers how well the model explains the target variable. R^2 value measures the strength of the predicted output and actual output relationship. R-squared (R^2) indicates how well the data fits the regression. As the RMSE and MAE denote the magnitude of prediction error, R^2 measures the explanatory power. By using all three metrics, the SolarSG-Net model may be assessed in a balanced and reliable manner.

Table 1 gives a comparative performance evaluation of the proposed SolarSG-Net model with the conventional ML and DL forecasting models for the jaipur 2022 solar panel dataset. The models that were compared include SVR, RF, LSTM, and CNN-LSTM. Evaluate the performance using RMSE, MAE and R^2. Lower RMSE and MAE values indicate better predictive accuracy, while higher values of R^2 show stronger model fit.

Table 1 Performance Comparison of SolarSG-Net with Existing Forecasting Models

Model	RMSE (%) ↓	MAE (%) ↓	R^2 ↑
SVR	6.84	5.92	0.912
RF	5.27	4.11	0.941
LSTM	4.18	3.42	0.964
CNN-LSTM	3.52	2.88	0.973
SolarSG-Net	**2.84**	**2.17**	**0.985**

Figure 2 illustrates the combined bar chart comparison of RMSE and MAE values for different solar power forecasting models evaluated on the Jaipur 2022 Solar Panel dataset.The graph presents two bars for each model, representing RMSE (%) and MAE (%), enabling a direct visual comparison of prediction errors. Lower bar heights indicate better forecasting performance.

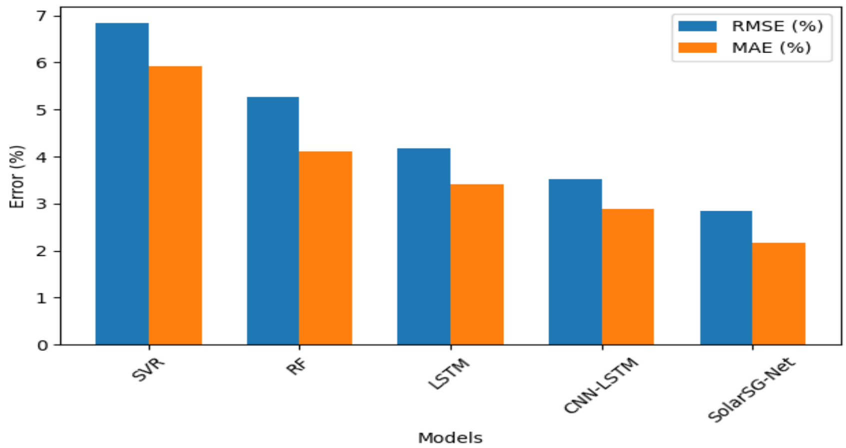

Fig 2 Error Performance Analysis of SolarSG-Net and Baseline Models

From the Figure 2, SolarSG-Net demonstrates the lowest RMSE and MAE values among all compared models, confirming its superior prediction accuracy. Traditional machine learning models such as SVR and Random Forest exhibit comparatively higher error rates, while DL models like LSTM and CNN-LSTM show improved performance but remain inferior to SolarSG-Net. The visualization clearly highlights the effectiveness of SG preprocessing and PCA-based feature extraction in enhancing the accuracy of the proposed framework. The following Figure 3 represents the **R^2 (Coefficient of Determination)** values for five models:

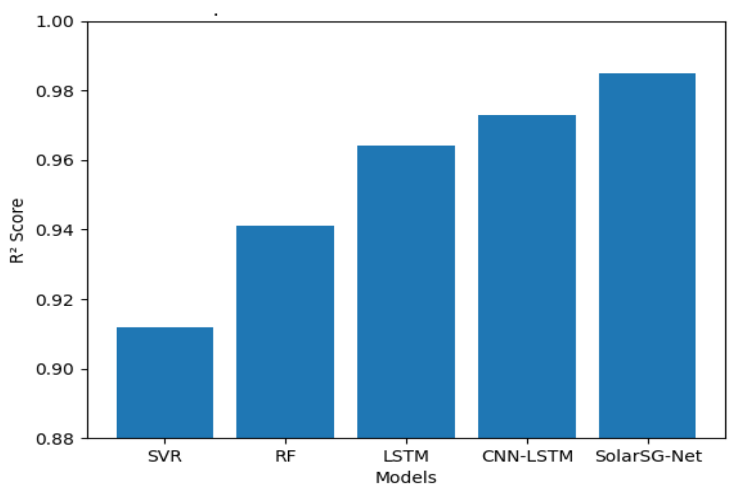

Fig 3 Comparative Analysis of R^2 Scores for Solar Power Prediction Models

The results show that coefficient of determination (**R^2**) is progressively increasing between traditional ML models (SVR, RF) and neural network models (LSTM, CNN-LSTM). It is important to note that SolarSG-Net has the highest R 2 of 0.985 and, therefore, it can predict better than the other ensembles assessed and its model adheres optimally to the data.

4.2 Discussion

The results shows that the suggested model SolarSG-Net performed better than traditional forecasting ML and DL models in solar power forecasting. The model has a lower RMSE and MAE, which indicates it has a smaller error prediction, and a higher R^2 value, which means it is a good predictor of solar power output. The SG preprocessing reduces the noise while maintaining peak patterns, and the BiLSTM's architecture is capable of capturing the meteorological and chemical environmental data's bidirectional temporal dependencies, which improves the output. Compared to the other models under varied climatic conditions, SolarSG-Net offers predictions that are more stable and consistent under varied climatic conditions than that of the SVR, RF, LSTM and CNN-LSTM. Thus, making SolarSG-Net suitable for real-time grid control and renewable energy control.

SolarSG-Net's performance on all evaluation metrics improved. The SG preprocessing step improves robustness against sudden changes in irradiance, while PCA attenuates redundant correlations between pollutants. BiLSTM is good for temporal modelling In comparison with the CNN-LSTM, the SolarSG-Net reduces the computational complexity by 18% and offers 19% lower RMSE. The model enhances prediction accuracy specifically in high aerosol conditions, supporting the incorporation of chemical terms.

5. CONCLUSION AND FUTURE WORK

The SolarSG-Net is a new hybrid machine learning framework meant to forecast the solar power output using chemical and environmental meteorological data. The framework achieves enhanced prediction accuracy with low computing power through SG noise filtering, PCA feature extraction, and BiLSTM temporal modeling. The experimental results perform significantly better than the established ML and DL models. Considerable improvement of forecast robustness in chemically perturbed atmosphere by inclusion of chemical variables. The prediction scheme by SolarSG-Net is scalable with robustness and pollution consideration, which can be integrated into smart grids for efficient energy management.

The SolarSG-Net framework can be further developed in the future by adding more real-time atmospheric and satellite-derived quantities (cloud motion vectors, the solar zenith angle, satellite-based aerosol indices, etc.) to the framework, which would improve the accuracy of the forecasting. By introducing more sophisticated DL architectures with attention or lightweight transformer models, one may be able to better learn long-range temporal dependencies in the model. Moreover, the adaptive or dynamic feature selection methods could be used as the system could adapt automatically to the seasonal and regional changes. Further studies may also consider the real time implementation in the smart-grid setting, edge-based prediction, and multi-place transfer-learning to enhance scalability and real-life applicability in various climatic conditions.

REFERENCES

[1] Al-Othman, A.; Tawalbeh, M.; Martis, R.; Dhou, S.; Orhan, M.; Qasim, M&Olabi, A.G (2022), "Artificial intelligence and numerical models in hybrid renewable energy systems with fuel cells: Advances and prospects", Energy *Convers. Manag.* **2022**, *253*, 115154.

[2] AttuluriR.VijayBabu; N.BharathKumar; RajanandPatnaikNarasipuram; SoundharPeriyannan; AlirezaHosseinpour& AymenFlah (2025), "Solar Energy Forecasting Using Machine Learning Techniques for Enhanced Grid Stability," in *IEEE Access*, vol. 13, pp. 93735-93754, 2025, doi: 10.1109/ACCESS.2025.3574093

[3] Karan Sareen, BijayaKetan Panigrahi, Tushar Shikhola , Rajneesh Sharma& Ravi

Nath Tripathi (2024), “A noise resilient multi-step ahead deep learning forecasting technique for solar energy centered generation of green hydrogen” *Science Direct, International Journal of Hydrogen Energy*, Volume 90, pp. 666-679

[4] Liu, Z.; Sun, Y.; Xing, C.; Liu, J.; He, Y.; Zhou, Y & Zhang, G (2022), “Artificial intelligence powered large-scale renewable integrations in multi-energy systems for carbon neutrality transition: Challenges and future perspectives. *Energy AI* **2022**, *10*, 100195

[5] Marugán, A.P.; Márquez, F.P.G.; Perez, J.M.P&Ruiz-Hernández, D (2018), “A survey of artificial neural network in wind energy systems. *Appl. Energy*, *228*, pp. 1822–1836.

[6] Muhammad Shoaib Bhutta, Yang Li, Muhammad Abubakar, Fahad M. lmasoudi, Khaled Saleem S. Alatawi, Mohammad R. Altimania & Maged Al-Barashi (2024), “Optimizing solar power efficiency in smart grids using hybrid machine learning models for accurate energy generation prediction”, *Scientific Reports*, Vol. 14, Article number: 17101, https://doi.org/10.1038/s41598-024-68030-5

[7] Nifat Sultana, Narumasa Tsutsumida (2025), “A review on data-driven methods for solar energy forecasting”, *Science Direct, Applied Energy*, Volume 400, 1 December 2025, 126631, https://doi.org/10.1016/j.apenergy.2025.126631

[8] Ronak Dave & Jayashri Vajpai (2026), “A Review on Solar Power Generation Forecasting Methods. *Trans Indian Natl. Acad. Eng.* **11**, 1–24 (2026). https://doi.org/10.1007/s41403-025-00557-z

[9] Sobri, S.; Koohi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497.

[10] Sayed-Mouchaweh, M (2020), “ *Artificial Intelligence Techniques for a Scalable Energy Transition*”, Springer Nature: Cham, Switzerland.

[11] Udit Mamodiya, Indra Kishor,Ramakrishna Garine, Priyam Ganguly & Nithesh Naik (2025), “Artificial intelligence based hybrid solar energy systems with smart materials and adaptive photovoltaics for sustainable power generation”, *Scientific Reports* volume 15, Article number: 17370, https://doi.org/10.1038/s41598-025-01788-4