FORECASTING SOLAR ELECTRICITY GENERATION POTENTIAL ON 18 MAJOR RESERVOIR SURFACES IN SRI LANKA USING LONG SHORT-TERM MEMORY (LSTM) MODEL

S.D.A.V.S. Premathilaka*, R.D. Yapa and R. Punchi-Manage

Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka

Sri Lanka's energy sector remains heavily reliant on fossil fuels, with coal and oil generating about 50% of total electricity as of 2023. Despite significant solar irradiance potential, solar energy contributes just 5% to the national grid. A major constraint to large-scale solar expansion is land-use conflict with agriculture, urban development, and ecological conservation. Floating Solar Photovoltaic (*FPV*) systems, deployed on reservoir surfaces, offer a sustainable alternative by utilizing underutilized water bodies without competing for land. This study focuses on forecasting the monthly solar energy potential across 18 major reservoirs in Sri Lanka, considering a 1 m² panel area for each site, using a Long Short-Term Memory (*LSTM*) model. The results of this study will help assess the feasibility of deploying FPV systems. Solar energy generation potential was derived by adjusting measured irradiance for environmental factors such as cloud cover, precipitation, atmospheric pressure, and surface reflectivity, combined with panel efficiency based on a 1 m² panel area. A separate *LSTM* model was trained for each site using a 12-month input sequence to predict the subsequent month's solar energy output. Data were standardized and split into training (80%) and testing (20%) subsets. Each model incorporated dropout and early stopping to mitigate overfitting, and performance was evaluated using Root Mean Squared Error (*RMSE)*, Mean Absolute Error (*MAE*), and Mean Absolute Percentage Error (*MAPE*). Historical data revealed strong seasonality, with average daily outputs ranging from 1–4 kWh/m², and peaks exceeding 5 kWh/m² in certain months. February to April were generally the most productive months. *RMSE* ranged from 0.64 to 0.89, with *MAPE* values below 0.57, indicating relatively accurate forecasts. Site-specific variability influenced model performance. Future studies could explore integrating climate change projections and reservoir-specific operational data to improve long-term forecasting accuracy and support strategic planning for FPV system deployment.

Keywords: Floating Solar Photovoltaics, Long Short-Term Memory, Solar Energy, Forecasting, Reservoir

* *Corresponding Author: akilapr@eng.pdn.ac.lk*

FORECASTING SOLAR ELECTRICITY GENERATION POTENTIAL ON 18 MAJOR RESERVOIR SURFACES IN SRI LANKA USING LONG SHORT-TERM MEMORY (LSTM) MODEL

S.D.A.V.S. Premathilaka*, R.D. Yapa and R. Punchi-Manage

Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka

INTRODUCTION

Sri Lanka's energy sector remains heavily reliant on conventional fossil fuels. In 2023, the energy mix consisted of 30% coal, 20% thermal oil, 38% hydro, and only 12% from Non-Conventional Renewable Energy (*NCRE*) sources. This heavy dependence raises both environmental and energy security concerns. Despite its substantial solar potential of 1,247–2,106 kWh/m² (Wijesena & Amarasinghe, 2018), solar energy contributes just 5% to the mix. A key barrier to large-scale solar deployment is land-use conflict, as it competes with agriculture, urban development, and ecological conservation (Capellán-Pérez et al., 2017). Floating solar photovoltaics (*FPV*), installed on reservoirs, offer an innovative solution by utilizing underused water surfaces and avoiding land competition. The FPV systems also improve efficiency through the cooling effect of water and increase shading, while minimizing water evaporation and inhibiting algae growth (Sahu et al., 2016). This study aims to explore the feasibility of FPV deployment in Sri Lanka by forecasting monthly solar energy potential across 18 major reservoirs using a Long Short-Term Memory (*LSTM*) model. The LSTMs are suited for time series prediction due to their ability to capture long-term dependencies, handle autocorrelated input sequences, and avoid vanishing gradient issues typical in traditional Recurrent Neural Networks (*RNN*).

METHODOLOGY

For this study, we obtained monthly meteorological and solar irradiance data from January 2010 to December 2022 from NASA's POWER Data Access Viewer for 18 major reservoirs in Sri Lanka (See **Figure 2** for names). The solar generation potential was then estimated using the equations below.
Solar energy generation can be estimated using the Equation (1) (Bamisile, 2025).

$$E_{solar} = A \times G_{eff} \times \eta_{eff} \tag{1}$$

where, A is the panel area (m^2), G_{eff} is the effective solar radiation after environmental adjustments, and η_{eff} is the efficiency of the solar panel.

Effective solar radiation is given by the equation (2).

$$G_{eff} = G_{mea} \times (1 - CCo) \times (1 - PCo) \times \frac{P_{ref}}{P_{act}} \times (1 - SAl) \quad (2)$$

where G_{mea} is irradiance (*W/m²*), CCo is cloud coverage, PCo is precipitation correction, P_{ref} is the reference atmospheric pressure, P_{act} and SAl are the actual surface pressure and surface albedo, respectively.
Wind plays an important role in cooling solar panels, which directly affects their efficiency. The panel temperature (T_{panel}) can be estimated using the equation (3).

$$T_{\mathrm{panel}} = T_{\mathrm{ambient}} - k \times WS \quad (3)$$

where T_{ambient} is the ambient temperature (i.e., Temperature at 2 Meters above the ground (°C)), WS is the wind speed (i.e., Wind Speed at 10 Meters above the ground (m/s)), k is the cooling coefficient (typically 0.5–1.0 °C per m/s).
The efficiency of solar panels decreases as their temperature rises above a reference temperature (T_{ref}). This relationship is captured in Equation (4).

$$\eta_{\mathrm{eff}} = \eta_0 \times \left(1 - \alpha \times \left(T_{panel} - T_{\mathrm{ref}}\right)\right) \quad (4)$$

Here η_0 is the nominal efficiency of the panel (i.e., 0.18), α is the temperature coefficient (i.e., 0.004–0.005 per °C), and T_{ref} is the reference temperature (usually 25°C) (Jathar, 2023).

The methodology for processing the collected data and applying the equations is summarized in **Figure 1**. For each reservoir, the raw time series data of solar energy output was first normalized and then split into training and test sets using an 80:20 ratio, preserving temporal order. Following this, sequences of 12 months were generated separately for each set using a sliding window approach, with the subsequent month's energy output as the prediction target. A separate LSTM model was built for each reservoir to capture site-specific temporal patterns.

Each model comprised a single-layer LSTM with 50 units, implemented using the *keras3* library (Kalinowski et al., 2024) in R, and incorporated dropout regularization and early stopping to mitigate overfitting. The input layer received a sequence of E_{solar} values over 12-time steps, with input dimensions of 32 samples, and 1 feature. The model learns temporal patterns through memory cells and gating mechanisms. The models were trained to minimize mean squared error (*MSE*). Performance was evaluated using three standard regression metrics: Root Mean Squared Error (*RMSE*), Mean Absolute Error (*MAE*), and Mean Absolute Percentage Error (*MAPE*).

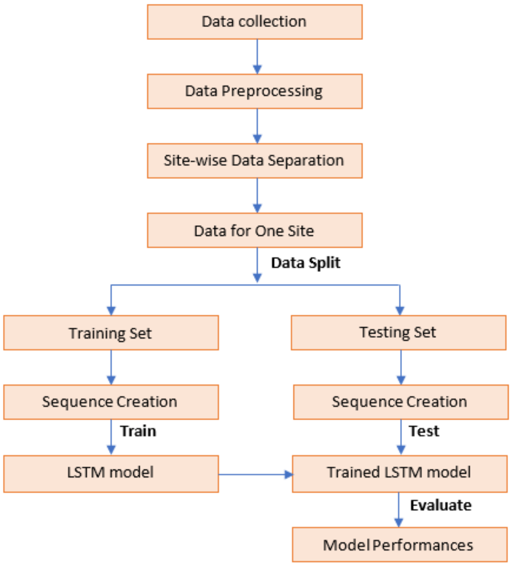

Figure 1: Methodology of the Study

RESULTS AND DISCUSSION

The solar energy potential for each reservoir was assessed based on a 1 m² area. Historical data (January 2010 to December 2022) show strong seasonality in solar energy production. Average daily energy output ranged between 1–4 kWh/m², with certain months exceeding 5 kWh/m² for specific reservoirs. The highest energy production was observed, in most cases, during February, March, and April.

The LSTM models exhibited varied performance across the 18 reservoir sites, reflecting the heterogeneity of site-specific temporal patterns and meteorological influences. **Table 1** summarizes the key evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) for each site.

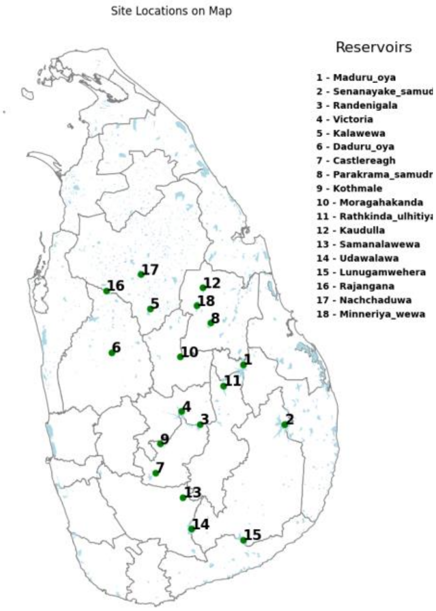

Figure 2: Reservoir Sites

Table 1: LSTM Model Results

SITE_ID	RMSE	MAE	MAPE
1	0.848982	0.723663	0.542252
2	0.893181	0.761157	0.551959
3	0.828649	0.698362	0.51008
4	0.787276	0.658797	0.499623
5	0.828956	0.667727	0.498181
6	0.743387	0.617407	0.484539
7	0.682667	0.579637	0.48024
8	0.736786	0.613298	0.470567
9	0.830304	0.695845	0.509021
10	0.773971	0.641064	0.46962
11	0.808424	0.709786	0.528446
12	0.86528	0.693574	0.488427
13	0.639623	0.529324	0.456416
14	0.648643	0.552815	0.503716
15	0.874179	0.759031	0.566486
16	0.780354	0.64509	0.471043
17	0.819289	0.660022	0.492801
18	0.800262	0.665191	0.488049

The RMSE values ranged from 0.64 to 0.89, indicating moderate absolute prediction errors. Sites 13 and 14 showed the lowest RMSE values, suggesting higher predictive accuracy, while sites 2 and 15 showed higher errors, potentially due to increased variability in solar output or data irregularities. The MAPE values remained under 0.57 across all sites, indicating acceptable relative forecasting performance. Sites 13, 8 and 16 achieved the lowest MAPE scores, highlighting them as more stable and predictable. **Figure 3** presents predicted versus actual outputs, with the actual values shown in blue and the predicted values represented by a dashed line, enabling visual assessment of temporal alignment. Overall, the results affirm the potential of LSTM models for solar energy forecasting.

The results indicate that Victoria Reservoir receives approximately 1.78 kWh of solar energy per square meter per day, which translates to about 53.4 kWh per square meter per month. With a surface area of 24 km², even utilizing a small portion for Floating Solar Photovoltaics (*FPV*) could generate significant solar electricity. Since Victoria Reservoir already operates a hydroelectric power system, integrating FPV could complement the existing generation without requiring additional land. This combined approach offers a promising and sustainable way to increase renewable energy production while minimizing environmental impact.

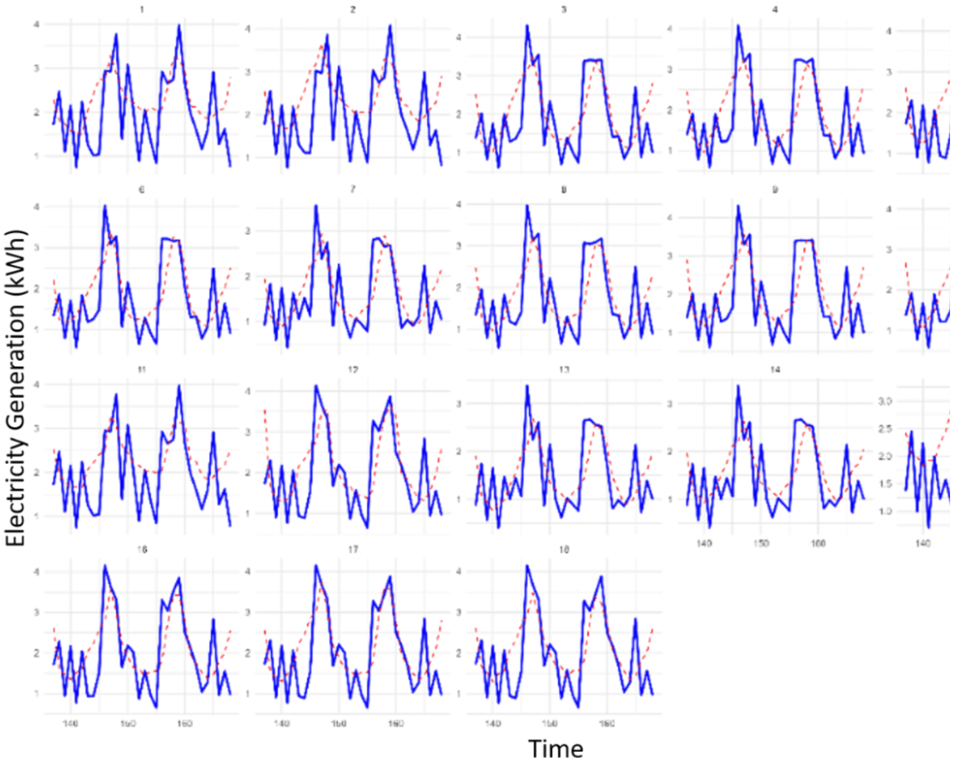

Figure 3: LSTM Predictions of Electricity Generation (kWh) using Solar Energy for All Sites.

CONCLUSIONS/RECOMMENDATIONS

The LSTM models demonstrate moderate accuracy in forecasting solar energy across diverse reservoir sites, with varying performance by location. The LSTM is effective at capturing nonlinear and complex temporal dependencies in the data. The results show that these reservoir sites have significant potential to contribute to electricity generation using solar energy. Future work may consider incorporating additional meteorological features to further enhance predictive performance and generalizability across all sites. Additionally, considering economic, environmental, and social factors will enhance the feasibility assessment of solar energy deployment.

REFERENCES

Journal Articles

Bamisile, O., Acen, C., Cai, D., Huang, Q., & Staffell, I. (2025). *The environmentalfactors affecting solar photovoltaic output. Renewable and Sustainable Energy Reviews*, 208, 115073.

Capell´an-P´erez, I., De Castro, C., & Arto, I. (2017). *Assessing vulnerabilities and limits in the transition to renewable energies: Land requirements under 100% solarenergy scenarios. Renewable and Sustainable Energy Reviews*, 77, 760–782.

Jathar, L. D., et al. (2023). *Comprehensive review of environmental factors influencing the performance of photovoltaic panels: Concern over emissions at various phases throughout the lifecycle. Environmental Pollution*, 326, 121474.

Kalinowski, T., Allaire, J., & Chollet, F. (2024, February). Keras3: R interface to 'keras'.

Sahu, A., Yadav, N., & Sudhakar, K. (2016). *Floating photovoltaic power plant: A review. Renewable and Sustainable Energy Reviews*, 66, 815–824.

Wijesena, G.H.D., & Amarasinghe, A. R. (2018). *Solar energy and its role in Sri Lanka. International Journal of Engineering Trends and Technology - IJETT.*