CS3103 – Machine Learning
Project Report

Group ID: 25GK26

1 Title and Group Details

Project Title: Renewable Energy Production Prediction using Hybrid Machine Learning Models
Group ID: 25GK26

Table 1: Group Members and Their Responsibilities

Name	Roll Number	Role and Responsibilities
Juhi Sahni	2301CS88	Defined overall roadmap and ensured integration. Implemented CNN+LSTM hybrid model and oversaw version control.
Saniya Prakash	2301CS49	Data preprocessing: outlier detection, normalization, and feature engineering. Implemented XGBoost.
Mihika	2301CS31	Built predictive models: implemented Linear Regression and Random Forest.
Shefali Bishnoi	2301CS87	Implemented SVR with multiple kernels and LSTM model. Prepared research documentation.

2 Introduction

Rapid fossil fuel depletion is causing a global energy crisis that calls for an immediate switch to renewable energy sources like solar and wind. However, grid stability and energy management are severely hampered by these resources' intrinsic intermittency and variability brought on by shifting environmental conditions. Therefore, accurate forecasting of renewable energy production is essential for reliable power supply, energy trading, and efficient grid operation.

With an emphasis on both photovoltaic (PV) and wind power generation, this project creates a thorough machine learning framework for forecasting the production of renewable energy. We want to develop reliable prediction models that can capture intricate relationships between environmental factors and energy output by utilising historical meteorological data and cutting-edge machine learning techniques.

3 Brief Related Work

Several machine learning techniques have been investigated in the past for the prediction of renewable energy. In renewable energy prediction tasks, vanishing gradient issues are successfully handled by Long Short-Term Memory (LSTM) networks, which have shown remarkable performance in time-series forecasting applications [2].

Traditional and ensemble approaches to solar energy forecasting have been compared. According to research by Hastomo et al. [3], XGBoost performed noticeably better than Linear Regression, with MAE values of 38.08 and 80.23, respectively. Similarly, in solar energy potential prediction tasks, Random Forest classifiers achieved an impressive 99.82% accuracy [4].

Although these studies offer insightful information, their primary focus is on predicting a single energy source. By creating a thorough framework for multi-output regression of simultaneous PV and wind production using hybrid architectures that take into account both temporal and spatial dependencies in the data, our study fills this gap.

4 Methodology

4.1 Dataset Description

The dataset [1] comprises 38,880 samples collected at 5-minute intervals, featuring comprehensive meteorological parameters including solar irradiance (DHI, DNI, GHI), wind speed, humidity, temperature, and corresponding energy production data for both PV and wind generation. Temporal features such as season and day of week are included, covering a complete annual cycle to enable robust modeling of renewable energy patterns.

4.2 Proposed Architecture

Our suggested CNN-LSTM hybrid architecture combines the advantages of long short-term memory networks for temporal sequence modelling and convolutional neural networks for spatial feature extraction. The temporal dependencies in patterns of energy production as well as the spatial correlations between meteorological features are both well captured by this method.

The mathematical formulation of our hybrid model can be represented as:

$$\hat{Y} = f_{LSTM}(f_{CNN}(X)) \tag{1}$$

where X represents the input feature matrix and $\hat{Y}$ contains the predicted PV and wind production values.

The convolutional operation extracts spatial features:

$$H_{conv} = \sigma(W_{conv} * X + b_{conv}) \tag{2}$$

The LSTM component processes temporal sequences through its gating mechanism:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{3}$$

$$h_t = o_t \odot \tanh(C_t) \tag{4}$$

In order to normalise the dataset, MinMaxScaler was used for preprocessing, which included feature engineering and careful handling of missing values. To ensure robust evaluation, the data was divided into training (70%) and test (30%) sets.

5 Key Results

Table 2: Performance Comparison of Different Models

Model	**PV R^2**	**Wind R^2**	**PV MAE**	**Wind MAE**	**Best Configuration**
Linear Regression	0.903	0.481	784.2	692.4	-
SVR (RBF)	0.911	0.554	652.0	615.3	RBF Kernel
Random Forest	0.983	0.946	198.8	142.2	-
XGBoost	0.971	0.837	318.1	348.6	-
LSTM	0.980	0.962	506.8	188.2	Window=48, tanh/hard_sigmoid
CNN-LSTM	**0.983**	**0.965**	**367.7**	**178.5**	Window=48, tanh/sigmoid/elu

The suggested CNN-LSTM hybrid architecture performs better, as shown by our experimental results. The hybrid model obtained excellent R^2 scores of 0.983 for PV prediction and 0.965 for wind prediction, with corresponding MAE values of 367.7 and 178.5, respectively, as indicated in Table 2.

With R^2 values of 0.983 and 0.946 for PV and wind prediction, respectively, the Random Forest model demonstrated excellent performance and obtained the lowest MAE for PV production (198.8). With an MAE of 178.5, the CNN-LSTM model, on the other hand, performed better across both prediction tasks. It was especially good at forecasting wind production.

With R^2 scores of 0.980 and 0.962 for PV and wind prediction, respectively, the LSTM model likewise demonstrated strong performance; however, its MAE values were marginally higher than those of the CNN-LSTM architecture. By achieving strong results with R^2 values of 0.971 and 0.837 for PV and wind prediction, respectively, XGBoost outperformed the conventional Linear Regression and SVR models, demonstrating the benefits of deep learning techniques for this challenging prediction task.

The CNN-LSTM architecture's superior and well-balanced performance was a result of its capacity to capture temporal dependencies through LSTM units and spatial relationships in meteorological data through convolutional layers. With a window size of 48 time steps and particular activation functions (tanh for CNN, sigmoid for LSTM gates, and elu for dense layers), the model showed strong generalization across various seasonal patterns and meteorological conditions.

6 Conclusion

A hybrid CNN-LSTM architecture for predicting the production of renewable energy was successfully developed and assessed in this project. The suggested model produces better prediction accuracy for both PV and wind energy generation by successfully capturing both spatial and temporal dependencies in meteorological data.

The thorough comparison with several baseline models shows how well deep learning techniques work for forecasting renewable energy. With R^2 values of 0.983 for PV prediction and 0.965 for wind prediction, along with MAE values of 367.7 and 178.5, respectively, the CNN-LSTM hybrid architecture demonstrated notable advancements over conventional machine learning techniques such as SVR and Linear Regression.

References

[1] S. R. Ortega, P. Castro-Correa, S. Sepúlveda-Mora, and J. Castro-Correa, "Renewable energy and electricity demand time series dataset with exogenous variables at 5-minute interval," *Mendeley Data*, vol. 1, 2023.

[2] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 1–44, 2020.

[3] W. Hastomo, A. Digdoyo, A. Satyo, and D. Arif, "Enhancing solar energy efficiency: Predictive modeling with XGBoost and linear regression," *JIKA (Jurnal Informatika)*, vol. 9, no. 1, pp. 66–74, 2025.

[4] N.P, M.Sari, and K.Kusnanto, "Predicting the potential of renewable solar energy based on weather data in Indonesia using the random forest method," *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 4, pp. 1771–1781, 2025.