

Advanced machine learning techniques for short-term solar irradiance forecasting

Pratima Kumari*

pratimatomar93@gmail.com
India Institute of Technology
Roorkee, Uttarakhand, India

Durga Toshniwal

India Institute of Technology
Roorkee, Uttarakhand, India

Abstract

The ever-growing population, their dependence on electricity, and awareness of the tremendous environmental impact of burning fossil fuels has expanded the renewable energy portfolios across the globe. Solar energy is arising as the most promising alternative to the fossil fuels. Among all the techniques that transforms solar energy to electricity, solar photovoltaic (PV) is becoming the most popular due to its simplicity and cheap maintenance. Although the solar energy has the advantage of being limitless and clean over conventional resources, it brings along several challenges. The PV power output is highly volatile as it depends on several meteorological factors, including solar irradiance, temperature, cloud cover, rainfall, etc. Solar energy is also an intermittent energy source as it only exists during day time. The uncertain and intermittent nature of the solar energy is the main hindrance in its reliable market penetration. The variability of solar power output affects the grid balance system and increases their operational costs. Therefore, with the increased installation of PV plants across the globe, accurate forecasting models are highly desirable for the successful integration of solar energy to the grid and proper functioning of energy industry. Present thesis is focused on developing advanced machine learning techniques by employing different type of input features to enhance the forecasting accuracy.

Keywords: Machine learning, solar energy, forecasting, artificial neural network, long short term memory

ACM Reference Format:

Pratima Kumari and Durga Toshniwal. 2021. Advanced machine learning techniques for short-term solar irradiance forecasting. In *Proceedings of International Conference on AI-ML Systems (AI-ML Systems'21)*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AI-ML Systems'21, October 21–24, 2021, Bengaluru, India

© 2021 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 Introduction

The era of abundance is approaching the end faster than ever. The Earth's resources are limited and demand a wise use to ensure the future sustainability of life. Fossil fuels are one of the most vital and necessary earth resources. The ever-growing population, rapid industrialization, urbanization and other development activities primarily depend on electricity, which has caused a substantial consumption of fossil fuels in past few decades [10]. The recent advancements in tools and technologies have also promoted the extraction of energy from different renewable energy sources, including wind, solar, geothermal, etc [13]. Solar energy has received much attention from industrial and research communities due to its clean and abundant nature [1]. Although solar energy has the advantage of being limitless and clean over conventional resources, it brings along several challenges in reliable market penetration [6].

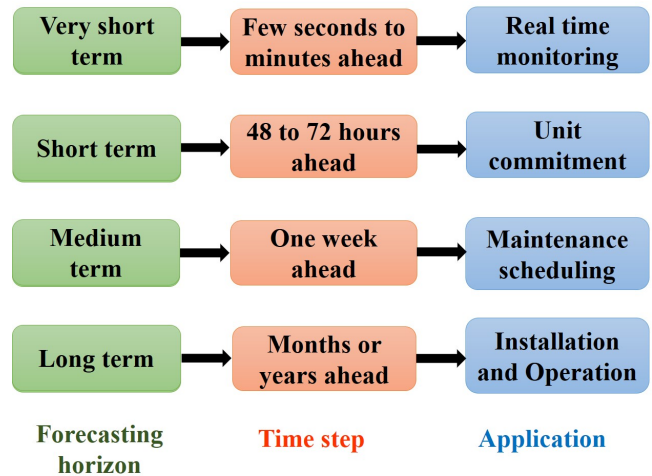


Figure 1. Forecasting horizon and the time step with their applications.

Solar energy is highly intermittent and volatile in nature, which depends on several meteorological factors, including solar irradiance, temperature, cloud cover, rainfall, etc. [4]. Consequently, accurate solar energy prediction is highly required for the successful integration of solar power to the grid. Usually, solar energy industries avoid sharing of PV data due to their privacy terms and policies [7]. Therefore,

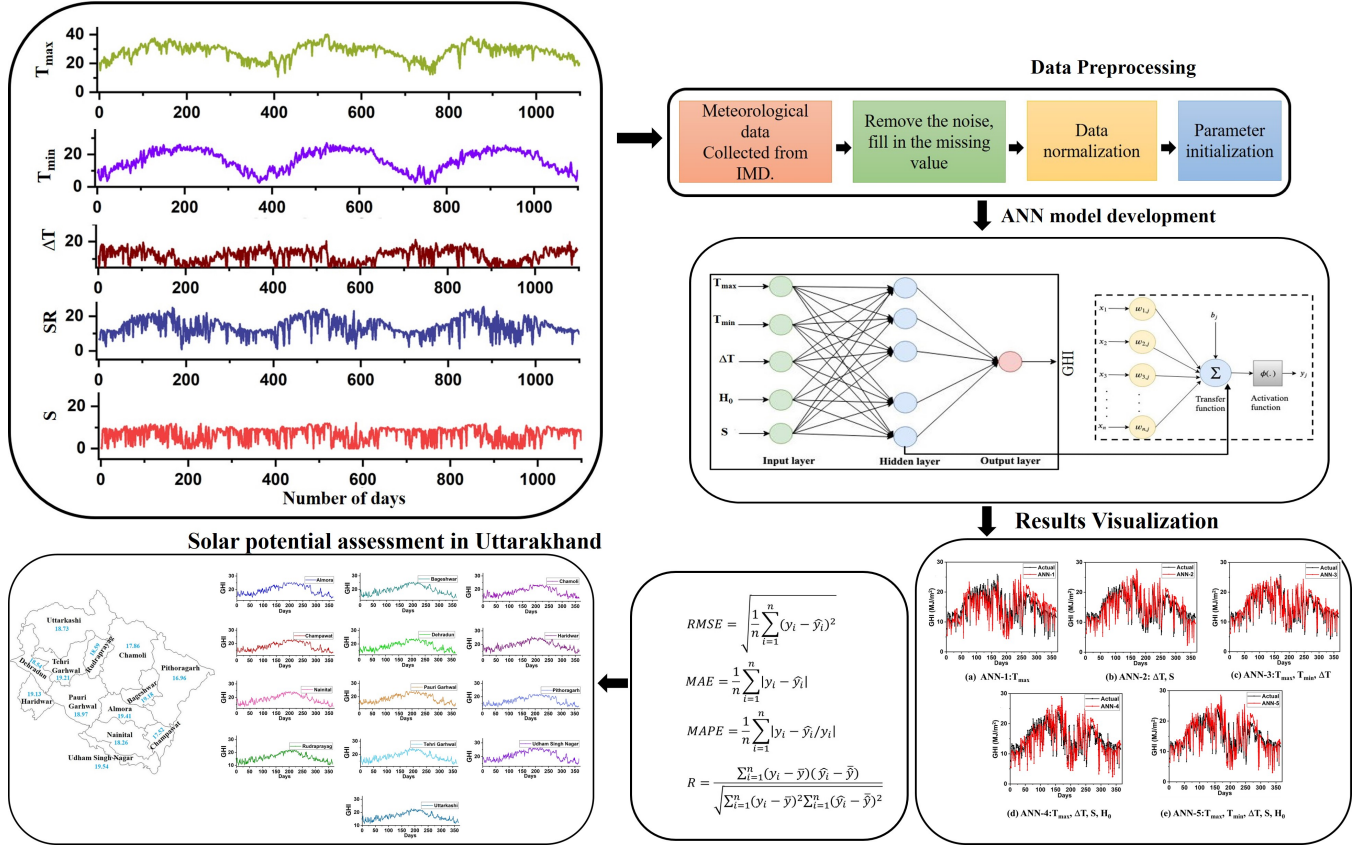


Figure 2. The block-diagram of the proposed framework.

researchers are more focused towards Global Horizontal Irradiance (GHI) forecasting. GHI forecasting plays significant role in various solar energy applications, such as demand and supply balancing, fault detection, load dispatching, maintenance scheduling, site selection, etc as shown in Fig. 1 [2]. For this reason, the research on developing solar irradiance forecasting models has gained significant attention from solar industry as well as scientific community in past two decades [3]. With the development of artificial intelligence techniques, machine learning-based solar irradiance forecasting models provide more promising performance than physical and statistical methods. Several machine learning models, such as artificial neural network [8], support vector machine [5], extreme learning machine [14] are extensively applied for solar irradiance forecasting to handle the nonlinear relationship between input and output variables.

2 Contributions

This thesis focuses on using machine learning, deep learning and ensemble learning approaches for short-term, i.e. hourly and daily solar irradiance forecasting. We analysed the limitations of the state-of-the-art literature in this domain and proposed new techniques to address the identified limitations and enhance the prediction accuracy. The main

research contributions of this thesis can be summarized as follows:

2.1 Selection of important meteorological parameters

We applied Artificial neural networks (ANNs) for daily solar irradiance forecasting [9]. Fig. 2 represents the flowchart of the proposed work. As using a large number of meteorological variables for model development increases the computational cost unnecessarily, it is beneficial to identify the most influential parameters. To conduct this study, a mountainous state of Uttarakhand, India is selected in this work. Initially, the ANN based GHI prediction models are developed with different combinations of meteorological variables, which includes minimum temperature (T_{min}), maximum temperature (T_{max}), temperature difference (ΔT), GHI, extraterrestrial radiation (H_0), and bright sunshine hours (S). We developed five types of ANN models (ANN-1 to ANN-5) with 32 input combinations for daily GHI prediction. The models are trained and tested on the data of one (i.e. Dehradun) out of thirteen districts of Uttarakhand. The best performing ANN model is selected from the 32 ANN models, which came out to be ANN-3 trained with T_{max} , T_{min} , ΔT . Further, this

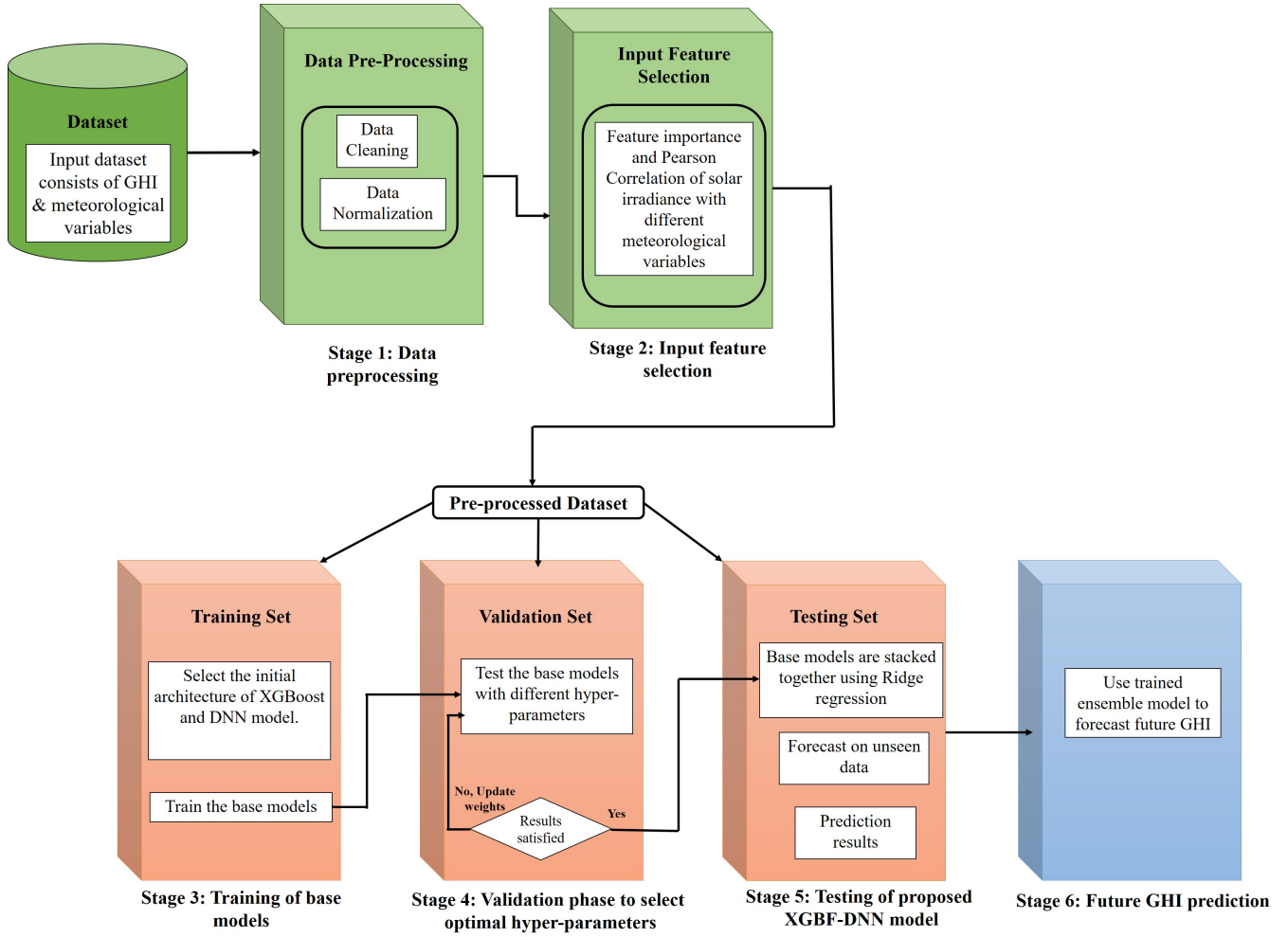


Figure 3. The schematic block diagram of proposed framework.

best ANN model is utilized for solar potential forecasting of twelve remaining districts of the Uttarakhand.

2.2 Ensemble learning based prediction model

We developed a novel ensemble model (XGBF-DNN) for hourly GHI forecast, which integrates extreme gradient boosting forest and deep neural networks [11]. The framework of the proposed methodology is demonstrated in Fig. 3. In order to eliminate over-fitting, ridge regression is employed to integrate the base models. The diversity of base models is also ensured in the proposed framework, as it is considered as the key of efficient ensemble models. Further, the feature selection is also performed and temperature, clear-sky index, relative humidity, and hour of the day are considered as the most important input features. To validate the performance of the proposed ensemble model, the model is assessed on three locations of India with different climate types, which includes New Delhi, Gangtok and Jaipur. Moreover, in order

to give a profound understanding of the model's characteristics, a seasonal analysis is also conducted.

2.3 A deep hybrid spatio-temporal features based prediction model

We proposed a novel deep learning based hybrid for hourly GHI forecasting, named as LSTM-CNN [12]. Proposed model integrates long short term memory (LSTM) and convolutional neural network (CNN) model to extract the spatio-temporal features from the data. The training of the model is done using the meteorological data of 23 sites in California state, USA. The input features used to train the model includes relative humidity, temperature, cloud cover, precipitation, pressure, etc. Initially, the data is pre-processed and organized properly in two forms. Since, the future values are highly influenced by the historical data, the historical GHI data is prepared for LSTM model. Moreover, the meteorological data of nearby locations also influences the solar irradiance of target location. Therefore, the meteorological

data of neighbour locations is also considered to extract the spatial information from it. For this reason, the meteorological data of target and its neighbour locations is utilized by CNN model. The proposed hybrid LSTM-CNN model firstly uses LSTM to extract the temporal features from historical time-series of solar irradiance data, followed by CNN, which extracts the spatial features from the correlation matrix of several meteorological variables of target and its neighbour location. The performance of the proposed model is rigorously examined for an year, different seasons (summer, winter, spring, autumn) and different sky conditions (sunny, mixed, cloudy).

3 Future work

In future, we propose to apply several advanced machine learning and deep learning-based GHI forecasting models for different smart cities of India. For the proposed work, the dataset of 21 cities, namely Agartala, Ahemdabad, Bhopal, Bhubneshwar, Chennai, Coimbatore, Davangere, Diu, Guwahati, Imphal, Indore, Jaipur, Kakinada, Kochi, Ludhiana, New Delhi, Pune, Solapur, Surat, Udaipur and Vishakhapatnam located in different climatic zones of India, selected under "Smart cities mission" are used for training and testing of the models. Based on the prediction accuracy of the developed models for considered cities, we can select the most appropriate cities for installation of solar power plants on a large scale. Alternative non-conventional energy resource projects such as wind energy, tidal energy should be preferred at sites where harnessing solar energy is not profitable due to unfavorable weather conditions or inefficient future prediction.

Acknowledgments

The first author is grateful to Ministry of Education, Government of India for the doctoral fellowship. The authors would like to express their gratitude towards Director, IITR for providing the research facilities.

References

- [1] Razin Ahmed, V Sreeram, Y Mishra, and MD Arif. 2020. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews* 124 (2020), 109792.
- [2] Kelsey Fahy, Michael Stadler, Zachary K Pecanak, and Jan Kleissl. 2019. Input data reduction for microgrid sizing and energy cost modeling: Representative days and demand charges. *Journal of Renewable and Sustainable Energy* 11, 6 (2019), 065301.
- [3] Mawloud Guermoui, Farid Melgani, Kacem Gairaa, and Mohamed Lamine Mekhalfi. 2020. A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production* 258 (2020), 120357.
- [4] Yan Hao and Chengshi Tian. 2019. A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. *Applied energy* 238 (2019), 368–383.
- [5] He Jiang and Yao Dong. 2017. Forecast of hourly global horizontal irradiance based on structured kernel support vector machine: A case study of Tibet area in China. *Energy Conversion and Management* 142 (2017), 307–321.
- [6] Kazım Kaba, Mehmet Sarıgül, Mutlu Avcı, and H Mustafa Kandirmaz. 2018. Estimation of daily global solar radiation using deep learning model. *Energy* 162 (2018), 126–135.
- [7] Pratima Kumari and Durga Toshniwal. 2019. Hourly solar irradiance prediction from satellite data using lstm. (2019).
- [8] Pratima Kumari and Durga Toshniwal. 2020. Real-time estimation of COVID-19 cases using machine learning and mathematical models-The case of India. In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 369–374.
- [9] Pratima Kumari and Durga Toshniwal. 2021. Analysis of ANN-based daily global horizontal irradiance prediction models with different meteorological parameters: a case study of mountainous region of India. *International Journal of Green Energy* (2021), 1–20.
- [10] Pratima Kumari and Durga Toshniwal. 2021. Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production* (2021), 128566.
- [11] Pratima Kumari and Durga Toshniwal. 2021. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *Journal of Cleaner Production* 279 (2021), 123285.
- [12] Pratima Kumari and Durga Toshniwal. 2021. Long short term memory-convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Applied Energy* 295 (2021), 117061.
- [13] Pratima Kumari and Rajesh Wadhvani. 2018. Wind power prediction using klm algorithm. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 154–161.
- [14] Shahaboddin Shamsirband, Kasra Mohammadi, Lip Yee, Dalibor Petković, and Ali Mostafaeipour. 2015. A comparative evaluation for identifying the suitability of extreme learning machine to predict horizontal global solar radiation. *Renewable and sustainable energy reviews* 52 (2015), 1031–1042.