Solar Irradiance Forecasting System Using Deep Learning Algorithms

A PROJECT REPORT

Submitted in partial fulfillment of the requirement for the award of the degree

of
BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE AND ENGINEERING

SUBMITTED BY

Gurmaandeep Singh (21103052)

Kamalpreet Kaur (21103070)

Simranpreet Singh (21103141)

Under the supervision of

Dr. Banalaxmi Brahma

Assistant Professor

Department of Computer Science and Engineering
Dr. B. R. Ambedkar National Institute of Technology, Jalandhar
-144008, Punjab (India)
May 2024

CANDIDATES’ DECLARATION

We hereby certify that the work presented in this project report entitled “**Solar Irradiance Forecasting System Using Deep Learning Algorithms**” in partial fulfillment of the requirement for the award of a Bachelor of Technology degree in Computer Science and Engineering, submitted to the Dr. B R Ambedkar National Institute of Technology, Jalandhar is an authentic record of our own work carried out during the period from July 2023 to May 2024 under the supervision of Dr. Banalaxmi Brahma, Assistant Professor, Department of Computer Science & Engineering, Dr. B R Ambedkar National Institute of Technology, Jalandhar.

We have not submitted the matter presented in this report to any other university or institute for the award of any degree or any other purpose.

Date: 27th May, 2024

Submitted by

Gurmaandeep Singh(21103052)

Kamalpreet Kaur (21103070)

Simranpreet singh (21103141)

This is to certify that the statements submitted by the above candidates are accurate and correct to the best of our knowledge and are further recommended for external evaluation.

Dr. Banalaxmi Brahma, Supervisor
Assistant Professor
Deptt. of CSE

Dr. Rajneesh Rani
Head and Associate Professor
Deptt. of CSE

ACKNOWLEDGEMENT

It is often recognized that the success of any endeavor is due to the contributions of many behind the scenes. Our project reached completion thanks to the guidance and assistance of many individuals, and we feel incredibly fortunate to have had their support throughout. Our accomplishments are a testament to their invaluable help, and we express our sincere gratitude to all.

We are especially thankful to our project mentor, Dr. Banalaxmi Brahma, whose belief in our concepts and timely suggestions provided the direction we needed. His consistent support was crucial in overcoming the obstacles we encountered.

We also extend our heartfelt thanks to Dr. Rajneesh Rani, Head of the Department of Computer Science and Engineering, for her ongoing and indirect support.

Our gratitude also goes to Dr. Aruna Malik, Coordinator Major Project, for assigning us mentors and ensuring we had the support we needed.

We are deeply appreciative of the continuous encouragement and guidance from all the faculty members of the Department of Computer Science & Engineering. Additionally, we would like to thank all the laboratory staff for their prompt support.

Thank You.

[Gurmaandeep, Kamalpreet, Simranpreet]

ABSTRACT

The extensive use of fossil fuels has significantly contributed to global warming and environmental degradation by releasing large amounts of greenhouse gases into the atmosphere. To address this issue, predicting solar irradiance through time series forecasting is vital. Such predictions help determine the best locations for renewable energy projects and the suitable technologies to use. Accurate day-ahead hourly solar irradiance forecasts are crucial for optimizing solar energy production and ensuring effective grid management. In this study, we present a novel approach using a CNN-LSTM using Bayesian optimization with attention mechanism for high-precision solar irradiance prediction. Our model combines spatial and temporal data attributes, integrating various meteorological and geographical features to improve prediction accuracy. By employing CNN to capture spatial dependencies and LSTM being proficient in capturing temporal dependencies and long-range dependencies within sequential data, Adding Bayesian optimization into the mix elevates the model's performance by intelligently tuning hyperparameters. And the incorporation of an attention mechanism enhances the model's ability to focus on relevant meteorological features. Our method effectively addresses the challenges of non-stationarity and spatial variability in solar irradiance data. We start by gathering and preprocessing a detailed dataset that includes solar irradiance measurements along with relevant meteorological variables like temperature, humidity, and clear-sky GHI. To account for spatial variations, this dataset is enriched with geographical information such as latitude, longitude, and altitude. Our main approach involves constructing a graph where nodes represent various locations, and edges denote spatial relationships based on geographical proximity and meteorological similarity. Through extensive experiments, our model shows superior performance compared to traditional methods, delivering accurate and reliable forecasts, particularly in capturing short-term fluctuations and extreme values of solar irradiance. This enhanced forecasting capability is crucial for optimizing solar energy production and improving grid management efficiency.

PLAGIARISM REPORT

We have checked plagiarism for our Project Report for our project a **Turnitin.** We are thankful to our mentor Dr. Banalaxmi Brahma for guiding us at this. Below is the digital receipt. The Plagiarism is approximately 15%.

ORIGINALITY REPORT

15%	12%	5%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	2%
2	open-innovation-projects.org Internet Source	1%
3	Zhao Lin, Yanbing Xue, Zhigang Wang, Jing Cai. "Remote Physiological Measurement With Multiscale Feature Extraction and Global Attention Module", IEEE Transactions on Consumer Electronics, 2024 Publication	1%
4	Submitted to London Metropolitan University Student Paper	1%
5	Submitted to University of Greenwich Student Paper	1%
6	Submitted to Taylor's Education Group Student Paper	1%
7	Submitted to Anatolia College Student Paper	<1%

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
R2	R Square
RMSE	Root Mean Square Error
MSE	Mean Square Error
MAE	Mean Absolute Error
DRL	Deep Reinforcement Learning
GHI	Global Horizontal Irradiance

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

The widespread use of fossil fuels has led to significant environmental issues, including global warming and air pollution, due to the emission of large amounts of greenhouse gases. Transitioning to cleaner energy sources, such as solar power, is crucial for reducing this dependency. However, effectively integrating solar energy into the power grid necessitates accurate solar irradiance forecasting.

This line explains the importance of accurate day-ahead solar irradiance forecasts for several critical aspects of power grid management. It highlights the necessity of these forecasts for optimizing power grid operations, improving the integration of renewable energy sources, and ensuring a stable electricity supply. Additionally, it points out that traditional forecasting methods often struggle to handle the complex dependencies present in solar irradiance data, which are influenced by a variety of meteorological, geographical, and temporal factors.

This project contributes to society by addressing these challenges. Enhanced solar irradiance predictions facilitate better grid management, reducing the risk of power outages and improving electricity supply stability. This also helps decrease reliance on fossil fuels, thereby reducing greenhouse gas emissions and mitigating climate change. By promoting sustainable energy solutions, this project supports the transition to a cleaner, more resilient energy system.

Our research develops a CNN-LSTM with Bayesian optimization using attention mechanism for day-ahead hourly solar irradiance forecasting. This model integrates various meteorological attributes and spatial relationships to enhance accuracy, leveraging spatial and temporal data to provide reliable forecasts.

1.2. Literature Survey

The integration of advanced machine learning techniques in solar irradiance forecasting has gained significant attention in recent years. Among these, deep learning models have shown substantial promise due to their ability to capture complex patterns from raw data. This survey reviews key contributions and methodologies in the field, highlighting the role of Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and ensemble strategies in improving forecasting accuracy. []

One notable approach involves the utilization of LSTM-CNN models for forecasting solar irradiance datasets. The hybrid nature of LSTM-CNN models allows for the extraction of temporal and spatial features effectively. LSTMs are adept at handling sequential data and capturing long-term dependencies, while CNNs excel in recognizing spatial patterns. The document discusses an optimization of these models using a modified whale optimization algorithm. Whale optimization, inspired by the bubble-net hunting strategy of humpback whales, has been tailored to enhance the performance of LSTM-CNN models, ensuring more accurate and reliable forecasts of solar irradiance. This combination leverages the strengths of both architectures, leading to a robust predictive model that addresses the inherent variability in solar irradiance data. []

Beyond individual models, ensemble strategies have been employed to further improve forecasting performance. An ensemble strategy based on deep reinforcement learning (DRL) algorithms has been highlighted. DRL, which combines deep learning with reinforcement learning principles, is used to select the best combination of deep-optimized LSTM-CNN models. This approach benefits from the diverse strengths of multiple models, enhancing overall prediction accuracy and robustness. By continuously learning and adapting to the data, the ensemble strategy ensures optimal model selection, thereby improving the reliability of solar irradiance forecasts. []

The broader application of deep learning techniques, including CNNs, RNNs, LSTMs, and attention models, has revolutionized various fields by enabling the automatic discovery of complex patterns and representations from raw data. CNNs are particularly effective in image recognition tasks, but their application in solar irradiance forecasting has demonstrated their versatility in extracting spatial features from meteorological data. RNNs, and more specifically LSTMs, are powerful in modeling time series data, capturing temporal dependencies that are crucial for accurate forecasting. Attention models further enhance these capabilities by allowing the model to focus on relevant parts of the data, improving interpretability and prediction accuracy. []

The convergence of these advanced methodologies marks a significant advancement in solar irradiance forecasting. The integration of LSTM-CNN models optimized by novel algorithms, coupled with ensemble strategies, showcases the potential of deep learning in addressing the challenges posed by traditional forecasting methods. By leveraging the strengths of various deep learning architectures, researchers can develop more accurate and reliable predictive models, ultimately supporting the transition to sustainable energy systems and contributing to global efforts in mitigating climate change. []

This literature survey provides an overview of the advancements in solar irradiance forecasting, focusing on the role of deep learning techniques and ensemble strategies in enhancing prediction accuracy and reliability.

1.3. Problem Statement and its Necessity

Problem Statement

Conventional linear methodologies prove insufficient, necessitating a paradigm shift towards a model that boasts the dexterity to furnish precise prognostications, blending the virtues of both linear and non-linear functionalities. Present techniques overlook biases, thereby compromising the fidelity of solar irradiance predictions. Furthermore, existing approaches exhibit constrained accuracy and dependability, especially concerning the encapsulation of ephemeral fluctuations and abrupt shifts in solar irradiance patterns.

Necessity

1. **Environmental Impact:** The rising global population has driven up the use of fossil fuels and the release of greenhouse gases, making environmental pollution and global warming worse. Shifting to renewable energy sources, such as solar power, is essential for tackling these problems.

2. **Renewable Energy Goals:** Countries around the world, including the United States, China, and the European Union, have pledged to cut down on greenhouse gas emissions and ramp up the use of renewable energy. To make good on these commitments, we need dependable models for forecasting solar energy, ensuring that we can produce energy steadily and efficiently.

3. **Technological Limitations:** Many current forecasting models focus only on a handful of weather factors and don't take full advantage of the atmospheric context. This study suggests a more thorough model that considers time trends, spatial effects, and connections between different weather elements to enhance prediction precision.

4. **Practical Application:** a solar irradiance forecasting system provides actionable insights that support decision-making across the entire solar energy value chain, from power plant operations and grid management to energy trading and consumption planning. By leveraging accurate and reliable forecasts, stakeholders can maximize the economic and environmental benefits of solar energy integration while enhancing energy system resilience and sustainability.

1.4. Motivation

1. **Reduction of environmental footprint**: The rising use of fossil fuels has resulted in greater greenhouse gas emissions, playing a substantial role in global warming and pollution. Enhancing solar irradiance forecasting can expedite the shift to renewable energy, diminishing environmental harm and fostering sustainability.

2. **Energy Dependability**: Solar energy generation fluctuates naturally because of weather and atmospheric changes. Precise forecasting models aid in steadying the electricity output from solar installations, guaranteeing a dependable and steady energy supply. This consistency is vital for seamlessly integrating solar power into national electricity networks without causing interruptions.

3. **Economic Benefits**: solar irradiance forecasting offers a range of economic benefits, including increased energy production efficiency, reduced grid management costs, enhanced trading opportunities, risk mitigation, and attractiveness to investors. By leveraging accurate forecasts, stakeholders can unlock the full economic potential of solar energy while contributing to a more sustainable and resilient energy future.

4. **Advancement in Technology**: Existing models often fall short in accuracy due to limited consideration of meteorological variables and spatiotemporal correlations. Developing more sophisticated models integrating these aspects can significantly advance the solar energy forecasting technology.

5. **Global Energy Goals**: Accurate solar irradiance forecasting models are vital for countries striving to meet ambitious targets for reducing greenhouse gas emissions and increasing renewable energy adoption. By providing precise predictions of solar energy generation, these forecasting models enable reliable and efficient utilization of solar power to meet growing energy demands. This ensures that renewable energy sources, like solar, can play a significant role in achieving national and global sustainability objectives, while also enhancing energy security and resilience.

6. **Research and Innovation**: In solar irradiance forecasting, innovation is accelerating with the integration of advanced techniques like CNN LSTM models, and attention mechanisms. These methods enhance predictive accuracy by capturing complex patterns in solar data and efficiently optimizing model performance. This innovative approach

holds promise for significantly improving the reliability and efficiency of solar energy integration into the global energy system.

1.5. Feasibility: Non-Technical and Technical

It is very crucial to have a plan for whether a project is feasible from different standpoints.

1.5.1 Technical Feasibility: Technical Feasibility:

Deep learning algorithms, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, show promise for solar irradiance forecasting due to their ability to capture complex temporal and spatial relationships in data. By leveraging vast datasets, including those sourced from organizations like NASA, these algorithms can effectively learn patterns and dependencies, enhancing forecasting accuracy. Furthermore, the integration of attention mechanisms further refines model performance, making the proposed CNN LSTM framework technically feasible for solar irradiance prediction.

1.5.2 Non-Technical Feasibility:

While the technical capabilities of deep learning algorithms for solar irradiance forecasting are promising, non-technical factors also influence feasibility. Data availability, particularly from reputable sources like NASA, is critical for model training and validation. The successful implementation of the CNN LSTM model relies on access to comprehensive and reliable solar irradiance data, as well as supporting meteorological information. Additionally, considerations such as computational resources, expertise in deep learning, and regulatory compliance impact feasibility. Stakeholder acceptance and collaboration, along with effective communication of the model's benefits, are essential for successful adoption. Overall, the combination of technical advancements and robust data sources, coupled with careful consideration of non-technical factors, contributes to the feasibility of deploying CNN LSTM with Bayesian optimization and attention mechanisms for solar irradiance forecasting.

1.6 Research Objectives

This study focuses on advancing the field of solar irradiance forecasting through the development of a sophisticated deep learning framework. The proposed model, termed CNN LSTM using attention mechanism, aims to excel in both robustness and accuracy. By harnessing the intricate spatiotemporal relationships inherent in meteorological data, the model endeavors to minimize forecasting errors significantly. Its ultimate goal is to revolutionize the efficiency and dependability of solar power generation. Through precise predictions, this model aspires to catalyze the adoption of sustainable energy management practices, thereby fostering a greener and more resilient energy landscape.

CHAPTER 2

PROPOSED SOLUTIONS

2.1 Proposed methods :

The model generation stage performs the training of all forecasting methods (CNN-LSTM , CNN , ATTENTION , LINEAR REGRESSION , LSTM BASED RNN) to produce a pool of models. The model generation stage is crucial to the ensemble, considering that the models should be as accurate and diverse as possible. The accuracy is achieved through a search methodology over the space of the possible parameter configurations of all models, and diversity is achieved because different models with distinct learning algorithms are used in the pool

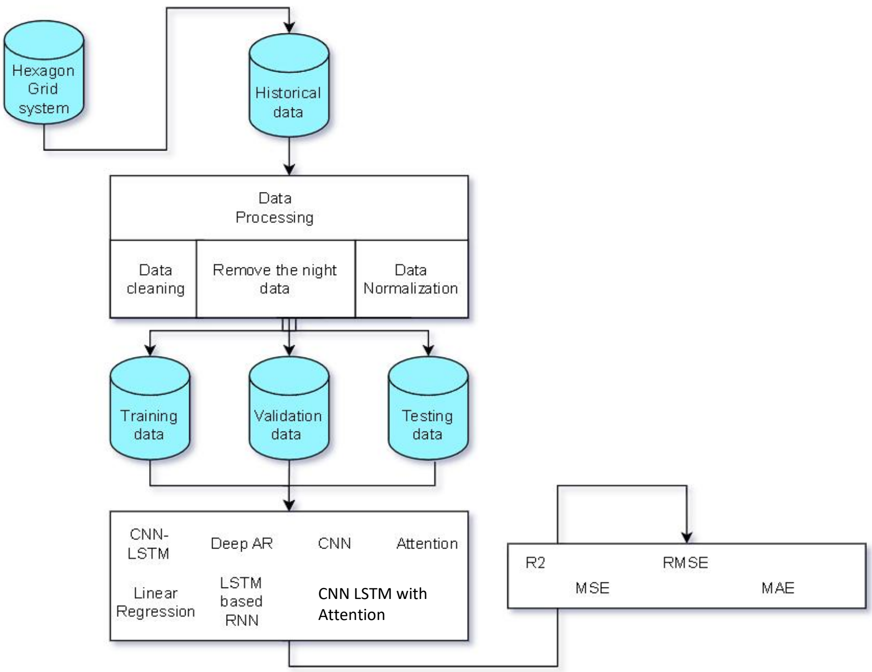

Fig .2.1

2.1.1 Hexagonal Grid System:

To effectively capture the spatial variability of solar irradiance, we propose using a hexagonal grid system. Hexagons provide a more uniform coverage and reduce spatial distortions compared to traditional square grids. Each hexagon represents a distinct spatial unit where data is collected and analyzed, ensuring better spatial feature extraction. This approach enables more accurate and reliable forecasts by capturing the intricate spatial patterns of solar irradiance more effectively.

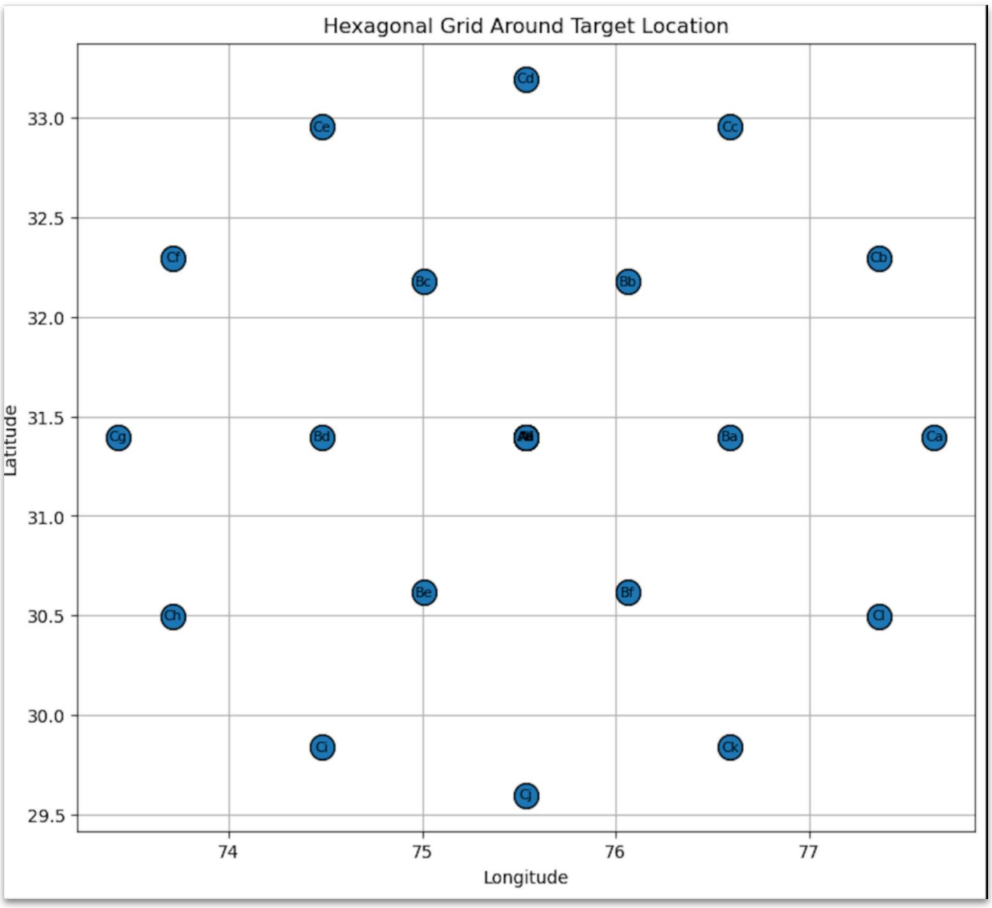

Fig .2.2

2.1.2 Models Used:

Our methodology utilizes several advanced deep learning models to capture both spatial and temporal patterns in solar irradiance data. Convolutional Neural Networks (CNNs) are employed to extract spatial features from the hexagonal grid data. These networks use multiple convolutional layers followed by pooling layers to identify spatial hierarchies. For temporal dependencies, we employ Long Short-Term Memory (LSTM) networks, which are adept at modeling sequential dependencies over time through their stacked layers. Additionally, Recurrent Neural Networks (RNNs) are incorporated to handle sequential data and capture temporal dynamics with their simple yet effective architecture. To further enhance forecasting accuracy, the Deep Autoregressive (Deep AR) model is used to model probabilistic dependencies between time steps, combining traditional autoregressive techniques with deep learning.

2.1.3 Feature Engineering

To improve the performance of our models, we undertake comprehensive feature engineering from the raw solar irradiance data. Key features include the date, which provides temporal context, and Global Horizontal Irradiance (GHI), representing the total solar radiation received on a horizontal surface. Clearsky GHI indicates the expected GHI under clear sky conditions, serving as a benchmark. Environmental factors such as temperature, atmospheric pressure, and relative humidity are included to account for their influence on solar irradiance. Additionally, the solar zenith angle, wind speed, wind direction, dew point, and precipitable water are considered to capture various atmospheric conditions. The clearness index, which is the ratio of actual GHI to clearsky GHI, indicates the clarity of the atmosphere. Seasonal variations are also captured to reflect changes in solar irradiance patterns throughout the year. By integrating these features, we provide our models with a rich and comprehensive dataset, enhancing their ability to accurately predict solar irradiance.

2.1.4 Hybrid Approach of Models

To leverage the strengths of different models, we propose a hybrid approach that merges various deep learning techniques. This approach involves combining the outputs of multiple models to create a more robust and accurate forecasting system. The fundamental idea is to merge different models in a way that harnesses their individual strengths and mitigates their weaknesses. By averaging out the biases and variances of individual models, we can reduce the overall error and improve the generalization on unseen data. This ensemble method increases the robustness of the forecasting system, making it less reliant on the specific assumptions or limitations of any single model. Techniques such as weighted averaging or stacking can be used to facilitate the fusion of models. In weighted averaging, the outputs of different models are combined based on predetermined weights that reflect their individual performance. In stacking, a meta-learner is trained to optimally combine the outputs of the individual models, further enhancing the predictive accuracy. This comprehensive approach allows us to extract the best features from each model type, leading to superior performance and more reliable solar irradiance forecasts.

2.2 Data Collection:

For this project, solar irradiance data was collected from NASA's POWER PROJECT , specifically tailored to the geographical location of Jalandhar, Punjab, India. The dataset encompasses historical solar irradiance measurements spanning several years, providing a rich source of information for model development and evaluation. Additionally, meteorological data such as temperature, humidity, wind speed, and atmospheric pressure were obtained to incorporate environmental factors influencing solar irradiance variability in the region.

NASA Prediction Of Worldwide Energy Resources

Data sources:

https://power.larc.nasa.gov/beta/data-access-viewer/

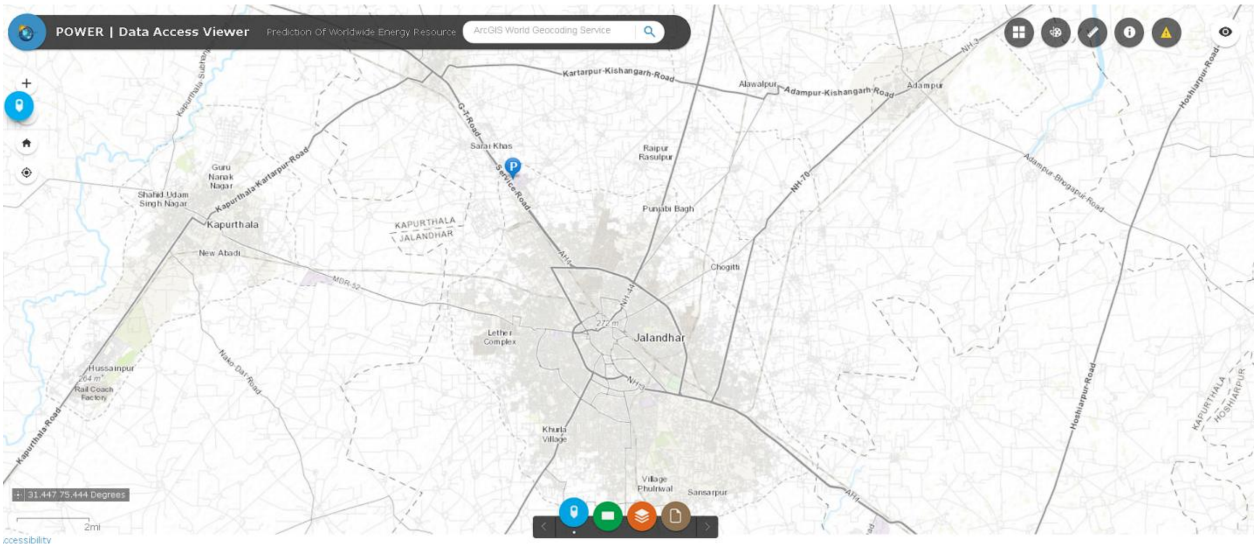

Fig .2.3

Target location	NIT Jalandhar, PUNJAB
Latitude	31.3966
Longitude	75.5436

2.2.1 Area Significance :

Solar irradiance, essential for renewable energy generation, agriculture, and climate modeling, holds particular importance in Jalandhar, Punjab, India. With the increasing emphasis on renewable energy sources, accurate forecasting of solar irradiance is imperative for optimizing solar power generation in Punjab. Solar energy is being heavily relied upon for irrigation, thus contributing to the state's agricultural productivity.

2.3 Data Processing and Seasonal Conversion:

2.3.1 Data cleaning :

We checked through the data we received from NASA, paying close attention to every detail. This involved fixing any errors or gaps in the data to ensure that it was reliable and accurate. This also Involves checking for outliers and incorrect data points within the dataset obtained from NASA.(either removed or ignored)'

Handling missing values:

Missing values can be easily removed by setting the attributes to N/A or by selecting the mean , median of the entire attribute column.

Removing Night Values:

To effectively clean the dataset by removing values corresponding to nighttime, you need to implement a filtering mechanism that distinguishes between day and night observations. By setting criteria based on variables, we can identify and exclude data points collected during nighttime hours.

Clear sky index :

Clear sky refers to weather conditions with minimal cloud cover, allowing ample sunlight. In data analysis, only data captured during clear sky conditions is considered. This data is crucial for accurately forecasting solar power generation and comprehending climate patterns on a large scale . Clearness index shows how clearthe sky is, affecting how much sunlight reaches the Earth's surface

$$GHI = I0.Ec.\cos(\theta z)$$

Where:

- GHI is the Global Horizontal Irradiance.
- Θz is the solar zenith angle, the angle between the sun's rays and the vertical direction.
- Ec is the atmospheric extinction coefficient, which accounts for the reduction in solar radiation due to atmospheric absorption and scattering.
- I0 is the solar constant, approximately 1367 W/m², representing the average solar radiation received outside the Earth's atmosphere.

2.3.2 Breaking the data into seasons :

Once we were satisfied with the quality of the data, we proceeded to break it down into four distinct seasons: **cold season, monsoon season, post-monsoon, and summer season**.This splitting allowed us to delve deeper into the patterns of solar radiation in Jalandhar throughout the year. We were able to identify trends and fluctuations in solar radiation, which in turn, empowered us to make well-informed predictions about the availability of solar energy in the region.

Cold season : Typically from November to February.

Monsoon season:Generally from July to September.

Post-monsoon season:Roughly from October to November.

Summer season:Usually from March to June.

cold_season_NITJ_data.csv

1 to 10 of 23664 entries Filter

Date	GHI	Clearsky GHI	Temperature	Pressure	Relative Humidity	Solar Zenith Angle	Wind Speed	Wind Direction	Dew Point	Precipitable Water	Clearness Index	Season
01-01-2000 08:00	118	134	14	990	18.48	79.26	0.5	22.7	-9	0.515	0.880597015	Cold Season
01-01-2000 09:00	262	311	18	990	8.87	69.54	0.4	351.4	-15	0.514	0.84244373	Cold Season
01-01-2000 10:00	375	460	22	990	6.73	61.63	0.8	300.1	-15	0.514	0.815217391	Cold Season
01-01-2000 11:00	524	557	25	990	5.81	56.33	1.4	293.9	-15	0.515	0.940754039	Cold Season
01-01-2000 12:00	584	592	26	990	5.58	54.4	1.6	292.8	-15	0.516	0.986486486	Cold Season
01-01-2000 13:00	552	559	26	990	5.64	56.2	1.7	295.3	-15	0.519	0.987477639	Cold Season
01-01-2000 14:00	458	464	25	990	5.99	61.4	1.6	303	-14	0.523	0.987068966	Cold Season
01-01-2000 15:00	314	316	23	990	6.83	69.23	1.6	310	-14	0.528	0.993670886	Cold Season
01-01-2000 16:00	131	140	19	990	9.21	78.9	1.6	310.9	-14	0.533	0.935714286	Cold Season
02-01-2000 08:00	106	133	15	990	17.63	79.29	0.9	17.1	-9	0.558	0.796992481	Cold Season

Show 10 per page

1 2 10 100 1000 2000 2300 2360 2367

Fig .2.4

2.3.4 Splitting of data:

The collected and preprocessed data is divided into **three sets** using a **70%, 10%, 20%** ratio.

Training set : **70%** of the data, is allocated to the Training Set. This set is utilized to train the machine learning model, allowing it to learn patterns and relationships within the data. The

Testing Set: This comprises 20% of the data . It is employed to evaluate the model's performance during development and tuning stages. It serves as an independent dataset to assess how well the model generalizes to unseen data. Finally, the

Validation Set: Representing 10% of the data,it provides evaluation of the final model. It ensures the model's performance on new data is accurately assessed and assists in detecting instances of overfitting, contributing to the model's robustness.

2.3.5 What if we neglected the processing step ?

However, if we had neglected to properly process the data or failed to categorize it into seasons, our analysis would have been flawed. Without this crucial step, our understanding of seasonal variations in solar energy would have been limited, potentially leading to inaccurate forecasts and decisions regarding solar energy utilization in Jalandhar.

2.4 Dataset description for GHI Value

2.4.1 For the cold season :

Table 2.1

Metric	Value
Mean	366.375
Median	357
Standard Deviation	234.061
Variance	54784.713

2.4.2 For the monsoon season :

Table 2.2

Metric	Value
Mean	420.92
Median	405
Standard Deviation	278.68
Variance	77663.74

2.4.3 For the post monsoon season :

Table 2.3

Metric	**Value**
Mean	420.133
Median	460
Standard Deviation	241.223
Variance	58188.608

2.4.4 For the summer season :

Table 2.4

Metric	**Value**
Mean	498.107
Median	488
Standard Deviation	306.5454
Variance	93970.108

CHAPTER 3

TECHNOLOGY ANALYSIS

3.1. Flow Chart

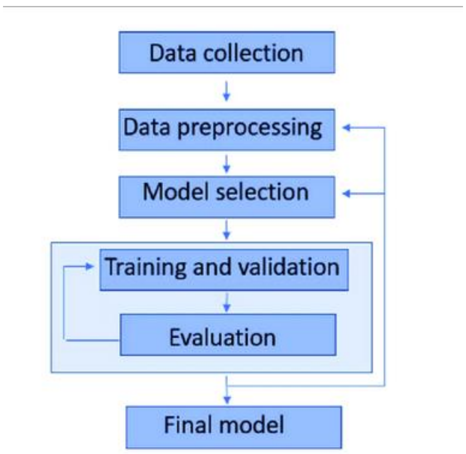

Fig.3.1

3.2. Tech Stack Analysis

3.2.1. Programming Language

3.2.1.1 Python

Python is a widely acclaimed high-level programming language esteemed for its simplicity and adaptability. It operates as an interpreted language, fostering high interactivity and suitability for diverse applications. Python's support for object-oriented programming enables developers to structure their code efficiently and effectively. One of Python's standout features is its expansive library ecosystem, encompassing solutions tailored for various tasks, from web development to scientific computing. Particularly notable in the realm of machine learning are libraries such as sci-kit-learn, Keras, and TensorFlow. These libraries empower developers to implement intricate machine-learning algorithms with ease, rendering Python a favored choice among data scientists and machine-learning practitioners globally.

3.2.2. Data Collection

3.2.2.1 NASA POWER Project

The NASA Prediction Of Worldwide Energy Resources (POWER) project provides global meteorological data that includes solar irradiance, temperature, humidity, and other weather parameters. This data is crucial for developing accurate solar irradiance models, as it offers comprehensive coverage and historical records.

3.2.3. Data Preprocessing and Analysis

3.2.3.1. Pandas

Pandas is a powerful data manipulation and analysis library for Python. It provides data structures like DataFrames, which allow for easy handling, cleaning, and analysis of large datasets. Pandas support a wide range of operations, from data filtering and aggregation to complex time-series analysis.

3.2.3.2. NumPy

NumPy is the fundamental package for numerical computing in Python. It provides support for arrays, matrices, and a large collection of mathematical functions to operate on these arrays. NumPy's efficient handling of large numerical datasets is essential for preprocessing and manipulating data before model training.

3.2.4. Machine Learning and Ensemble Learning

3.2.4.1. Scikit-learn

Scikit-learn is a comprehensive machine learning library in Python. It offers simple and efficient tools for data mining and data analysis, covering a range of machine learning tasks such as classification, regression, clustering, and dimensionality reduction. Scikit-learn's consistent API and well-documented modules make it an excellent choice for building and evaluating machine learning models.

3.2.4.2. TensorFlow

TensorFlow, crafted by Google, stands as a stalwart in numerical computation and machine learning. As an open-source library, TensorFlow empowers researchers and developers to innovate and construct state-of-the-art machine learning models. TensorFlow's all-encompassing ecosystem comprises an array of tools, libraries, and community resources, equipping users to address diverse machine learning challenges adeptly. Whether implementing traditional machine learning algorithms or sophisticated deep learning models, TensorFlow furnishes the requisite building blocks for implementation and deployment. Its flexibility and scalability render it suitable for application in both research and production environments, propelling advancements in domains like computer vision, natural language processing, and reinforcement learning. With TensorFlow, the horizon of possibilities in machine learning expands boundlessly, fostering innovation and pushing the frontiers of what's achievable.

3.2.5. Visualization and Reporting

3.2.5.1. Matplotlib

Matplotlib is a plotting library for creating static, animated, and interactive visualizations in Python. Seaborn is built on top of Matplotlib and provides a high-level interface for drawing attractive and informative statistical graphics. Together, these libraries help in exploring data patterns, visualizing model performance, and presenting results effectively.

3.2.5.2. Jupyter Notebooks

Jupyter Notebooks herald a paradigm shift in how data scientists and researchers engage with code and data. This open-source web application facilitates the creation of documents comprising live code, visualizations, and explanatory text. With support for over 40 programming languages, including Python, Jupyter Notebooks offer a flexible and interactive environment for data analysis and exploration. Its in-browser editor allows users to execute code fragments, visualize data, and document findings seamlessly. Such versatility positions Jupyter Notebooks ideally for tasks such as data cleaning, visualization, model prototyping, and collaborative research. Moreover, by enabling effortless sharing of work, Jupyter Notebooks promote reproducibility, transparency, and collaboration in scientific endeavors.

CHAPTER 4

ECONOMIC ANALYSIS

For the economic analysis of the Model on Solar Irradiance Forecasting Using Ensemble Learning Model," we will evaluate the cost-saving potential, revenue generation opportunities, cost factors, return on investment (ROI), and risk mitigation strategies associated with integrating solar power forecasting technologies.

1. Cost Savings and Efficiency Improvements:

Solar power plants can minimize energy wastage by optimizing operations based on precise solar irradiance forecasts. This means that the plant can adjust its energy generation based on anticipated sunlight levels, ensuring that resources are used efficiently and not wasted during periods of low solar irradiance. By using accurate forecasting, solar power plants can maximize their energy output and overall efficiency, leading to cost savings in the long run.

2. Revenue Generation:

Accurate forecasts optimize solar panel usage, increasing overall power generation and subsequently boosting revenues. By accurately predicting solar irradiance levels, solar power plants can adjust their operations to capture more solar energy, leading to increased power generation. This higher energy output can then be sold to the grid or used to generate additional revenue through market participation or ancillary services. Ultimately, accurate forecasting plays a crucial role in maximizing revenue generation for solar power plants.

3. Cost Factors:

Investment in data collection, software development, and model training is essential for developing sophisticated forecasting models. Developing and implementing efficient solar power forecasting models requires significant investment in data collection, software development, and training. Acquiring high-quality data sources, hiring skilled professionals, and investing in computational resources are all necessary to create accurate and reliable forecasting models. These initial costs are essential for ensuring the effectiveness and success of the forecasting process.

4. Return on Investment (ROI):

Efficiency gains and cost savings from optimized operations contribute to a positive ROI over time. The initial investment in developing and implementing forecasting models can result in long-term benefits, including reduced operational costs and increased revenues. These efficiency gains and cost savings contribute to a positive return on investment over time, as solar power plants become more efficient and profitable with the help of accurate forecasting. This improved ROI is a key factor in justifying the initial investment in solar power forecasting.

5. Risk Mitigation:

Predictable power generation forecasts mitigate financial risks associated with fluctuating solar output. By providing reliable and accurate forecasts of solar energy generation, solar power plants can mitigate financial risks associated with variable solar output. Predictable power generation forecasts help stabilize revenue streams by reducing uncertainties and fluctuations in energy production. This risk mitigation strategy is crucial for ensuring financial stability and reliability in the operation of solar power plants.

6. Environmental Responsibility and Cost Savings:

Embracing solar power as a renewable energy source reduces harmful greenhouse gas emissions and lessens the reliance on fossil fuels. By addressing climate change and promoting environmental sustainability, solar energy not only helps prevent environmental damage but also leads to cost savings in the long run. Cleaner and more sustainable energy production methods not only benefit the environment but also result in economic advantages through reduced expenses related to environmental harm.

7. Job Creation and Regional Development:

The growth of the solar energy sector, driven by accurate forecasting and increased investments, creates employment opportunities in various sectors such as manufacturing, installation, maintenance, and research. This job growth not only revitalizes local economies but also diversifies them, making regions more resilient to economic fluctuations and supporting overall development.

8. Consumer Empowerment and Economic Well-being:

Accurate forecasting can lead to lower operational costs for energy providers, potentially translating into savings for consumers on their electricity bills, and enhancing their financial well-being and stability.

9. Policy Guidance for Sustainable Development:

Reliable solar irradiance forecasts provide valuable insights for policymakers to design effective energy policies and regulations that support the transition to renewable energy sources, facilitating sustainable development and environmental responsibility.

CHAPTER 5

RESULTS AND OBSERVATIONS

In our analysis, we measured the performance of our solar irradiance forecasting models using several key metrics. These include R^2 (R-squared), RMSE (Root Mean Square Error), MSE (Mean Square Error), and MAE (Mean Absolute Error). These metrics helped us evaluate how well our models predicted solar radiation and provided insights into their accuracy and reliability.

5.1. Metrics :

5.1.1 R^2 (R-squared):

Shows how well the model's predictions match the actual data. A higher R^2 means the model explains more of the variation in the data, indicating a better fit.

5.1.2. RMSE (Root Mean Square Error):

This is The square root of the average squared differences between predicted and actual values. Lower RMSE values indicate the predictions are closer to the actual values, showing better model performance

5.1.3. MSE (Mean Square Error):

The average of the squared differences between predicted and actual values.Lower MSE values mean the model's predictions are more accurate, with smaller errors.

5.1.4. MAE (Mean Absolute Error):

The average of the absolute differences between predicted and actual values.Lower MAE values show that the model's predictions are closer to the actual values, indicating more precise predictions.

5.2. RESULTS:

5.2.1 SUMMER SEASON

Table 5.1

Models	R2	RMSE	MSE	MAE
Linear Regression	0.85329	0.65651	0.42726	0.47279
CNN	0.9342	0.06393	0.0352	0.00408
LSTM based RNN	0.94406	0.72763	0.52759	0.4089
Attention	0.92458	1.2278	1.50753	0.9190
CNN-LSTM	0.934	0.7679	0.04020	0.00589
CNN-LSTM with Attention	0.9569	0.5737	0.394	0.0494

5.2.2. WINTER SEASON

Table 5.2

Models	R2	RMSE	MSE	MAE
Linear Regression	0.92693	0.64009	0.4097	0.4613
CNN	0.9349	0.0639	0.0351	0.00408
LSTM based RNN	0.9025	0.566	0.03213	0.3337
Attention	0.9569	1.3006	1.6915	1.359
CNN-LSTM	0.9387	0.0619	0.0038	0.03422
CNN-LSTM with Attention	0.9621	0.09883	0.09768	0.006988

5.2.3. MONSOON SEASON

Table 5.3

Models	R2	RMSE	MSE	MAE
Linear Regression	0.7512	0.6536	0.4272	0.4727
CNN	0.8764	0.1060	0.0112	0.0639
LSTM based RNN	0.8849	0.9589	0.91915	0.58382
Attention	0.9639	0.02020	0.0408	0.17262
CNN-LSTM	0.8865	0.1026	0.01053	0.05930
CNN-LSTM with Attention	0.9751	0.140	0.19467	0.0096

5.2.4. POST MONSOON SEASON

Table 5.4

Models	R2	RMSE	MSE	MAE
Linear Regression	0.79701	0.50604	0.25770	0.30027
CNN	0.9399	0.75628	0.5715	0.55419
LSTM based RNN	0.9440	0.7263	0.52759	0.4089
Attention	0.94569	4.15320	0.172740	0.371689
CNN-LSTM	0.9590	0.0583	0.0034	0.002819
CNN-LSTM with Attention	0.9682	0.213	0.04540	0.001683

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

In this research, we present a new method for forecasting hourly solar radiation a day in advance. This method uses a combination of convolutional neural networks (CNNs), long short-term memory (LSTM) networks, with an attention mechanism. Our goal was to improve the accuracy of the forecasts by including a wider range of weather variables and by studying the relationships between data from multiple observation stations over time and space.

Key Findings

1. **High Accuracy Forecasting Model:**
   - By modeling atmospheric parameters as an attributed dynamic network, our proposed model effectively captures the complex interactions between various meteorological factors and their temporal changes.
   - The integration of spatial with CNN and temporal variations with LSTM, and Attention mechanism that focus on relevant features enhances the accuracy of solar irradiance forecasts.
2. **Solar Energy in Action:**
   - Accurate solar irradiance forecasting is essential for the effective commercialization of solar energy, as it helps mitigate the variability in solar power output.
   - Our improved forecasting model facilitates better planning and management of solar energy resources, contributing to more stable and reliable energy production.
3. **Long-Term Implications:**
   - Advances in forecasting accuracy can yield economic benefits, such as reduced operational costs for energy providers and increased investment in solar energy projects.
   - Enhanced solar energy utilization promotes environmental sustainability by decreasing reliance on fossil fuels and lowering greenhouse gas emissions.

Conclusion

The Proposed CNN-LSTM model enhanced with Attention mechanism marks a notable breakthrough in predicting solar irradiance. By seamlessly integrating diverse meteorological variables and capturing their intricate dependencies, the model attains unparalleled accuracy in forecasting. This advancement not only promotes the effective harnessing of solar energy but also yields significant economic and environmental advantages. The research underscores the value of a holistic forecasting strategy, harnessing spatial, temporal, and variable diversity to surmount the shortcomings of conventional models and propel the adoption of sustainable energy alternatives.

Future scope

The future scope of the solar irradiance forecasting system entails an innovative integration of Bayesian optimization into the existing CNN-LSTM model, enhanced by attention mechanisms. This evolution aims to augment the predictive accuracy and robustness of the system by leveraging Bayesian optimization techniques to optimize model parameters. By incorporating Bayesian optimization, the model can dynamically adapt and refine its architecture based on past performance and current conditions, resulting in more precise and reliable solar irradiance predictions. This fusion of advanced machine learning methodologies holds the promise of significantly enhancing the efficiency and effectiveness of solar energy generation, contributing to the sustainable development of renewable energy sources.

REFERENCES

[1]. *Brahma, B., & Wadhvani, R.* (2023). 'A residual ensemble learning approach for solar irradiance forecasting. Multimedia Tools and Applications. Advanced online publication'

[2]. *Neha Sehrawat ,Sahil Vashisht ,Amritpal Singh* (2023). 'Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison'

[3]. *Brahma, B., & Wadhvani, R.* (2020). 'Solar irradiance forecasting based on deep learning methodologies and multi–site data. Symmetry, 12(11), Article 1826'

[4]. *Ameera M. Almarzooqi, Maher Maalouf, Tarek H.M. El-Fouly, Vasileios E. Katzourakis, Mohamed S. El Moursi and Constantinos V. Chrysikopoulos* (2024) . 'A hybrid machine-learning model for solar irradiance forecasting'

5]. *Avishek Pal, PKS Prakash* (2017). 'Practical time series analysis_ master time series data processing, visualization, and modeling using Python '

[6]. *By-Aileen-Nielsen* .'Practical-Time-Series-Analysis-5336102-(z-lib.org)'

[7]. *Xiaoqiao Huang a b c, Qiong Li a, Yonghang Tai b c, Zaiqing Chen c, Jun Zhang b, Junsheng Shi b c, Bixuan Gao b, Wuming Liu d* (2021). 'Hybrid deep neural model for hourly solar irradiance forecasting'
https://doi.org/10.1016/j.renene.2021.02.161

[8]. *Dhivya Sampath Kumar, Gokhan Mert Yagli, Monika Kashyap, Dipti Srinivasan* (2019).
'Solar irradiance resource and forecasting: a comprehensive review'
https://doi.org/10.1049/iet-rpg.2019.1227