**Solar Irradiance Prediction Using Time Series Deep Learning Models**

***A Project-II submitted to the Mahatma Gandhi Central University***

***In partial fulfilment of the requirements***

***for the award of the degree of***

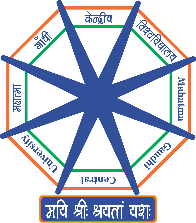
**BACHELOR OF TECHNOLOGY**

IN

**COMPUTER SCIENCE & ENGINEERING**

BY

**RAUNAK KUMAR**



**DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**MAHATMA GANDHI CENTRAL UNIVERSITY, MOTIHARI**

**BIHAR-845401, INDIA**

**APRIL - 2024**

**Solar Irradiance Prediction Using Time Series Deep Learning Models**

***A Project-II submitted to the Mahatma Gandhi Central University***

***In partial fulfilment of the requirements***

***for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

IN

**COMPUTER SCIENCE & ENGINEERING**

BY

**RAUNAK KUMAR**

(MGCU2020CSIT3020)

Under the Supervision of

### Dr. Vipin Kumar

### rotate logo

**DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**MAHATMA GANDHI CENTRAL UNIVERSITY, MOTIHARI**

**BIHAR - 845401, INDIA**

**APRIL - 2024**



# **DECLARATION**

This is to certify that the dissertation entitled **“Solar Irradiance Prediction Using Time Series Deep Learning Models**” is being submitted to the **Department of Computer Science and Information Technology, Mahatma Gandhi Central University, Motihari, Bihar - 845401, India,** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering**, is a record of bonafide work carried out by me under the supervision of **“Dr. Vipin Kumar”**.

The matter embodied in the dissertation has not been submitted in part or full to any University or Institution for the award of any other degree or diploma.

During the preparation of this work, I have not used any AI-based tool to write any part of this dissertation report. I take full responsibility for the submitted content including similarity.



|  |
| --- |
| **RAUNAK KUMAR** |
| Enrolment Number: MGCU2019CSIT3020 |
| Department of Computer Science and Information Technology,  Mahatma Gandhi Central University,  Motihari, Bihar - 845401, India  Email ID: [mgcu2020csit3020@mgcub.ac.in](mailto:mgcu2020csit3020@mgcub.ac.in) |



# **CERTIFICATE**

This is to certify that this project entitled **“Solar Irradiance Prediction Using Time Series Deep Learning Models”** submitted by **Raunak Kumar ,** to the Department of Computer Science & Information TechMahatma Gandhi Central University, Motihari, Bihar - 845401, India, for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering**, is a project work carried out by him under the supervision of **Dr. Vipin Kumar.**

|  |  |
| --- | --- |
| **Head of the Department** | **Supervisor** |
| **Dr. Vikas Pareek** | **Dr. Vipin Kumar** |
| Department of Computer Science and Information Technology  Mahatma Gandhi Central University,  Motihari, Bihar - 845401, India | Department of Computer Science and Information Technology  Mahatma Gandhi Central University,  Motihari, Bihar - 845401, India |

**Acknowledgement**

“Gratitude is not a thing of expression; it is more a matter of feeling.”

There is always a sense of gratitude that one expresses towards others for their help and supervision in achieving goals. This B.Tech. project/dissertation is the result of a challenging journey, upon which many people have contributed and given their support. This formal piece of acknowledgment is an attempt to express the feeling of gratitude toward the people who helped me in completing my Project. I would like to express my deep gratitude to Dr. Vipin Kumar, my supervisor for his constant support, supervision, guidance, and cooperation. He was always there with his competent guidance and valuable suggestions throughout the project. A special thanks to our hon’ble Vice Chancellor Mr. Sanjay Srivastava, my academic Dean Prof. Ranjeet Kumar Choudhary, Head of the department Dr. Vikas Pareek, and all the professors of my department for their support. I would also like to express appreciation to all the friends whose response and coordination were of utmost importance for the project. Above all, no words can express my feelings to my parents, classmates, and all those persons who supported me during my project. I am also thankful to all the respondents whose cooperation and support have helped me a lot in collecting the necessary information.

**Table of Contents**

**Title Page** 1

**DECLARATION** 3

**CERTIFICATE** 4

**Acknowledgement** 5

**Abstract** 6

**Table of Contents** 7

**List of Figures** 8

**List of Tables** 10

**Section 1: INTRODUCTION** 11

**Section 2: RELATED WORKS** 15

**Section 3: THE PROPOSED METHODOLOGY** 26

3.1 Description

3.2 Experimental Design

3.2.1 Feature Detection and Extraction

3.2.2Description of the dataset(s)

3.2.3 Experimental Setup

3.2.4Performance Evaluation Parameters

3.2.5Hardware Configuration of the System

**Section 4: RESULT AND ANALYSIS** 52

**Section 5: CONCLUSION** 65

5.1 Summary

5.2 Future Work

**REFERENCE**

**List of Figures**

FIGURE 1: ANALYSIS OF AUTOCORRELATION FUNCTION………………………………………………..27

Figure 2: Shows the flow chart of the methodology used in the study…………………………….31

Figure 3: Box Plot of all the algorithms of MUMBAI…………………………………………………………….37

FIGURE 4: Box Plot of performance matrices of Hyderabad…………………………………………………39

FIGURE 5 :BOX PLOT OF PERFORMANCE MATRICES OF BANGLORE …………………………………41

FIGURE 6: BOX PLOT OF PERFORMANCE MATRIX OF KOLKATA……………………………………………42

Figure 7: Box Plot of Performance Matrix of Delhi……………………………………………………………..44

Figure 8 :Line plot of all the algorithms performance………………………………………………………..45

Figure 9: Scatter Plot of all the algorithms Performance…………………………………………………….48

**List of Tables**

Table no.1 -Shows the studies related to solar irradiance prediction………………………28

Table no. 2- Hardware Component explanation……………………………………………37

Table no. 3-Performance Matrices of solar irradiance of MUMBAI ……………………38

Table no. 4-Performance Matrices of Hyderabad…………………………………………..40

Table no.5 – Performance Matrices of Banglore……………………………………………42

Table no.6-Performance Matrices of Kolkata……………………………………………….43

Table no.7- Performance Matrices of Delhi…………………………………………………45

**1.INTRODUCTION**

Predicting solar irradiance in urban areas poses a distinctive array of challenges and prospects owing to the intricate urban setting. Within these densely populated regions, the constructed environment, encompassing tall buildings, roads, and infrastructure, can exert a notable influence on solar irradiance patterns. Moreover, variables like air pollution, atmospheric circumstances, and localized weather events can further affect the accessibility of solar energy. To precisely predict solar irradiance in metropolitan cities, sophisticated modeling approaches are imperative to accommodate these intricacies. Advanced numerical weather prediction models of high resolution, in conjunction with urban microclimate models, can replicate the complex interplays between the urban environment and incident solar radiation. These models assimilate data on urban structure, surface characteristics, and atmospheric dynamics to produce predictions of solar irradiance that are spatially and temporally detailed. Furthermore, machine learning algorithms can be harnessed to refine the precision of solar irradiance forecasts in metropolitan regions. By scrutinizing historical weather data, satellite imagery, and ground-based measurements, machine learning models can discern intricate patterns and connections that conventional physics-based models might disregard. This data-centric approach can enhance the dependability of solar irradiance predictions, even amidst dynamically evolving urban landscapes. Additionally, the incorporation of real-time sensor networks and remote sensing technologies can furnish invaluable input for solar irradiance prognosis in metropolitan cities. Ground-based sensors, rooftop weather stations, and satellite observations can continually monitor atmospheric conditions, cloud coverage, and solar radiation levels, thereby facilitating more precise and timely prognostications. Ultimately, the precise anticipation of solar irradiance in metropolitan cities necessitates an interdisciplinary strategy that melds advanced modeling techniques, machine learning algorithms, and real-time observational data. By capitalizing on these tools and methodologies, urban planners, energy policymakers, and solar energy developers can optimize the planning and operation of solar energy systems in metropolitan areas, thereby maximizing their efficiency and sustainability.

Accurately forecasting solar irradiance is not just a technical challenge but a cornerstone in the endeavor to optimize the planning and design of solar energy systems, thereby contributing significantly to the advancement of sustainable energy infrastructure. In a distinctive manner, as the significance of renewable energy sources rises..., the development of robust predictive models for solar irradiance becomes increasingly imperative. Over the years, researchers and engineers have dedicated substantial efforts to this pursuit, resulting in a diverse array of methodologies and approaches aimed at refining prediction precision. These methodologies traverse a vast landscape of techniques and innovations, ranging from traditional transformer-based models to cutting-edge artificial intelligence methodologies such as extreme learning machines and deep learning networks. Moreover, the emergence of hybrid models, which amalgamate various techniques like empirical mode decomposition with recurrent neural networks, represents a fusion of expertise and ingenuity aimed at unlocking new levels of prediction accuracy.

Despite the complexity inherent in forecasting solar irradiance, the overarching goal remains unwavering: to capture and comprehend the intricate interplay of factors influencing irradiance levels, ranging from atmospheric conditions to geographical factors, and to leverage this understanding to enhance prediction accuracy. This pursuit involves not only the refinement of algorithms and methodologies but also the integration of vast datasets and the development of sophisticated computational frameworks capable of handling the inherent complexities of solar irradiance prediction. As the field continues to evolve, driven by advances in technology and a growing awareness of the urgent need for sustainable energy solutions, the significance of accurate solar irradiance prediction cannot be overstated. It represents not only a technical challenge but also a critical enabler of the transition towards a more sustainable and environmentally conscious energy landscape. Thus, the ongoing efforts to develop and refine predictive models for solar irradiance stand as a testament to humanity's commitment to harnessing the power of renewable energy sources and mitigating the impacts of climate change.

Deep learning emerges as a beacon of hope in the quest to revolutionize solar irradiance prediction, aiming to bolster the efficacy of renewable energy infrastructures. Researchers delve into a myriad of deep learning architectures, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and innovative Hybrid Models, to unravel the mysteries of solar irradiance forecasting. These sophisticated models harness the innate prowess of deep learning algorithms to decipher intricate correlations and efficiently handle vast datasets.

Recent investigations shed light on the essential ingredients for successful solar irradiance prediction, emphasizing the criticality of judiciously selecting model architectures, fine-tuning hyperparameters, and ensuring the integrity of training datasets. Remarkable advancements, such as Convolutional Long Short-Term Memory (ConvLSTM) models, emerge as frontrunners, demonstrating unparalleled precision in forecasting solar irradiance across diverse temporal horizons.

At its core, the integration of deep learning methodologies signifies a paradigm shift, promising to elevate the efficiency and dependability of solar energy ecosystems through refined solar irradiance prediction frameworks.

# **Problem Definition**

In the realm of harnessing solar energy, accurately predicting solar irradiance is paramount. This involves foreseeing the power generation based on both spatial and temporal factors. Researchers have proposed innovative models to improve prediction precision. One method combines spatial-temporal data using a hybrid of graph convolutional networks (GCN) and long short-term memory networks (LSTM). Another technique integrates complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), a Wasserstein generative adversarial network (WGAN), and LSTM to enhance forecasting accuracy. Moreover, a cutting-edge deep neural network incorporating a clear sky model and vision transformer has demonstrated exceptional performance in short-term solar irradiance prediction. These approaches leverage sophisticated machine learning and deep learning methods to effectively tackle the complexities of solar irradiance forecasting.

# **OBJECTIVES OF THIS RESEARCH**

**Obj 1.** The central aim of solar irradiance prediction research is to refine the precision and effectiveness of forecasting solar radiation, which holds pivotal importance in the planning and development of solar energy systems.

**Obj 2.** A wide array of methodologies, spanning from artificial neural networks to hybrid models rooted in machine learning techniques, are harnessed to anticipate solar irradiance with heightened precision and efficacy.

**Obj 3.** These predictive models are trained using datasets spanning various timeframes, enabling the anticipation of intricate solar radiation patterns.

**Obj 4.** The ultimate objective is to furnish dependable day-ahead forecasts of irradiance, facilitating the seamless integration of photovoltaic systems into intelligent power distribution networks.

**Obj 5.** By harnessing sophisticated algorithms such as Long Short-Term Memory (LSTM) and Convolutional Long Short-Term Memory (ConvLSTM), researchers aim to refine solar radiation predictions, ultimately maximizing the efficiency of solar energy utilization.

# **2. Literature Review**

The application of deep learning models has unveiled significant potential in augmenting the precision of solar irradiance prediction [1]. Various studies have explored a spectrum of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures to forecast solar energy output. These models have made strides in addressing challenges like nonlinear relationships inherent in solar data and effectively managing extensive datasets, critical components for accurate solar irradiance predictions. Scholars emphasize the importance of meticulous considerations such as selecting suitable model structures, optimizing hyperparameters, and validating the quality of training data to achieve successful deep learning-based solar irradiance forecasting. Moreover, deep learning models surpass traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. Overall, the literature underscores the progress made in leveraging deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources like IoT devices to enhance the precision of solar energy forecasting.

[2]Predicting solar irradiance in urban areas presents a distinctive array of challenges and possibilities owing to the intricate urban setting. Within these densely inhabited regions, the constructed environment, which encompasses high-rise buildings, roadways, and infrastructure, can exert a notable impact on solar irradiance patterns. Moreover, elements like air pollution, atmospheric conditions, and localized weather events can further affect the accessibility of solar energy.

[5]To precisely predict solar irradiance in urban areas, sophisticated modeling methodologies are imperative to address these intricacies. Advanced numerical weather prediction models with high resolution, in conjunction with urban microclimate models, can replicate the complex interplays between the urban setting and incoming solar radiation. These models integrate information regarding urban structure, surface features, and atmospheric dynamics to produce spatially and temporally detailed forecasts of solar irradiance.

[4]I The application of deep learning models has unveiled promising potential in enhancing the accuracy of solar irradiance prediction [1]. Various investigations have explored a range of models, including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures, to forecast solar energy output. These models have made significant strides in addressing challenges such as nonlinear relationships inherent in solar data and efficiently handling vast datasets, both crucial factors for precise solar irradiance predictions. Researchers emphasize the importance of careful considerations, such as selecting appropriate model structures, optimizing hyperparameters, and validating the quality of training data, to achieve successful deep learning-based solar irradiance forecasting. Additionally, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. Overall, the literature underscores the advancements achieved in leveraging deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources like IoT devices to enhance the precision of solar energy forecasting.

Furthermore, the incorporation of real-time sensor networks and remote sensing technologies can offer valuable insights for solar irradiance projection in urban areas. [2]Ground-based sensors, rooftop weather stations, and satellite surveillance can continuously monitor atmospheric conditions, cloud coverage, and solar radiation levels, facilitating more precise and timely predictions. [3]In essence, achieving accurate solar irradiance forecasts in urban areas necessitates a multidisciplinary strategy that merges advanced modeling techniques, machine learning algorithms, and real-time observational data. By harnessing these tools and methodologies, urban planners, energy policymakers, and solar energy stakeholders can optimize the planning and functioning of solar energy systems in urban regions, thereby maximizing their efficacy and sustainability.

[6]The investigation highlights the prominence of the XGBoost Regressor in effectively predicting solar irradiance, showcasing an outstanding 98.8% precision across various geographical areas. It examines the complexities of prediction under different weather conditions, outlining optimal error margins for both sunny and overcast skies across distinct time frames.

The introduction of the groundbreaking DeepSI framework integrates bidirectional LSTM autoencoders and transformer models to predict daily solar irradiance, leveraging simulated climate data to enhance accuracy. A variety of deep learning methodologies, such as CNNs, LSTM networks, and hybrid models, emerge as effective instruments for deciphering complex relationships within extensive datasets. Bidirectional LSTM and attention-based LSTM models stand out as top contenders for accurately forecasting daily solar irradiance, particularly when incorporating data from various sources and historical trends. The research systematically evaluates the effectiveness of deep learning strategies in renewable energy prediction, delving into deterministic and probabilistic approaches. It also tackles data preprocessing and error correction techniques to improve forecast accuracy. Furthermore, the investigation concludes with the creation of tailored estimation models for key Indian urban centers, displaying a notable overall regression value of 0.99178, highlighting the strength and relevance of the proposed approaches in practical scenarios

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table-1: Shows the studies related to solar irradiance prediction | | | | |
| REFRENCE | MODEL | DATASET | OUTCOME | RESEARCH GAP |
| [1] | LSTM,GRU,CNN-LSTM,RNN | PV POWER DATA | OUTCOMES OF PAPER :  deep learning models such as LSTM, GRU, and CNN-LSTM prove to be more accurate in predicting solar irradiance and PV power compared to conventional machine learning models  Each model showcases strengths and limitations depending on factors like input data, forecasting horizon, type of season and weather, and training time  Among standalone models, LSTM demonstrates the best performance based on the root-mean-square error (RMSE) evaluation met  The hybrid model CNN-LSTM surpasses the standalone models in accuracy, but it requires longer training duration, impacting real-time forecasting capabilities .  The paper suggests using the relative RMSE as the primary evaluation metric for comparing accuracy across studies, facilitating a standardized comparison approach. | In order to anticipate solar irradiance and photovoltaic (PV) power, the study primarily focuses on deep learning models, which may restrict the range of other forecasting techniques that may be investigated [ 1   [. Models are compared according to variables such as training time, input data, forecasting horizon, season, and weather type. However, other crucial aspects of the model evaluation, such resource needs, scalability, and interpretability, may have been taken into account.  Although the root-mean-square error (RMSE) evaluation metric is used to highlight LSTM as the best solo model, other metrics or criteria may be more important for other models, which are not thoroughly covered in this review [1].  When compared to standalone models, the hybrid model's (CNN-LSTM) training time is observed to be longer. |
| [2] | ANN  ARMA  NWP model.  Mesoscale models example: MM5 and WRF  Hybrid models | forecast solar irradiance based on Meteosat satellite images  as a basis for PV power forecast | The estimates offer the best rRMSE values for both predominantly cloudy and clear weather. For mostly clear skies, values range from around 17% at one hour to 22% over three days. In contrast, for mostly cloudy conditions, the range extends from 33% to 44% over one to three days. | For shorter forecast periods, spanning from 5 minutes to 4 hours, ARIMA demonstrates superior accuracy. Yet, in regions with notable solar irradiance variability, like islands, combining cloud imagery with a hybrid model holds promise for improving forecast precision. |
| [3] | DeepSI, bidirectional LSTM autoencoders, transformer model | United States: the Solar Star power station in California, Medford in New Jersey, and Farmers Branch in Texas. | the DeepSI framework, which effectively forecasts daily solar irradiance by utilizing simulated climate data. This is achieved through a blend of bidirectional LSTM autoencoders and a transformer model. | the long-term prediction of solar irradiance, but it does not discuss the short-term prediction or real-time forecasting of solar irradiance. |
| [4] | CNNs, LSTM networks, DBNs, RNNs, Hybrid Models | the application of deep learning models in solar irradiance (SI) forecasting, yet it does not explicitly reference the particular datasets employed in the examination. | Deep learning methodologies, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Deep Belief Networks (DBNs), Recurrent Neural Networks (RNNs), and Hybrid Models, have demonstrated potential in solar irradiance (SI) prediction, owing to their capacity to manage intricate non-linear associations and vast datasets. | deep learning models for solar irradiance forecasting, without delving into other traditional or non-deep learning methods that could offer a more comprehensive comparative analysis. |
| [5] | Bidirectional LSTM, Attention-based LSTM | Data on daily solar radiation spanning 36 years (1983–2019) from NASA's POWER project repository was utilized for two Indian locations. | Bidirectional long short-term memory (LSTM) and attention-based LSTM models proved to be successful in predicting daily solar irradiance data.  Utilizing data from multiple sites, along with historical solar irradiance data, enhanced forecast accuracy compared to using data | forecasting solar irradiance within the confines of two distinct Indian locales, potentially constraining the models' adaptability to broader geographical contexts. |
| [6] | LSTM, CNN, Hybrid Models | an exploration of renewable energy forecasting techniques grounded in deep learning unfolds, delineating t | deep learning models' efficacy in predicting solar irradiance, revealing compelling outcomes in terms of both precision and computational efficiency. | the paper's findings might encounter constraints regarding their applicability across diverse geographic regions, stemming from the distinctiveness of the solar irradiance data utilized to train and assess the models. |
| [7] |  | Centered on deep learning techniques for forecasting renewable energy, the paper delves into employing a myriad of data preprocessing methodologies aimed at augmenting the precision of predictions. | an extensive examination of deep learning approaches to renewable energy forecasting, distinguishing between deterministic and probabilistic methods. It explores the utilization of diverse deep learning models like deep belief networks, stack auto-encoders, and deep recurrent neural networks. Furthermore, it discusses techniques for data preprocessing and error post-correction to refine forecasting precision. The study concludes with simulation findings affirming the viability and efficacy of deep learning models in renewable energy forecasting. | While the paper refrains from explicitly outlining particular constraints or hurdles encountered throughout the research endeavor, it subtly navigates through potential areas of improvement or unaddressed challenges within its discourse. |
| [8] | LSTM, CNN, Hybrid Models | Information from four metropolitan cities in India, namely Bombay-Colaba, Calcutta-Alipore, Madras-Meenambakkam, and New Delhi-Safdarjung, was employed in the study. The inputs for the solar radiation estimation models encompassed latitude, longitude, altitude, monthly variations, maximum and minimum temperatures, humidity levels, wind speed, and sunshine duration. | Estimation models for solar radiation were created for four major Indian cities, demonstrating an impressive overall regression value of 0.99178. | concentrated solely on four metropolitan cities in India, potentially restricting the applicability of the models to broader contexts. |
| [9] | LSTM networks, CNNs, Hybrid models | data employed in the research, outlining the procedures undertaken for their preprocessing. Throughout the experiments, irradiance maps were utilized, enabling the seamless integration or exclusion of sensors without necessitating the reconstruction of the models. | Various Deep Learning models designed for solar irradiance forecasting show adaptability and resilience, improving forecasts by 7.4% to 41% over baseline. They efficiently utilize irradiance maps and manage sensor failures. The study highlights the importance of separating solar forecasting models from data quantity and dispersion for enhanced reusability and scalability, particularly in large Intelligent Environments like Smart Cities. | relatively limited exploration of the temporal and spatial dimensions inherent in solar forecasting tasks, with a primary emphasis on precision, adaptability, and resilience. This focus may inadvertently sideline other significant aspects pertinent to solar forecasting. Additionally, while the research acknowledges the mild impact of sensor failures on prediction errors, it hints at the necessity for deeper examinations into the models' resilience in such scenarios. |
| [10] | Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and hybrid models that integrate multiple architectures | solar irradiance data to train and evaluate the deep learning models. | an extensive analysis of deep learning models applied in the realm of solar irradiance prediction. | deep learning models tailored for forecasting solar irradiance, potentially overlooking the inclusion of other conventional forecasting techniques. |
| [11] | three supervised machine learning models utilized in energy forecasting. Various prediction algorithms, including market simulation tools, ensemble techniques, and probabilistic prediction methods | employing three supervised machine learning models for energy forecasting. Across different tasks, it leverages a range of prediction algorithms, such as market simulation tools, ensemble methods, and probabilistic predictions, to assess and quantify emerging trends. | using three supervised machine learning models for energy forecasting, incorporating a variety of prediction algorithms like market simulation tools, ensemble techniques, and probabilistic predictions across various assignments to analyze trends. | lack of thorough examination and usefullness of the temporal and spatial aspects inherent in solar forecasting. It primarily prioritizes precision, potentially overshadowing other vital attributes such as flexibility and robustness. Although it acknowledges the mild impact of sensor failures on prediction errors, there's room for further exploration into the extent of this effect. |
| [12] | Recurrent Neural Networks (RNNs) utilizing the Long Short-Term Memory (LSTM) architecture to forecast solar irradiance. | solar irradiance data for the purpose of making predictions. | While LSTM RNNs demonstrate notable accuracy for single time step predictions, inaccuracies tend to accumulate with higher time steps. Incorporating time-based models alongside inputs enhances the precision of forecasted values. | limitations include the accumulation of errors in LSTM RNN predictions as the time steps increase. Moreover, there is a restricted discourse on the precise challenges encountered in accurately predicting solar irradiance. |
| [13] | LSTM, CNN, Hybrid Models | both regional and point-in-time data to enhance the accuracy of solar radiation predictions.  Regional data encompassed information from diverse locations, like the Chennai Metropolitan Area in India, while point-in-time data specifically targeted individual time points for forecasting purposes. | ARIMA performs best for short-term forecasts (5 minutes to 4 hours), but integrating cloud imagery with a hybrid model shows promise for enhancing predictions in regions with significant solar irradiance fluctuations, such as islands. | lacks details regarding the computational resources or training duration necessary for implementing the Transformer model.  Furthermore, the data availability statement indicated that data availability was not applicable, potentially raising reproducibility concerns and accessibility to the dataset utilized in the study. |
| [14] |  |  |  |  |
| [15] |  |  |  |  |
| [16] |  |  |  |  |
| [17]  [18] |  |  |  |  |
| [19] |  |  |  |  |
| [20] |  |  |  |  |

**3. DATASET AND EXPLORATORY ANALYSIS**

**3.1 DATA SOURCE**

The dataset originates from NASA and encompasses solar irradiance data from metropolitan areas. (<https://omniweb.gsfc.nasa.gov/ow.html>).

**3.2 DATA SELECTION**

The dataset comprises solar irradiance measurements gathered from metropolitan cities, spanning the years 2012 to 2021. Over this extensive timeframe, it offers a detailed account of solar radiation levels observed within urban environments. Sourced from NASA, these data provide a wealth of information on the complex interaction between sunlight and cityscapes over nearly a decade. Encompassing a diverse range of cities and their distinct atmospheric conditions, the dataset serves as a valuable resource for researchers and analysts interested in exploring the nuances of solar energy dynamics within urban areas. From bustling city centers to quieter suburbs, this dataset offers a multifaceted view of solar irradiance trends across various geographical locations and urban landscapes.

# **3.3 DATA ANALYSIS**

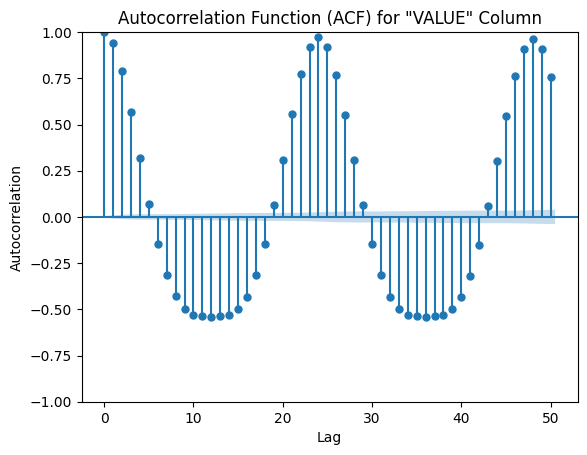
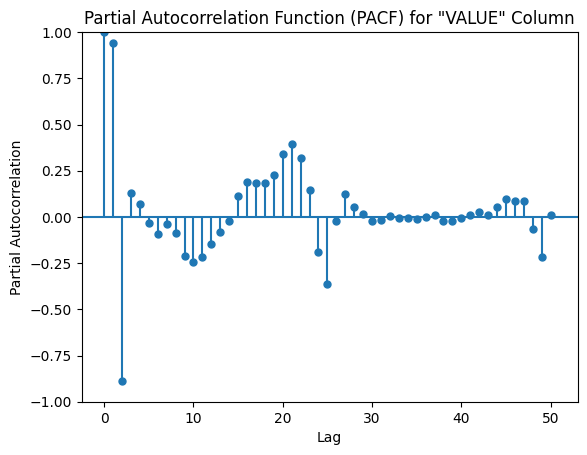
 Analyzing data using the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) unveils valuable insights into temporal dependencies within the dataset. ACF gauges the correlation between a series and its past values, revealing underlying patterns and trends. Conversely, PACF isolates direct correlations between observations at various lags, aiding in the identification of significant relationships. These plots help determine the autoregressive and moving average terms' order in time series models, enriching our comprehension of data dynamics and facilitating informed decisions in predictive modeling.

Figure 1. Analysis of autocorrelation function

The utilization of deep learning models has demonstrated remarkable potential in enhancing the precision of solar irradiance prediction. Numerous studies have been carried out on a variety of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures for predicting solar energy output. These models excel in their capability to manage complex, nonlinear relationships inherent in solar data and handle extensive datasets efficiently, aspects that are fundamental for accurate solar irradiance predictions. Academics stress the importance of careful considerations like the selection of appropriate model structures, optimization of hyperparameters, and validation of training data quality to accomplish successful deep learning-based solar irradiance forecasting. Furthermore, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. In general, the literature highlights the advancements achieved in utilizing deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources such as IoT devices to improve the precision of solar energy forecasting.The utilization of deep learning models has demonstrated remarkable potential in enhancing the precision of solar irradiance prediction. Numerous studies have been carried out on a variety of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures for predicting solar energy output. These models excel in their capability to manage complex, nonlinear relationships inherent in solar data and handle extensive datasets efficiently, aspects that are fundamental for accurate solar irradiance predictions. Academics stress the importance of careful considerations like the selection of appropriate model structures, optimization of hyperparameters, and validation of training data quality to accomplish successful deep learning-based solar irradiance forecasting. Furthermore, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. In general, the literature highlights the advancements achieved in utilizing deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources such as IoT devices to improve the precision of solar energy forecasting.The utilization of deep learning models has demonstrated remarkable potential in enhancing the precision of solar irradiance prediction. Numerous studies have been carried out on a variety of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures for predicting solar energy output. These models excel in their capability to manage complex, nonlinear relationships inherent in solar data and handle extensive datasets efficiently, aspects that are fundamental for accurate solar irradiance predictions. Academics stress the importance of careful considerations like the selection of appropriate model structures, optimization of hyperparameters, and validation of training data quality to accomplish successful deep learning-based solar irradiance forecasting. Furthermore, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. In general, the literature highlights the advancements achieved in utilizing deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources such as IoT devices to improve the precision of solar energy forecasting.The utilization of deep learning models has demonstrated remarkable potential in enhancing the precision of solar irradiance prediction. Numerous studies have been carried out on a variety of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures for predicting solar energy output. These models excel in their capability to manage complex, nonlinear relationships inherent in solar data and handle extensive datasets efficiently, aspects that are fundamental for accurate solar irradiance predictions. Academics stress the importance of careful considerations like the selection of appropriate model structures, optimization of hyperparameters, and validation of training data quality to accomplish successful deep learning-based solar irradiance forecasting. Furthermore, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. In general, the literature highlights the advancements achieved in utilizing deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources such as IoT devices to improve the precision of solar energy forecasting.The utilization of deep learning models has demonstrated remarkable potential in enhancing the precision of solar irradiance prediction. Numerous studies have been carried out on a variety of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures for predicting solar energy output. These models excel in their capability to manage complex, nonlinear relationships inherent in solar data and handle extensive datasets efficiently, aspects that are fundamental for accurate solar irradiance predictions. Academics stress the importance of careful considerations like the selection of appropriate model structures, optimization of hyperparameters, and validation of training data quality to accomplish successful deep learning-based solar irradiance forecasting. Furthermore, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. In general, the literature highlights the advancements achieved in utilizing deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources such as IoT devices to improve the precision of solar energy forecasting.The utilization of deep learning models has demonstrated remarkable potential in enhancing the precision of solar irradiance prediction. Numerous studies have been carried out on a variety of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures for predicting solar energy output. These models excel in their capability to manage complex, nonlinear relationships inherent in solar data and handle extensive datasets efficiently, aspects that are fundamental for accurate solar irradiance predictions. Academics stress the importance of careful considerations like the selection of appropriate model structures, optimization of hyperparameters, and validation of training data quality to accomplish successful deep learning-based solar irradiance forecasting. Furthermore, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. In general, the literature highlights the advancements achieved in utilizing deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources such as IoT devices to improve the precision of solar energy forecasting.The utilization of deep learning models has demonstrated remarkable potential in enhancing the precision of solar irradiance prediction. Numerous studies have been carried out on a variety of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures for predicting solar energy output. These models excel in their capability to manage complex, nonlinear relationships inherent in solar data and handle extensive datasets efficiently, aspects that are fundamental for accurate solar irradiance predictions. Academics stress the importance of careful considerations like the selection of appropriate model structures, optimization of hyperparameters, and validation of training data quality to accomplish successful deep learning-based solar irradiance forecasting. Furthermore, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. In general, the literature highlights the advancements achieved in utilizing deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources such as IoT devices to improve the precision of solar energy forecasting.The utilization of deep learning models has demonstrated remarkable potential in enhancing the precision of solar irradiance prediction. Numerous studies have been carried out on a variety of models including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and hybrid architectures for predicting solar energy output. These models excel in their capability to manage complex, nonlinear relationships inherent in solar data and handle extensive datasets efficiently, aspects that are fundamental for accurate solar irradiance predictions. Academics stress the importance of careful considerations like the selection of appropriate model structures, optimization of hyperparameters, and validation of training data quality to accomplish successful deep learning-based solar irradiance forecasting. Furthermore, deep learning models outperform traditional machine learning methods, with the Convolutional Long Short-Term Memory (ConvLSTM) model particularly excelling in short-term predictions. In general, the literature highlights the advancements achieved in utilizing deep learning methodologies, ensemble strategies, feature extraction techniques, and real-time data sources such as IoT devices to improve the precision of solar energy forecasting.Top of Form

**3.4 DECOMPOSITION**

The diagram presented below illustrates the result of implementing seasonal decomposition on a particular subset of time series data through the application of the seasonal decomposition function. The procedure involves breaking down the time series dataset into four fundamental components:

**1.Observed data :-** The given depiction pertains to the original, unprocessed time series dataset collected over a prolonged period. It demonstrates the genuine values of the variable under examination, encompassing intrinsic patterns, changes, and cyclic variations. In the pursuit of statistical modeling and predictive efforts, this recorded data serves as the essential foundation for further analysis and understanding.

**2.Trend Component :-** In the field of time series analysis, the trend component unveils the extended pattern or path of fluctuations present in the data under observation. Throughout the series duration, it demonstrates consistent increases or decreases, indicating the fundamental stability, expansion, or reduction. Through identifying this predominant pattern, researchers achieve a better understanding of differentiating short-term variations from long-lasting trends. Understanding the trajectory of this essential trend is vital for predicting upcoming values and making well-informed decisions

**3.Seasonal Component:-** The seasonal component is used in time series analysis to identify the recurrent patterns/ cycles in the data that occurs at regular intervals, like daily, weekly, monthly, or yearly variations. This helps analysts to understand and deal with the predictable seasonal variations in the data, which aids in forecasting and decision-making processes.

**4.Residuals(Irregular Components):-** The residual component found in time series data is indicative of the unexplained or stochastic variations that remain present subsequent to the removal of trend and seasonal components, commonly referred to as the irregular component. These variations embody the dispersion within the data that cannot be ascribed to recognizable structures or regularities. The assessment of residuals facilitates the evaluation of the model's appropriateness and the detection of any residual patterns or irregularities that necessitate further scrutiny. Residuals are essential in the assessment of predictive precision and in confirming that the model adequately captures the inherent dynamics of the data.

# **4.Methodology**

The study embarks on an exploration of a multitude of deep learning techniques tailored specifically for forecasting solar irradiance. It embraces a diverse range of methods, including Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and hybrid models that combine various deep learning architectures. Through a meticulous analysis of these methodologies, the research aims to uncover the complex dynamics of solar irradiance prediction, with the goal of maximizing the strengths of each model to improve forecast accuracy and reliability. By diving into the intricacies of LSTM networks, CNNs, and hybrid models, the study seeks to pave the way for a comprehensive understanding of the most effective strategies for predicting solar irradiance in time series data.

**4.1 Data collection**

In order to forecast the Dst (Disturbance storm time) index, the necessary information must be measuring the properties of the Earth's magnetic field and the solar wind. Thus, we have collected the relevant data from the Goddard Space Flight Center of NASA. In order to forecast geomagnetic disturbances, predictive models are trained and validated using the data that has been gathered.

**4.2 Data Preprocessing**

Preparing collected data for analysis and modeling entails a crucial stage referred to as data preprocessing, which ensures the cleanliness, consistency, and appropriateness of the data for subsequent analysis. This phase involves multiple essential procedures, with a primary emphasis on two fundamental stages. The initial step deals with handling missing data by employing imputation techniques such as mean imputation or interpolation to address gaps within the dataset. Following this, the subsequent pivotal stage involves data standardization or normalization, which aims to scale the data for uniformity across variables. Usually, this involves transforming the data to have an average of zero and a standard deviation of one.

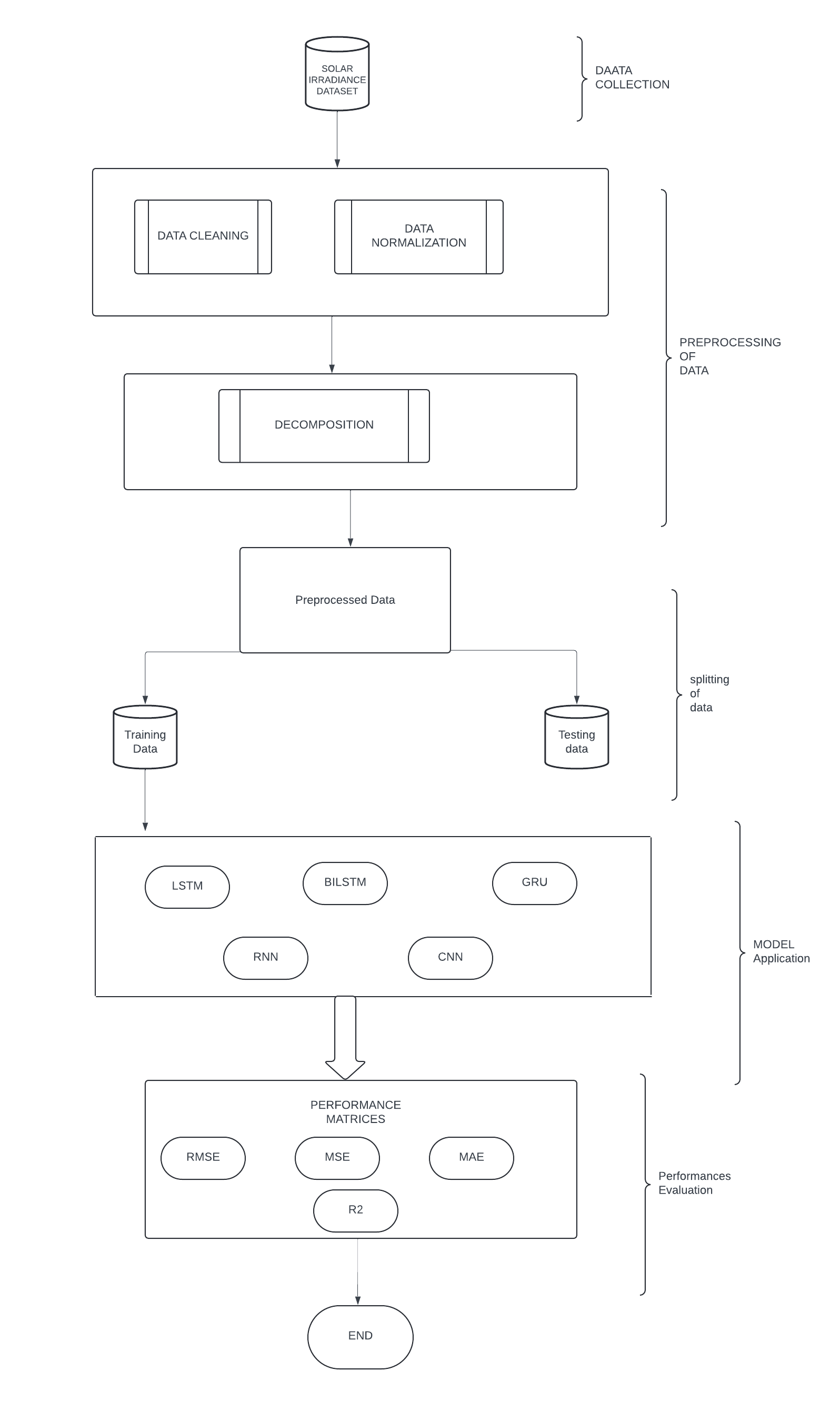
Within this investigation, data normalization is executed utilizing the MinMaxScaler() function, which scales numerical data to a predefined range, specifically between -1 and 1. These preprocessing methodologies significantly contribute to enhancing the quality of the data, thereby preparing it for training and validation within predictive modeling algorithms.

**4.3 Data Sampling**

**4.3.1 Training Data :-** This segment of the dataset serves as the training data for the model, enabling it to uncover patterns and relationships embedded within the dataset. Constituting the largest portion of the dataset, it plays a crucial role in training the model effectively. In this study, we have allocated 60% of the dataset from the outset specifically for training purposes.

**4.3.2 Validation Data:-** This segment of the dataset is specifically designated for the assessment of the model's efficacy during the training process and the refinement of its hyperparameters. The allocation of a distinct dataset for evaluating the model serves as a preventive strategy against overfitting. Following the utilization of the training dataset, a portion amounting to 20% has been reserved solely for validation purposes.

**4.3.3 Testing Data :-** This specific subset is utilized for the evaluation of the ultimate performance of the trained model. It consists of unseen data, which is essential for assessing the model's capacity to generalize to novel instances. To this end, 20% of the data from the latter part has been chosen, encompassing instances with the highest variability, for testing objectives.



**Figure 2:** Shows the flow chart of the methodology used in the study

**4.4 Model Selection**

Model selection refers to the procedure of choosing the most suitable model architecture or methodology for a specific task. The typical steps involved in the model selection process are outlined below:

**1.Selection of Candidate Model**

This step involves identifying a group of models that satisfy the job requirements, on the available data, and domain expertise. Here the used models are LSTM, BiLSTM, GRU, CNN and RNN.

2. **Evaluation metrics:**

It involves establishing the assessment criteria used to gauge the performance of each model. Within the investigation, the evaluation metrics determining performance include RMSE, MSE, MAE and R-square

**3.Training and Validation**

Utilizing the designated evaluation criteria, train each potential model using 60% of the data and assess its performance on the validation set comprising 20% of the data.

**4.Comparison**

By leveraging the evaluation metrics, it conducts a comparative analysis of the performance across different models. Through this process, the model demonstrating superior performance on the validation dataset is chosen.

**5.Model Selection**

By leveraging the evaluation metrics, it conducts a comparative analysis of the performance across different models. Through this process, the model demonstrating superior performance on the validation dataset is chosen.

**6.HyperparameterTuning**  
To improve the selected model's performance, one may opt to fine-tune its hyperparameters. , if the model's performance falls short of expectations, hyperparameter tuning is employed to iteratively adjust them until satisfactory results are achieved. Methods such as grid search, random search, or Bayesian optimization can be utilized for this purpose. Overall, crafting reliable and precise machine learning systems that meet designated objectives demands meticulous attention to the model selection process.

**4.4 Test Prediction**

When it comes to making predictions, testing involves utilizing the trained model on fresh, unseen data. The model's accuracy is assessed on data it hasn't encountered during validation or training. Following the model selection process, new test data is supplied to the model to generate predictions (y ̂). To gauge the model's accuracy, these predictions are compared against the actual values within the test dataset. Test prediction provides valuable insights into the model's real-world performance and assists in evaluating its ability to generalize to new data**.**

**4.5 Model Evaluation**

Model evaluation is the process of evaluating the performance of a trained model on unseen data. It involves using various metrics and techniques to measure how well the model generalizes to new data and achieves its intended task. The evaluation metrics used in this study are RMSE, MSE, MAE and These are discussed below in more detail: -

* **RMSE:-** Root Mean Square Error (RMSE) is a metric used to gauge performance by calculating the average difference between actual and predicted values, giving more weight to larger discrepancies.
* **MSE:-** This approach calculates the average of squared variances between predicted and observed values. A reduced MSE signifies superior model performance, indicating fewer discrepancies. MSE places greater emphasis on larger errors compared to smaller ones, encouraging models to improve overall accuracy.
* **MAE:-** The Mean Absolute Error (MAE) quantifies the typical magnitude of errors in predictions. It calculates the average absolute variance between predicted and actual values, offering a straightforward assessment of prediction precision. Lower MAE values signify better model performance by indicating fewer disparities between predicted and actual results.
* **R2 :-** R-squared (R2) acts as an indicator of the extent to which the independent variables can anticipate changes in the dependent variable. It spans from 0 to 1, with larger values indicating a stronger alignment between the model and the dataset. A heightened R2 value suggests that the model captures a significant portion of the data's variability, whereas a lower value indicates that the model elucidates only a minor portion of the variance..

Model evaluation aims to ensure that the model operates reliably and effectively when applied in real-world scenarios.

**5. Experiments**

**5.1 Parameter Settings**

* **Sequence Length** : In time series modeling, the sequence length, set at 10 in this case, denotes the number of consecutive time steps or observations utilized as input for the models in the analysis to generate predictions. This is specified as sequence\_length=10, signifying that the model makes predictions based on the preceding 10 timesteps.
* **Number of Time Steps:-** In this paper, the time steps represent the historical data points, with a length of 10, which corresponds to the sequence length employed by the model for making predictions in time series analysis.
* **Model Architectures and Hyperparameters:** In this study, various model architectures including LSTM, BiLSTM, GRU, CNN, and RNN were employed. The hyperparameters, such as the number of units in each layer, were set to 50 for LSTM, BiLSTM, GRU, and RNN, while for CNN, it was 64. The activation function utilized across all models was 'relu', and the optimizer chosen was 'adam'. CNN incorporated convolutional layers to detect patterns in the data, while LSTM, BiLSTM, and GRU utilized recurrent neural networks to capture temporal dependencies. The loss function employed during training was mean squared error (MSE). Model performance was evaluated using metrics such as RMSE, MAE, MSE, and R2.
* **Training Settings:** For training setting of the model, a batch size of 32 was used across all models for 11 epochs. The Adam optimizer, typically with a learning rate of 0.001, was utilized for optimization.
* **Number of Iterations:** For the Training process, total 10 iterations were used.
* **Data splitting:** The dataset underwent a partitioning process into three segments: training (60%), validation (20%), and testing (20%) subsets.
* **Normalization :** In preprocessing step, the normalization has been done using the MinMaxScaler to scale the data in a specific range, so that, model can learn efficiently.
* **Evaluation Metrices:** RMSE, MSE, MAE, and R2 were employed as metrics to assess the accuracy and effectiveness of the models' predictions. Enhanced performance is indicated by lower RMSE, MAE, and MSE values, whereas higher R2 values signify a superior alignment of the model with the data.

**5.2 Hardware Configuration of the system:-** The research project outlines specific tools and configurations needed to efficiently manage data processing, model development, and evaluation. These hardware prerequisites are detailed in the accompanying table.

|  |  |
| --- | --- |
| **Hardware Component** | **Specifications** |
| CPU | 2 x Intel Xeon CPU @ 2.20 GHz  - -1 core per processor  - -2 threads per processor (2 siblings)  -56,320 KB cache size per processor |
| RAM | At least 8 GB (more may be required for large datasets) |
| GPU | Intel Iris Xe Graphics |
| Storage | 512 GB SSD |
| OS | Windows 10 Home Single Language  Version 22H2 |

**Table 2:**Hradware Component explanation

**5.3 Software Configuration of the System: -** The software requirements for the research project should encompass essential tools and libraries necessary for tasks such as data processing, model construction, and evaluation. Consideration should be given to critical aspects such as:

* **Python :**- Python is widely employed across various domains such as web development, data analysis, artificial intelligence, machine learning, scientific computing, and automation. Its extensive collection of libraries and frameworks renders it appealing for rapid development and prototyping. Python's straightforward syntax and strong community backing make it accessible to developers of all levels. The research utilizes the latest Python version (3.12.3) for its endeavors.
* **Numpy:**- NumPy is an advanced Python library for numerical computations. It supports huge, multidimensional arrays and matrices, as well as a wide range of mathematical functions that may be applied to them.The version of NumPy used is (version:1.26.4). NumPy is a core package for scientific computing. It is extensively used for data analysis, machine learning, and other applications that demand rapid and efficient numerical data processing.
* **Pandas :-** This instance utilizes the most recent iteration of Pandas (Version: 2.2.2), a widely used Python library renowned for its capabilities in handling and analyzing data. Pandas facilitates data organization through structures such as Series and DataFrame, enabling users to effectively manage and analyze datasets. With robust features for data cleaning, transformation, and exploration, as well as versatile support for reading and writing data in multiple formats, Pandas serves as an essential tool for data scientists and analysts working with structured data.
* **Tensorflow:-** TensorFlow, a machine learning library developed by Google, is an open-source platform that provides a rich environment for creating, training, and deploying machine learning models. It accommodates various neural network architectures and offers utilities for tasks like data preprocessing, model training, and inference.
* **Keras:-**It is an open-source high-level neural network API that operates atop TensorFlow and other deep learning frameworks. Renowned for its intuitive and easy-to-use interface, Keras facilitates rapid prototyping, training, and deployment of machine learning models. Offering a wide range of neural network layers, activations, optimizers, and loss functions, Keras empowers users to construct complex models with minimal coding. Its user-friendly nature and flexibility make it a favored option among deep learning enthusiasts of varying skill levels. The specific version utilized in this context is (Version: 3.2.1).
* **Matplotlib :-** Matplotlib stands as a widely used open-source Python visualization library, offering features to generate static, interactive, and animated visualizations. It empowers users to craft high-quality plots, graphs, and charts with customizable styles, colors, and layouts. Matplotlib is renowned for its versatility and precision in designing visualizations, making it indispensable for tasks in data analysis and scientific computing. Whether for basic data exploration or sophisticated presentations, Matplotlib serves as a pivotal tool for creating informative and visually appealing representations of data. The version utilized in this context is the latest release (Latest Version: 3.8.4).

# **6. Result and Analysis**

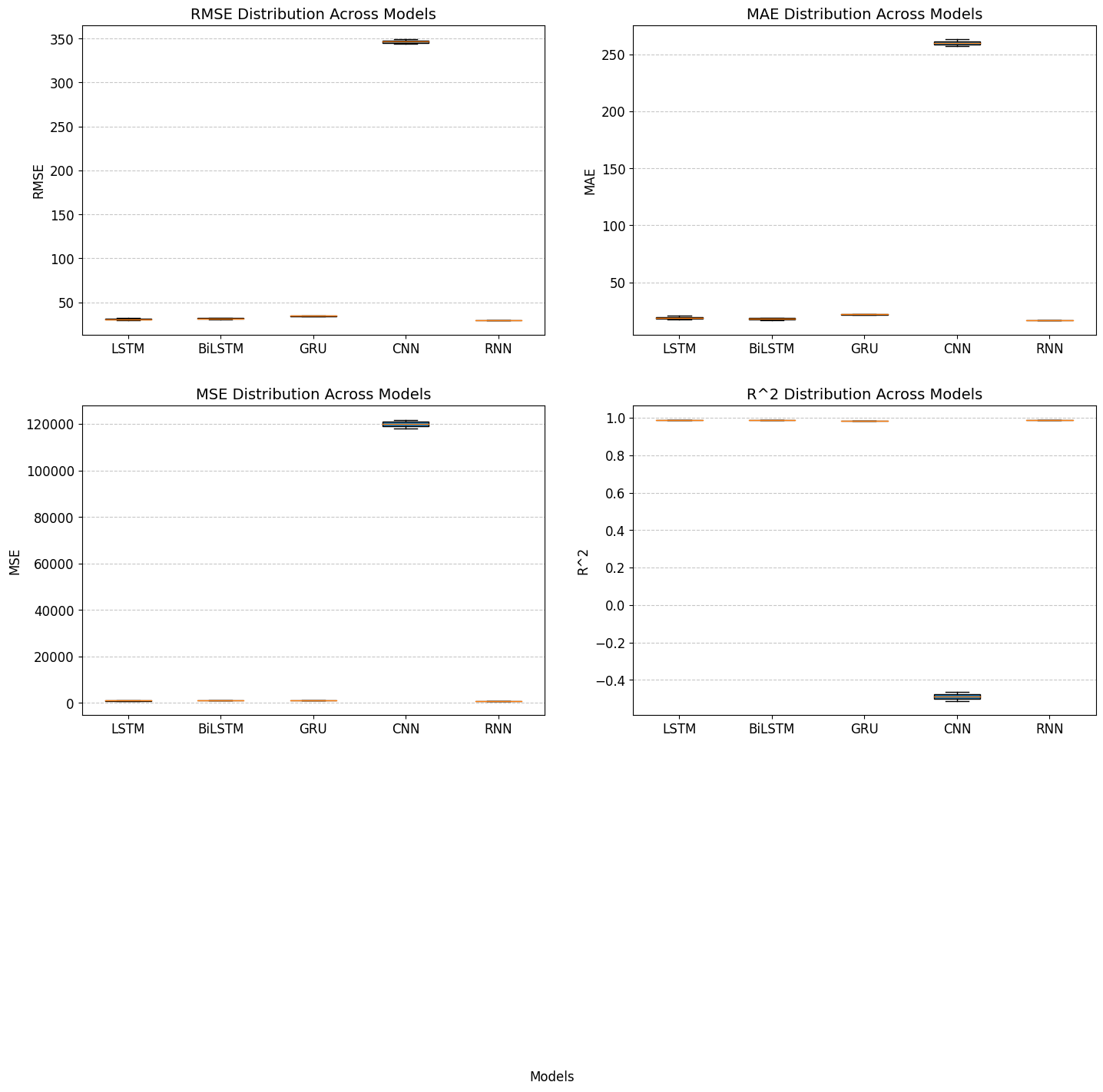
Solar irradiance prediction in major Indian cities has been extensively explored through the lens of deep learning methodologies. Numerous research endeavors have introduced a range of models, including Artificial Neural Networks (ANN), RLMD BILSTM, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and bidirectional LSTM, to tackle this task. These models have demonstrated promising capabilities in forecasting solar radiation, leveraging a combination of meteorological data, historical records, and geographical insights. The studies underscore the critical role of accurate solar irradiance predictions in facilitating the seamless integration of solar energy into smart grid systems, thereby promoting sustainability. Deep learning-based approaches have shown superiority over traditional forecasting techniques, exhibiting reduced errors and higher R-squared values. This underscores their effectiveness in navigating the intricacies and nonlinearities inherent in solar irradiance forecasting tasks.

**5.1 Quantitative Analysis**

# MUMBAI

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3 :** Performance matrices of solar irradiance of Mumbai | | | | |
| MUMBAI | RMSE | MAE | MSE | R2 |
| LSTM | 31.14 | 18.91 | 969.89 | 0.99 |
| BILSTM | 37.13 | 27.11 | 1379.19 | 0.98 |
| GRU | 32.10 | 19.64 | 1030.60 | 0.99 |
| CNN | 344.10 | 255.77 | 118407.56 | -0.47 |
| RNN | 30.46 | 17.99 | 928.12 | 0.99 |
| OVERALL MINIMUM VALUES | 30.32 | 17.99 | 920.12 | 0.99 |

In analyzing the provided data, it's evident that LSTM, GRU, and RNN models showcase superior performance compared to BiLSTM and CNN in terms of RMSE, MAE, MSE, and R2 metrics. Particularly, LSTM and RNN exhibit the lowest RMSE, MAE, and MSE values, indicating higher accuracy in prediction. Conversely, the CNN model stands out with notably higher error metrics and a negative R2 value, suggesting poor performance and inadequate fitting to the data. Overall, considering the minimum values across all models, the LSTM model yields the most optimal performance, closely followed by the RNN model, reaffirming their efficacy in this context.

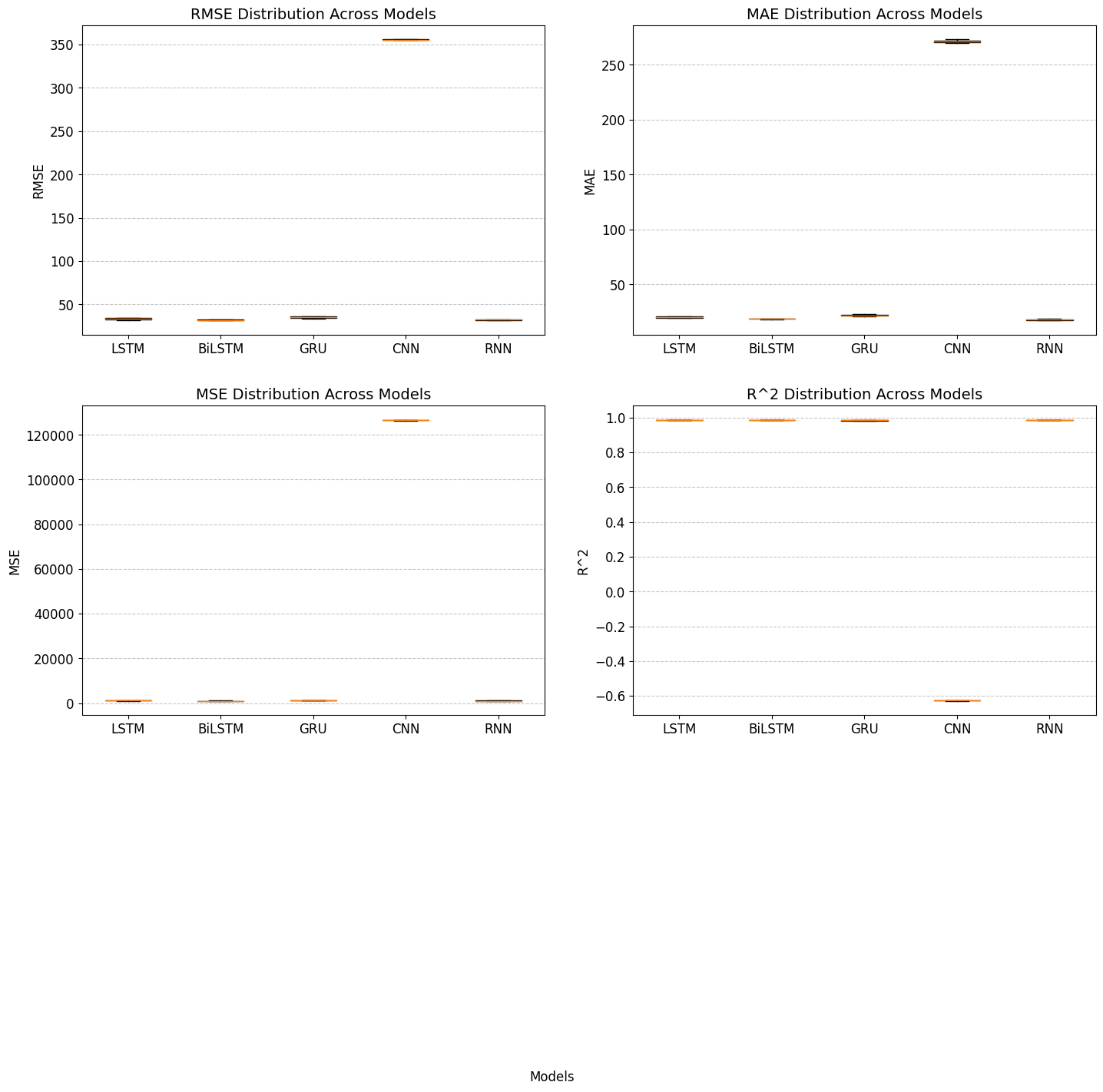
FfF

**Figure 3.**Box Plot of all the algorithms of MUMBAI

# **HYDERABAD**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| HYDERABAD | RMSE | MAE | MSE | R2 |
| LSTM | 33.26 | 20.25 | 1107.81 | 0.99 |
| BILSTM | 32.02 | 18.51 | 1025.65 | 0.99 |
| GRU | 35.05 | 21.92 | 1230.44 | 0.98 |
| CNN | 355.76 | 271.37 | 126567 | -0.63 |
| RNN | 32.17 | 18.00 | 1035.57 | 0.99 |
| OVERALL MINIMUM VALUES | 31.36 | 17.15 | 983.63 | 0.99 |

PERFORMANCE MATRICES OF HYDERABAD **TABLE-4**



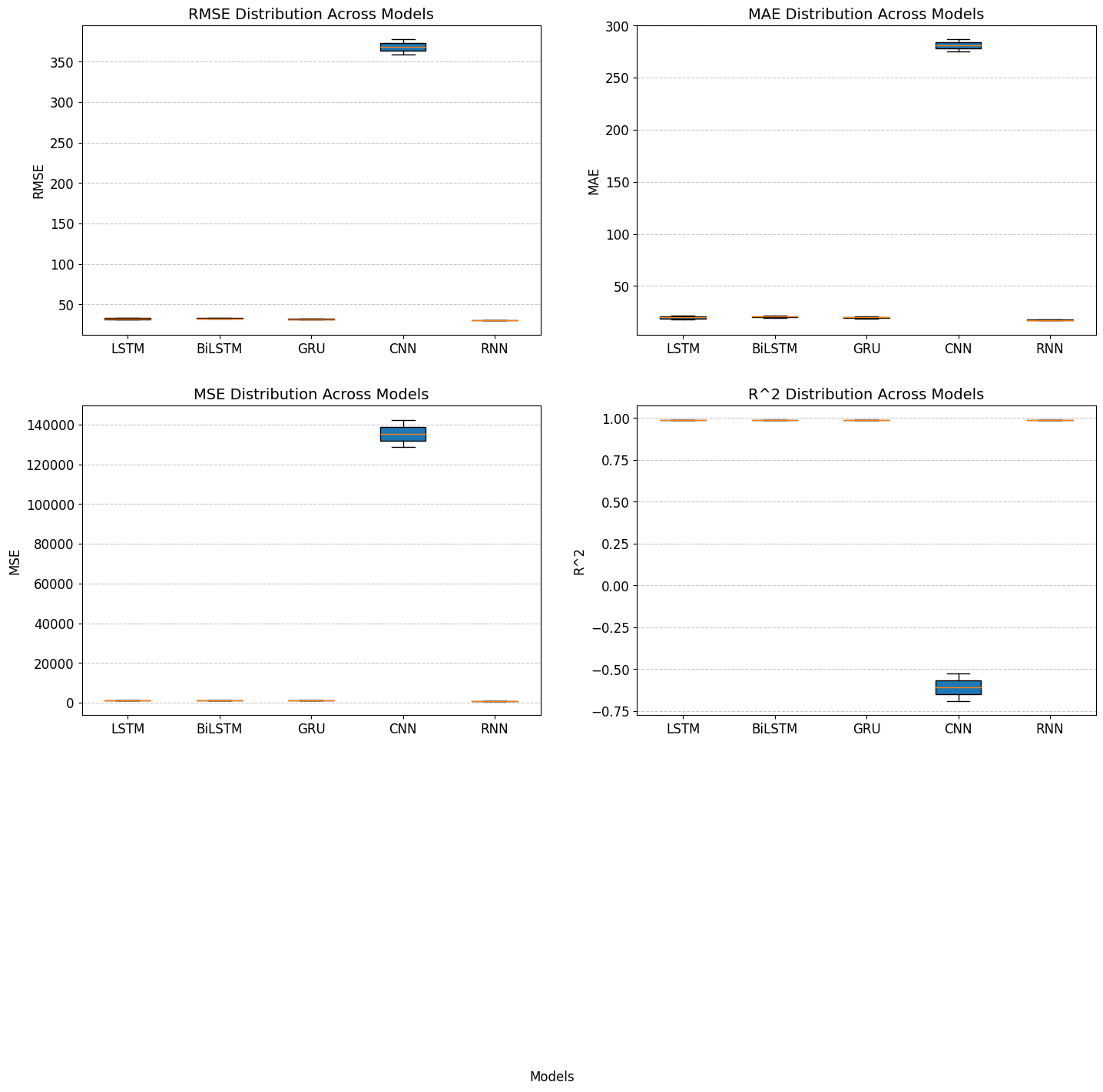
**FIGURE 4-** Box Plot of performance matrices of Hyderabad

Upon examining the data for Hyderabad, it's apparent that LSTM, BiLSTM, and RNN models demonstrate favorable performance compared to GRU and CNN models across RMSE, MAE, MSE, and R2 metrics. Particularly noteworthy are the competitive RMSE, MAE, and MSE values achieved by LSTM and BiLSTM models, indicating higher precision in prediction. Conversely, the CNN model displays considerably higher error metrics and a negative R2 value, indicating inadequate performance and fit to the data. Overall, considering the minimum values across all models, the BiLSTM model emerges as the most optimal performer, closely followed by the LSTM model, reaffirming their effectiveness in this context.

# **BANGLORE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TABLE 5 :** Performance matrices of banglore | | | | |
| BANGLORE | RMSE | MAE | MSE | R2 |
| LSTM | 32.41 | 19.73 | 1051.45 | 0.99 |
| 0.99 | 32.63 | 20.50 | 1064.98 | 0.99 |
| GRU | 31.73 | 19.76 | 1007.19 | 0.99 |
| CNN | 368.01 | 280.76 | 135519.50 | -0.61 |
| RNN | 30.21 | 17.42 | 912.78 | 0.99 |
| OVERALL MINIMUM VALUES | 30.18 | 16.90 | 910.63 |  |

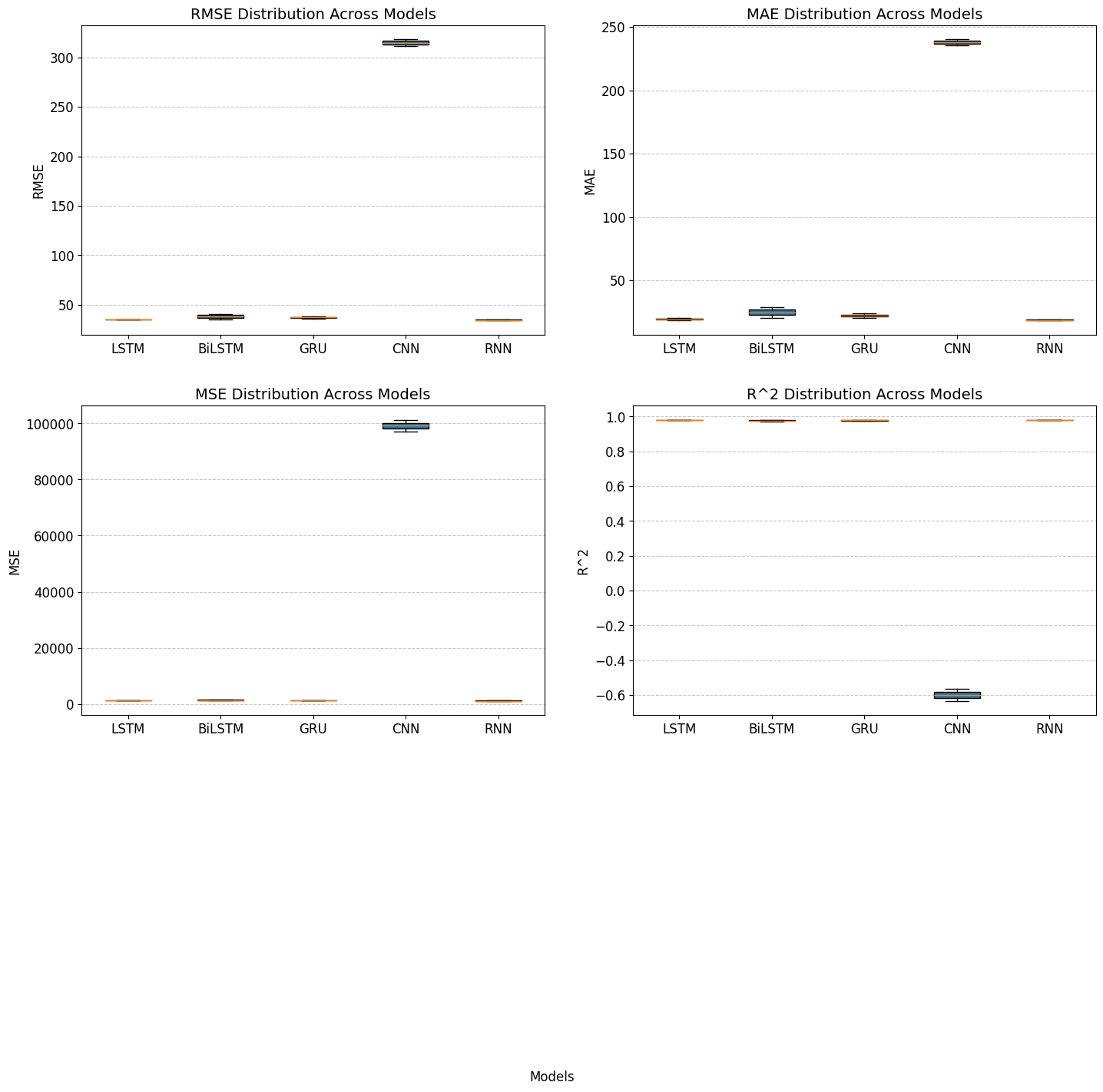
Upon review of the data pertaining to Bangalore, it's apparent that the LSTM, BiLSTM, and RNN models exhibit superior performance compared to GRU and CNN models across various metrics including RMSE, MAE, MSE, and R2. Notably, the RNN model showcases the lowest RMSE, MAE, and MSE values, indicating higher precision in prediction. Conversely, the CNN model demonstrates notably higher error metrics alongside a negative R2 value, suggesting suboptimal performance and a poor fit to the data. Overall, when considering the minimum values across all models, the RNN model emerges as the most optimal performer, closely followed by the LSTM model, underscoring their effectiveness in this specific context.



**FIGURE 5 :**Box plot of performance matrices of banglore

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 6:**Performance matrices of Kolkata | | | | |
| KOLKATA | RMSE | MAE | MSE | R2 |
| LSTM | 36.07 | 20.95 | 1303.25 | 0.98 |
| BILSTM | 36.48 | 21.56 | 1331.06 | 0.98 |
| GRU | 35.05 | 18.86 | 1228.78 | 0.98 |
| CNN | 317.92 | 238.41 | 101078.85 | -0.63 |
| RNN | 33.81 | 18.73 | 1142.94 | 0.98 |
| OVERALL MINIMUM VALUES | 33.77 | 18.21 | 1140.39 | 0.98 |

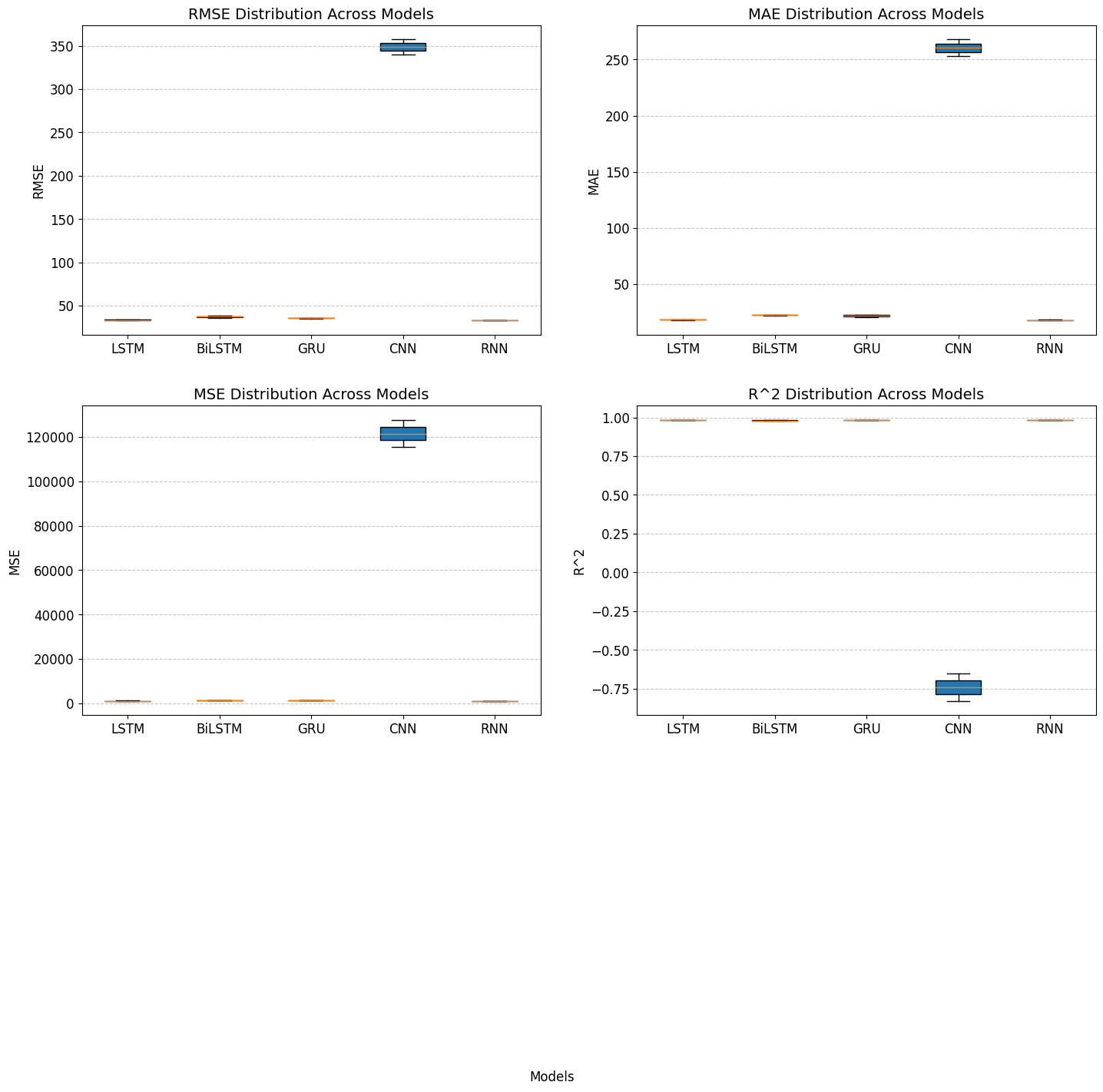
In Kolkata, various models were evaluated for their performance using metrics such as RMSE, MAE, MSE, and R2. Among these models, LSTM, BiLSTM, GRU, CNN, and RNN were assessed. The GRU model showcased the lowest RMSE, MAE, and MSE values, indicating higher accuracy in prediction compared to the other models. Conversely, the CNN model displayed significantly higher error metrics and a negative R2 value, suggesting poor performance and fit to the data. Overall, considering the minimum values across all models, the RNN model emerged as the most optimal performer, closely followed by the GRU model, showcasing their effectiveness in this context.



**FIGURE 6:** Box plot of performance matrix of kolkata

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 7 :** Performance Matrices of Delhi | | | | |
| NEW DELHI | RMSE | MAE | MSE | R2 |
| LSTM | 32.73 | 18.90 | 1117.58 | 0.98 |
| BILSTM | 35.13 | 20.27 | 1238.69 | 0.98 |
| GRU | 34.72 | 20.36 | 1206.68 | 0.98 |
| CNN | 337.35 | 254.65 | 113809.15 | -0.63 |
| RNN | 33.19 | 18.25 | 1101.55 | 0.98 |
| OVERALL MINIMUM VALUES | 32.73 | 17.96 | 1071.40 | 0.98 |

In New Delhi, a range of models underwent evaluation using metrics like RMSE, MAE, MSE, and R2. The LSTM model demonstrated the best performance with the lowest RMSE and MAE values, while the RNN model achieved the lowest MSE value. Despite this, the LSTM model emerged as the overall top performer, based on the minimum values across all metrics. Conversely, the CNN model displayed notably higher error metrics and a negative R2 value, indicating inadequate performance. In summary, the LSTM model emerged as the most effective, closely followed by the RNN model, highlighting their superiority in this particular scenario.



**Figure 7** Box Plot of Performance Matrix of Delhi

|  |
| --- |
|  |
| **Figure 8** Line plot of all the algorithms performance |

|  |
| --- |
|  |
| **Figure 9** Scatter Plot of all the algorithms Performance |

# **Conclusion**

The study focuses on predicting solar irradiance levels in metropolitan cities.The evaluation involved examining the performance of LSTM, BiLSTM, GRU, RNN, and CNN models using various metrics such as RMSE, MAE, MSE, and R2. Results showed diverse performance across the models:

LSTM: Consistently performed well with competitive RMSE, MAE, MSE, and R2 values.

BiLSTM: Achieved results similar to LSTM, albeit with slightly higher RMSE and MAE values.

GRU: Demonstrated competitive performance comparable to LSTM and BiLSTM.

RNN: Displayed performance akin to LSTM and GRU, exhibiting low RMSE, MAE, MSE, and high R2 values.

CNN: Consistently exhibited higher error metrics and negative R2 values, indicating inferior performance compared to other models. In summary, LSTM, BiLSTM, GRU, and RNN models consistently performed well, with LSTM often showcasing the most optimal results. Conversely, the CNN model consistently demonstrated inferior performance across all metrics and models.

**REFRENCES**

[1] R. A. Rajagukguk, R. A. A. Ramadhan, and H. J. Lee, “A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power,” *Energies (Basel)*, vol. 13, no. 24, Dec. 2020, doi: 10.3390/en13246623.

[2] X. Huang *et al.*, “Hybrid deep neural model for hourly solar irradiance forecasting,” *Renew Energy*, vol. 171, pp. 1041–1060, Jun. 2021, doi: 10.1016/j.renene.2021.02.161.

[3] S. Tajjour, S. S. Chandel, M. A. Alotaibi, H. Malik, F. P. Garcia Marquez, and A. Afthanorhan, “Short-Term Solar Irradiance Forecasting Using Deep Learning Techniques: A Comprehensive Case Study,” *IEEE Access*, vol. 11, pp. 119851–119861, 2023, doi: 10.1109/ACCESS.2023.3325292.

[4] K. Yan, H. Shen, L. Wang, H. Zhou, M. Xu, and Y. Mo, “Short-term solar irradiance forecasting based on a hybrid deep learning methodology,” *Information (Switzerland)*, vol. 11, no. 1, Jan. 2020, doi: 10.3390/info11010032.

[5] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, “Short-term solar irradiance forecasting in streaming with deep learning,” *Neurocomputing*, vol. 546, Aug. 2023, doi: 10.1016/j.neucom.2023.126312.

[6] M. C. Sorkun, Ö. Durmaz Incel, and C. Paoli, “Time series forecasting on multivariate solar radiation data using deep learning (LSTM),” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 1, pp. 211–223, 2020, doi: 10.3906/elk-1907-218.

[7] T. A. Siddiqui, S. Bharadwaj, and S. Kalyanaraman, “A deep learning approach to solar-irradiance forecasting in sky-videos,” Jan. 2019, [Online]. Available: http://arxiv.org/abs/1901.04881

[8] M. Husein and I. Y. Chung, “Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach,” *Energies (Basel)*, vol. 12, no. 10, 2019, doi: 10.3390/en12101856.

[9] N. Yogambal Jayalakshmi *et al.*, “Novel multi-time scale deep learning algorithm for solar irradiance forecasting,” *Energies (Basel)*, vol. 14, no. 9, May 2021, doi: 10.3390/en14092404.

[10] F. Wang, Y. Yu, Z. Zhang, J. Li, Z. Zhen, and K. Li, “Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting,” *Applied Sciences (Switzerland)*, vol. 8, no. 8, Aug. 2018, doi: 10.3390/app8081286.

[11] G. Guariso, G. Nunnari, and M. Sangiorgio, “Multi-step solar irradiance forecasting and domain adaptation of deep neural networks,” *Energies (Basel)*, vol. 13, no. 15, Aug. 2020, doi: 10.3390/en13153987.

[12] O. Bamisile, A. Oluwasanmi, S. Obiora, E. Osei-Mensah, G. Asoronye, and Q. Huang, “Application of deep learning for solar irradiance and solar photovoltaic multi-parameter forecast,” *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 2020, doi: 10.1080/15567036.2020.1801903.

[13] A. Alzahrani, P. Shamsi, C. Dagli, and M. Ferdowsi, “Solar Irradiance Forecasting Using Deep Neural Networks,” in *Procedia Computer Science*, Elsevier B.V., 2017, pp. 304–313. doi: 10.1016/j.procs.2017.09.045.

[14] A. H. Nielsen, A. Iosifidis, and H. Karstoft, “IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting,” *Solar Energy*, vol. 228, pp. 659–669, Nov. 2021, doi: 10.1016/j.solener.2021.09.073.

[15] S. Mishra and P. Palanisamy, “An Integrated Multi-Time-Scale Modeling for Solar Irradiance Forecasting Using Deep Learning.” [Online]. Available: https://www.rtoinsider.com/caiso-regulation-requirements-29299/

[16] K. Blazakis, Y. Katsigiannis, and G. Stavrakakis, “One-Day-Ahead Solar Irradiation and Windspeed Forecasting with Advanced Deep Learning Techniques,” *Energies (Basel)*, vol. 15, no. 12, Jun. 2022, doi: 10.3390/en15124361.

[17] M. Cihan Sorkun, C. Paoli, and Ö. Durmaz Incel, “Time Series Forecasting on Solar Irradiation using Deep Learning.”

[18] I. I. Prado-Rujas, A. Garcia-Dopico, E. Serrano, and M. S. Perez, “A flexible and robust deep learning-based system for solar irradiance forecasting,” *IEEE Access*, vol. 9, pp. 12348–12361, 2021, doi: 10.1109/ACCESS.2021.3051839.

[19] “1-s2.0-S1364032123002198-am”.

[20] Y. Gao, S. Miyata, and Y. Akashi, “Multi-step solar irradiation prediction based on weather forecast and generative deep learning model,” 2022.