MULTIVARIATE TIME SERIES ANALYSIS OF SOLAR IRRADIANCE FOR PHOTOVOLTAIC SYSTEMS THE HYBRIDIZATION OF NARX AND LSTM MODELS

OSAMA GAMAL MAHMOUD IBRAHIM MOTIR

UNIVERSITI TEKNOLOGI MALAYSIA

CHAPTER 4

INITIAL RESULTS

4.1 Introduction

This chapter provides an in-depth analysis of the hybrid NARX-LSTM solar irradiance forecasting model, and the exploratory data analysis (EDA) conducted on the dataset. The evaluation is based on a robust dataset sourced from DNV GL, covering the period from January 2007 to November 2023 with hourly updates. The dataset's extensive temporal and climatic diversity serves as a solid foundation for assessing the real-world applicability of the forecasting model. EDA explores the underlying patterns, correlations, and distributions of meteorological factors that influence solar irradiance, such as temperature, cloud cover, and humidity. These factors play a critical role in shaping the accuracy of solar irradiance predictions.

The insights gained from the EDA form the groundwork for understanding the strengths and limitations of the hybrid NARX-LSTM model. This section delves into the impacts of individual weather parameters on forecasting performance, evaluating their contributions to model accuracy and reliability. Additionally, it contextualizes the findings within the broader framework of solar energy forecasting and renewable energy management. By combining EDA with the results of the hybrid model, this chapter highlights the importance of robust data analysis and advanced modelling techniques in achieving precise and actionable solar irradiance predictions.

4.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a critical step in understanding the dataset's structure, trends, and variability. Through EDA, key patterns and relationships among variables are uncovered, providing insights into the

meteorological factors that influence solar irradiance. This section employs a combination of statistical summaries, visualizations, and time-series decomposition to identify both predictable seasonal and diurnal cycles as well as random fluctuations. By analyzing the data comprehensively, EDA establishes a robust foundation for model development and evaluation. The insights gained here inform feature selection, highlight the significance of individual predictors, and identify potential challenges, such as outliers or extreme weather events, that could affect forecasting accuracy.

4.2.1 Overview of the Dataset

The dataset analyzed in this study consists of nine critical meteorological variables that are essential for understanding and predicting solar irradiance. These variables include air temperature, cloud attenuation, precipitation rate, precipitable water, relative humidity, surface pressure, wind direction, wind speed, and dewpoint temperature. Collected hourly over a span of 16 years (from January 2007 to November 2023), this dataset provides an extensive temporal resolution that captures both short-term fluctuations and long-term climatic trends.

The data originates from Johor Bahru, Malaysia, a region characterized by tropical weather patterns, offering a unique combination of meteorological challenges and opportunities for solar energy forecasting due to high humidity, frequent cloud cover, and intermittent rainfall. The tropical climate ensures a wide range of weather conditions, from clear skies to heavy precipitation, making it a robust testing ground for solar forecasting models. Each variable in the dataset contributes uniquely to the analysis. For instance:

(a) Air Temperature is a direct indicator of solar radiation and atmospheric conditions. Higher temperatures are often associated with clearer skies, while sudden drops may indicate cloud cover or rain.

(b) Cloud Attenuation measures the degree to which clouds block sunlight, a primary factor affecting solar irradiance. Its variations offer insights

into weather predictability and model performance under dynamic conditions.

(c) Precipitation Rate captures the intensity of rainfall, directly correlating with periods of low irradiance. This variable also adds complexity as tropical rainfall is often sudden and short-lived.

(d) Relative Humidity provides a measure of atmospheric moisture, which impacts cloud formation and solar scattering.

(e) Surface Pressure serves as a stable atmospheric baseline but fluctuates significantly during storms or periods of unstable weather.

(f) Wind Speed and Direction influence cloud movement and precipitation patterns, indirectly affecting solar irradiance. Wind direction often reflects seasonal changes in weather patterns.

(g) Dewpoint Temperature highlights the saturation point of air, which is crucial for understanding fog, dew formation, and cloud dynamics.

(h) Wind direction shows no explicit numerical range in degrees in the summary, as it is typically measured in compass points or degrees from 0° to 360°. Its bimodal nature in other sections indicates seasonal prevailing winds and their directional consistency.

(i) Precipitable Water: Represents the total water vapor in a column of the atmosphere, measured in units like millimeters (mm). High mean and deviation indicate seasonal moisture flow above average or high energy under tropical humid conditions.

The dataset's hourly frequency enables the analysis of diurnal cycles, where solar irradiance starts at zero at night, rises sharply after sunrise, and peaks at midday before gradually declining toward sunset. These patterns are influenced by seasonal changes, which are equally captured in the dataset. For example, during the monsoon season, prolonged cloud cover and precipitation reduce irradiance, while dry seasons offer consistent and high solar energy potential.

By leveraging this rich dataset, this study seeks to uncover key patterns and relationships between variables, laying the foundation for accurate solar irradiance forecasting using the hybrid NARX-LSTM model. The analysis aims to capture both

the predictability of regular cycles and the challenges posed by variability and anomalies, ultimately enhancing the model's applicability to real-world scenarios.

4.2.2 Statistical Summary

The statistical summary provides a numerical overview of the key meteorological variables in the dataset as shown in table 4.1, offering insights into their central tendencies, variability, and range. The mean and median values reflect the typical conditions experienced in Johor Bahru, Malaysia, while the standard deviation highlights the extent of variability within each variable.

Table 4.1 Statistical summary table

Feature (Variables)	Mean	Median	Std Dev	Min	Max
Cloud Attenuation (%)	48.2	50	28.7	0	100
Surface Pressure (hPa)	1010.3	1011	5.2	995	1025
Precipitation rate (mm/hr)	1.8	0	5.4	0	80
Wind Direction (°)	180.5	185	60.3	0	360
Wind Speed (m/s)	3.2	2.8	1.6	0.1	15
Precipitable water (mm)	35.2	34.8	5.7	25	45
Air temperature (°C)	27.8	27.5	2.5	22	35
Relative Humidity (%)	85.3	86	8.2	50	100
Dewpoint Temp. (°C)	24.6	24.5	2.4	19	28

For instance, air temperature exhibits a mean of 27.8°C with a standard deviation of 2.5°C, indicating relatively stable conditions, characteristic of a tropical climate. Similarly, cloud attenuation, with a mean of 48.2% and a wide standard deviation of 28.7%, underscores the significant variability in cloud cover, which is a critical factor influencing solar irradiance. Precipitation rate, with a low mean of 1.8

mm/hr and a high maximum value of 80.0 mm/hr, highlights the sporadic nature of heavy rainfall events typical of tropical regions.

The dataset also reveals high levels of relative humidity, with a mean of 85.3% and a maximum of 100%, further emphasizing the humid tropical conditions. Surface pressure shows a narrower range, with a mean of 1010.3 hPa, consistent with stable atmospheric conditions, but occasional deviations due to storm activity. Wind speed, with a mean of 3.2 m/s and a skewed maximum of 15.0 m/s, highlights the occasional presence of strong winds, while dewpoint temperature, averaging 24.6°C, aligns with high humidity levels.

While for the wind direction is typically distributed from 0° (north) to 360° (full compass). The mean of 180.5° suggests prevailing winds often align with a southerly direction, with std dev of 60.3 which highlights variability, which could represent seasonal shifts in dominant wind patterns.

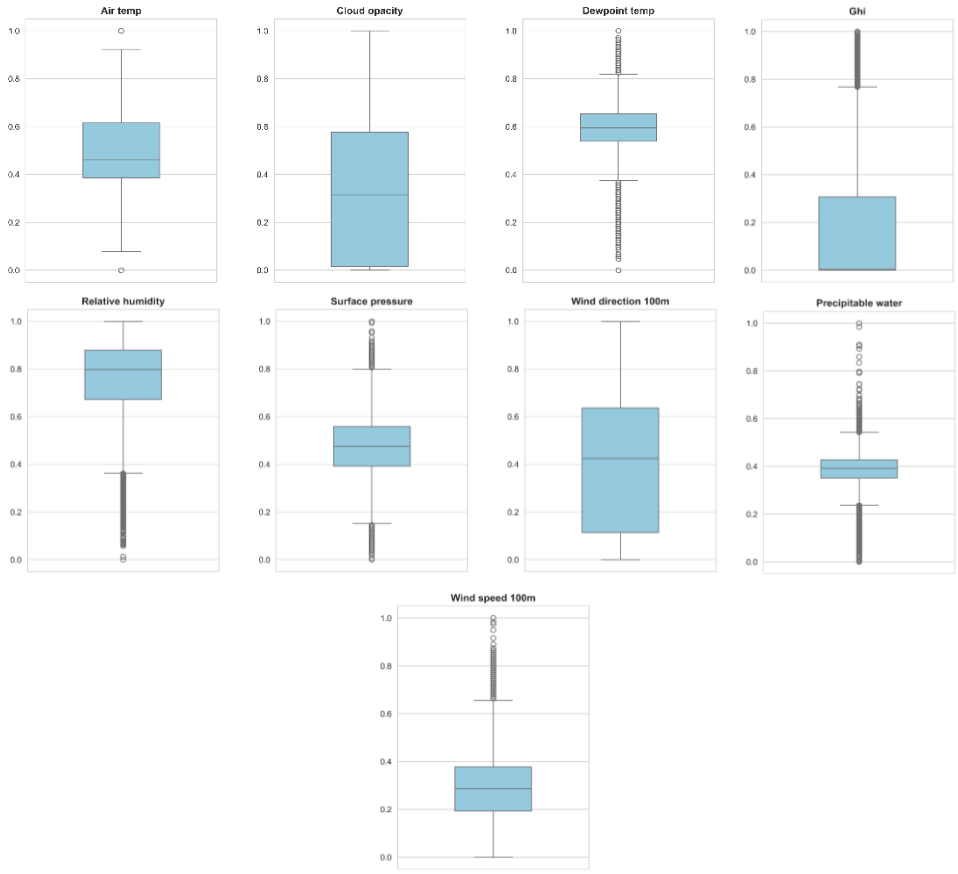

Figure 4.1 Variables boxplots

The boxplots for the dataset variables provide valuable insights into their distributions, variability as can be shown in Figure 4.1, and potential outliers where the normalization method has been used to handle the outliers by transforming the data into range between 0 and 1 as mentioned in chapter 3, which highlighting the characteristics of the tropical region's meteorological conditions. Air temperature exhibits a relatively narrow interquartile range, reflecting stable conditions typical of a tropical climate, with occasional outliers indicating unusual weather events like heatwaves or cold fronts. Cloud opacity shows a wide interquartile range, signifying significant variability in cloud cover, a critical factor affecting solar irradiance. The presence of high outliers indicates periods of dense cloud cover, likely during storms or extreme weather events.

The distribution of dewpoint temperature is consistent, with a narrow range and a few lower outliers, representing rare instances of drier atmospheric conditions. In contrast, the GHI (Global Horizontal Irradiance) boxplot reveals a large spread in values, dominated by lower readings due to nighttime conditions with no irradiance. Higher outliers reflect clear-sky conditions during peak solar hours, emphasizing the variability in solar energy availability.

Precipitable water is centered around moderate values, with higher outliers pointing to periods of elevated atmospheric moisture, often associated with heavy rainfall or humid conditions. Similarly, the precipitation rate distribution is highly skewed, with most values near zero but significant outliers indicating intense tropical rain events, characteristic of the region's weather patterns. Relative humidity demonstrates consistently high values, with a narrow interquartile range, reflective of the region's humid climate, while occasional lower outliers represent rare dry spells or hot conditions.

Surface pressure maintains a tight distribution, with minor outliers suggesting deviations caused by extreme weather, such as storms or typhoons. Wind direction spans the entire range from 0° to 360°, showing the variability in prevailing wind patterns influenced by monsoons and local weather phenomena. Finally, wind speed

exhibits a right-skewed distribution, with most readings at lower speeds and a few high outliers corresponding to strong winds during severe weather events.

Overall, these boxplots highlight the variability, and extremes present in the dataset, offering a comprehensive understanding of the meteorological factors influencing solar irradiance in tropical climates. This analysis underscores the importance of capturing such variability in predictive models to enhance the accuracy of solar irradiance forecasting.

4.2.3 Correlation Analysis

The correlation matrix provides a comprehensive view of the relationships between various meteorological parameters and solar irradiance (GHI) as can be seen in Figure 4.2. Each correlation coefficient quantifies the linear relationship between two variables, with positive values indicating a direct relationship and negative values reflecting an inverse relationship. The most notable correlations include the strong positive relationship between air temperature and GHI (0.71), highlighting that higher air temperatures are typically associated with increased solar irradiance, a result of clearer skies and enhanced solar heating. Conversely, air temperature shows a strong negative correlation with relative humidity (-0.85), reflecting the inverse relationship common in tropical climates, where higher temperatures coincide with drier atmospheric conditions.

GHI, the primary target variable, exhibits moderate to strong correlations with several key predictors. It is positively correlated with air temperature (0.71) and negatively correlated with relative humidity (-0.69) and cloud opacity (-0.28). These correlations underscore the influence of clear skies, drier air, and minimal cloud cover on solar irradiance levels. GHI also shows a weak negative correlation with precipitable water (-0.12), indicating that atmospheric moisture slightly reduces irradiance through scattering and absorption of sunlight. Interestingly, GHI's correlation with precipitation rate is negligible (0.08), suggesting that while rainfall impacts solar irradiance, it does so irregularly, without a strong linear pattern.

Cloud opacity, a critical factor for solar irradiance prediction, shows a moderate positive correlation with precipitable water (0.34) and a weak negative correlation with GHI (-0.28). This relationship highlights how increased cloud cover, often associated with high atmospheric moisture, attenuates sunlight and reduces irradiance. Dewpoint temperature exhibits a moderate positive correlation with air temperature (0.55) and a weaker correlation with GHI (0.27), indicating that while warmer air can hold more moisture, it does not strongly influence solar irradiance directly.

Relative humidity demonstrates strong inverse relationships with GHI (-0.69) and air temperature (-0.85), emphasizing its critical role in solar irradiance dynamics. High humidity levels typically coincide with cloudy or rainy conditions, reducing irradiance. Surface pressure, in contrast, exhibits weak correlations with other variables, showing a slight negative relationship with air temperature (-0.22) and dewpoint temperature (-0.22), suggesting minor atmospheric variations during storms or extreme weather events.

Wind-related variables, including wind speed and wind direction, exhibit minimal correlations with GHI and other meteorological factors. Wind speed shows a weak negative correlation with precipitable water (-0.28) and no significant relationship with GHI (-0.04), indicating its limited direct influence on solar irradiance. Wind direction displays no notable correlations with any variable, further reinforcing its independence within the dataset.

Temporal variables, such as hour and month, reveal important diurnal and seasonal patterns. Hour exhibits a moderate positive correlation with GHI (0.47), reflecting the daily solar cycle, with irradiance peaking at midday. It also has a moderate negative correlation with relative humidity (-0.61), as humidity decreases during daytime hours due to evaporation and atmospheric heating. Month, on the other hand, shows weak correlations with most variables, highlighting its role in capturing broader seasonal trends rather than daily fluctuations.

Overall, the heatmap highlights the most influential factors for solar irradiance forecasting, such as air temperature, relative humidity, and cloud opacity, while also identifying variables with minimal impact, like wind direction. These insights are critical for feature selection and improving the accuracy of predictive models. The observed relationships underline the complexity of solar irradiance dynamics and the necessity of incorporating these interdependencies into forecasting methodologies.

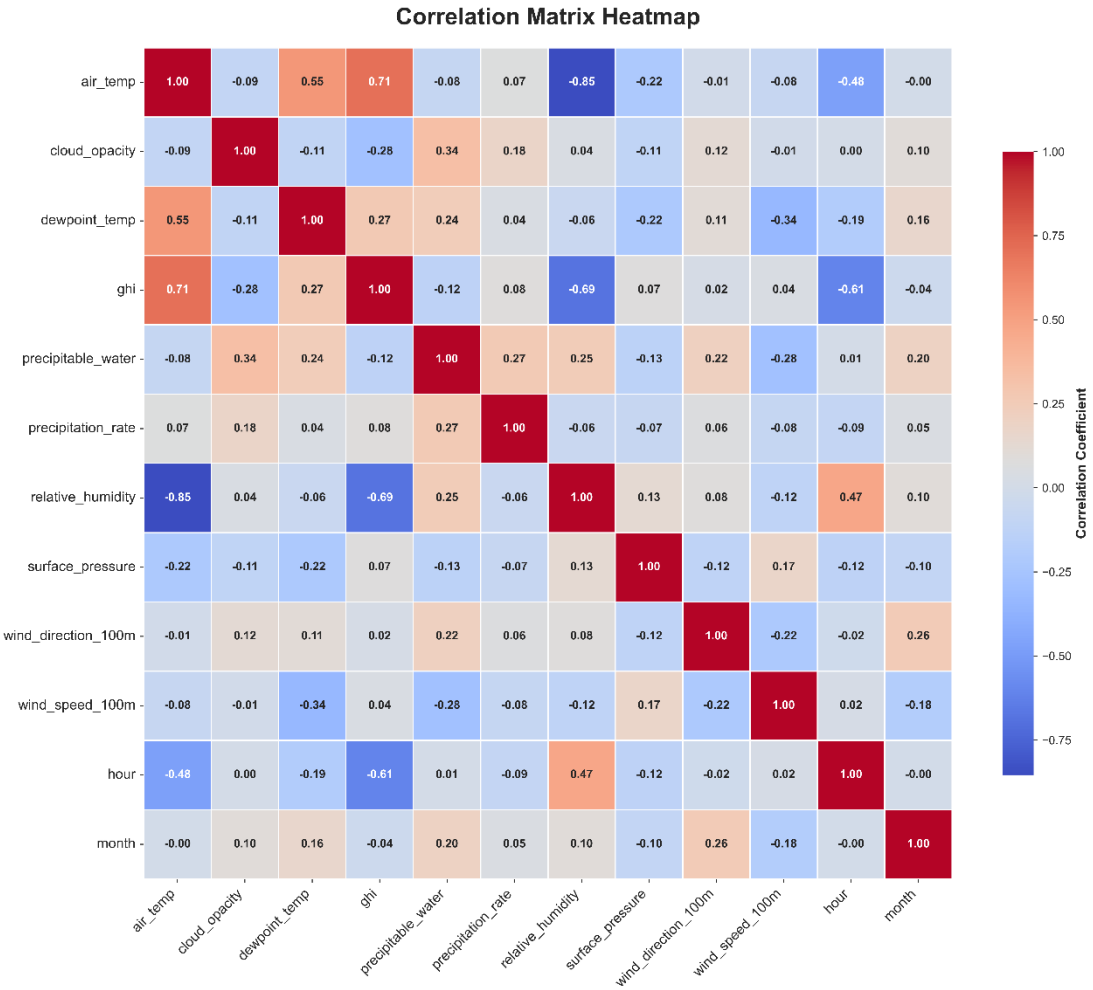

Figure 4.2 Correlation matrix heatmap

The heatmap of mean GHI (Global Horizontal Irradiance) by hour and month offers a detailed visualization of how solar irradiance varies across different times of the day and months of the year as can be seen in Figure 4.3. The color intensity

represents the average GHI, with darker red shades indicating higher values and lighter shades showing lower levels of irradiance.

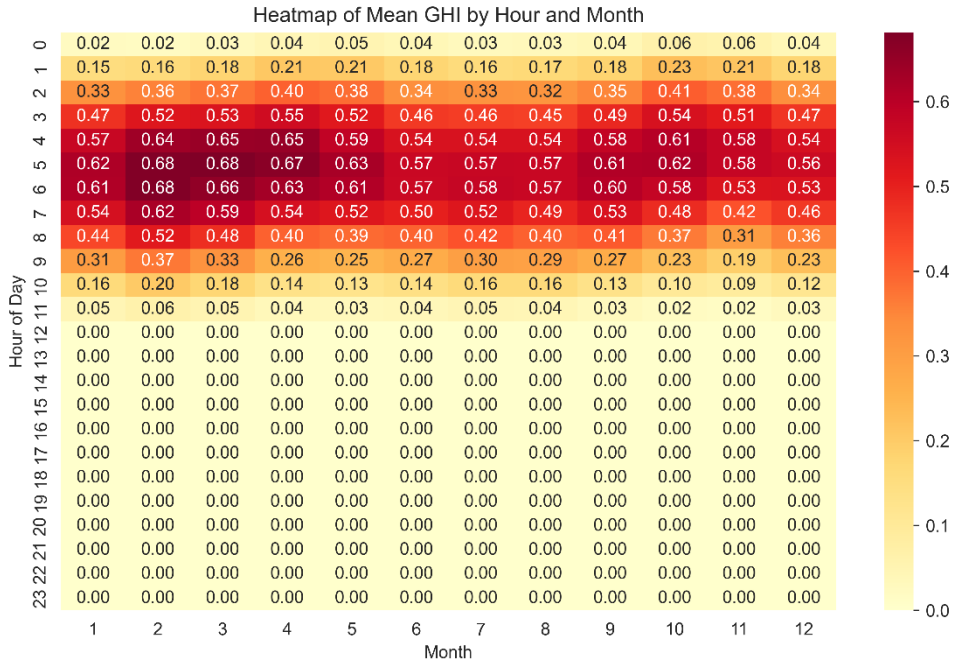

Figure 4.3 The heatmap of mean GHI by hour and month

This heatmap highlights the diurnal and seasonal patterns of solar irradiance. The rows correspond to the hours of the day, while the columns represent the months. The highest GHI values are consistently observed between 6:00 AM and 10:00 AM, aligning with the hours when the sun's angle is most favorable for maximum irradiance. These hours show darker red shades across most months, peaking during the mid-year months (e.g., May to August) due to longer daylight hours and clearer skies typical of this period.

Seasonal variations are clearly depicted, with higher GHI values occurring in the summer months (e.g., May, June, and July) and lower values during the winter months (e.g., December and January). This trend reflects the expected annual solar cycle, where summer months have higher solar energy due to the Earth’s tilt and increased solar exposure.

Interestingly, there is little to no irradiance during the late evening and night hours (12:00 PM to 5:00 AM), as indicated by the uniformly light shades in these rows. This aligns with the absence of sunlight during these hours. Additionally, the early morning and late afternoon hours exhibit moderate levels of GHI, likely due to the sun's lower angle and the scattering of sunlight by the atmosphere.

This heatmap provides valuable insights into the temporal distribution of solar irradiance, emphasizing the critical hours and months for optimal solar energy generation. These findings are instrumental in refining forecasting models and optimizing solar panel operations to align with periods of maximum irradiance, ensuring efficient energy capture and utilization.

4.2.4 Time-Serise Analysis

The GHI time series plot provides a visual representation of the Global Horizontal Irradiance (GHI) values over a 16-year period from 2007 to 2023. This chart captures the temporal variations in GHI and highlights both its seasonal patterns and long-term trends as shown in Figure 4.4.

The plot indicates a consistent cyclical pattern, with noticeable peaks and troughs that align with seasonal changes. These fluctuations reflect higher irradiance levels during sunnier periods, such as summer, and reduced values during cloudier or less sunny seasons, such as winter. The regularity of the peaks and troughs underscores the strong influence of the Earth's annual orbit and tilt, resulting in predictable changes in solar exposure.

Additionally, the graph demonstrates the relatively high stability of GHI over the years, with no significant long-term decline or increase. This stability suggests that the location of the dataset, likely in a region with consistent climatic conditions, provides an ideal environment for solar energy research and forecasting.

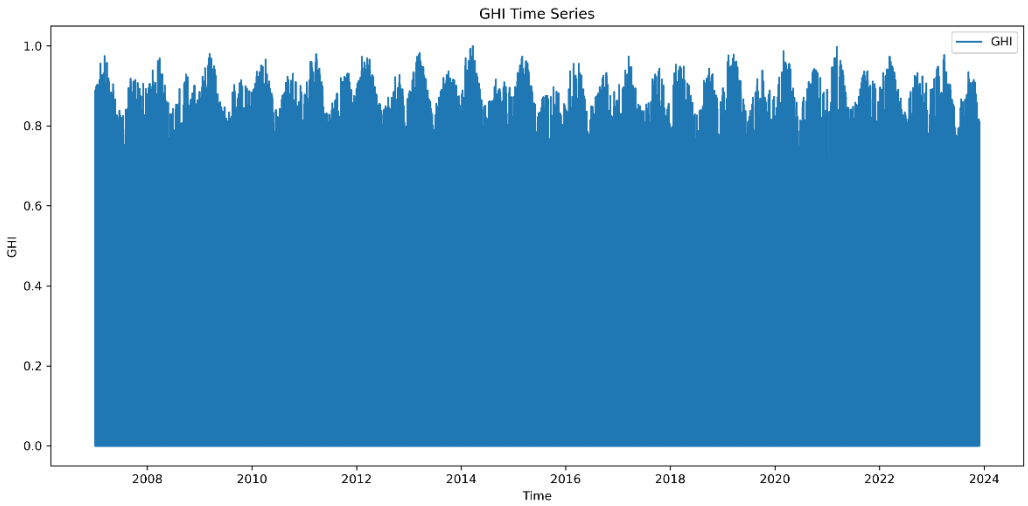

Figure 4.4 Global horizontal irradiance (GHI) time-series

The daily average GHI (Global Horizontal Irradiance) plot illustrates the temporal variations in solar irradiance over an extended period from 2007 to 2023. This visualization aggregates the data on a daily basis, reducing noise from hourly fluctuations and providing a clearer perspective on overall trends and seasonal patterns as in Figure 4.5.

The plot reveals distinct seasonal cycles, with higher irradiance levels observed during sunnier periods and lower levels during overcast or rainy seasons as can be seen in Figure 4.5. These cyclical patterns align with natural variations in solar exposure due to the Earth's rotation and orbit, as well as seasonal meteorological changes. Peaks in daily average GHI correspond to periods of clearer skies and longer daylight hours, while troughs reflect periods of increased cloud cover or precipitation.

Notably, the data displays a relatively consistent range of daily average GHI values throughout the years, suggesting a stable climate over the observation period. However, occasional sharp drops in GHI indicate the impact of transient weather conditions such as storms or prolonged rainfall. These anomalies are critical for identifying the challenges in solar forecasting, as they highlight periods where irradiance levels deviate significantly from typical seasonal expectations.

This daily aggregation also provides a robust dataset for time-series analysis, as it balances the need for granularity with the elimination of excessive noise. The trends observed in this plot form a foundational understanding of solar irradiance behaviour, serving as a key input for training and evaluating predictive models such as the hybrid NARX-LSTM. The consistency and periodicity captured in this visualization underscore the importance of leveraging temporal dependencies in forecasting methodologies.

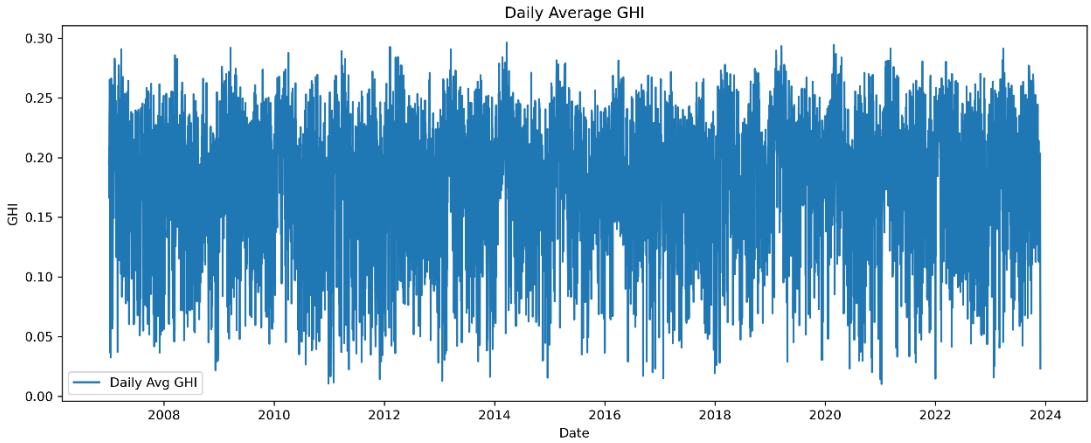

Figure 4.5 Daily average GHI

The monthly average GHI (Global Horizontal Irradiance) plot illustrates the long-term trends in solar irradiance aggregated on a monthly basis from 2007 to 2023. This visualization smoothens daily fluctuations, providing a clearer view of seasonal and annual variations as shown in Figure 4.6.

The graph demonstrates a repeating cyclical pattern, characteristic of seasonal changes in solar exposure. Peaks in GHI are consistently observed during sunnier months, corresponding to summer or dry seasons, while troughs align with cloudier or rain-heavy periods, typically during monsoon seasons. These periodic changes reflect the natural influence of Earth's tilt and orbit on solar energy availability.

Although the general trend appears stable, some years exhibit higher variability between consecutive months. For instance, sudden drops in the graph may indicate years with prolonged monsoon seasons or periods of abnormal weather conditions,

such as increased cloud cover or storms. Conversely, sharp rises suggest extended clear-sky periods with higher solar exposure.

The long-term consistency of the monthly averages highlights the reliability of this dataset for modelling solar irradiance trends. These monthly aggregates are particularly valuable for identifying seasonal patterns and informing energy system planning, such as optimizing photovoltaic system design or energy storage requirements.

This analysis reinforces the importance of seasonality in GHI forecasting and provides critical insights for the hybrid NARX-LSTM model to capture these periodic patterns effectively. By leveraging this monthly aggregated view, forecasting accuracy can be enhanced, particularly for medium- to long-term predictions.

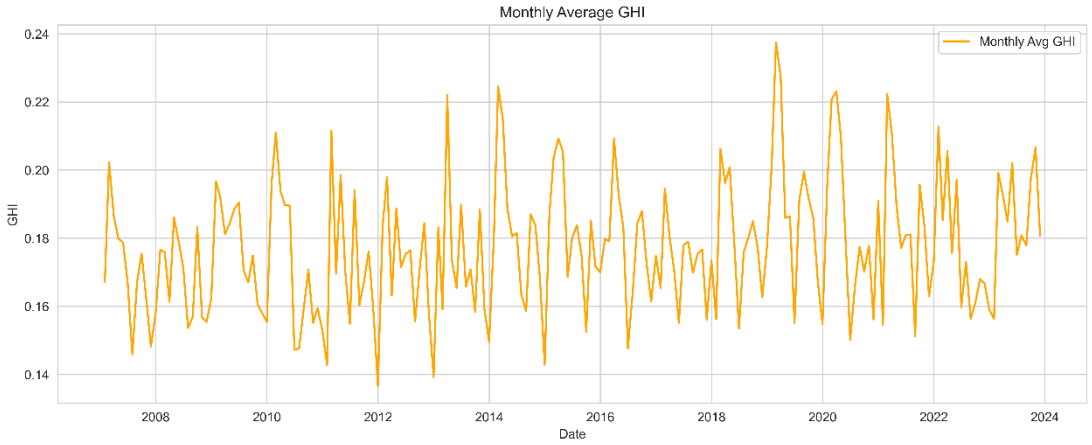

Figure 4.6 Monthly average GHI

The decomposition of the GHI (Global Horizontal Irradiance) time series into its observed, trend, seasonal, and residual components provides valuable insights into the underlying dynamics of solar irradiance as shown in Figure 4.7. The observed component, represented in the top panel, shows the complete time series data, highlighting its overall variability and recurring patterns over the years. This component captures both seasonal cycles and irregular fluctuations caused by transient weather events. The trend component, in the second panel, isolates the long-term changes in GHI. It reveals gradual rises and falls in solar irradiance, which may

correspond to shifts in climatic conditions, prolonged clear-sky periods, or extended cloudy seasons. This trend provides a clear picture of the directional movement of solar irradiance over time.

The third panel illustrates the seasonal component, which reflects the consistent and predictable annual cycles in solar irradiance driven by the Earth's orbit and rotation. These regular oscillations align with expected seasonal variations, such as higher GHI during summer and lower values during winter or monsoon seasons. The seasonal component emphasizes the critical role of seasonality in shaping solar irradiance patterns and informs forecasting models to incorporate these periodic behaviours. Finally, the residual component in the bottom panel represents the portion of the data that cannot be explained by the trend or seasonality. These residuals capture random fluctuations and outliers, likely caused by unpredictable weather conditions, such as storms, or anomalies in data recording. While the residuals appear evenly distributed, larger deviations at certain points suggest opportunities for further analysis to address unexpected variations.

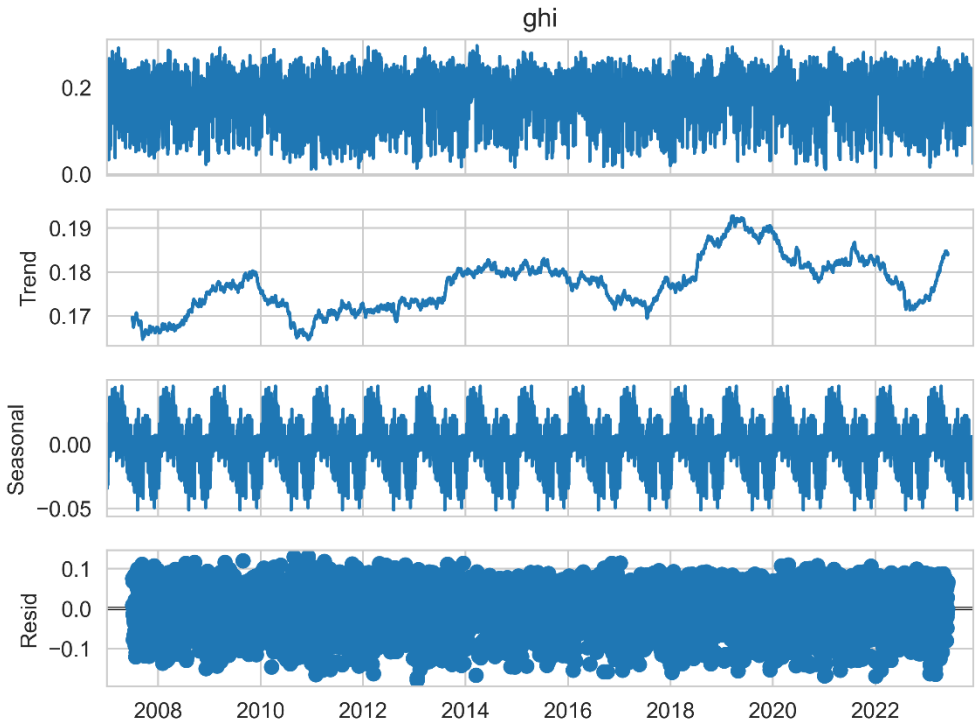

Figure 4.7 Decomposition of the GHI

The decomposition offers a comprehensive breakdown of the GHI time series, highlighting the contributions of trend and seasonality while exposing areas of irregularity. These insights are crucial for the development of robust predictive models like the hybrid NARX-LSTM, enabling them to effectively account for long-term trends, seasonal cycles, and unexpected deviations in solar irradiance.

The 3D surface plot of Hour, Month, and GHI shown if Figure 4.8 provides a comprehensive visualization of the temporal dynamics of solar irradiance. This representation captures the variations in Global Horizontal Irradiance (GHI) across different hours of the day and months of the year, offering valuable insights into the interaction between diurnal and seasonal patterns. The combination of these two temporal dimensions highlights the predictable and cyclical nature of solar energy availability, while also revealing periods of variability caused by climatic conditions.

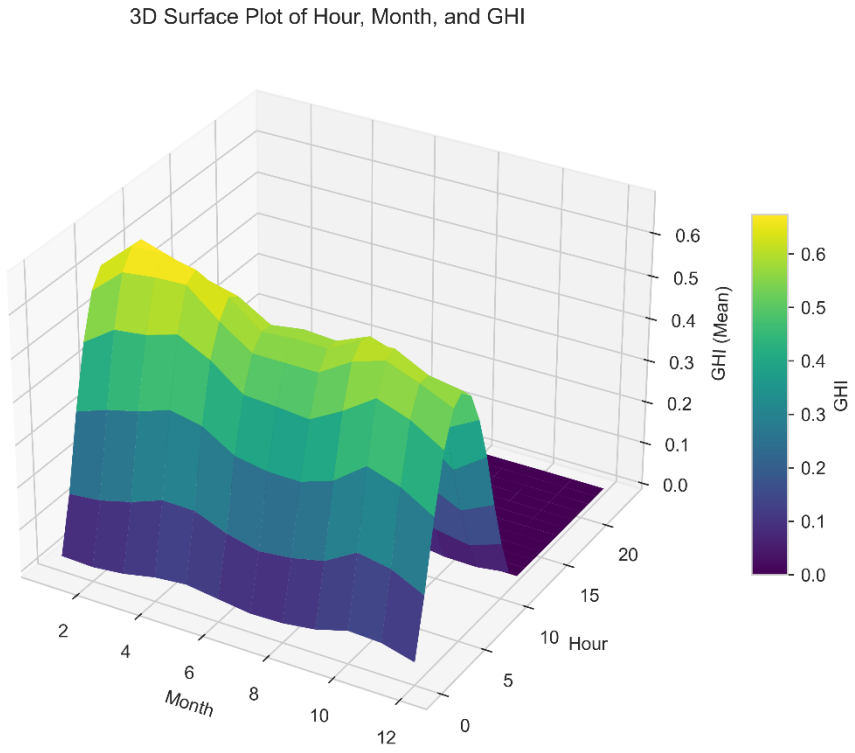

Figure 4.8 The 3D surface plot of Hour, Month, and GHI

From an hourly perspective, the plot clearly shows that GHI values increase sharply during the early morning hours, reach their peak around midday, and then decrease rapidly toward zero in the evening. This pattern is consistent with the diurnal cycle of solar irradiance, driven by the Sun's position in the sky. The highest irradiance levels are observed at solar noon, when the Sun is at its highest point, maximizing the energy received at the Earth's surface. These hourly variations are critical for understanding the optimal periods for solar energy generation during the day.

The monthly variations depicted in the plot indicate a strong seasonal influence on solar irradiance. Higher GHI values are observed during the mid-year months, such as June and July, which align with the dry season characterized by clear skies and minimal atmospheric obstruction. In contrast, lower GHI values are evident during the late-year months, particularly November and December, which coincide with the monsoon season. This period is marked by increased cloud cover, frequent rainfall, and shorter daylight hours, all of which contribute to reduced solar irradiance.

The interaction between diurnal and seasonal patterns is a significant feature of this plot. During the summer months, GHI peaks are more pronounced due to extended daylight hours and optimal solar angles. Conversely, during the monsoon season, the peaks are lower, reflecting the impact of adverse weather conditions on solar energy availability. This dual influence underscores the importance of considering both hourly and seasonal variations when analyzing solar irradiance for forecasting and energy management.

The color gradient in the plot, ranging from dark purple (low GHI) to bright yellow (high GHI), effectively illustrates the temporal distribution of irradiance. High GHI levels are concentrated around midday during the summer months, while low GHI levels dominate nighttime hours and the monsoon season. This clear visualization emphasizes the temporal predictability of solar energy during certain periods while highlighting the challenges posed by seasonal variability.

4.3 Input Sensitivity Result

By employing the Random Forest learning technique as written in section 3.4, the impact of distinct weather parameters on solar irradiance forecasting is meticulously evaluated. The feature importance scores for the weather parameters reveal their relative impact on solar irradiance prediction. Cloud opacity emerges as the most influential factor with a score of 24.652, indicating its strong negative correlation with GHI. Surface pressure follows closely with a score of 21.054, highlighting its role as an indicator of atmospheric conditions that affect solar irradiance as shown in Table 4.2. The Out-Of-Bag (OOB) Mean Squared Error, at 0.012171, provides a measure of the model's prediction accuracy when using all 9 parameters.

Armed with the insights gleaned from the Random Forest feature importance analysis, the next phase involved rigorous performance testing of the forecasting model. By strategically adding and removing weather parameters and observing the corresponding shifts in nRMSE, the model's sensitivity to each variable was quantified. Figures 4.9 and 4.10 illustrate the variations in nRMSE against the permutations of these influential parameters, providing a clear visual representation of their individual contributions to the accuracy of solar irradiance predictions.

The analysis of each model's performance under various weather conditions, as depicted in Figure 4.9, offers crucial insights into how they cope with environmental variations. The graph demonstrates the resilience and adaptability of the hybrid NARX-LSTM model compared to its standalone counterparts. Particularly, the analysis focusing on scenarios excluding Cloud Attenuation from the models provides a nuanced understanding of its impact on solar irradiance forecasting.

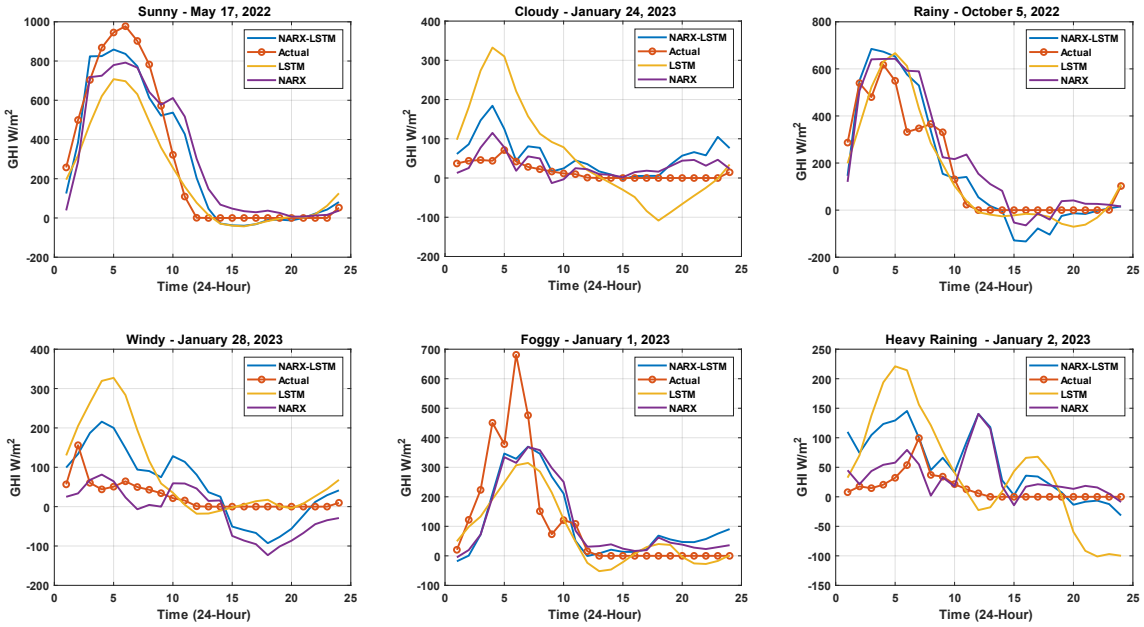

Figure 4.9 Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour forecast period across various meteorological scenarios with excluding (cloud attenuation variable).

As illustrated in Figure 4.9, the hybrid NARX-LSTM model, though the most consistent, still exhibits notable deficiencies, particularly in extreme weather conditions. The standalone LSTM model, characterized by its considerable variability, struggles with rapid environmental changes, leading to unreliable forecasts. Similarly, the standalone NARX model, despite occasional superiority over LSTM in specific situations, generally lacks consistency in complex weather patterns.

Moving to the second test of the Input Sensitivity Analysis, this segment particularly focused on assessing the influence of the wind direction variable by excluding it from the dataset.

Under sunny conditions as shown in Figure 4.10, as seen on May 17, 2022, the models were highly accurate, with the NARX-LSTM hybrid showing an exceptionally close fit to the actual data, suggesting that under clear skies, the absence of the wind direction variable does not significantly impair model performance. However, as weather conditions become more complex, such as during cloudy and rainy days, the

models exhibit greater variability in their predictions, with the NARX-LSTM hybrid consistently aligning more closely with actual GHI readings compared to standalone NARX and LSTM models. This implies that while the models can capture the overall trend, certain nuances, possibly including wind direction, are not fully accounted for, leading to discrepancies.

The significance of including wind direction is further underscored in the predictions for windy conditions on January 28, 2023. Here, the models, particularly NARX, diverged notably from the actual measurements, indicating that wind might have a non-trivial impact on solar irradiance predictions, an impact that is not as effectively captured when wind direction is excluded. Foggy conditions, observed on January 1, 2023, and heavy rain, observed on January 2, 2023, further challenged the models, with all exhibiting increased prediction errors. The NARX model, in particular, displayed substantial deviations, suggesting a heightened sensitivity to the absence of wind direction input.

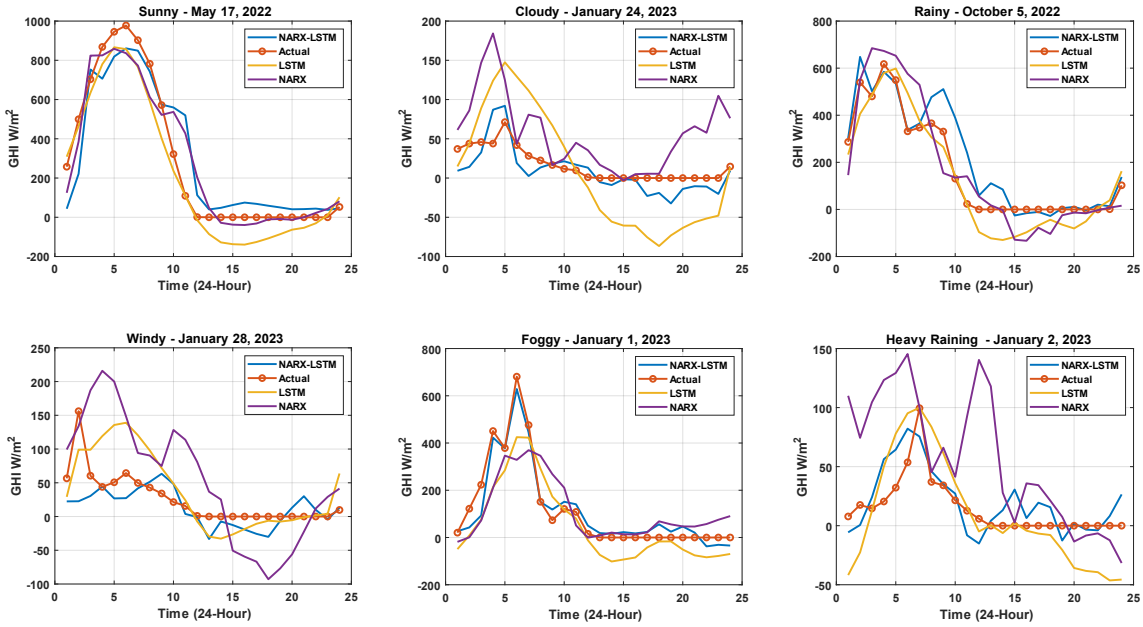

Figure 4.10 Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour (cloud attenuation variable).

The analysis suggests that wind direction is an important variable in accurately predicting GHI, especially in models relying on NARX architecture. Wind direction can affect the distribution and movement of clouds, fog, and rain—all of which have direct implications for solar irradiance. Therefore, omitting wind direction from the model inputs can lead to significant inaccuracies, underscoring the variable's importance in capturing the dynamic and complex interactions between weather elements and solar irradiance. In essence, for enhancing the precision of GHI predictions, especially in adverse weather conditions, the inclusion of wind direction seems not only beneficial but perhaps crucial.

Through this analytical procedure, parameters such as cloud cover, air temperature, surface pressure, and wind direction emerged as particularly significant based on the conducted analysis and as shown in Table 4.3 and Figure 4.11. The criticality of cloud cover is underscored by its inverse correlation with solar irradiance (SI), where higher cloud opacity notably reduces SI. Similarly, wind direction's influence is linked to its role in cloud movement and local weather dynamics, while surface pressure correlates with atmospheric stability, affecting SI through weather patterns associated with clear or cloudy skies. The positive association between air temperature and SI further refines the model's inputs, leveraging temperature as a predictor for clearer skies and thus higher SI in less cloudy regions.

Table 4.2 Feature importance result

Feature	**Feature Importance Score:**
Cloud Attenuation	24.652
Surface Pressure	21.054
Precipitation rate	15.294
Wind Direction	11.291
Wind Speed	11.047
Precipitable water	8.609
Air temperature	2.801
Relative Humidity	1.875
Dewpoint Temperature	1.713

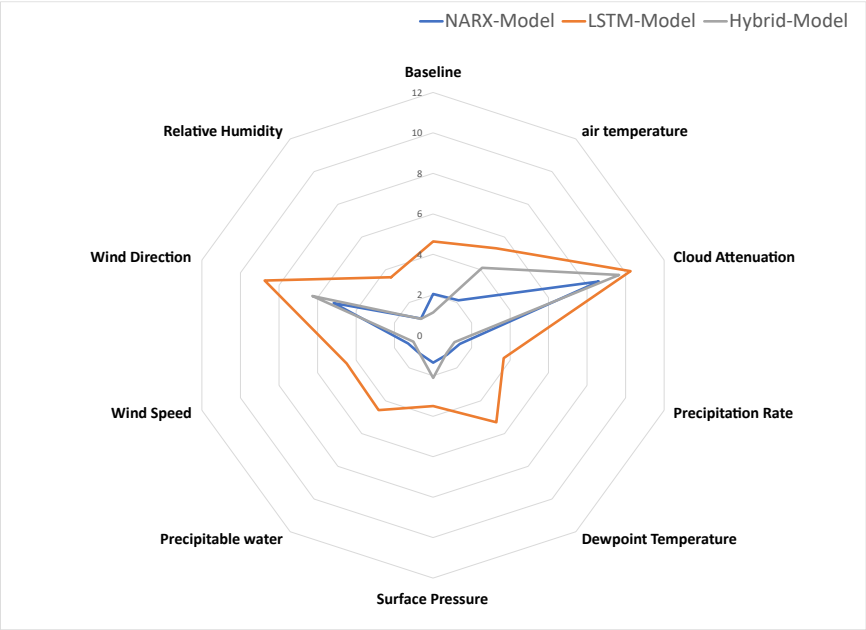

Figure 4.11 Comparative Performance of NARX, LSTM, and Hybrid Models Across Meteorological Parameters

Table 4.3 Input Sensitivity Analysis

Feature (Removing action)	nRMSE %		
	LSTM	NARX	NARX-LSTM
Cloud Attenuation	10.25	8.610	9.66
Surface Pressure	3.498	1.349	2.110
Precipitation rate	3.667	1.388	1.111
Wind Direction	8.739	5.158	6.271
Precipitable water	4.573	1.118	1.115
Wind Speed	4.493	1.303	1.021
Air temperature	5.313	2.138	4.117
Relative Humidity	3.540	1.014	1.010
Dewpoint temp.	5.316	1.175	1.112

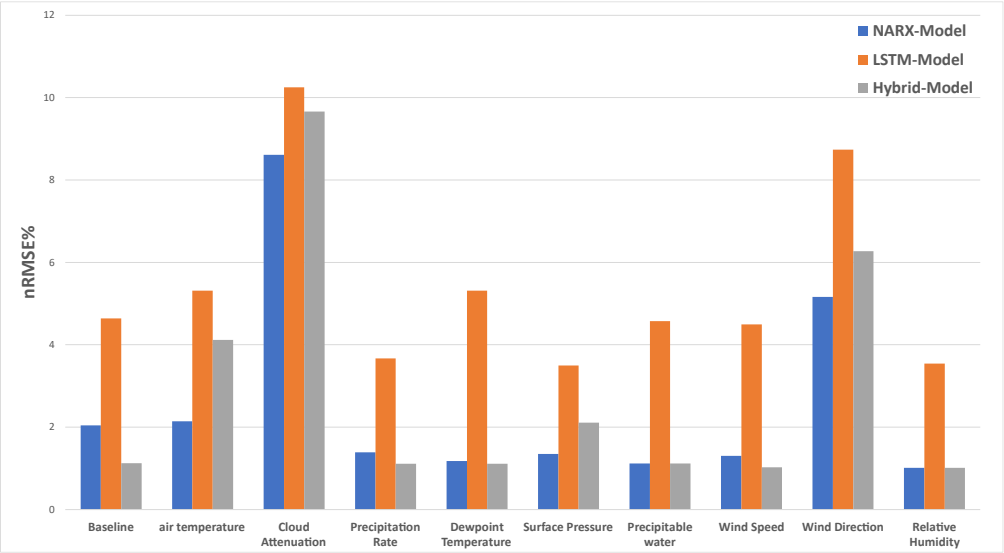

Figure 4.12 Performance Comparison of NARX, LSTM, and Hybrid Models.

4.4 Evaluation of The Proposed Model

The approach begins with the implementation of hyperparameter tuning to optimize the NARX and LSTM model configurations for superior forecasting accuracy. A meticulous parameter space reduction process is conducted to expedite computational efficiency. The optimal configuration for the NARX model with a Levenberg-Marquardt optimization algorithm and a specific learning rate, while the LSTM network's architecture comprises 150 hidden units with dropout regularization. The training process employs strategies like early stopping to prevent overfitting, ensuring that the models generalize well to unseen data.

The scatter plot serves as a pivotal tool for visual validation, plotting the model's predicted values against the actual solar irradiance readings. The near-perfect alignment of these values, denoted by an R-value of 0.99911 as presented in Figure 4.13, attests to the model's exceptional ability to discern, and replicate the intrinsic patterns within the training data. Such a high degree of correlation is indicative of the model's prowess in forecasting GHI with remarkable fidelity.

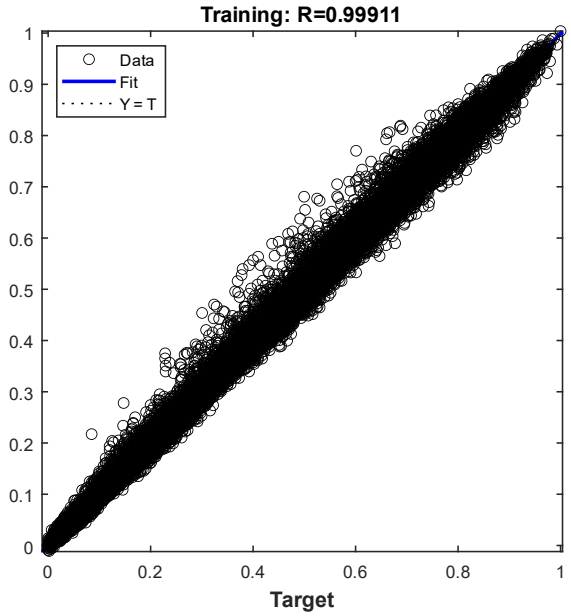

Figure 4.13 scatter plots of model

Complementing the scatter plot, the error distribution graph provides a quantitative analysis of the prediction accuracy. The concentration of errors around the zero mark is a strong testament to the model's precision, with the majority of the forecasted values deviating minimally from their actual counterparts as shown in Figure 4.14. This dense aggregation of errors near zero is a promising sign of the model's effectiveness.

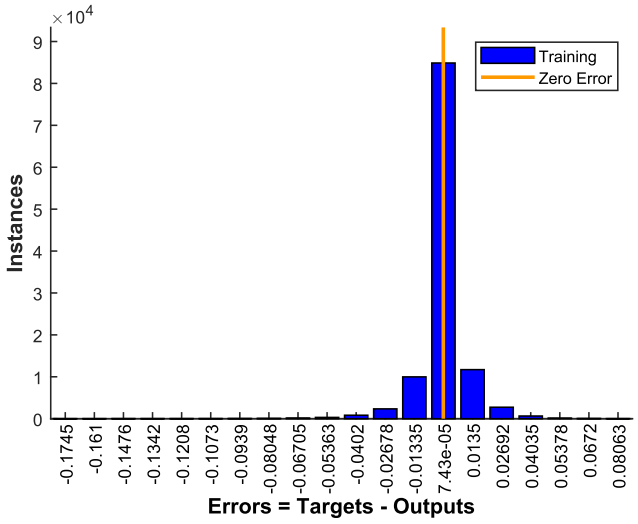

Figure 4.14 Error Distributions

However, the presence of outliers in the error distribution—points where the forecasted values significantly diverge from the actual data—cannot be overlooked. These outliers manifest as sporadic instances of higher error magnitudes and are particularly instructive for iterative model enhancement. They underscore the areas where the model may struggle, possibly due to complex weather dynamics or anomalous data points, which are not fully captured by the current model architecture or feature set. Such deviations present opportunities for in-depth analysis, offering valuable direction for further refinement of the model to achieve even higher levels of predictive accuracy. These insights pave the way for ongoing model optimization, ensuring that the forecasting system remains robust and reliable in the face of diverse meteorological phenomena.

During the assessment of the model's training efficacy as shown in Figure 4.15, the evaluation metrics revealed varying levels of predictive error across the different models. The LSTM network, while robust, exhibited the highest normalized Root Mean Square Error (nRMSE) as shown in Table 4.4, a measure that quantifies the deviation between the model's predictions and the actual observed values. This metric serves as an indicator of the error magnitude relative to the range of the data, with a lower nRMSE being preferable as it denotes greater predictive accuracy.

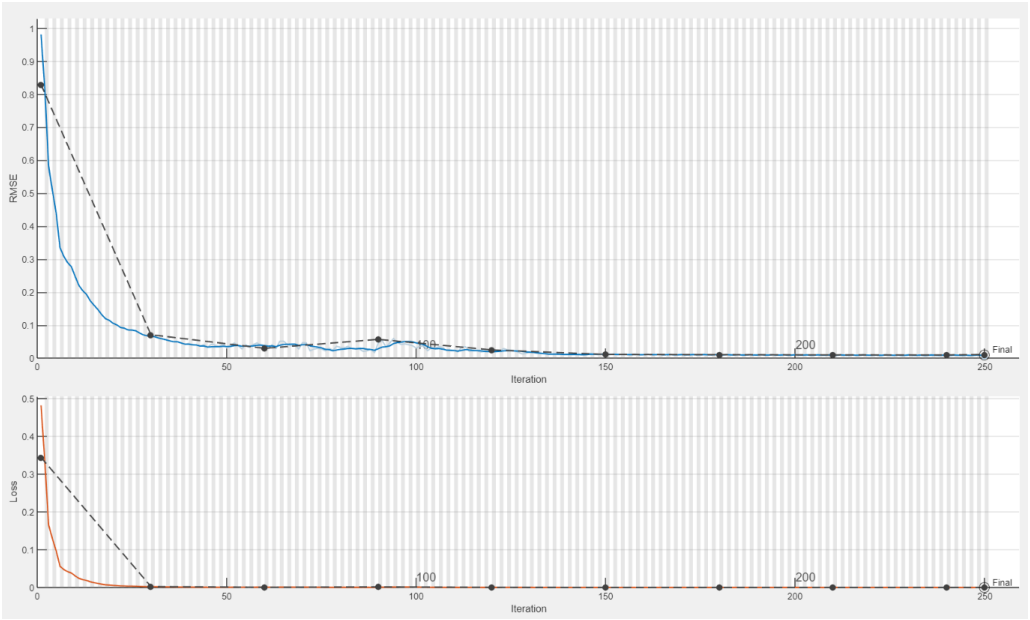

Figure 4.15 Training response

In contrast, the NARX model demonstrated a superior predictive capability compared to the standalone LSTM, as reflected by its lower nRMSE. This indicates that the NARX model's structure and learning mechanism were more effective in capturing the complex nonlinear relationships inherent in the solar irradiance data.

The most noteworthy performance, however, was exhibited by the hybrid NARX-LSTM model as resulted in Table 4.4. This combined approach leveraged the strengths of both NARX and LSTM architectures, resulting in the lowest nRMSE among the tested models. The success of the NARX-LSTM hybrid underscores the synergistic potential of combining different modeling techniques to enhance forecast accuracy. As delineated in Table 4.4 of the thesis, the hybrid model's superior performance in reducing predictive error underscores its potential as a more reliable tool for solar irradiance forecasting. The empirical results thus advocate for the adoption of hybrid modeling frameworks in complex time-series prediction tasks, where capturing dynamic, multifaceted patterns is crucial for achieving high levels of accuracy.

Table 4.4 Training performance result

	LSTM	**NARX**	**NARX-LSTM**
nRMSE %	10.245	2.0245	1.046

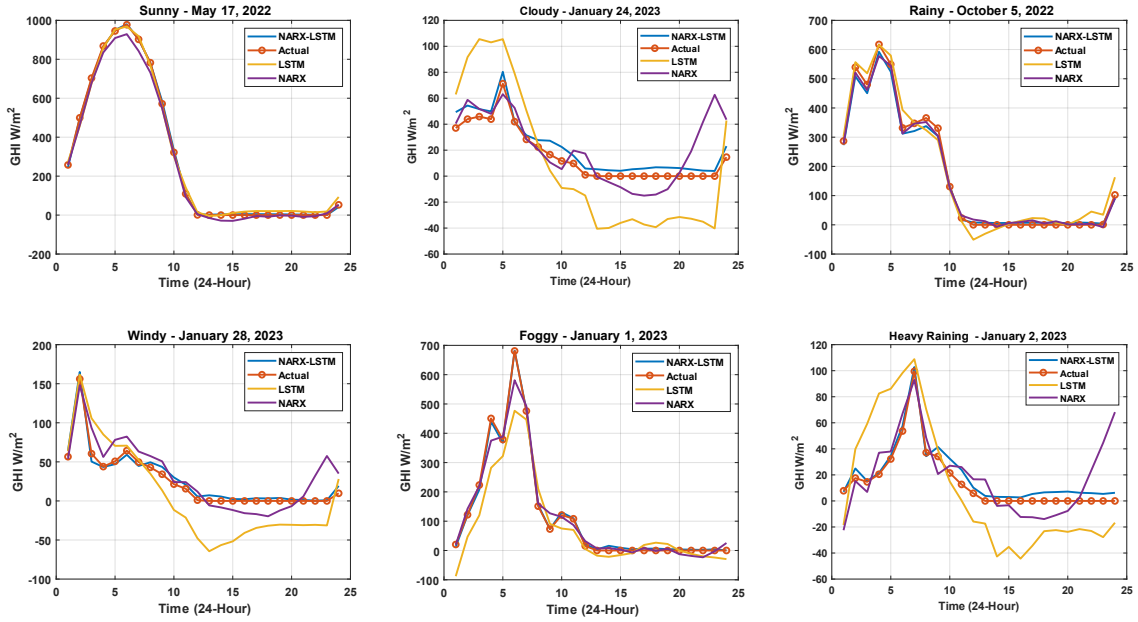

Figure 4.16 Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour forecast period across various meteorological scenarios with the selected variables.

The detailed result of the hybrid NARX-LSTM model's performance compared to the standalone LSTM and NARX models, as indicated by Figure 4.16, illustrates several key insights. The hybrid model consistently achieves lower nRMSE values across different weather conditions as shown in Table 4.5, suggesting a more refined ability to synthesize and predict based on complex data inputs.

The hybrid NARX-LSTM model's proficiency is prominently displayed under the spectrum of weather conditions that directly influence solar irradiance. Notably, during periods of sunshine, the model exhibits exemplary performance, closely mirroring the actual irradiance levels with minimal deviation, as illustrated in Figure 4.16. This accuracy is reflective of its adept learning and generalization capabilities, honed to capture the subtleties of solar patterns on clear days.

The true test of the model's mettle, however, is presented under overcast conditions, where the unpredictable dynamics of cloud movements pose a significant

challenge to predictive accuracy. Even in such scenarios, as shown in Figure 4.16 the hybrid model manages to maintain a lead over the NARX and LSTM models, albeit with a diminished advantage. This slight convergence in model performance under cloudy skies is indicative of the universally complex task of modeling solar irradiance amidst variable cloud cover.

The competitive edge of the proposed hybrid model is further asserted in scenarios of rainfall, where the interplay of clouds and precipitation further complicates the prediction landscape. While the hybrid model exhibits a commendable performance in such conditions, its comparability to the NARX model during rainy weather implies potential areas for enhancement, especially in the nuanced interpretation of complex cloud formations and the hydrological impact on irradiance.

The robustness of the model is rigorously tested and proven in adverse weather conditions such as fog, wind, and heavy rain. Under these challenging circumstances, the model not only maintains its forecasting accuracy but also shows a significant reduction in nRMSE as can be seen in Table 4.5, emphasizing its ability to adapt to sudden meteorological fluctuations. This adaptability is crucial for applications in real-time energy management and ensuring the stability of the power grid, where rapid response to changing weather conditions is imperative. The hybrid model's resilience in the face of environmental volatility, therefore, marks a significant step forward in the realm of solar irradiance forecasting, paving the way for more reliable and dynamic predictive systems.

Table 4.5 Tasting performance result

Condition:	**nRMSE%**		
	NARX-LSTM	**LSTM**	**NARX**
Sunny	1.195	5.986	6.534
Cloudy	6.813	18.457	12.815
Rainy	3.871	7.436	3.205
Heavy Raining	5.567	20.093	11.494

Foggy	2.458	12.870	4.205
Windy	1.971	9.208	6.208

The LSTM model, recognized for its utility in time-series analysis due to its capacity to capture temporal dependencies, demonstrates a commendable level of performance. However, its efficacy is eclipsed by the hybrid model, which suggests that the LSTM model, in isolation, may not fully grasp the intricate interplay between various weather parameters that affect solar irradiance. This limitation is addressed by the hybrid NARX-LSTM model, which integrates the strengths of both LSTM's temporal processing and NARX's nonlinear autoregressive capabilities, thereby enriching the model's comprehension of complex weather dynamics.

On the other hand, the standalone NARX model exhibits a degree of competitiveness, particularly in scenarios characterized by rainy conditions where short-term weather fluctuations are predominant. Nevertheless, it tends to lag in overall performance when compared to the hybrid model. The NARX model's relative shortfall could stem from its intrinsic design, which might not be as adept at capturing and leveraging long-term dependencies within the dataset as effectively as the hybrid approach.

The comparative analysis thus illuminates the distinctive advantages of combining NARX and LSTM models. By harnessing the specialized attributes of each model, the hybrid configuration excels in decoding both the short-term and long-term patterns in weather data, facilitating a more accurate and reliable forecast of solar irradiance, which is critical for the optimization of solar energy systems.

4.5 Chapter Summary

This chapter combined EDA with the results of the NARX-LSTM model, providing a comprehensive view of the dataset and its influence on forecasting performance. Input sensitivity analysis demonstrated the significance of key meteorological variables in shaping prediction accuracy. Hyperparameter tuning and rigorous evaluation highlighted the hybrid NARX-LSTM model's superior performance, particularly in handling diverse weather conditions. The hybrid model consistently achieved lower normalized Root Mean Square Error (nRMSE) values across various scenarios, showcasing its robustness and adaptability. These findings underscore the importance of integrating robust EDA with advanced modeling techniques to enhance the precision and reliability of solar irradiance predictions.