

# **Research Design and Analysis in Data Science MCST1043**

**Multivariate Time Series Analysis of Solar Irradiance  
For Photovoltaic Systems The Hybridization Of NARX  
And LSTM Models**

**Prepared for: Prof. Madya.Ts.Dr. Mohd Shahizan**

**Presented by:  
Osama Gamal Mahmoud**

## Contents

Introduction

Problem Statement

Objectives

Literature Reviews

Scope of Research

Methodology

- Data collection and Analysis steps
- The hybridization process

Results & Discussion

- Simulation Results

Conclusion

References

# Introduction

• • • •

- Solar Irradiance forecasting is a complex and dynamic problem, critical in the field of solar energy management. Despite advancements, accurately predicting daily solar irradiance remains a challenging task due to its susceptibility to fluctuating weather conditions [1].
- The accurate forecasting of GHI is vital for efficient operation and planning of photovoltaic (PV) systems, particularly within microgrids. These systems find diverse applications in modern energy solutions, from residential solar setups to large-scale renewable energy, making precision in forecasting a key factor for their success [2].
- The integration of ML models for solar irradiance forecasting has emerged as a benchmark in renewable energy research. The utilization of ML models represents a significant step in addressing the complexity of daily solar irradiance prediction [3].

# Problem Statement

The integration of renewable energy into the power grid is escalating globally, with countries like Malaysia targeting a 70% RE contribution by 2050. As the proportion of RE within the grid increases, so does the complexity of managing its variability. Solar energy, a key RE source, is particularly unpredictable due to its dependence on complex weather patterns. This unpredictability becomes particularly impactful in systems with significant solar energy contributions, where the inability to accurately forecast, solar irradiance can lead to inefficiencies, increased operational costs, and stability issues within the grid. Thus, robust and precise forecasting models are imperative to predict solar irradiance with high reliability, ensuring optimal grid performance and facilitating the seamless incorporation of solar energy.

# Research Objectives

1

To develop a hybrid machine learning model combining LSTM and NARX neural networks, focusing on improving the precision of daily solar irradiance predictions.

2

To analyze the influence of various weather parameters, determining their individual and collective impact on the forecasting model's performance.

3

To validate the model's performance across different weather conditions and days, ensuring robustness and adaptability in varying solar irradiance scenarios

# Literature Review

	<b>Tittle</b>	<b>Author</b>	<b>Method</b>	<b>Finding</b>
1	Predicting Solar Generation from Weather Forecasts Using Machine Learning	Navin Sharma, Pranshu Sharma, David Irwin, and Prashant Shenoy	Support Vector Machines (SVM)	To tackle the substantial variations in solar intensity, three models were trialed: Linear Least Squares Regression (LLSR), Support Vector Machine (SVM) with an RBF Kernel, and Past-Predicts-Future (PPF). The SVM model with RBF kernel yielded the best performance, exhibiting lower cross-validation RMS-Error (164 watts/m <sup>2</sup> ) and slightly higher prediction RMS-Error (163 watts/m <sup>2</sup> ), demonstrating its superior efficacy in predicting solar intensity amidst massive fluctuations.
2	Solar Irradiance Forecasting Using Deep Neural Networks	Ahmad Alzahranaia, Pourya Shamsia, Cihan Daglib, and Mehdi Ferdowsia	Deep recurrent neural networks (DRNNs)	By implementing a Deep Recurrent Neural Network (DRNN) with Long Short-Term Memory (LSTM) units, the predictive model achieved an average root mean square error (RMSE) of 0.0513 in training and 0.068 in testing. In comparison, the average RMSE of the FNN was 0.0746 in training and 0.086 in testing. The lower RMSE values of the DRNN indicate its superior performance in short-term solar irradiance prediction.
3	Solar Irradiance Forecasting in Remote Microgrids using Markov Switching Model	Ayush Shakya, Semhar Michael, and Christopher Saunders,	Markov Switching Model (MSM)	To manage solar irradiance variations, a Markov Switching Autoregressive Model was developed, using historical irradiance data. It achieved an annual RMSE of 106.8 W/m <sup>2</sup> and MAPE of 31.9% for 2011. Despite higher summer errors, this model provides a practical forecasting tool, especially valuable for remote locations without sophisticated forecasting infrastructure.

# Literature Review

	<b>Tittle</b>	<b>Author</b>	<b>Method</b>	<b>Finding</b>
4	Fuzzy Prediction Interval Models for Forecasting Renewable Resources and Loads in Microgrids	Doris Sáez, Fernanda Ávila, Daniel Olivares, Claudio Cañizares, and Luis Marín	fuzzy prediction interval	The fuzzy prediction interval models are implemented for microgrid energy management forecasting. With input data categorized into various subsets, the one-step-ahead (15 min) forecast achieved an RMSE of 1.0753, while the day-ahead forecast recorded an RMSE of 1.4165.
5	Deep Reinforcement Learning for Energy Microgrids Management Considering Flexible Energy Sources	Nikita Tomin, Alexey Zhukov, and Alexander Domyshev	Deep Reinforcement Learning	A Deep Q-Network (DQN) processed inputs with convolutions and layers. Despite solar irradiance variations, the model managed real-time distributed flexibility, efficiently adapting to unseen electricity demand and solar irradiance configurations, providing a robust solution for operating electricity microgrids in stochastic environments.

# Literature Summary

- Research in solar irradiance forecasting has seen a variety of approaches. Traditional methods, including statistical and rule-based models, often fall short in accurately predicting solar energy due to the intricate and dynamic nature of weather patterns affecting solar irradiance. Recognizing these limitations, recent studies have shifted towards more sophisticated techniques. This project specifically explores the integration of machine learning algorithms, notably a hybrid model combining Nonlinear Autoregressive with Exogenous Inputs (NARX) and Long Short-Term Memory (LSTM) networks, to enhance daily solar irradiance forecasting accuracy in microgrids.



# Scope of Research

01

The research will focus on constructing a hybrid model that combines the strengths of Nonlinear Autoregressive with Exogenous Inputs (NARX) and Long Short-Term Memory (LSTM) networks to forecast solar irradiance.

The model will integrate various weather parameters which are Air Temperature, Cloud Attenuation, Precipitation Rate, Dewpoint Temperature, Surface Pressure, Precipitable Water, Relative Humidity, Wind Speed, Wind Direction then emphasizing those with the most significant impact on solar irradiance, to enhance the forecasting accuracy for PV systems.

02

MATLAB software and Python will be employed to develop, simulate, and analyze the performance of the hybrid forecasting model.

03

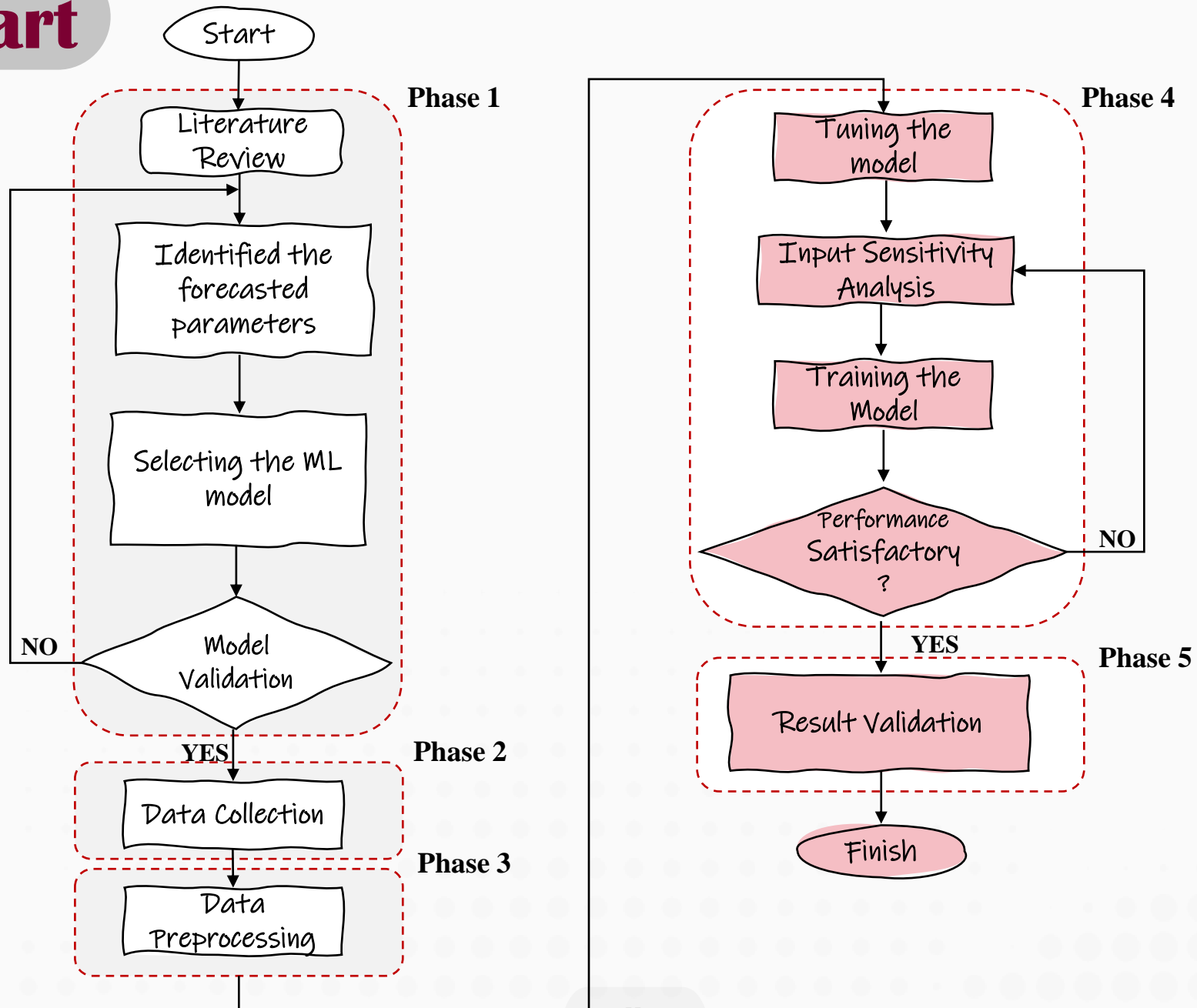
The forecasting - model's predictions will be validated against historical solar irradiance data that measured within Johor Bahru environments to assess performance and accuracy.

04

# Methodology



# Flow-Chart



# Data Collection

## Provider:

Det Norske Veritas, DNV  
Company



## Location

Johor Bahru, Johor, Malaysia



## Parameters Collected

- Solar irradiance
- air temperature
- cloud attenuation
- precipitation rate
- precipitable water
- relative humidity
- air pressure
- wind direction
- wind speed
- dewpoint temperature



## Data Source and Parameters

## Period

Jan 1, 2007, to November 29, 2023  
Consist of 148,225 index



## Frequency

Hourly readings (60 minutes)  
Each day have 24 readings

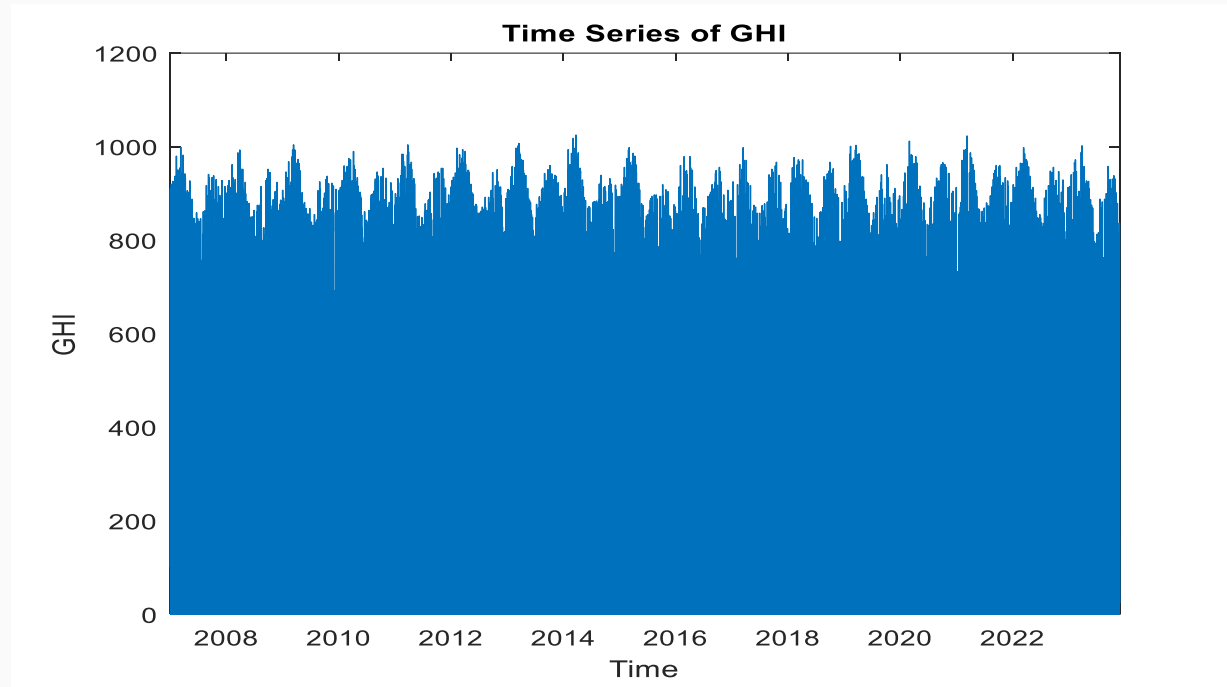


## Data Preprocessing Steps

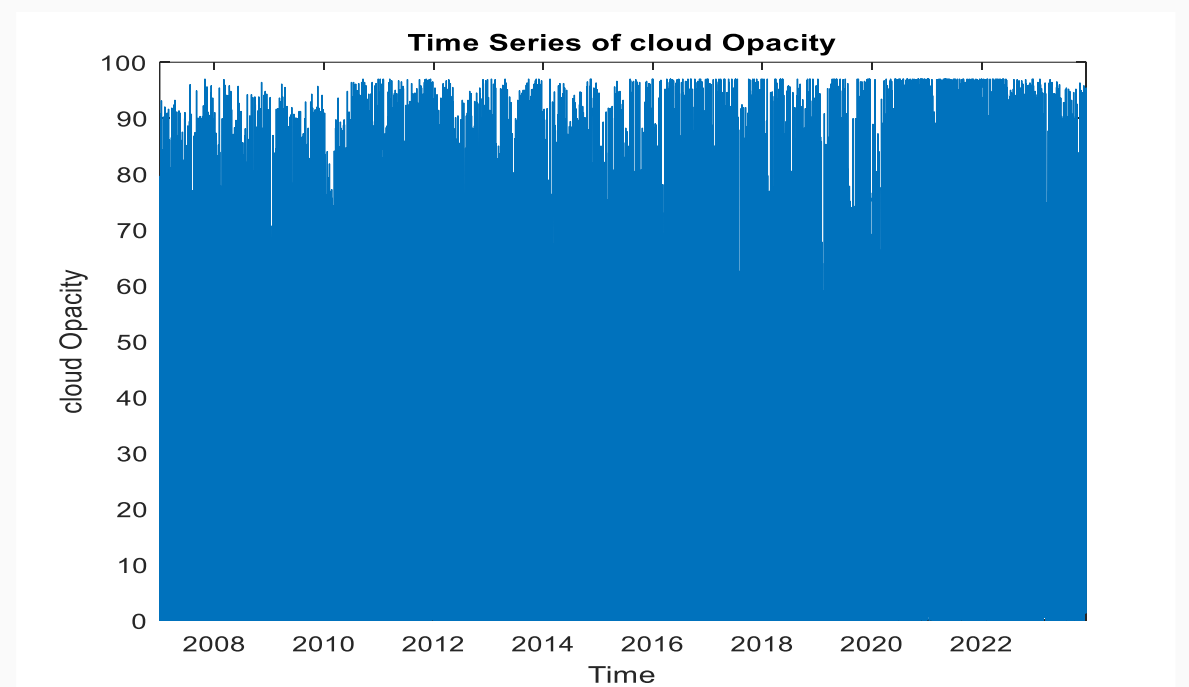
1. Cleaning: Removed outliers and corrected anomalies.
2. Normalization: Standardized feature scale to enhance model performance.
3. Time-Series Formatting: Converted data for compatibility with NARX and LSTM models.

# Time-series graph of Parameters

01

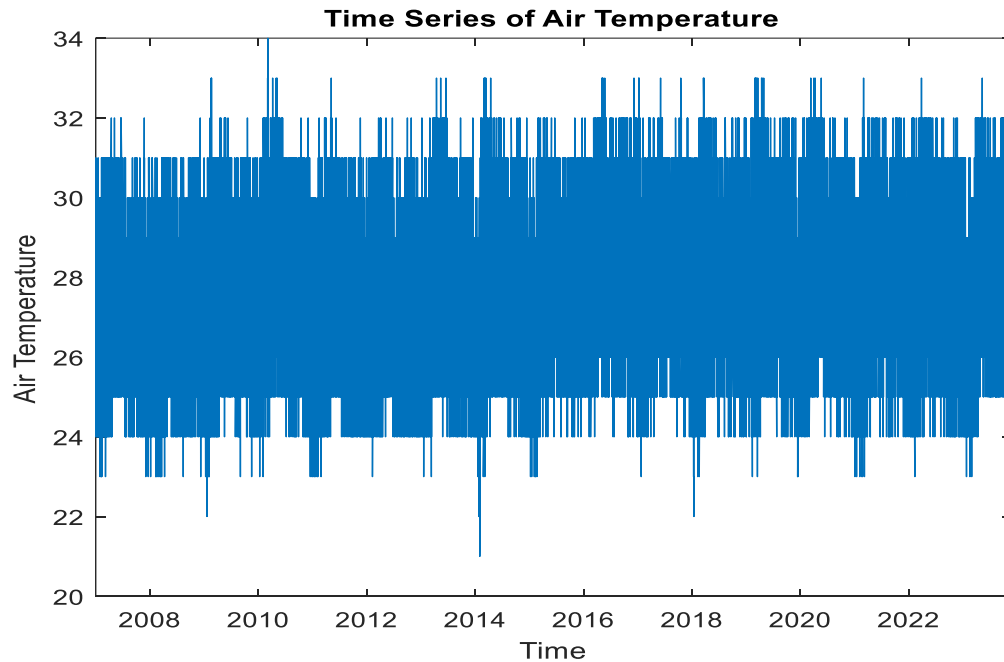


02

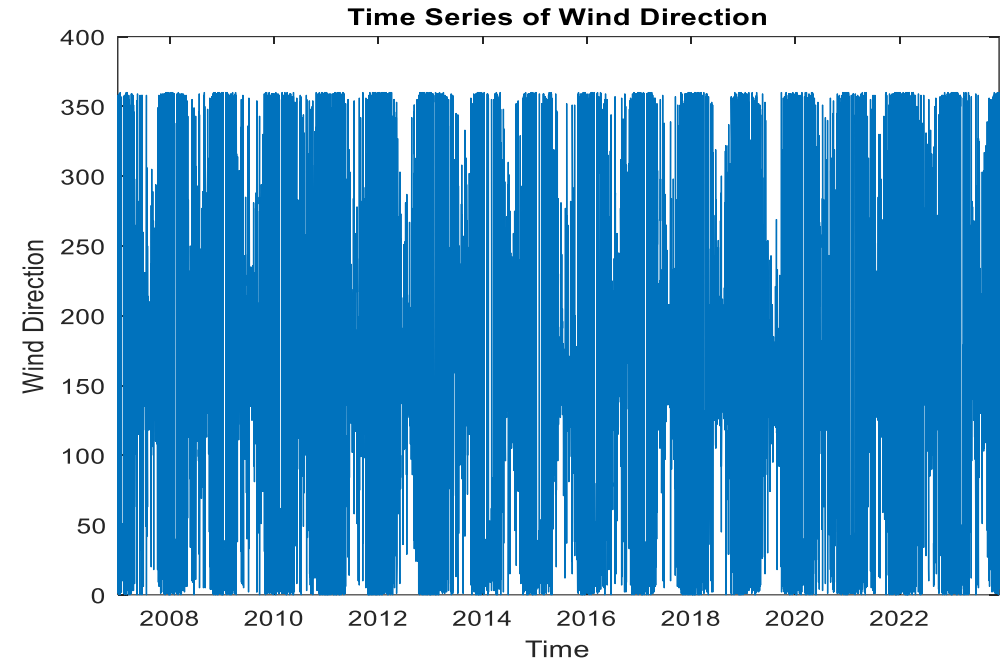


# Time-series graph of Parameters

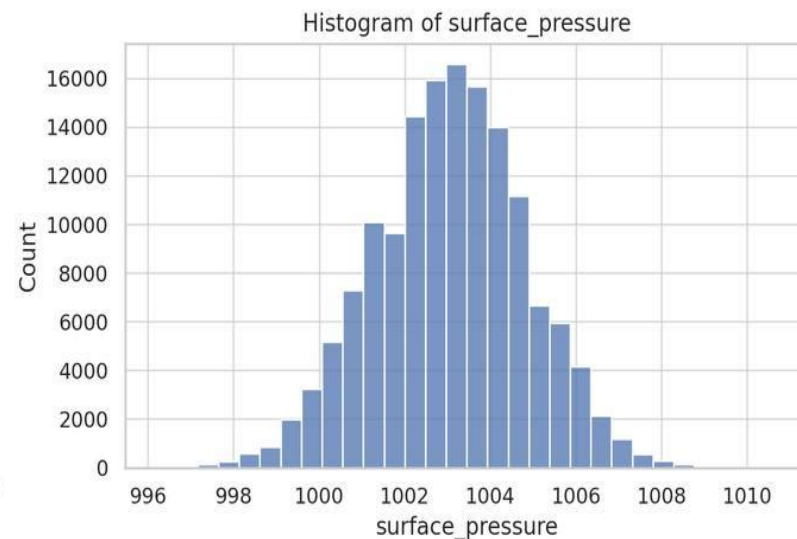
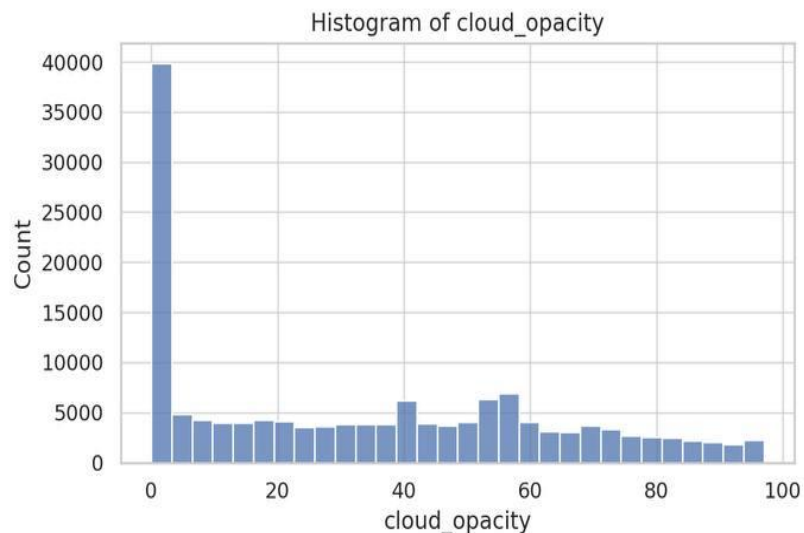
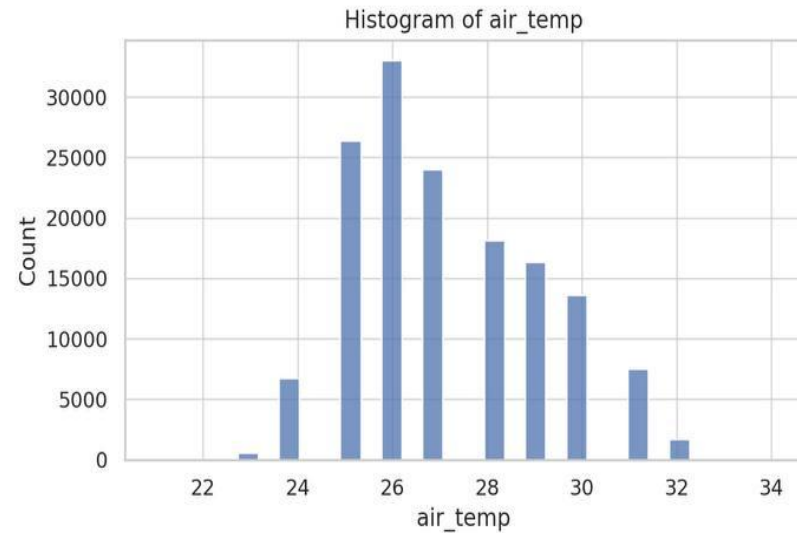
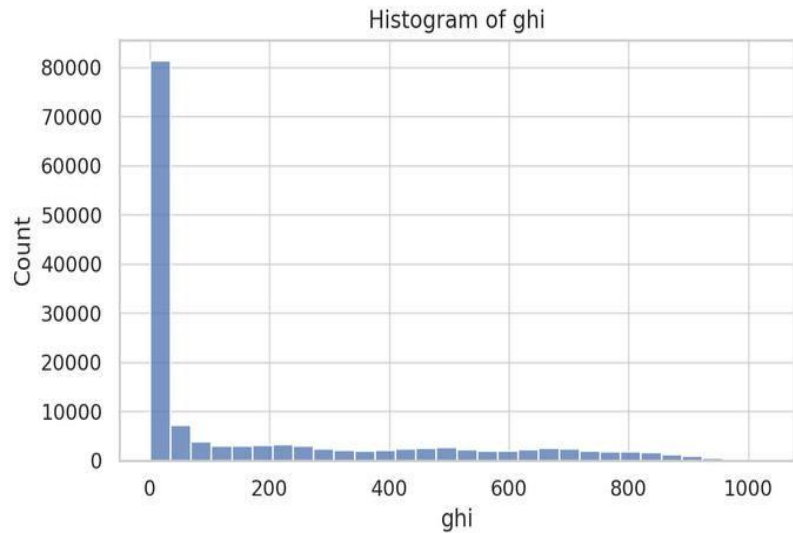
03



04

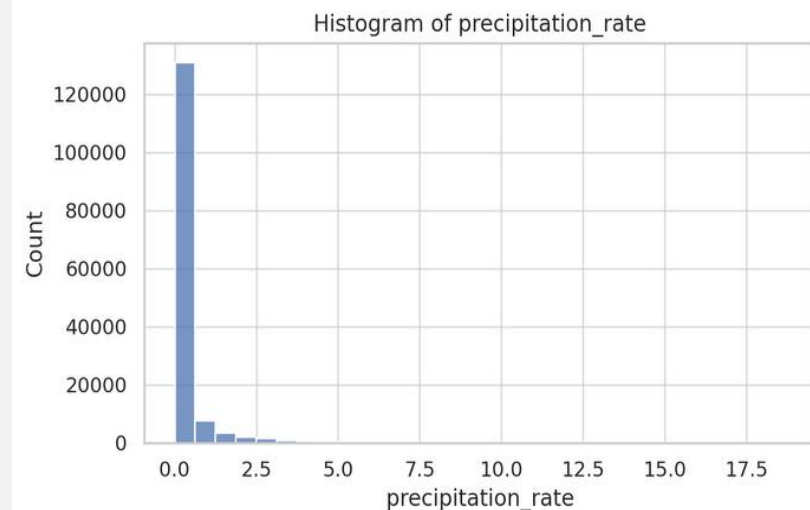
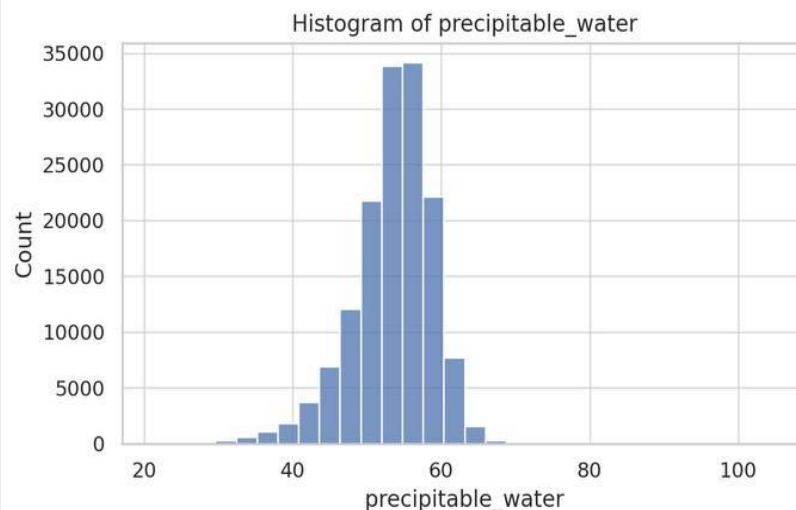
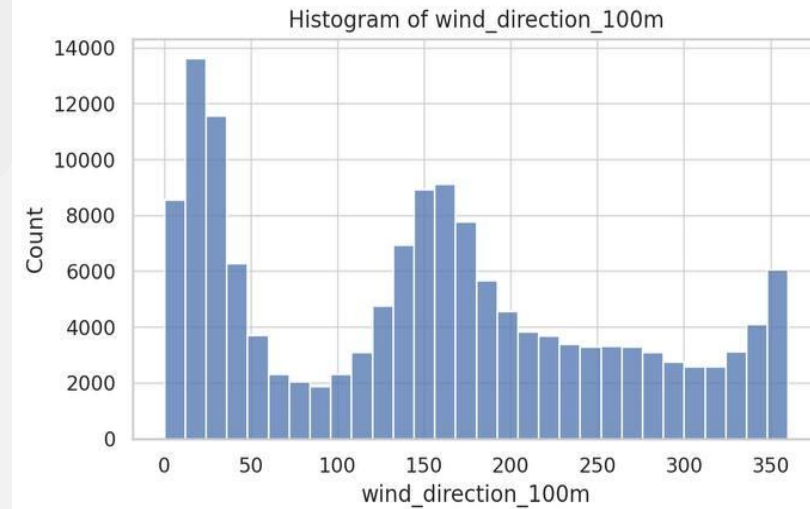
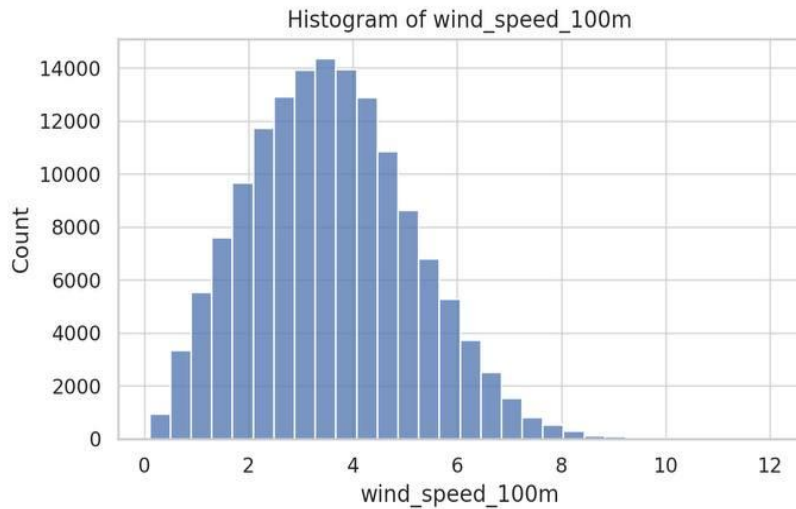


# Data Analysis [Distributions of variables]



- **GHI** shows a right-skewed distribution, indicating more frequent low irradiance values, which is typical due to night-time periods and cloudy days.
- **The air temperature** distribution appears multi-modal, possibly reflecting different times of day or seasons.
- **Cloud opacity** has a high frequency of low values with a long tail, suggesting many clear sky conditions with sporadic heavy cloudiness.
- **Surface pressure** is normally distributed, indicating consistent measurement with no extreme weather events affecting the readings.

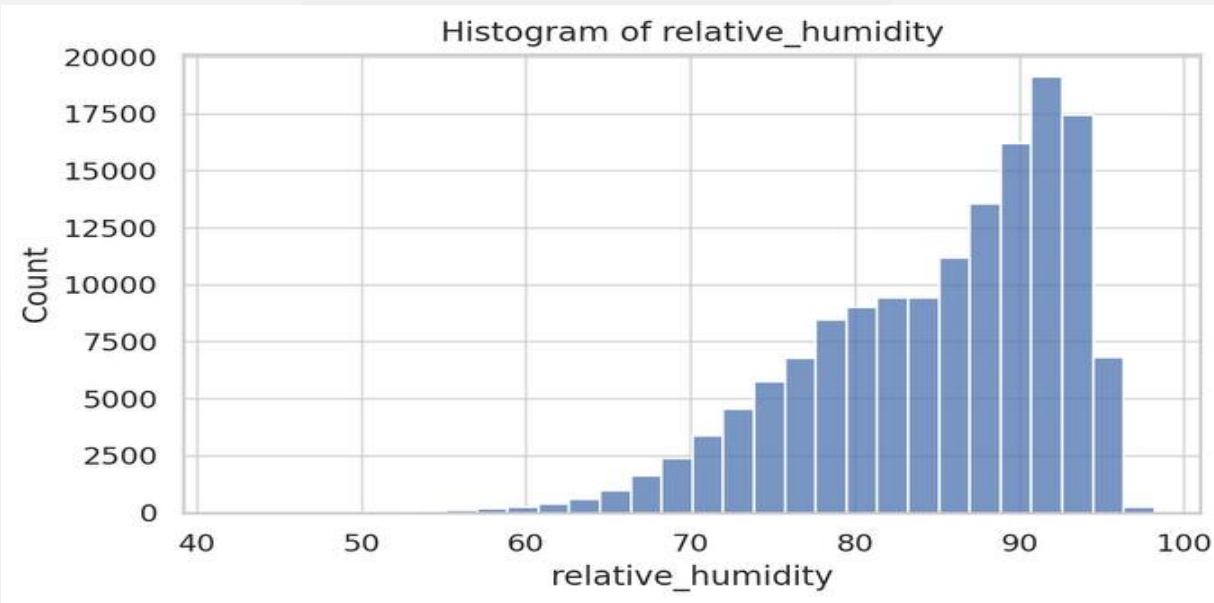
# Data Analysis [Distributions of variables]



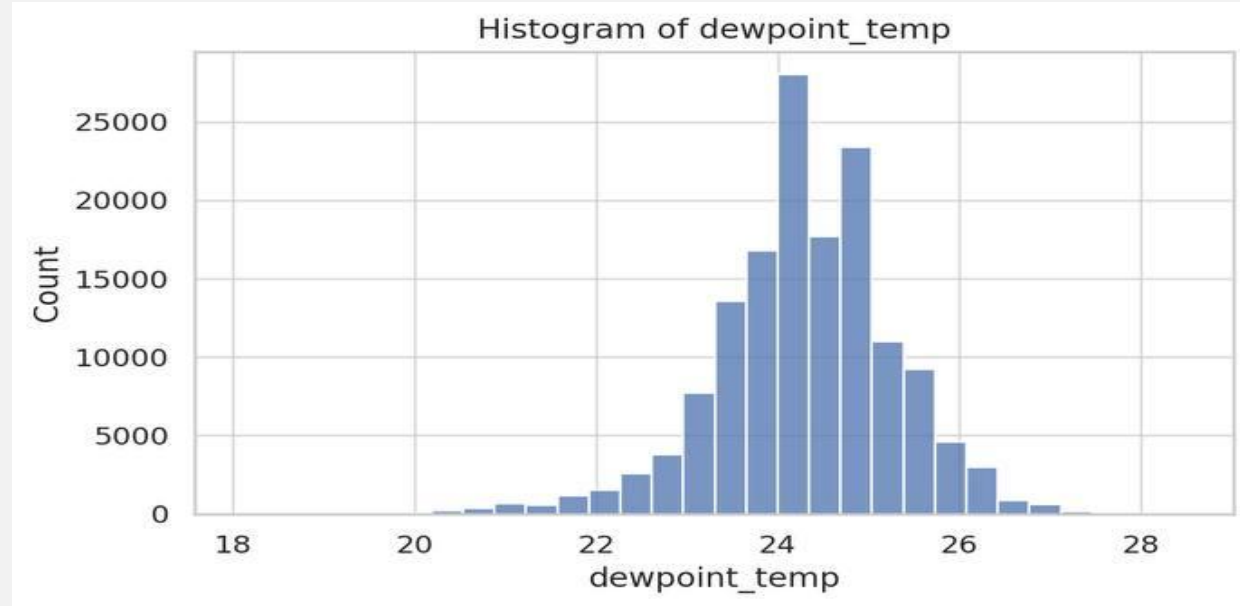
- **Wind speed** demonstrates a skewed distribution with a peak at lower speeds, indicating calm conditions are more common.
- **Wind direction** shows a multi-modal distribution, which may suggest prevailing wind patterns or the influence of local geography.
- **Precipitable water** is concentrated in lower values with a quick drop-off, which could imply a predominance of dry days.
- **The precipitation rate** is highly skewed, with most time periods experiencing no rain.



# Data Analysis [Distributions of variables]



- **Relative humidity** skews towards higher values, common in tropical climates that experience high levels of moisture in the air.



- **Dewpoint temperature** has a somewhat normal distribution with a slight skew, pointing to a stable range of moisture conditions with occasional extremes.

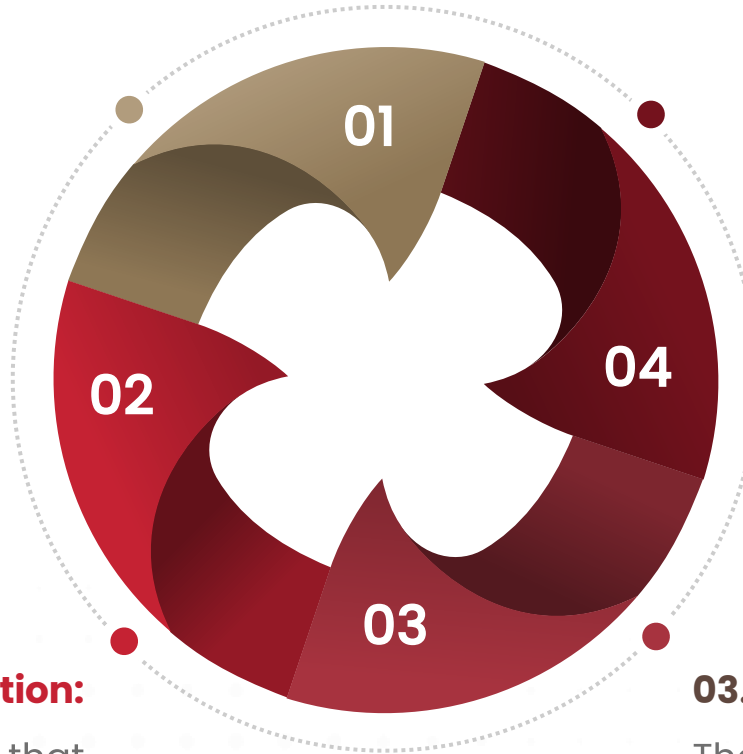
# Insights from Variable Distributions

## 01. Model Input Sensitivity

A skewed distribution in GHI shows the model is frequently encounter lower irradiance values, which it must accurately predict, especially for energy management during low-light conditions.

## 04. Predictive Feature Identification:

Variables with distinctive distributions that correlate well with GHI is identified as strong predictive features. For example, if cloud opacity has a strong inverse correlation with GHI, it is a key feature in the proposed model.



## 02. Data Preprocessing Requirements

Multimodal distributions or significant skewness in variables like air temperature and cloud opacity indicate that certain preprocessing step is need to be taken which is normalization to improve the model's ability to learn.

## 03. Feature Engineering Opportunities:

The distributions suggest opportunities for creating new features that will improve model performance. The time of day and Month can be inferred from the distribution of GHI values, which are an important feature for forecasting.

# Data Normalization

## For 24-h

air_temp	cloud_opa	dewpoint	ghi	precipitat	precipitat	relative_h	surface_p	wind_dire	wind_spe
25	59.1	23.3	101	53.6	0	91.7	1004.1	316	3.8
26	55	23.9	213	53.9	0	90.3	1004.8	313	3.4
27	56	24.2	296	54.3	0	86.2	1005.2	317	3.6
28	43	24.1	468	54.9	0	81	1005.2	324	4
29	7.6	24.1	832	55.8	0	76	1004.7	327	4.1
30	0	24.3	910	56.7	0.1	73.7	1003.9	328	4.2
29	9.8	23.6	766	57.3	1.9	73.8	1003.1	327	4.2
27	17.4	22.8	595	56.8	0.5	75.8	1002.1	335	4
27	47.8	23.3	285	56.3	0.3	80.5	1001.4	346	4
27	75.9	23.7	77	56.7	0.2	83.4	1001.6	343	3.9
26	49.7	23.6	41	57.4	0.2	84.9	1002.4	339	3.8
26	67	23.5	0	57.7	0.1	86.1	1003.2	340	3.7
26	69.3	23.5	0	58	0.1	87.2	1004.1	332	3.9
26	55.7	23.5	0	58.7	0.1	88.7	1004.9	317	4.6
25	50.9	23.5	0	59.3	0.1	90.9	1005.4	308	4.9
25	49.6	23.4	0	59.9	0	92.2	1005.6	301	4.7
25	36.3	23.4	0	60.3	0.1	92.8	1005.5	301	4.6
25	32.5	23.3	0	60.1	0.1	93	1004.9	299	4.6
25	47.7	23.3	0	59.8	0	93	1004.2	288	5
25	51.5	23.2	0	60	0	93	1003.5	282	5.5
24	46.1	23.3	0	60.2	0.1	93.6	1003.1	286	5.6
25	48	23.4	0	59.2	0.4	93.6	1003.3	300	5
24	32.6	23.2	0	58.3	0.2	93	1003.5	309	4.6



air_temp	cloud_opa	dewpoint	ghi	precipitat	precipitat	relative_h	surface_p	wind_dire	wind_spe
0.307692	0.609278	0.5	0.098537	0.386174	0	0.884342	0.544828	0.877778	0.310924
0.384615	0.56701	0.557692	0.207805	0.38975	0	0.859431	0.593103	0.869444	0.277311
0.461538	0.57732	0.586538	0.28878	0.394517	0	0.786477	0.62069	0.880556	0.294118
0.538462	0.443299	0.576923	0.456585	0.401669	0	0.69395	0.62069	0.9	0.327731
0.615385	0.078351	0.576923	0.811707	0.412396	0	0.604982	0.586207	0.908333	0.336134
0.692308	0	0.596154	0.887805	0.423123	0.005348	0.564057	0.531034	0.911111	0.344538
0.615385	0.101031	0.528846	0.747317	0.430274	0.101604	0.565836	0.475862	0.908333	0.344538
0.461538	0.179381	0.451923	0.580488	0.424315	0.026738	0.601423	0.406897	0.930556	0.327731
0.461538	0.492784	0.5	0.278049	0.418355	0.016043	0.685053	0.358621	0.961111	0.327731
0.461538	0.782474	0.538462	0.075122	0.423123	0.010695	0.736655	0.372414	0.952778	0.319328
0.384615	0.512371	0.528846	0.04	0.431466	0.010695	0.763345	0.427586	0.941667	0.310924
0.384615	0.690722	0.519231	0	0.435042	0.005348	0.784698	0.482759	0.944444	0.302521
0.384615	0.714433	0.519231	0	0.438617	0.005348	0.80427	0.544828	0.922222	0.319328
0.384615	0.574227	0.519231	0	0.446961	0.005348	0.830961	0.6	0.880556	0.378151
0.307692	0.524742	0.519231	0	0.454112	0.005348	0.870107	0.634483	0.855556	0.403361
0.307692	0.51134	0.509615	0	0.461263	0	0.893238	0.648276	0.836111	0.386555
0.307692	0.374227	0.509615	0	0.466031	0.005348	0.903915	0.641379	0.836111	0.378151
0.307692	0.335052	0.5	0	0.463647	0.005348	0.907473	0.6	0.830556	0.378151
0.307692	0.491753	0.5	0	0.460072	0	0.907473	0.551724	0.8	0.411765
0.307692	0.530928	0.490385	0	0.462455	0	0.907473	0.503448	0.783333	0.453782
0.230769	0.475258	0.5	0	0.464839	0.005348	0.918149	0.475862	0.794444	0.462185
0.307692	0.494845	0.509615	0	0.45292	0.02139	0.918149	0.489655	0.833333	0.411765
0.230769	0.336082	0.490385	0	0.442193	0.010695	0.907473	0.503448	0.858333	0.378151

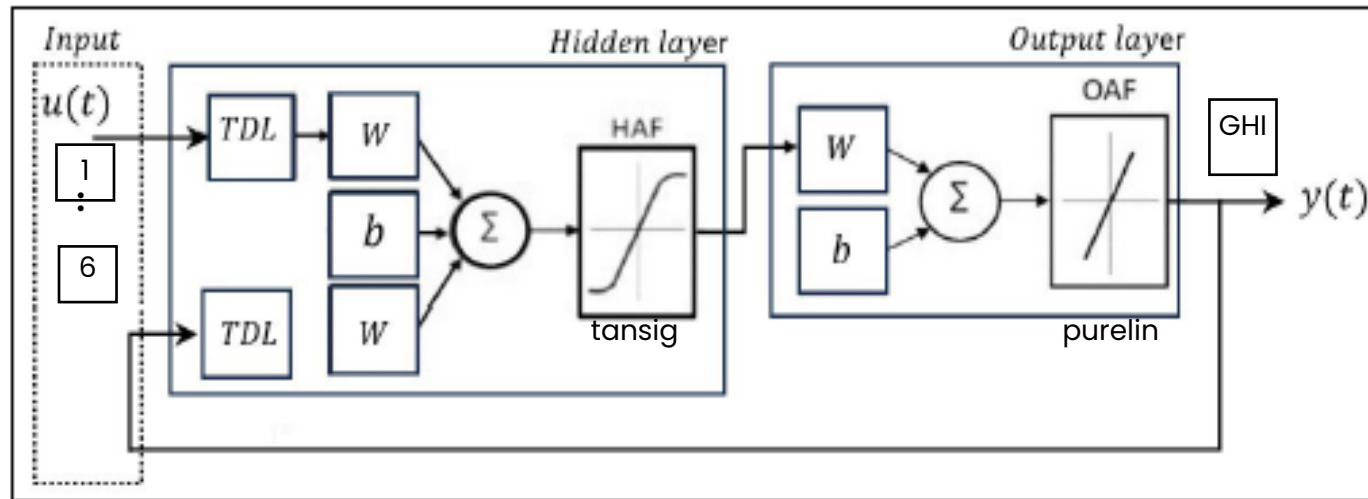
# Models Architecture

## Nonlinear Autoregressive with External Input, NARX Model:

- A type of neural network that uses past values of a target variable and past values of exogenous inputs to predict future values.
- Strengths: Excellent for capturing nonlinear relationships and temporal dynamics in time-series data.

### Network Flow:

Inputs are fed into the TDLs, then transformed in the hidden layer, and passed through the output layer to generate the GHI forecast.



# Nonlinear Autoregressive with External Input, NARX Model:

```
% The feedback delays and input delays for the NARX network
inputDelays = 1:24; % 24-hour lags as inputs
feedbackDelays = 1:24; % 24-hour feedback for the target series

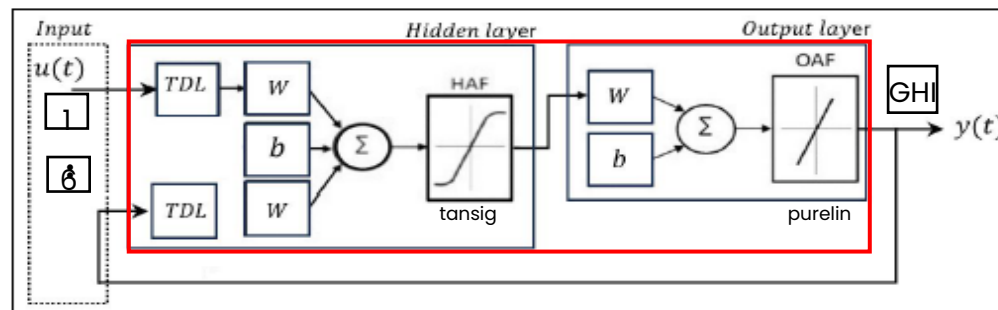
% the number of hidden neurons
hiddenLayerSize = 50;

% the NARX network
narx_net = narxnet(inputDelays, feedbackDelays, hiddenLayerSize);

% Prepare and set up the data division
narx_net.divideFcn = 'divideblock'; % Divide the data into blocks for training, validation, and testing
narx_net.divideParam.trainRatio = 70/100;
narx_net.divideParam.valRatio = 15/100;
narx_net.divideParam.testRatio = 15/100;

% Training
narx_net.trainFcn = 'trainlm'; % Levenberg-Marquardt optimization
narx_net.performFcn = 'mse'; % Mean Squared Error performance function
narx_net.trainParam.epochs = 1000;
narx_net.trainParam.min_grad = 1e-7;
narx_net.trainParam.max_fail = 6;

% Train the NARX network
[trainedNarxNet, tr] = train(narx_net, trainInputsNARX, trainTargetsNARX);
```

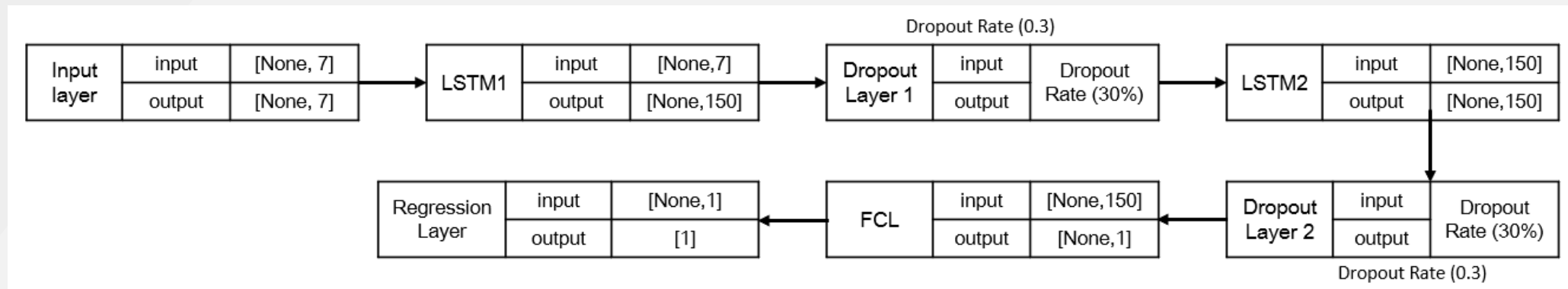


# Long Short-Term Memory, LSTM Model

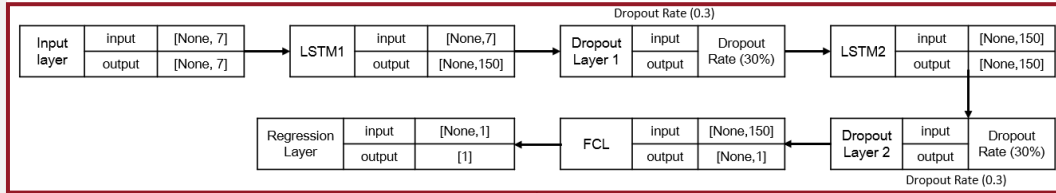
- An advanced type of recurrent neural network designed to remember long-term dependencies and overcome the vanishing gradient problem.
- Strengths: Its memory cells can store, modify, and access information over long sequences, making it ideal for predicting time-series data with long-range temporal dependencies.

## Model Flow:

Data flows from the input layer through LSTM layers with regularization by dropout, then through the fully connected layer, culminating in the regression output for GHI forecasting.



# Long Short-Term Memory, LSTM Model



**%% LSTM network**

```
numFeatures = size(trainInputs, 1);
numHiddenUnits = 150; % Number of hidden units
```

**% LSTM Model Configuration**

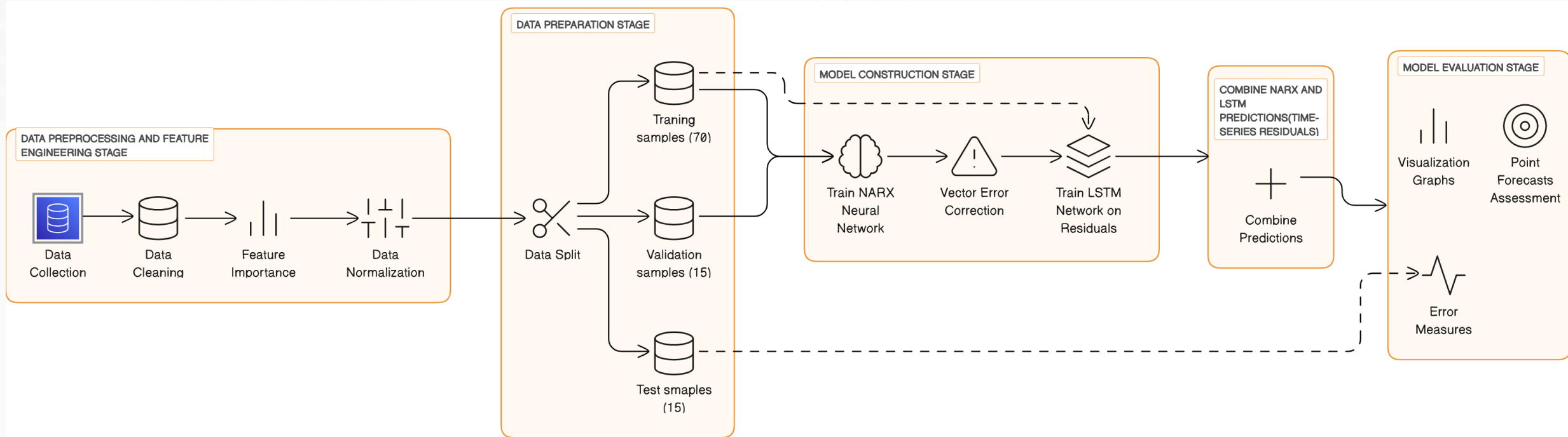
```
layers = [ ...
    sequenceInputLayer(numFeatures)
    lstmLayer(numHiddenUnits, 'OutputMode', 'sequence') % LSTM layer
    dropoutLayer(0.3) % Dropout layer with 30% dropout rate
    lstmLayer(numHiddenUnits, 'OutputMode', 'sequence') er
    dropoutLayer(0.3)
    fullyConnectedLayer(1)
    regressionLayer];
```

**% Training Options**

```
options = trainingOptions('adam', ...
    'MaxEpochs', 50, ... % number of epochs
    'MiniBatchSize', 64, ...
    'GradientThreshold', 1, ...
    'InitialLearnRate', 0.005, ...
    'LearnRateSchedule', 'piecewise', ...
    'LearnRateDropPeriod', 50, ...
    'LearnRateDropFactor', 0.2, ...
    'Shuffle', 'every-epoch', ...
    'Verbose', 0, ...
    'Plots', 'training-progress', ...
    'ValidationData', {valInputs, valResiduals}, ... % Include validation data
    'ValidationFrequency', 30, ... % How often to check validation
    'ValidationPatience', 5); % Early stopping patience
```

# Hybrid Model Integration

## NARX-LSTM architecture for GHI forecasting



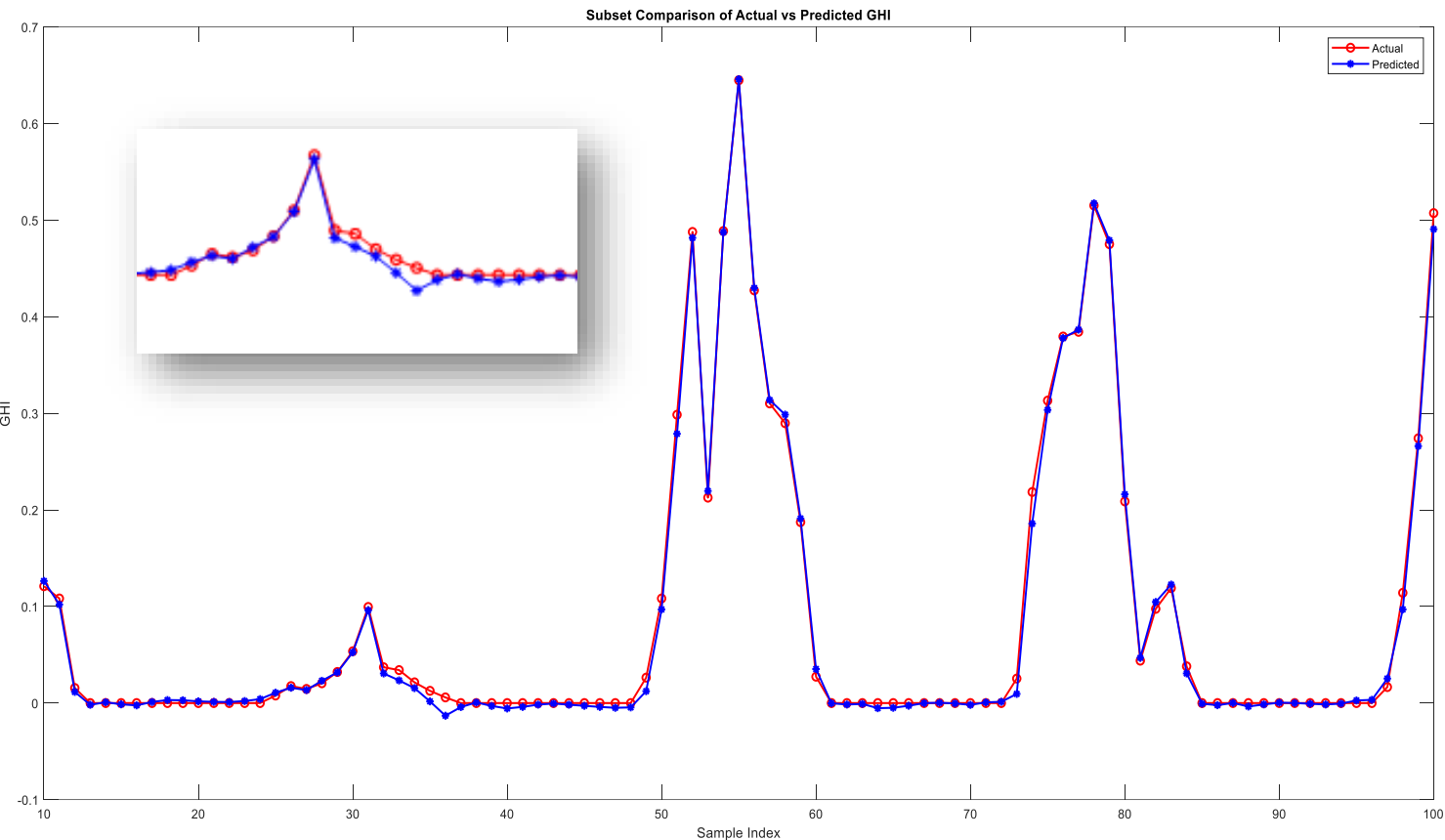


The impact of each variable on the forecasting model is methodically analyzed. Influential predictors are identified, and their contributions to prediction accuracy are detailed. Non-linear relationships and time-dependent factors are captured, with interdependencies between variables meticulously considered. This analysis is fundamental to refining the model's precision and unveiling critical insights into solar irradiance forecasting.

# **Impact of Variables On Forecasting Model**

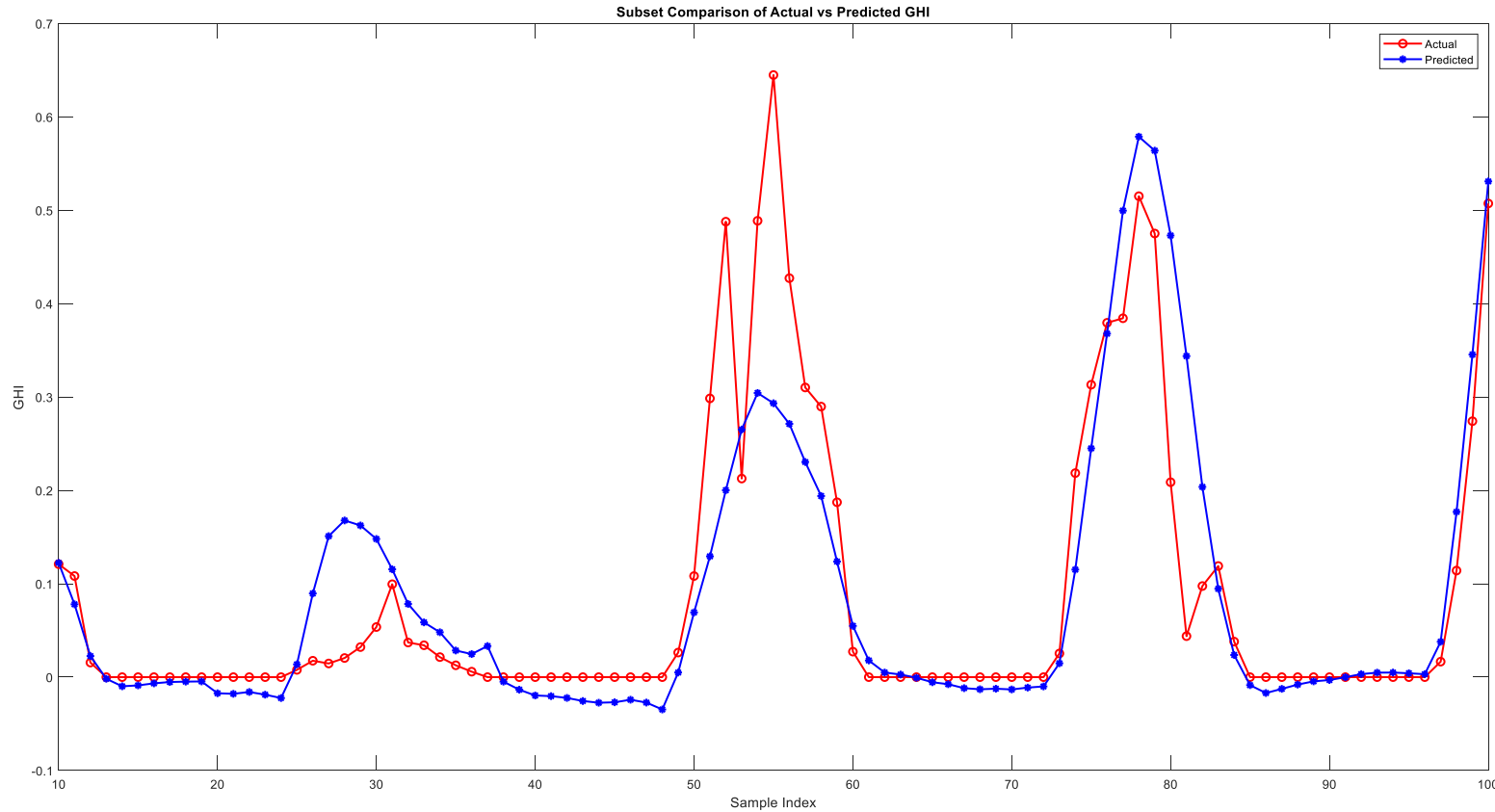
....

# Model Performance with all Variables



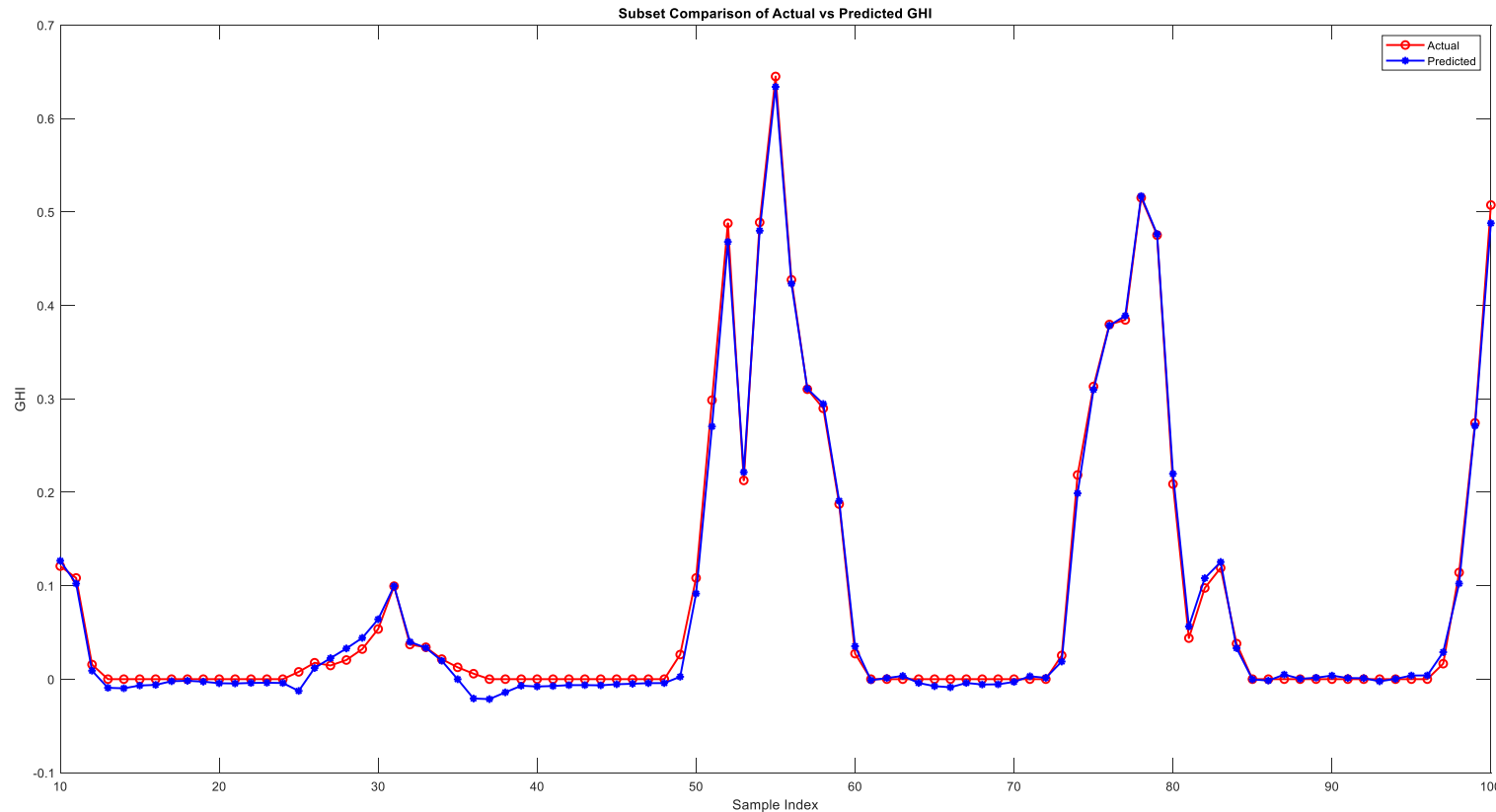
	MSE	RMSE	MAE
NARX-MODEL	0.00013012	0.011407	0.0067768
LSTM-MODEL	0.00215370	0.046408	0.035780
HYBRID-MODEL	0.00012658	0.011251	0.0064454

# Impact of Removing Cloud Attenuation



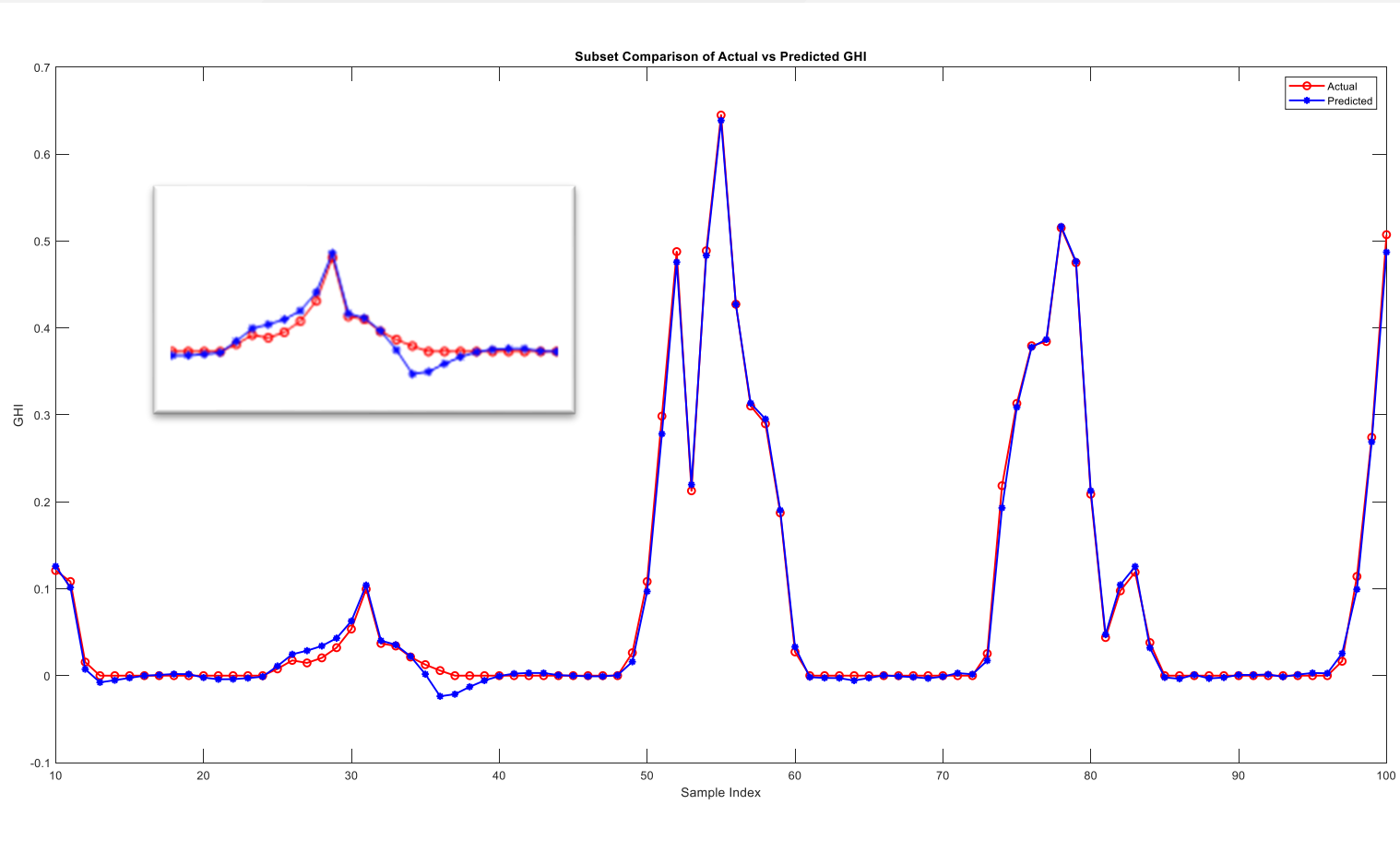
	MSE	RMSE	MAE
<b>NARX-MODEL</b>	0.0074136	<b>0.086102</b>	0.047974
<b>LSTM-MODEL</b>	0.0085655	<b>0.09255</b>	0.056052
<b>HYBRID-MODEL</b>	0.0075143	<b>0.086685</b>	0.045959

# Impact of Removing Wind Direction



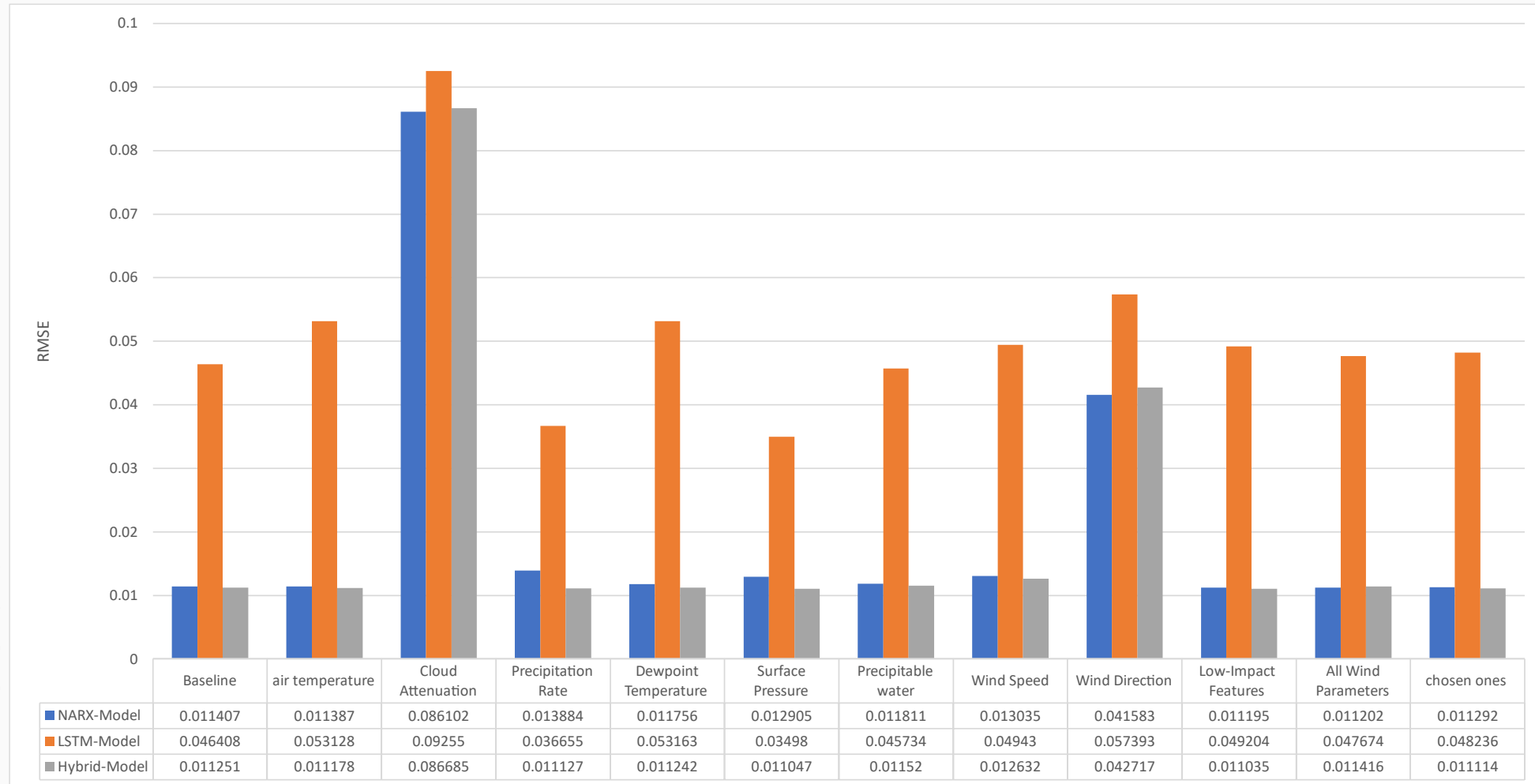
	MSE	RMSE	MAE
<b>NARX-MODEL</b>	0.0017292	<b>0.041583</b>	0.0069137
<b>LSTM-MODEL</b>	0.003294	<b>0.057393</b>	0.044592
<b>HYBRID-MODEL</b>	0.0018247	<b>0.042717</b>	0.0066581

# Impact of Adding Precipitable Rate



	MSE	RMSE	MAE
<b>NARX-MODEL</b>	0.00016654	<b>0.012905</b>	0.0085109
<b>LSTM-MODEL</b>	0.00122360	<b>0.034980</b>	0.0263340
<b>HYBRID-MODEL</b>	0.00012204	<b>0.011047</b>	0.0062957

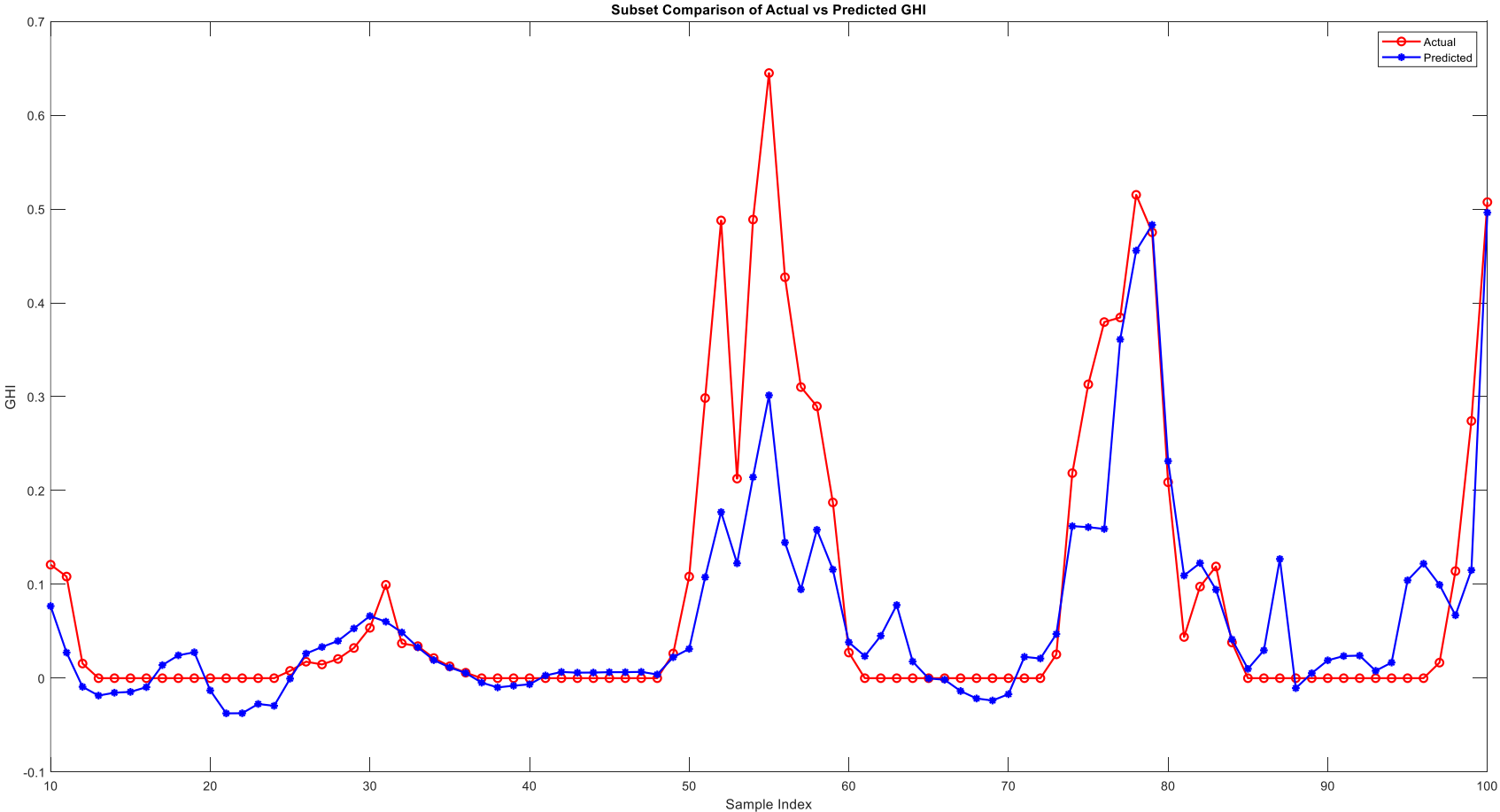
# Overall Results



# Overall Results



# unconsidered the Hourly and Monthly time features:

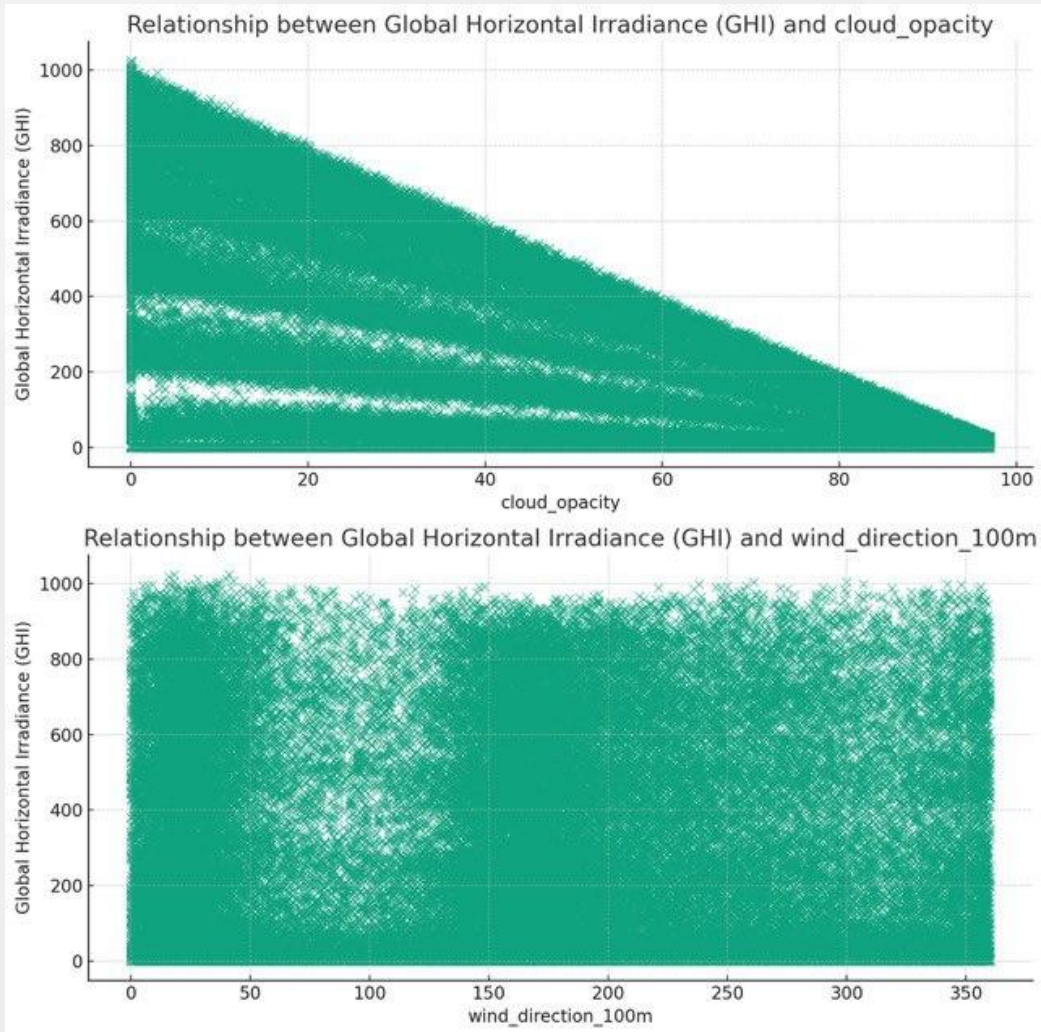


	MSE	RMSE
NARX-MODEL	0.013726	0.11716
LSTM-MODEL	0.0046952	0.068522
HYBRID-MODEL	0.0044567	0.066758



# The Final Input Variables

## [Cloud Attenuation, Surface pressure, Wind Direction and Air temperature]



### Cloud Opacity:

- The inverse correlation between cloud opacity and GHI confirms its critical role in solar irradiance prediction.
- Higher cloud opacity significantly diminishes GHI, highlighting the need for accurate cloud cover assessment in forecasting models.

### Wind Direction:

- Wind direction displays a diffuse relationship with GHI, which is attributed to its indirect influence through cloud movement and local weather patterns.
- This variable is more valuable when combined with cloud opacity weather parameters where it has a direct impact on cloud behavior.

# The Final Input Variables

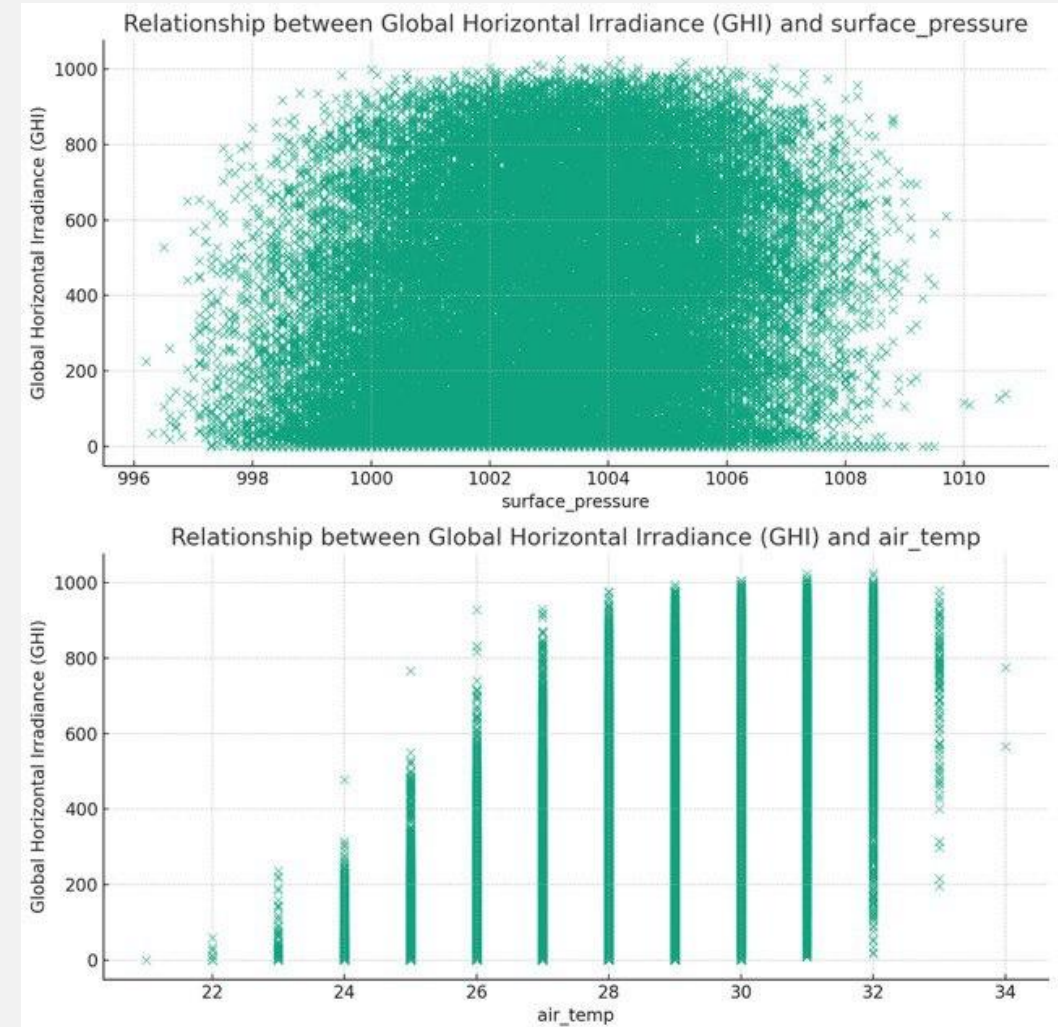
## [Cloud Attenuation, Surface pressure, Wind Direction and Air temperature]

### Surface Pressure:

- Surface pressure shows a more dispersed pattern with GHI, showing that while it has an impact, it is not as direct as cloud opacity
- Surface pressure serves as an indicator of atmospheric stability, with high readings often leading to clear skies and consequently higher GHI values.
- Low surface pressure, often associated with the monsoon seasons, can lead to increased cloudiness. This augmented cloud cover scatters and absorbs solar radiation, reducing the GHI and impacting the potential solar energy capture for PV systems..

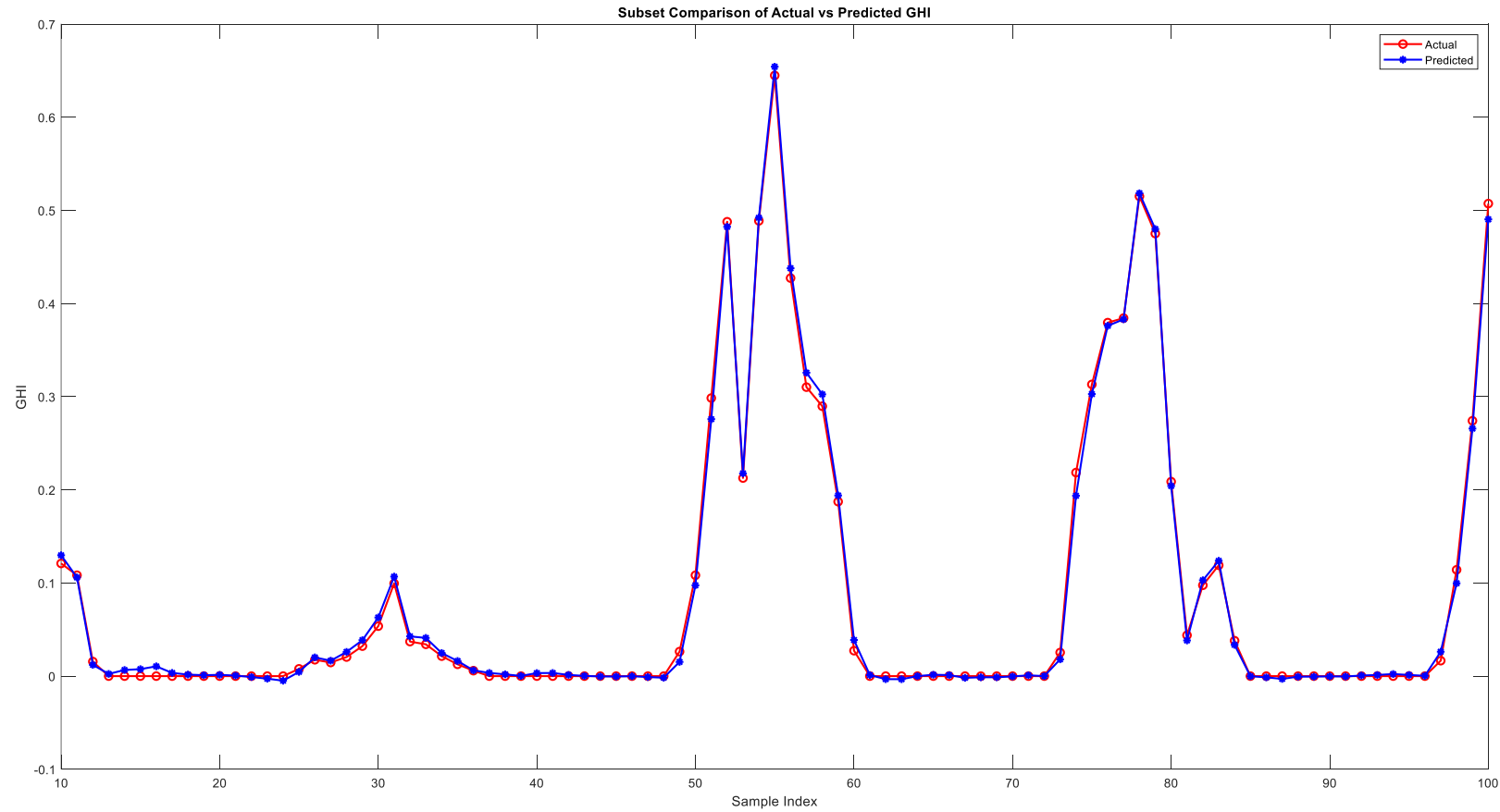
### Air Temperature:

- A positive trend between air temperature and GHI is evident, likely due to the association of higher temperatures with clear sky conditions.
- Temperature, as a proxy for sunshine duration, which is a strong predictor for GHI in regions with less cloud cover.

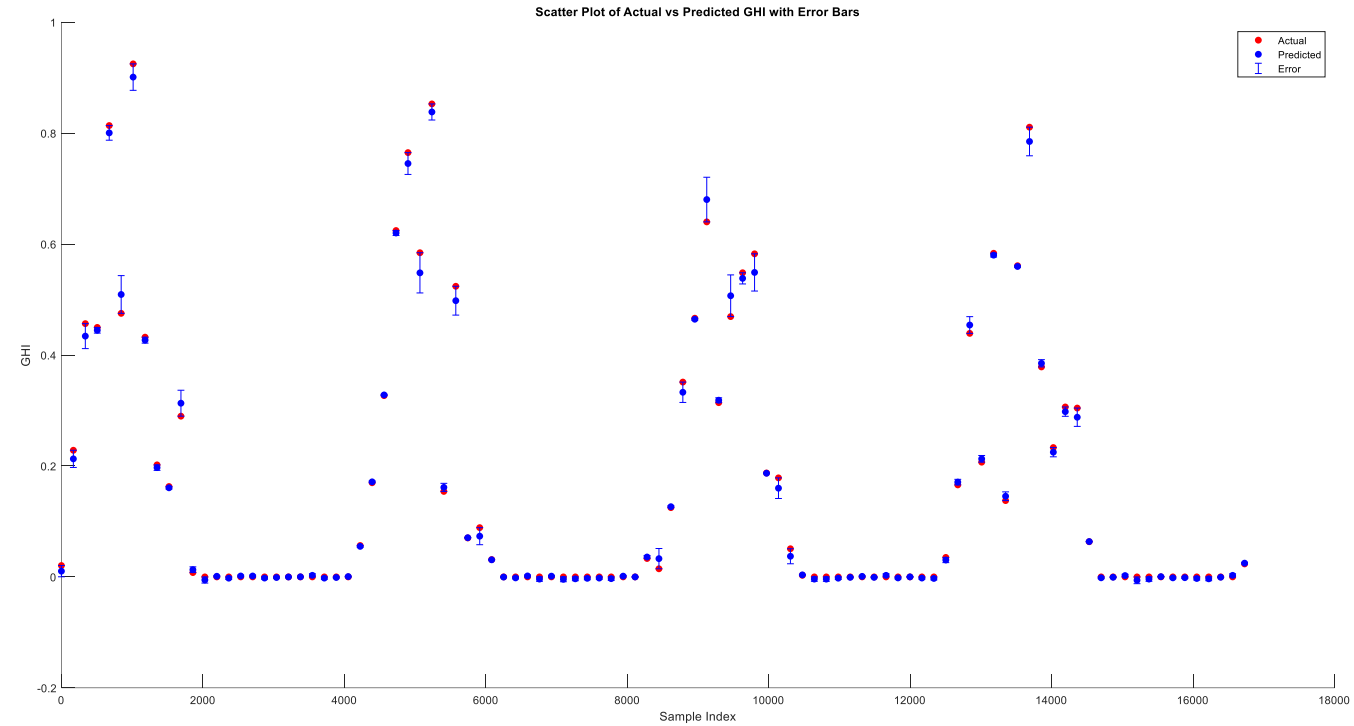


# Results & Discussion

# Simulation Results:



# Simulation Results:



## Hybrid Model Performance on Test Set

MSE

0.00012353

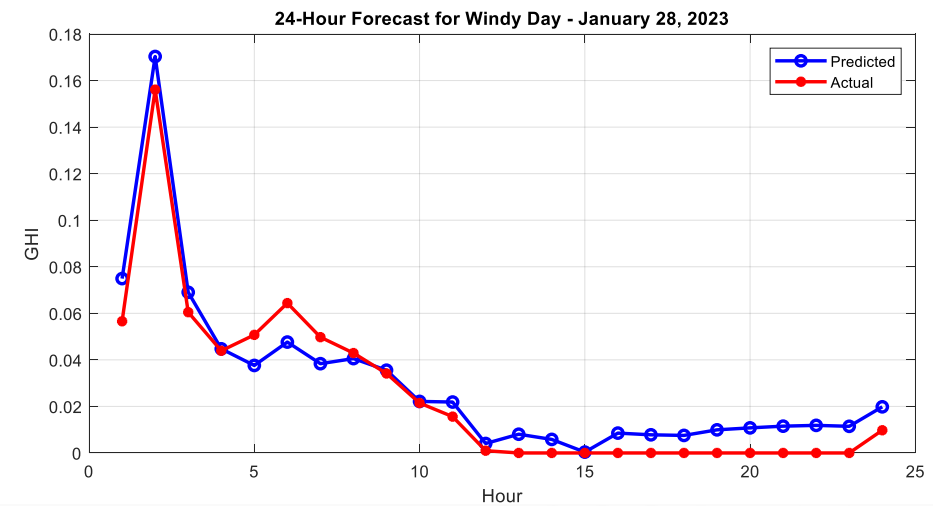
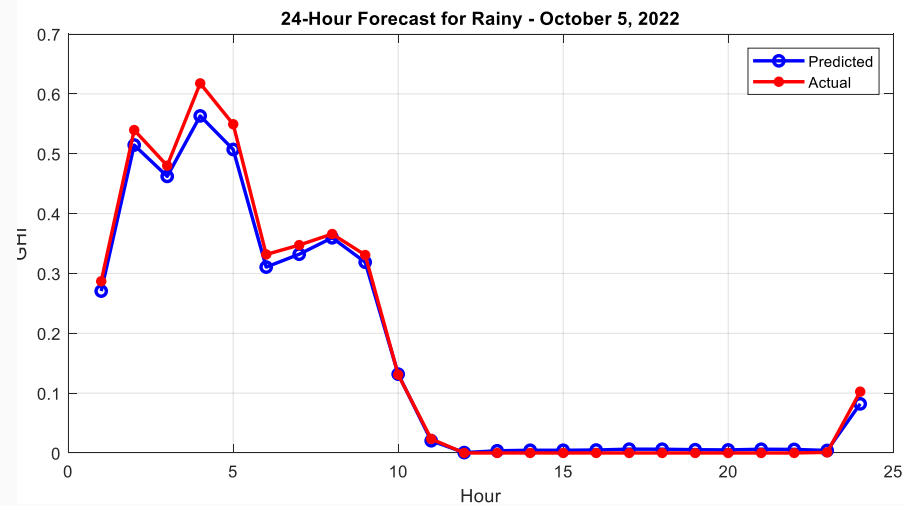
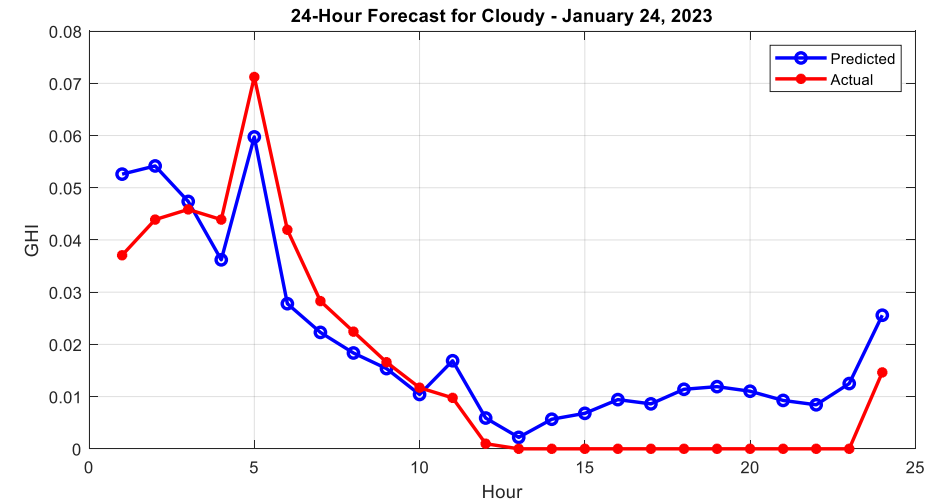
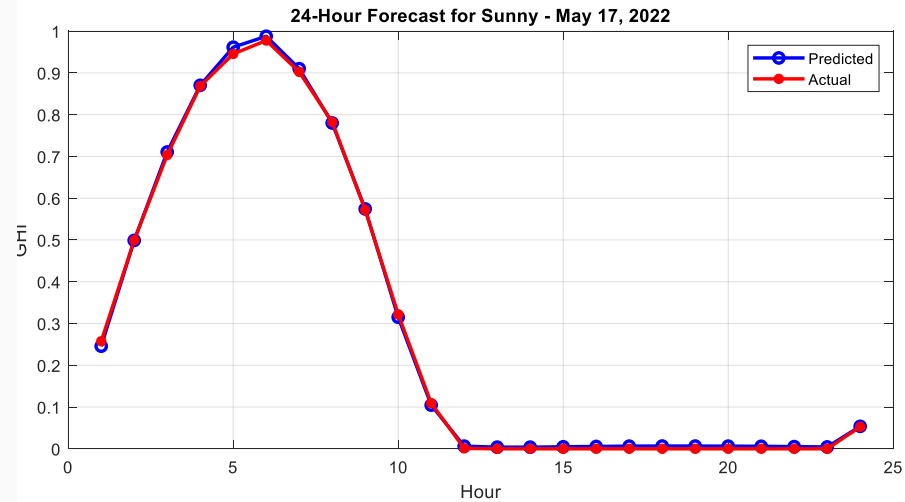
RMSE

0.0111140

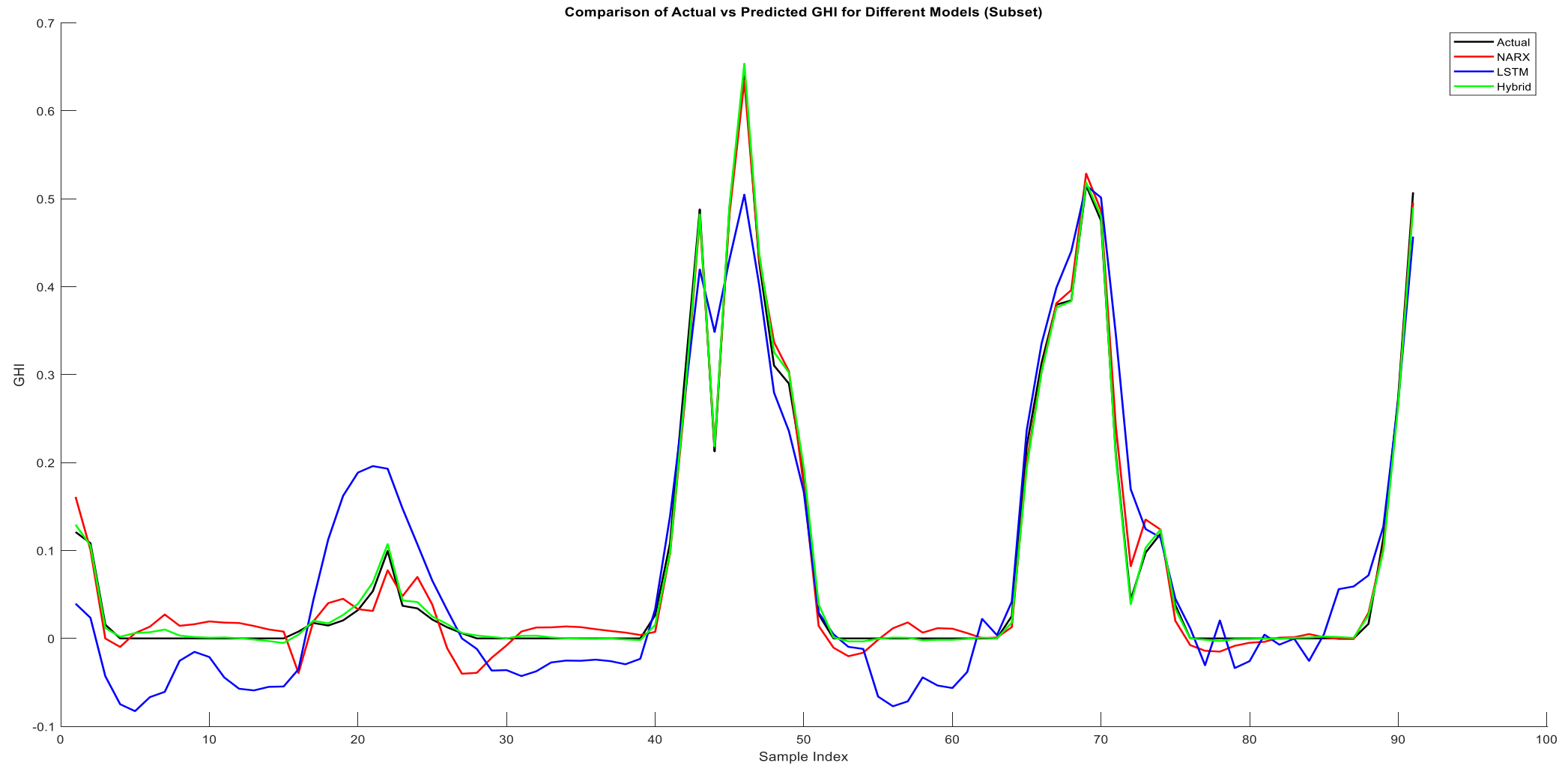
MAE

0.0064167

# Daily results for 24-hour ahead.



# Model Comparison



# Conclusion

- 1 For forecasting accuracy, actual historical weather data was collected from a comprehensive dataset covering diverse meteorological conditions. The hybrid NARX-LSTM model was meticulously calibrated using this data.
- 2 The hybrid model was successfully developed and demonstrated superior forecasting of solar irradiance, crucial for microgrid energy management.
- 3 Comparative results indicated that the hybrid NARX-LSTM model achieved enhanced predictive accuracy for GHI forecasting over different weather conditions and standalone models.



- [1] Y. Yoldaş, A. Önen, S. M. Mueen, A. V. Vasilakos, and İ. Alan, “Enhancing smart grid with microgrids: Challenges and opportunities,” *Renewable and Sustainable Energy Reviews*, vol. 72. Elsevier Ltd, pp. 205–214, 2017.
- [2] J. Ma and X. Ma, “A review of forecasting algorithms and energy management strategies for microgrids,” *Systems Science and Control Engineering*, vol. 6, no. 1, pp. 237–248, Jan. 2018.
- [3] I. Antonopoulos et al., “Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review,” *Renewable and Sustainable Energy Reviews*, vol. 130. Elsevier Ltd, Sep. 01, 2020.
- [4] F. Rodríguez, A. Fleetwood, A. Galarza, and L. Fontán, “Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control,” *Renew Energy*, vol. 126, pp. 855–864, Oct. 2018.
- [5] 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.
- [6] A. Shakya et al., “Solar Irradiance Forecasting in Remote Microgrids Using Markov Switching Model,” *IEEE Trans Sustain Energy*, vol. 8, no. 3, pp. 895–905, Jul. 2017.
- [7] D. Sáez, F. Ávila, D. Olivares, C. Cañizares, and L. Marín, “Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids,” *IEEE Trans Smart Grid*, vol. 6, no. 2, pp. 548–556, Mar. 2015.

# THANK YOU



[utm.my](http://utm.my)



[univteknologimalaysia](https://www.facebook.com/univteknologimalaysia)



[utmofficial](https://www.instagram.com/utmofficial)