MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

BÙI DUY LINH

STUDYING SHORT-TERM FORECASTING MODEL FOR POWER CAPACITY OF SOLAR POWER PLANTS USING RECURRENT NEURAL NETWORKS

SUMMARY OF ENERGY ENGINEERING DOCTORAL THESIS

HANOI – 2024

The thesis has been completed at the Graduate University of Science and Technology - Vietnam Academy of Science and Technology

Supervisor 1: Dr. Nguyen Quang Ninh

Supervisor 2: Dr. Doan Van Binh

Reviewer 1: ………………………………………………

Reviewer 2 ………………………………………………

Reviewer 3: ………………………………………………

The thesis is defended in front of the Thesis Committee at Vietnam Academy Of Science And Technology - Graduate University Of Science And Technology, at ….… hour……, date…… month….…year 2024

This thesis could be found at:

- The National Library of Vietnam
- The Library of Graduate University of Science and Technology, Vietnam Academy of Science and Technology

INTRODUCTION

1. The urgency of the thesis

- Currently, the contribution of solar power (SP) to the total power generation capacity of the Vietnamese power grid during peak hours is typically around 30% to 35%. However, during periods of low electricity demand, such as extended holidays, this contribution can rise to nearly 50%. Due to the highly volatile nature of SP output, primarily driven by varying weather conditions, this type of energy source poses significant challenges in power grid operation. Specifically, these challenges include: Maintaining a High Capacity Reserve: The power grid must maintain a substantial capacity reserve to ensure a stable power supply even when the power output from renewable energy sources fluctuates significantly and unpredictably; Inertia and Grid Stability: SPS installations primarily employ electronic power devices, which contribute very little to the grid's overall inertia. When a large portion of the grid's power comes from these sources, it can lead to reduced stability and increased risk of frequency and voltage fluctuations on the electrical grid.
- Therefore, to ensure the safety and optimization of operational activities, it is imperative to have highly accurate forecasts of the power output from these sources. Researching technologies to develop short-term power output forecasting models for solar power plants, especially utilizing recurrent neural network models, is a promising direction with a high degree of applicability in the context of the robust development of solar power sources today. The application of this technology involves not only data processing and training and testing to construct the best possible model but also encompasses the development of a feasible deployment process that can be practically implemented at industrial-scale solar power plants. This process

should be adaptable for real-world application, addressing the unique challenges presented by these renewable energy sources.

2. Research objectives

- Developing short-term power output forecasting models for solar power plants using long short-term memory (LSTM) neural networks.
- Proposing effective improvement strategies for model development.
- Creating a procedure and tools for short-term power output forecasting using LSTM networks.

3. Research methods

The research methods employed in this thesis encompass:

- Literature review for gathering scientific and practical information.
- Experimental approach
- Analysis and summary

4. Research scope

- *Forecasting framework:* The thesis concentrates on investigating suitable forecasting methods and conducting experiments for short-term forecasting.
- *Research subject*: The primary focus of the thesis is on researching short-term power output forecasting models for industrial-scale solar power plants.
- *Approach to forecasting problem:* The thesis emphasizes a direct forecasting approach for predicting the power output of solar power plants based on selected meteorological and other influencing factors.

5. Scientific and Practical Basis of the thesis

- Scientific Basis: This research builds upon the scientific foundations of the solar energy field, recurrent neural networks, and power forecasting methods. Existing studies have demonstrated that the application of deep learning algorithms can significantly enhance the accuracy of solar power output predictions.

- Practical Basis: The practical basis of this thesis stems from the growing demand for accurate short-term power output predictions for solar power plants in the Vietnamese electrical system and worldwide. The rapid development of solar energy in recent years has introduced challenges in controlling and managing the electrical grid, particularly in the short-term forecasting of solar power output.

6. Thesis's novel contributions

- The thesis successfully constructs a short-term power output forecasting model for solar power plants employing LSTM networks. The model has been rigorously tested and proven to be more accurate than traditional forecasting models for solar power plants in Vietnam. It provides superior forecasting quality compared to both domestic and international LSTM-based models published over the past five years.
- The thesis proposes a series of efficient solutions for enhancing the model development process. These solutions include: (1) Data preprocessing techniques utilizing the P/GHI ratio in conjunction with GHI clustering; (2) Training techniques involving the use of meteorological forecast data; (3) Utilization of clear sky solar radiation data as a substitute for time-based indicators. The research results have been formally published in international journals, comprising 2 SCI-Q1 papers and 2 Scopus-Q3, Q4 papers.
- Development of a comprehensive forecasting procedure and tool for short-term power output of solar power plants using LSTM networks, facilitating practical implementation and utilization of forecasting technology for solar power plants.

7. Thesis’s structure

The thesis is structured as follows:

- Introduction
- Chapter 1: Research Overview
- Chapter 2: Building short-term power output forecasting models for solar power plants using long short-term memory networks (LSTM)
- Chapter 3: Effective model improvement solutions and the development of forecasting procedure and tools
- Conclusion and future research directions
- List of published works
- References
- Appendices

CHAPTER 1: RESEARCH OVERVIEW

1.1. Introduction

This session will outline the necessary research areas that need to be explored to obtain a comprehensive perspective on the issue of forecasting the power output of solar power plants.

1.2. Power Output of Solar Power Plants and Influencing Factors

The factors that impact the power output of a solar power plant include:

- Meteorological factors (Solar radiation, temperature, wind speed, humidity, etc.)
- Panel orientation
- Load and Control mode
- Grid connection limitations

1.3. Classification of Solar Power Plant Power Output Forecasting Frameworks

- Long-Term Forecasting
- Medium-Term Forecasting
- Short-Term Forecasting

1.4. Model Performance Evaluation Methods

1.4.1. Evaluation via Performance Metrics

Various studies employ a range of criteria to assess the quality of forecasting models. Some commonly used criteria include MAE, PE, APE, MAPE, MSE, RMSE, nRMSE.

1.4.2. Evaluation via Error Distribution Charts

Error distribution charts, often referred to as error histograms, provide a comprehensive and visual representation of the likelihood of a specific error value occurring.

1.5. Short-Term Forecasting Methods for Power Capacity of Solar Power Plant

Presentation of research on methods: Physics, Statistics, Machine Learning, Combined; Approaches: Indirect, direct.

Long Short-Term Memory (LSTM) Networks

Long Short-Term Memory, is a specialized type of recurrent neural network capable of learning both short-term and long-term dependencies. LSTM networks are designed to address the issue of vanishing gradients that can occur in traditional recurrent neural networks.

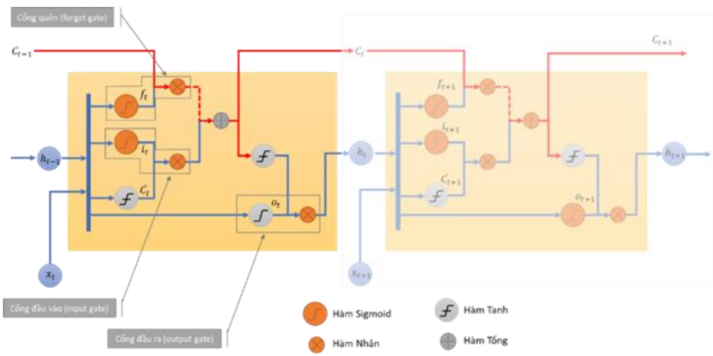

Figure 1.13. Structure of a Sequence of LSTM Blocks

1.6. Research status

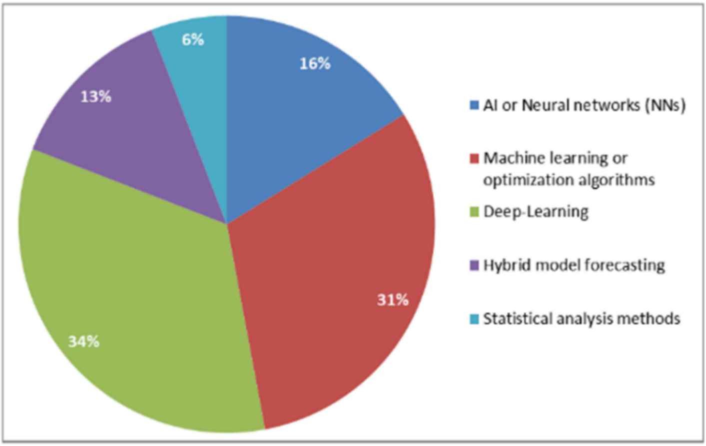

Figure 1.14: Proportion of Methods Used in Solar Power Output Forecasting Studies. Source: Tsai (2023)

In the 2023 study by Tsai and colleagues, which was published in 2023, the authors compiled 70 research studies published in reputable journals from 2020 to 2023 on the topic of solar power output forecasting. As depicted in Figure 1.14, the conclusion drawn is that deep learning-based forecasting methods are currently the predominant approach, constituting 34% of the studies. Among these methods, the most commonly used deep learning model is the LSTM, which has been demonstrated to be highly effective in short-term solar power output forecasting.

Table 1.4: Summary of Some Solar Power Output Forecasting Models Using LSTM Published from 2019

No	Model	Year	Rate Po-wer	Inter-val	Hori-zon	Step	Frame	MAPE (%)	nRMSE (%)
1	Wang et al.	2019	23,4 kWp (19,5 kW)	5 min	t	1	5 min	2,092	5,879

No	Model	Year	Rate Po-wer	Inter-val	Hori-zon	Step	Frame	MAPE (%)	nRMSE (%)
2	Ospina et al.	2019	12,6 MW	30 min	t+1	1	30 min	6,563	12,184
3	Zhou H et al.	2019	20 kW	7,5 min	t+1	1	7,5 min	4,000	6,950
				7,5 min	From t+1 to t+2	2	15 min	5,000	8,000
				7,5 min	From t+1 to t+4	4	30 min	6,100	9,050
				7,5 min	From t+1 to t+8	8	60 min	7,350	10,450
4	Wen et al.	2019	106,6 kWp	60 min	t	1	60 min	7,566	13,051
5	Harrou et al.	2020	9 MWp	15 min	t+1	1	15 min	8,930	
6	Zhou N et al.	2020	5,83 kWp (4,85 kW)	05 min	From t+1 to t+288	288	1 day	11,639	20,075
7	Zhang et al.	2020	6,41 kW	15 min	From t+1 to t+96	96	1 day	8,418	15,489
8	Park et al.	2021	500 kWp	60 min	t	1	60 min		13,200
9	Li et al.	2021	40 kW	60 min	t	1	60 min		6,400
10	Zhou H et al.	2021	20 kW	7,5 min	t+1	1	7,5 min	3,500	6,500
			20 kW	7,5 min	From t+1 to t+2	2	15 min	4,250	7,000
			20 kW	7,5 min	From t+1 to t+4	4	30 min	6,900	10,200

No	Model	Year	Rate Po-wer	Inter-val	Hori-zon	Step	Frame	MAPE (%)	nRMSE (%)
11	Liu et al.	2021	18,78 kW	05 min	t	1	05 min		4,886
12	Pombo et al.	2022	10 kW	05 min	From t+1 to t+60	60	5 hour		14,990
13	Suresh et al.	2022	317 kWp (265 kW)	60 min	From t+1 to t+24	24	1 day	4,800	9,046
14	Nguyễn Đức Tuyên et al.	2020	78 MW	60 min	From t+1 to t+24	24	1 day	3,743	5,402

1.7. Conclusion

Recent studies have demonstrated that regression neural networks, especially the LSTM model, are an effective tool in forecasting solar power capacity in the short term. Although the number of studies published in the world in the past 5 years is relatively large and diverse, publications from domestic authors are still limited. Part of the reason is that innovation has only really boomed in development in Vietnam from 2019 to now, leading to the lack of data sets to serve the implementation of research. In addition, the statistics in Section 1.6 have shown that new studies have only focused on power forecasts for small and medium-scale solar power systems from several tens of kW to several MW, while there have been few studies on forecasts for large-scale solar power plants. This will also make it difficult to implement the results of practical application because the dataset nature of large-scale solar power plants is very different when the meteorological datasets collected are representative of a very large area of the plant instead of reflecting relatively accurately as for systems small scale. The reason is that usually, according to regulations, for industrial-scale solar power plants, on average, from every 30 MW to 50 MW of installed capacity, there are from 1 to 2 meteorological measuring stations.

Therefore, this thesis will delve into research focused on using LSTM recurrent neural networks to build short-term power output forecasting models for industrial-scale solar power plants in Vietnam. The thesis not only applies the LSTM methodology but also proposes solutions to enhance the effectiveness of using LSTM for short-term power forecasting. Moreover, to facilitate further research and development of related techniques, this thesis deeply analyzes the proposed techniques based on detailed statistical data. This data serves as a basis for determining the model's parameters.

CHAPTER 2.
BUILDING A SHORT-TERM FORECASTING MODEL FOR POWER CAPACITY OF SOLAR POWER PLANTS USING LONG SHORT-TERM MEMORY NETWORKS

2.1. Introduction

2.2. Data Collection

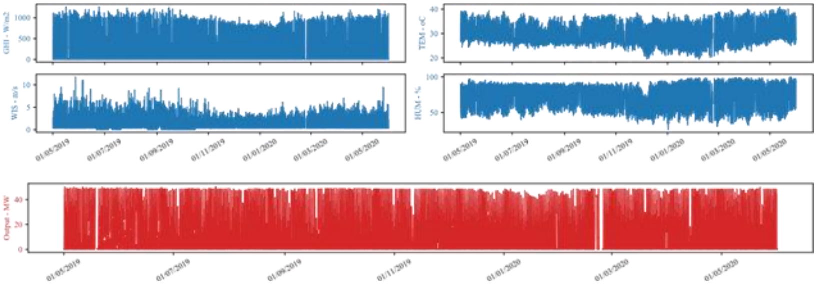

Figure 2.1. Historical Operation Data of the Plant

The dataset for experimentation was collected from a solar power plant located in the southern region of Vietnam, spanning from May 2019 to June 2020.

2.3. Experimental Environment

Presentation of the environment and tools used in experiments.

2.4. Data Preprocessing

The Interquartile Range (IQR) technique is used for outlier detection to prevent noise during model training and testing.

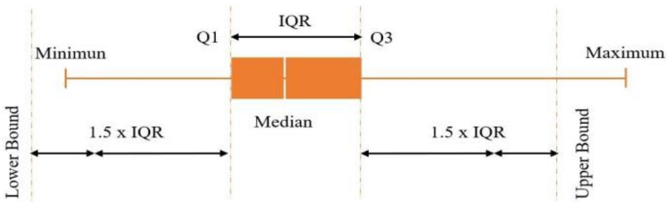

Figure 2.4. The IQR method for identifying outliers

The data preprocessing process based on the application of IQR involves the following steps:

- In Step 1, the original data is processed and filtered using IQR.
- In Step 2, missing data points are calculated by interpolating from nearby data points. However, this may lead to inaccurate data because of the large range of missing data.
- In Step 3, the data after Step 2 is further filtered using IQR. Reasonable data points are retained, while outliers are replaced with the median (Q2) of the data values within that time frame.

2.5. Building the LSTM Model and Comparative Models

The selected models for comparison include:

- Persistence model
- ARIMA model (4,0,0)
- MLP models (1-4 layers; 7, 50, 100 hidden nodes in each layer)

To choose optimal parameters for the LSTM model, the network configuration and training settings are selected as follows:

- Each input is a matrix with dimensions of 5 rows x 7 columns.
- The number of network layers varies from 1 to 4.
- The number of hidden nodes in each network layer ranges from 7 (corresponding to 7 input nodes) to 50 and 100 nodes.

2.6. Training the LSTM Model

Conduct training of LSTM models according to the stated configurations.

2.7. Comparison of Forecasting Results from Different Models

Table 2.6. Summary of Evaluation Metrics for Error Results on the Test Dataset for Different Models

Model	Layer (L) – Node (N)	MSE	RMSE	nRMSE	MAE	MAPE
		MW^2	MW	%	MW	%
Persistence		128.259	11.325	23.594	7.401	15.418
ARIMA		29.268	5.410	11.271	3.518	7.329
MLP	1L-100N	12.038	3.470	7.229	2.086	4.346
MLP	2L-100N	12.063	3.473	7.235	2.165	4.511
MLP	3L-100N	15.276	3.908	8.142	2.339	4.873
MLP	4L-100N	17.641	4.200	8.750	2.546	5.304
LSTM	1L-7N	12.951	3.599	7.498	2.094	4.362
LSTM	1L-50N	10.700	3.271	6.815	1.901	3.960
LSTM	1L-100N	12.126	3.482	7.254	2.037	4.244
LSTM	2L-7N	12.778	3.575	7.448	2.153	4.485
LSTM	2L-50N	10.425	3.229	6.727	1.831	3.815
LSTM	2L-100N	10.176	3.190	6.646	1.766	3.679
LSTM	3L-7N	10.880	3.299	6.873	1.928	4.017
LSTM	3L-50N	10.414	3.227	6.723	1.811	3.772
LSTM	3L-100N	9.223	3.037	6.327	1.694	3.529
LSTM	4L-7N	12.104	3.479	7.248	2.110	4.395
LSTM	4L-50N	10.085	3.176	6.617	1.803	3.757
LSTM	4L-100N	9.499	**3.082**	**6.421**	1.676	**3.491**

LSTM networks demonstrate an advantage in terms of accuracy on the test dataset compared to other traditional methods. The simplest LSTM network with a structure of one hidden layer and 7 hidden nodes yields results quite similar to the best-performing MLP network. By increasing the depth of the LSTM network (complexity), the best MAPE result (3.491%) is achieved from the 4-layer network with 100 nodes (LSTM-4L100N).

2.8. Training the Model Using Validation Set and Early Stopping Technique

Table 2.6. Optimal Training Results for the Model

Index	MAE	MAPE	MSE	RMSE	nRMSE
Unit	***MW***	***%***	***MW^2***	***MW***	***%***
Basic Training	1.676	3.491	9.499	3.082	6.421
Training with 10% Validation and Early Stopping	1.412	2.942	4.692	2.166	4.513

The model obtained from training with validation and early stopping resulted in improved forecast error:

- MAPE decreased from 3.491% to 2.942%, showing an improvement of approximately 16.0% compared to basic training.
- RMSE decreased from 3.082% to 2.166%, indicating an improvement of about 29.7% compared to basic training.

2.9. Compare with univariate model

Table 2.8. Compare models using multivariate and univariate inputs

Inputs		MSE	RMSE	nRMSE	MAE	MAPE
		MW^2	**MW**	**%**	**MW**	**%**
Multivariate		***4,692***	***2,166***	***4,513***	***1,412***	***2,942***
Uni-variate	**GHI**	57,268	7,568	15,767	4,892	10,192
	TEM	192,344	13,869	28,894	10,695	22,281
	HUM	299,621	17,31	36,063	13,491	28,106
	WIS	453,961	21,306	44,388	17,081	35,585

2.10. Forecast 01 next step

In previous sections, the thesis studied the t-step prediction model using inputs at t, t-1, t-2, t-3, t-4. The resulting model will be applied when forecast meteorological data is available to forecast capacity in future cycles. In this section, the thesis experiments with forecasting the next 01 step using only historical data. This model applies in the case when the deployment does not use forecast meteorological data.

Table 2.9. Power forecast error results for the t-cycle using past meteorological inputs (1) t-4 to t and (2) t-4 to t-1

No	**Meteo-rological inputs**	**Out-put**	**MAE**	**MAPE**	**MSE**	**RMSE**	**nRMSE**
			MW	***%***	***MW^2***	***MW***	***%***
1	t-4, t-3, t-2, t-1, t	t	1,412	2,942	4,692	2,166	4,513
2	t-4, t-3, t-2, t-1	t	3,143	6,549	23,664	4,865	10,140

2.11. Forecast multi-step

The study tested the forecasting of multiple output steps for the next 01 day (next 24 hours). Because the dataset has a resolution of 05 minutes, the next number of output steps to forecast is 288 steps. At each training step, the MAE loss function value is calculated from the next 288 forecast values simultaneously instead of just the next 1.

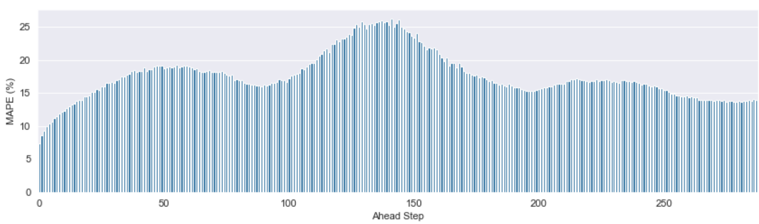

Figure 2.16. MAPE index results (%) on the test dataset of each forecast step in the multistep forecasting model

2.12. Evaluation and Conclusion

In order to evaluate the thesis, compare the model results achieved with some results from domestic and foreign publications related to the forecast of solar power using LSTM network from 2019 to now presented in Section 1.6 on summarizing the research situation.

Table 2.11. Compare study results with published models

Model	Year	Rate Po-wer	Inter-val	Hori-zon	Step	Fra-me	MAPE (%)	nRMSE (%)
Forecast for the current interval (with current interval meteorological data)								
Thesis's Model		**48 MW**	**05 min**	**t**	**1**	**05 min**	**2,942**	**4,513**
Liu et al.	2021	18,78 kW	05 min	t	1	05 min		4,886
Wang et al.	2019	23,4 kWp	05 min	t	1	05 min	2,092	5,879
Li et al.	2021	40 kW	60 min	t	1	60 min		6,400
Wen et al.	2019	106,6 kWp	60 min	t	1	60 min	7,566	13,051
Suresh et al.	2022	317 kWp	60 min	t	1	60 min	4,800	9,046
Park et al.	2021	500 kWp	60 min	t	1	60 min		13,200
Forecast for the next interval (only have meteorological data)								
Thesis's Model		**48 MW**	**05 min**	**t+1**	**1**	**05 min**	**6,549**	**10,140**
Zhou H et al.	2019	20 kW	7,5 min	t+1	1	7 ,5 min	4,000	6,950
Zhou H et al.	2021	20 kW	7,5 min	t+1	1	7,5 min	3,500	6,500
Harrou et al.	2020	9 MWp	15 min	t+1	1	15 min	8,930	

Model	Year	Rate Po-wer	Inter-val	Hori-zon	Step	Fra-me	MAPE (%)	nRMSE (%)
Ospina et al.	2019	12,6 MW	30 min	t+1	1	30 min	6,563	12,184
Forecast for the next day (only have meteorological data)								
Thesis's Model		**48 MW**	**05 min**	**From t+1 to t+288**	**288**	1 day	**13,902**	**19,190**
Zhou N et al.	2020	5,83 kWp	05 min	From t+1 to t+288	288	1 day	11,639	20,075
Zhang et al.	2020	6,41 kW	15 min	From t+1 to t+96	96	1 day	8,418	15,489
Nguyễn Đức Tuyên et al.	2020	78 MW	60 min	From t+1 to t+24	24	1 day	3,743	5,402

Table 2.11 presents a comparison of the calculation results from the model in the thesis with several published models according to the Normalized Root Squared Mean Error (nRMSE) and Mean Absolute Percentage Error (MAPE) over forecast timeframes.

CHAPTER 3.
IMPROVING MODEL EFFICIENCY AND DEVELOPING A FORECASTING PROCEDURE AND TOOL

3.1. Data preprocessing technique with the P/GHI ratio combined with GHI clustering

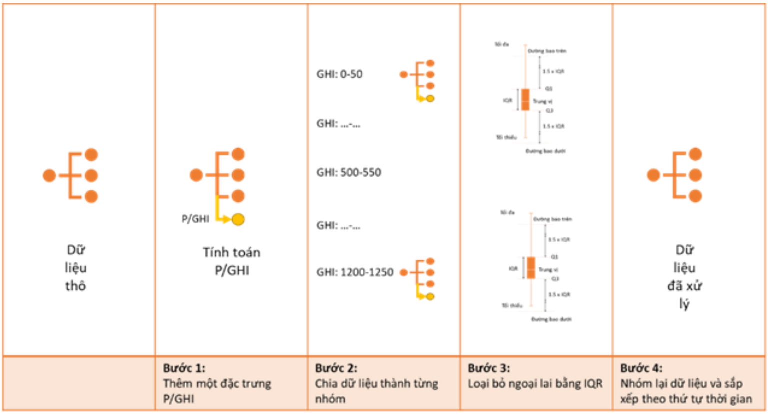

Figure 3.2. Description of the steps in the proposed data processing technique."

To overcome the challenges of handling outlier data in the solar power plant dataset, a study published in 2022 proposed a solution that combines the IQR method with the P/GHI ratio and divides the GHI range into appropriate segments for preprocessing the input data. The process is illustrated in Figure 3.2.

Bảng 3.1. Prediction Error Results on the Test Dataset with Pre- and Post-Processing Using the P/GHI and GHI Clustering Technique

Index	**MAE**	**MAPE**	**MSE**	**RMSE**
Dataset	**MW**	**%**	**MW^2**	**MW**
Before Processing	2,501	6,413	10,354	3,218
After Processing Using P/GHI with GHI Clustering Technique	1,602	4,109	4,706	2,169

Table 3.1 presents the forecast error results on the test dataset with the training dataset before and after processing with the P/GHI combined with GHI clustering technique. It can be observed that applying the proposed processing technique reduces the MAPE from 6.413% to 4.109% with a 36.2% improvement.

3.2. Training Technique Using Forecasted Weather Data in the Training and Forecasting Process

While the model obtained in Chapter 2 performs well with actual weather data, the use of weather forecast data significantly increases errors, with RMSE of 8.065 MW (2.62 times higher) and MAPE of 10.857% (3.11 times higher). From this, it can be observed that there is a substantial increase in errors when using weather forecast data, primarily due to the forecasting errors of meteorological data. These errors are unavoidable and primarily depend on the quality of the forecast provider.

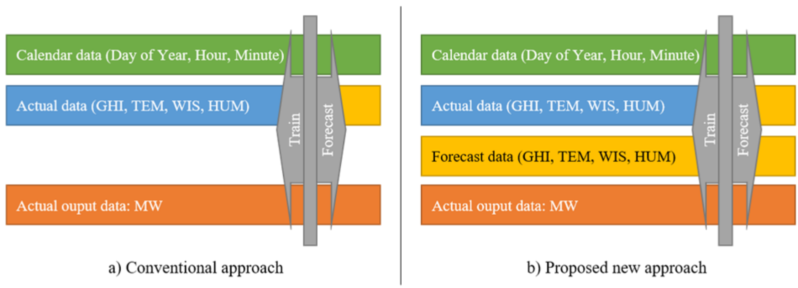

Figure 3.7. Differences in the input data structure of the proposed model.

The main idea of the solution is that if the weather data prediction algorithm is stable, then the weather data prediction error also follows a hidden pattern that the LSTM model can discover. Training the model solely on real operational data will not allow it to learn this pattern. However, incorporating historical weather forecast data during the training process may lead to a worse error when testing with real data, but it can improve the actual power output forecasting results.

Table 3.3. Synthesis compares the error of models with different inputs

(a) Errors on the test dataset

Model	Number of inputs	*MSE*	*RMSE*	*MAE*	*MAPE*
		MW^2	MW	MW	%
7I	7 (3 calendar inputs + 4 historical meteorological inputs)	9,499	3,082	1,676	3,491
7I+1F	8 (3 calendar inputs + 4 historical meteorological inputs + GHI forecast in the past)	12,482	3,533	1,983	4,132
7I+2F	9 (3 calendar inputs + 4 historical meteorological inputs + GHI, TEM forecast in the past)	10,829	3,291	1,909	3,978
7I+3F	10 (3 calendar inputs + 4 historical meteorological inputs + GHI, TEM, HUM forecast in the past)	11,490	3,390	1,887	3,932
7I+4F	11 (3 calendar inputs + 4 historical meteorological inputs + GHI, TEM, HUM, WIS forecast in the past)	12,093	3,477	1,946	4,054

(b) Errors on the test dataset when applying forecasts in practice (replacing actual meteorological inputs with forecast meteorological data)

Model	Số đầu vào	*MSE*	*RMSE*	*MAE*	*MAPE*
		MW^2	MW	MW	%
7I	7 (3 calendar inputs	**65,041**	**8,065**	**5,211**	**10,857**

	+ 4 historical meteorological inputs)				
7I+1F	8 (3 calendar inputs + 4 historical meteorological inputs + GHI forecast in the past)	67,119	8,193	5,182	10,796
7I+2F	9 (3 calendar inputs + 4 historical meteorological inputs + GHI, TEM forecast in the past)	**56,348**	**7,507**	**4,743**	**9,881**
7I+3F	10 (3 calendar inputs + 4 historical meteorological inputs + GHI, TEM, HUM forecast in the past)	61,027	7,812	4,921	10,253
7I+4F	11 (3 calendar inputs + 4 historical meteorological inputs + GHI, TEM, HUM, WIS forecast in the past)	66,762	8,171	5,184	10,801

The error test results reflect the effectiveness of the proposed model. The 7I model now gives a significant increase in error with MAPE of 10.857% and RMSE of 8.065%. The models of 8 inputs (7I+1F), 9 inputs (7I+2F), 10 inputs (7I+3F), 11 inputs (7I+4F), adding respectively WRITE, STAMP, HUM, WIS forecasts in the past all recorded an improvement in MAPE compared to the 7I model. Among them, the error of the 7I+2F model with MAPE of 9.881% achieved the most significant improvement of about 0.976% compared to the 7I model, the improvement was about 9%.

3.3. Technique of utilizing clear sky solar irradiance data as a replacement for time index inputs

The idea is to replace the dataset containing time indices (Day of the year, Hour of the day, Minute of the hour) with solar radiation data. Solar radiation is a potential input for improving the quality of forecasting models. This radiation is not derived from the plant's operational values, but it depends on the plant's geographical location and time of the year. Solar radiation can be pre-computed for any moment in the year, using available computational models and the plant's geographical coordinates. This value varies throughout the year and changes over the course of the day (hours and minutes). Therefore, it is a potential choice to replace all three values (Day of the year, Hour of the day, Minute of the hour).

Table 3.6. Comparison of forecast results on the test dataset

Model	**Frame**	**MAE**	**MAPE**	**MSE**	**RMSE**	**nRMSE**
		MW	***%***	***MW^2***	***MW***	***%***
Model using calendar indicator	Current interval	1,412	2,942	4,692	2,166	4,513
	Next interval	3,143	6,549	23,664	4,865	10,140
	Next day	6,673	13,902	84,835	9,211	19,190
Model using clearsky GHI indicator	Current interval	1,309	2,728	4,002	2,001	4,168
	Next interval	2,106	4,388	12,139	3,484	7,258
	Next day	5,909	12,310	80,966	8,988	18,745

3.4. Procedure and Tools for Short-Term Forecasting with LSTM

Drawing from the research process and the execution of experiments, the thesis presents a comprehensive procedure for addressing short-term power forecasting for a solar power plant using LSTM recurrent neural networks as follows:

- Step 1: Preprocessing the Input Data
- Step 2: Model Training
- Step 3: Forecasting

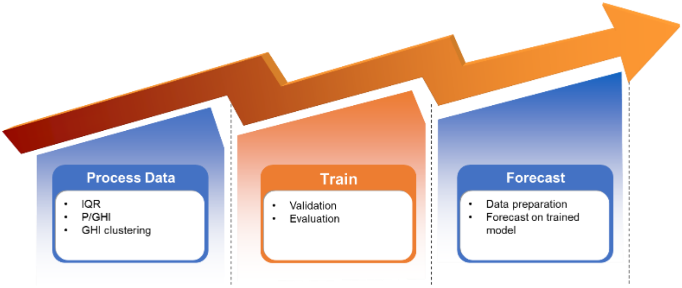

Figure 3.11. The process for handling the power forecasting problem for SPP

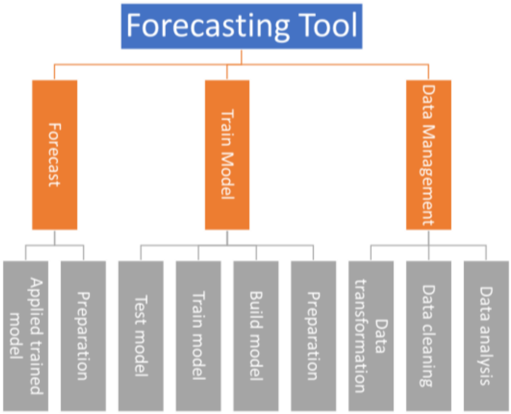

Figure 3.13. Main Functional Diagram of the Software Tool.

The main functional diagram of the software tool, as presented in Figure 3.13, consists of three primary functional blocks: (1) Data Management (2) Model Training (3) Forecasting.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

1. Conclusion

Through the research and experimentation process, the thesis on "***Research on Short-Term Power Forecasting Models for Solar Power Plants Using Recurrent Neural Networks***" has successfully completed all research objectives and achieved specific significant results as follows:

- Extensive data collection and synthesis of knowledge related to the field of solar power plant power forecasting, covering various aspects, including input data, forecasting frameworks, evaluation criteria, and forecasting methods. The analysis of research papers, both domestic and international, published in the last five years, has solidified the choice of investigating short-term power forecasting models for solar power plants using LSTM recurrent neural networks.
- Successfully experimented to build a short-term forecasting model of the generation capacity of a solar power plant using regression neural networks with an advanced variant of long-short memory networks (LSTMs). Through experimentation on solar plant data, the thesis demonstrated the efficiency of the 4-layer LSTM model with 100 hidden nodes per layer for MAPE error indicators of 3.491%, significantly better than empirical results from other popular comparison methods (MLP 4.346%, ARIMA 7.329%, Persistence 15.418%). The thesis model after optimal training (with 10% verifiable dataset and early stop technique) reduced MAPE error to 2.942%, nRMSE error to 4.513% for the current interval, achieving results comparable to the best models published in the past 5 years. For the forecasting framework for the next 01 interval, the MAPE error reached 6.549% and nRMSE reached 10.140%, equivalent to some published models with data resolution and plant scale conditions that are almost identical to the thesis data. For the forecast framework for the coming day, the error achieved by the thesis model is quite high, but when compared to other published models, the difference is not large and the main reason for the quality difference comes from

the difference in the size of the solar plant object of each study.

- Research has been conducted to propose innovative and experimental solutions that prove effective in building forecasting models including: (1) Data preprocessing techniques with P/GHI factors combined with GHI clustering reduce MAPE errors from 6.413% to 4.109% with an improvement of about 36.2% (2) Training techniques using meteorological data forecasting resulted in a reduction in MAPE error on the forecast dataset from 10.857% to 9.881%, an improvement of approximately 9.0% (3) The technique of using sky-radiation data in the substitution of time indicators resulted in a reduction in MAPE from 0.214% to 2.161% and nRMSE from 0.165% to 1.381% for forecasting frames.
- Development of a complete process and user-friendly software tool with a convenient interface for handling short-term power forecasting for solar power plants using LSTM recurrent neural networks.
- Formal publication of research results related to the thesis in international journals, including 2 SCI-Q1 papers and 2 Scopus-Q3, Q4 papers.

2. Future research direction

- An online training system and real-time training capabilities
- Advanced techniques to enhance model quality
- Further refinement of the procedure and tools
- Study of sky image application model to forecast in short-term

LIST OF PUBLISHED WORKS

International Journals: 4 publications

[1] **L. D. Bui**, N. Q. Nguyen, B. Van Doan, and E. R. Sanseverino, "Forecasting energy output of a solar power plant in curtailment condition based on LSTM using P/GHI coefficient and validation in training process, a case study in Vietnam," *Electric Power Systems Research*, vol. 213, p. 108706, Dec. 2022, doi: 10.1016/J.EPSR.2022.108706 (***IF = 3.818, 2022, SCI - Q1***).

[2] N. Q. Nguyen, **L. Duy Bui**, B. Van Doan, E. R. Sanseverino, D. Di Cara, and Q. D. Nguyen, "A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks, a case study in Vietnam," *Electric Power Systems Research*, vol. 199, p. 107427, 2021, doi: 10.1016/j.epsr.2021.107427 (***IF = 3.818, 2022, SCI - Q1***).

[3] **D. L. Bui**, Q. N. Nguyen, V. B. Doan, T. K. Pham, and D. D. Le, "Evaluating an Effectiveness of a Solar Power Plant Output Forecasting Model Based on LSTM Method Using Validation in Different Seasons of a Year in Vietnam," *GMSARN International Journal*, vol. 18, pp. 114–122, 2024 (***Scopus - Q4***).

[4] N. Quang, **L. Duy, B**. Van, and Q. Dinh, "Applying Artificial Intelligence in Forecasting the Output of Industrial Solar Power Plant in Vietnam," *EAI Endorsed Transactions on Energy Web*, p. 169166, Jul. 2021, doi: 10.4108/eai.29-3-2021.169166 (***Scopus – Q3***).

Patent: 01

The doctoral candidate participated in research and developed a patent for the "**LSTM Model Training Method for Solar Power Generation Forecast**" This patent has been submitted to the Intellectual Property Office and has been officially accepted (Document number 14874w/QD-SHTT dated March 31, 2023, issued by the Director of the Intellectual Property Office).