



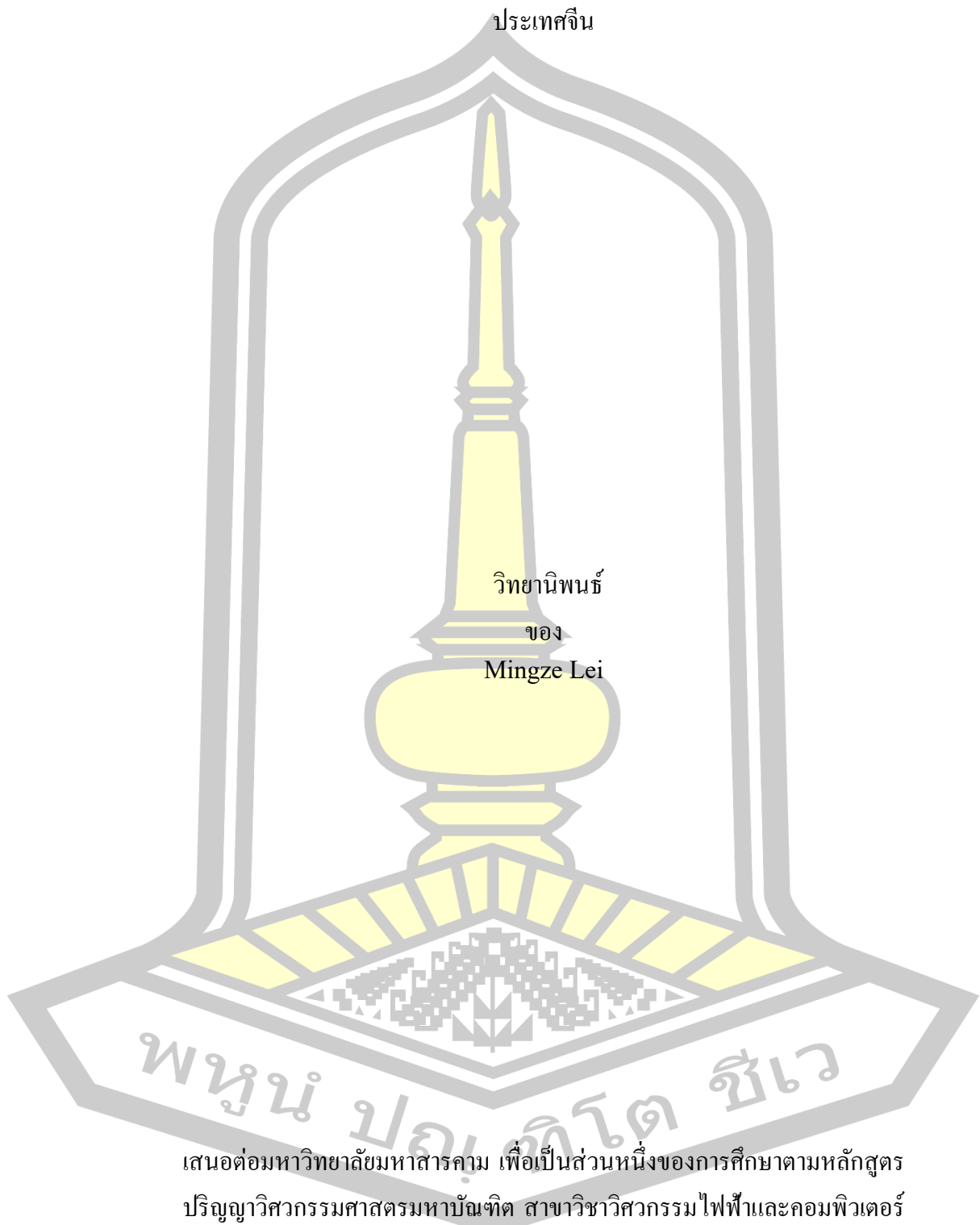
Solar Power Generation Prediction using LSTM Deep learning algorithm in Ningxia
Province of China

Mingze Lei

A Thesis Submitted in Partial Fulfillment of Requirements for
degree of Master of Engineering in Electrical and Computer Engineering
September 2025

Copyright of Mahasarakham University

การพยากรณ์กำลังการผลิตไฟฟ้าจากโซลาร์เซลล์โดยใช้วิธีการเรียนรู้เชิงลึกในมณฑลหนิงเซี่ย
ประเทศจีน



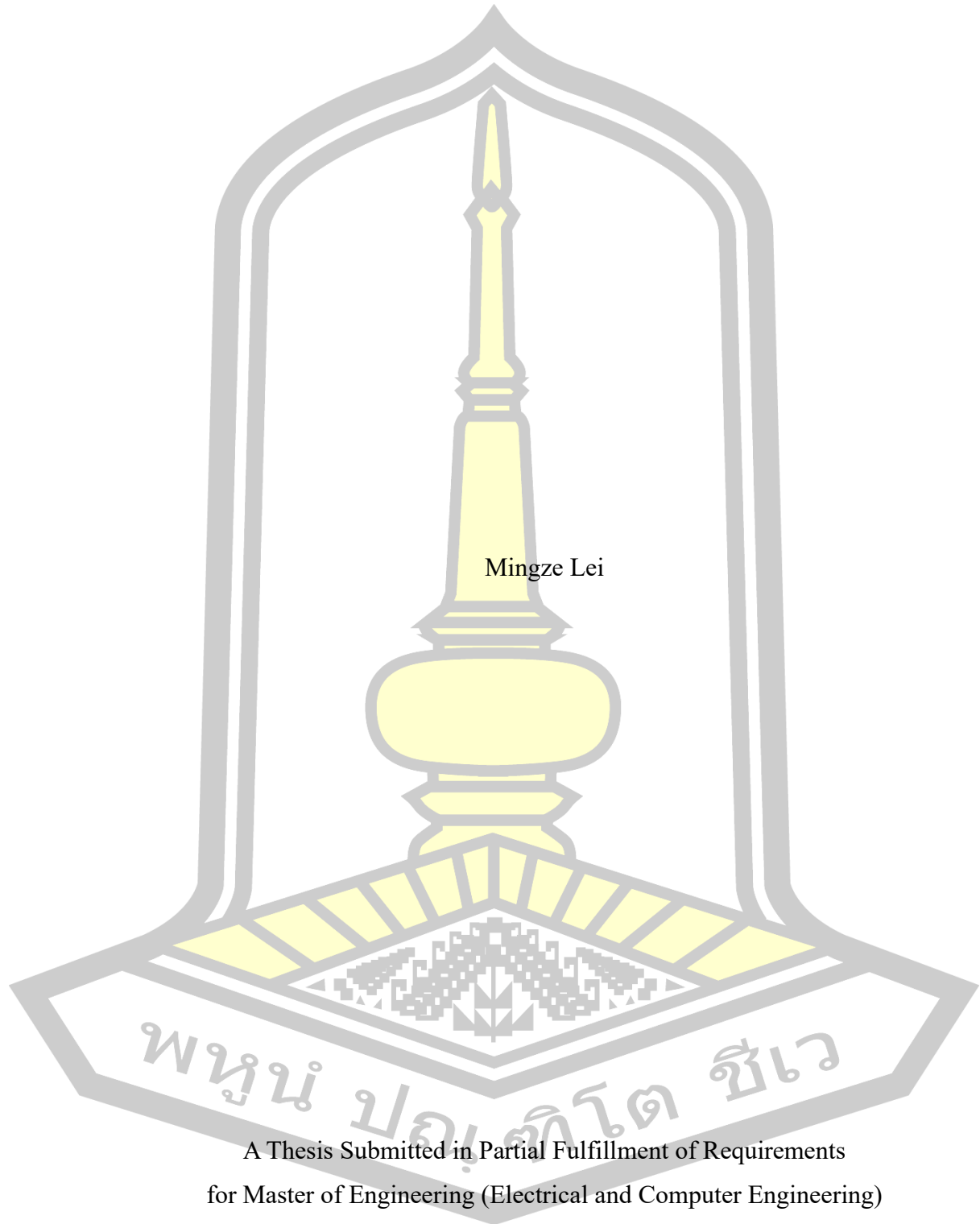
เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์

กันยายน 2568

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Solar Power Generation Prediction using LSTM Deep learning algorithm in Ningxia
Province of China

Mingze Lei



A Thesis Submitted in Partial Fulfillment of Requirements
for Master of Engineering (Electrical and Computer Engineering)

September 2025

Copyright of Mahasarakham University



The examining committee has unanimously approved this Thesis, submitted by Mr. Mingze Lei , as a partial fulfillment of the requirements for the Master of Engineering Electrical and Computer Engineering at Mahasarakham University

Examining Committee

Chairman

(Assoc. Prof. Anan Kruesubthaworn
, Ph.D.)

Advisor

(Asst. Prof. Buncha Wattana , Ph.D.)

Committee

(Prof. Worawat Sa-Ngiamvibool ,
Ph.D.)

Committee

(Asst. Prof. Chaiyong Soemphol ,
Ph.D.)

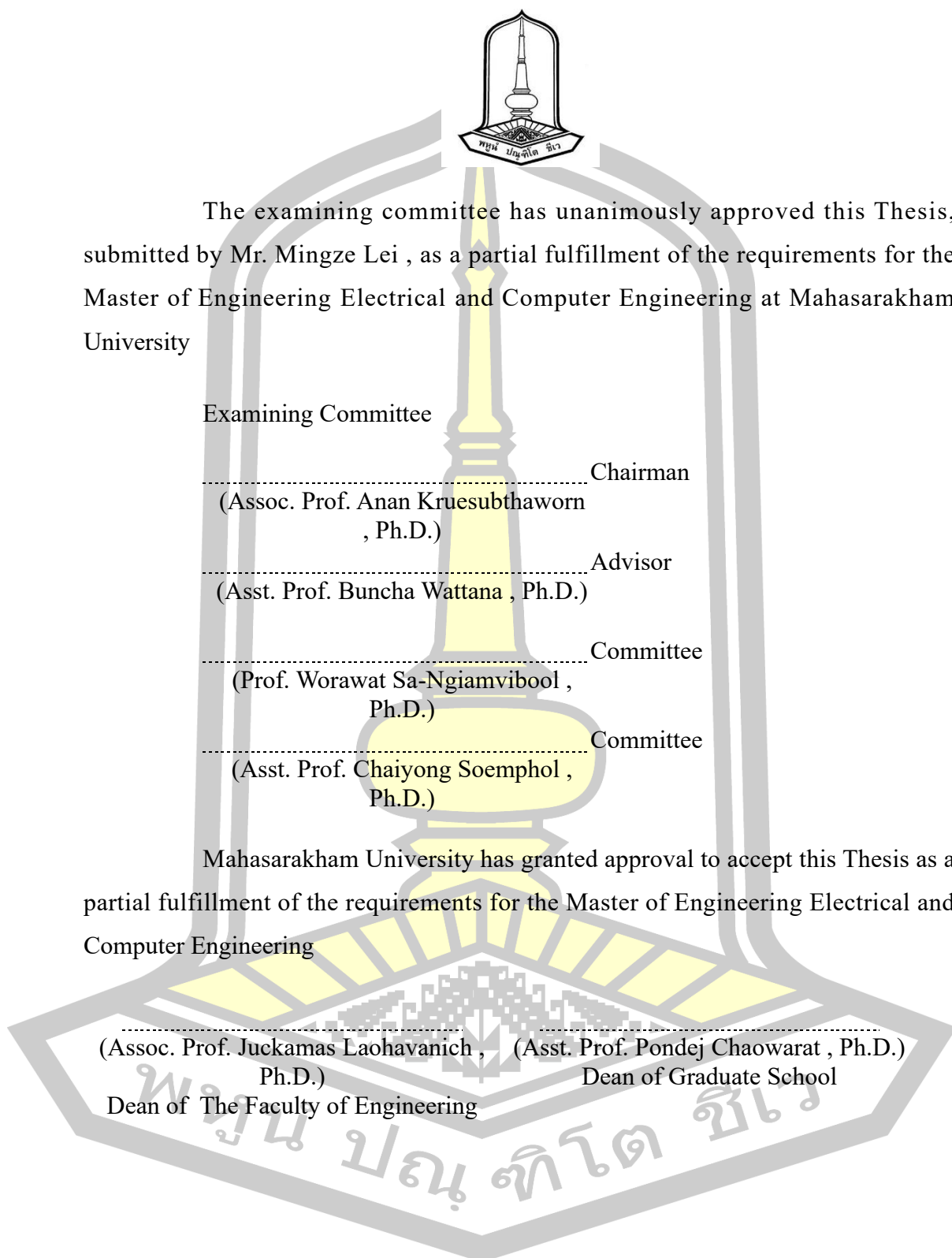
Mahasarakham University has granted approval to accept this Thesis as a partial fulfillment of the requirements for the Master of Engineering Electrical and Computer Engineering

(Assoc. Prof. Juckamas Laohavanich ,
Ph.D.)

Dean of The Faculty of Engineering

(Asst. Prof. Pondej Chaowarat , Ph.D.)

Dean of Graduate School



TITLE	Solar Power Generation Prediction using LSTM Deep learning algorithm in Ningxia Province of China		
AUTHOR	Mingze Lei		
ADVISORS	Assistant Professor Buncha Wattana , Ph.D.		
DEGREE	Master of Engineering	MAJOR	Electrical and Computer Engineering
UNIVERSITY	Maharakham University	YEAR	2025

ABSTRACT

The rapid expansion of photovoltaic (PV) power generation faces significant challenges due to the intermittent and stochastic characteristics of solar energy, which affect grid stability and energy management. Accurate forecasting of PV power output is crucial for optimizing grid operations and supporting the transition to clean energy. This paper proposes a deep learning approach based on Long Short-Term Memory (LSTM) networks to predict PV power generation in Ningxia Province, China. The model leverages historical power and meteorological data, which undergo comprehensive preprocessing, including outlier removal, normalization, and feature correlation analysis. The experiment is based on collecting data at 15-minute intervals, totaling 35,000 samples from a 1 MW photovoltaic power station in Ningxia for the entire year of 2023. The data include seven characteristic dimensions such as irradiance, temperature, and humidity. Comparative experiments involving Support Vector Machine (SVM), Convolutional Neural Network (CNN), and LSTM demonstrate that LSTM outperforms other methods with superior accuracy and robustness, achieving a coefficient of determination (R^2) of 0.9927. The results confirm LSTM's effectiveness in capturing temporal dependencies and nonlinear patterns in PV power data. This study provides valuable insights for enhancing photovoltaic grid integration and advancing intelligent power systems.

Keyword : Photovoltaic power generation, Deep learning algorithm, Long short-term memory, Power generation prediction

ACKNOWLEDGEMENTS

The authors would like to express gratitude to the Electrical and Computer Engineering Research Unit, Faculty of Engineering, Mahasarakham University and the College of Electrical Engineering, Hunan Mechanical & Electrical Polytechnic for the facility support.

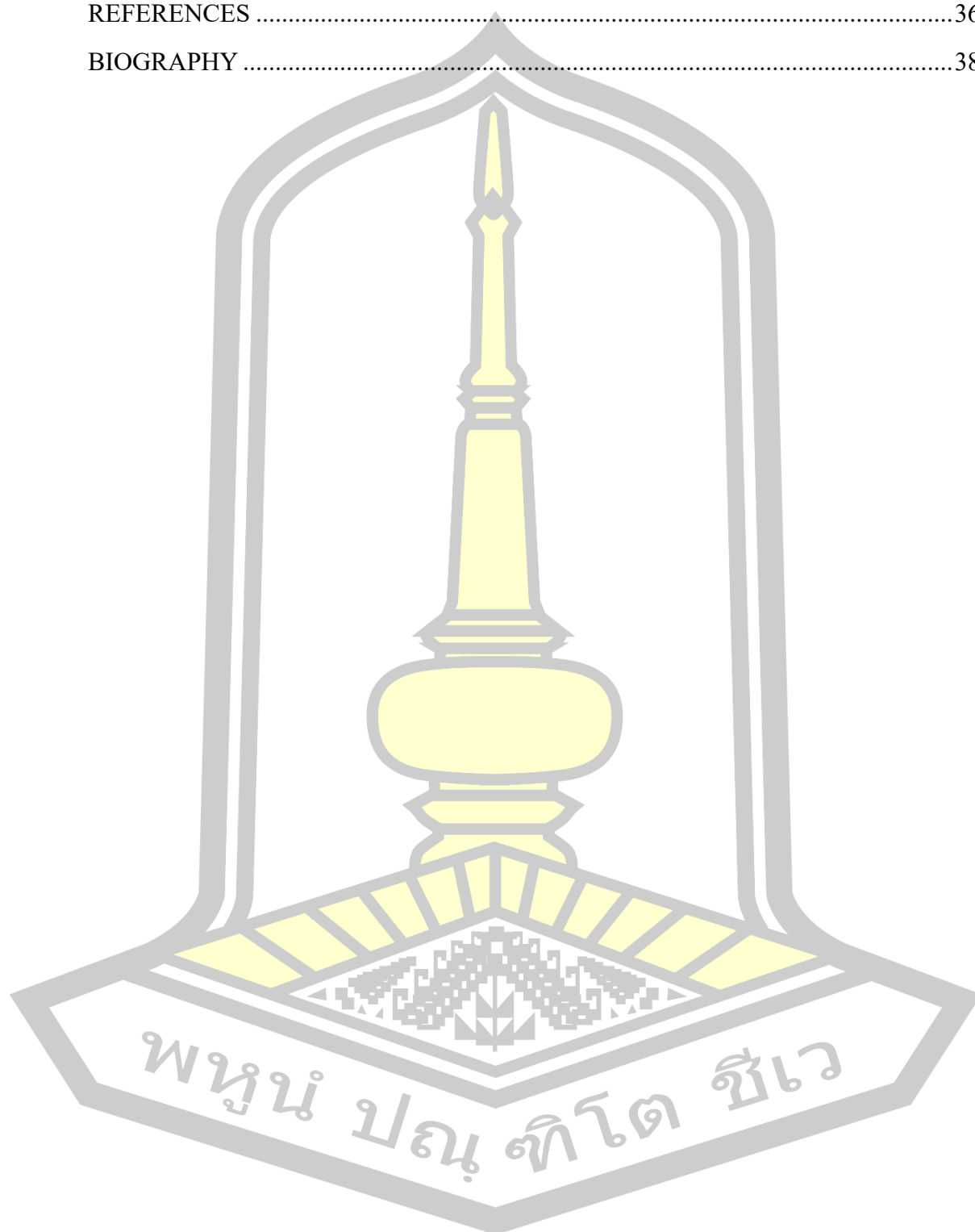
Mingze Lei



TABLE OF CONTENTS

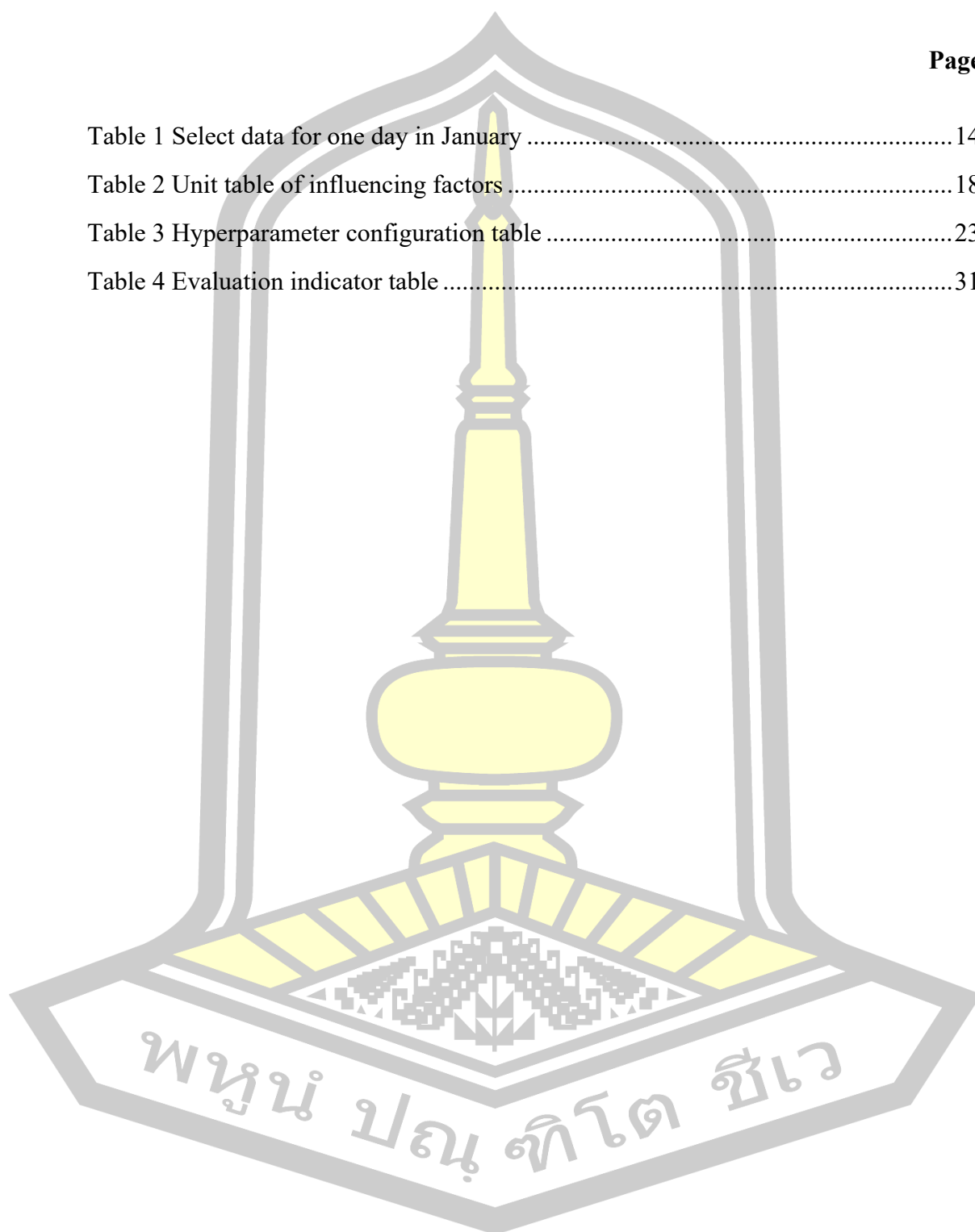
	Page
ABSTRACT.....	D
ACKNOWLEDGEMENTS.....	E
TABLE OF CONTENTS	F
List of tables.....	H
List of figures.....	I
Chapter 1: Introduction	1
1.1 Background and significance	1
1.2 Current research status at home and abroad	5
1.3 Chapter arrangement.....	7
Chapter 2 : Correlation Model Theory.....	8
2.1 Support Vector Machine	8
2.2 Convolutional Neural Networks	9
2.3 Long Short-Term Memory Neural Network.....	10
Chapter 3: Data Analysis	12
3.1 Data Source.....	12
3.2 Dataset Division.....	18
3.3 Data preprocessing.....	18
3.4 Data Correlation Analysis.....	20
3.5 Hyperparameter Configuration.....	22
3.6 Evaluation Criteria.....	23
Chapter 4: Results and Discussion.....	25
4.1 Introduction.....	25
4.2 Results Analysis.....	26
4.3 Model performance comparison experiment	30
Chapter 5: Conclusion.....	33

REFERENCES	35
REFERENCES	36
BIOGRAPHY	38



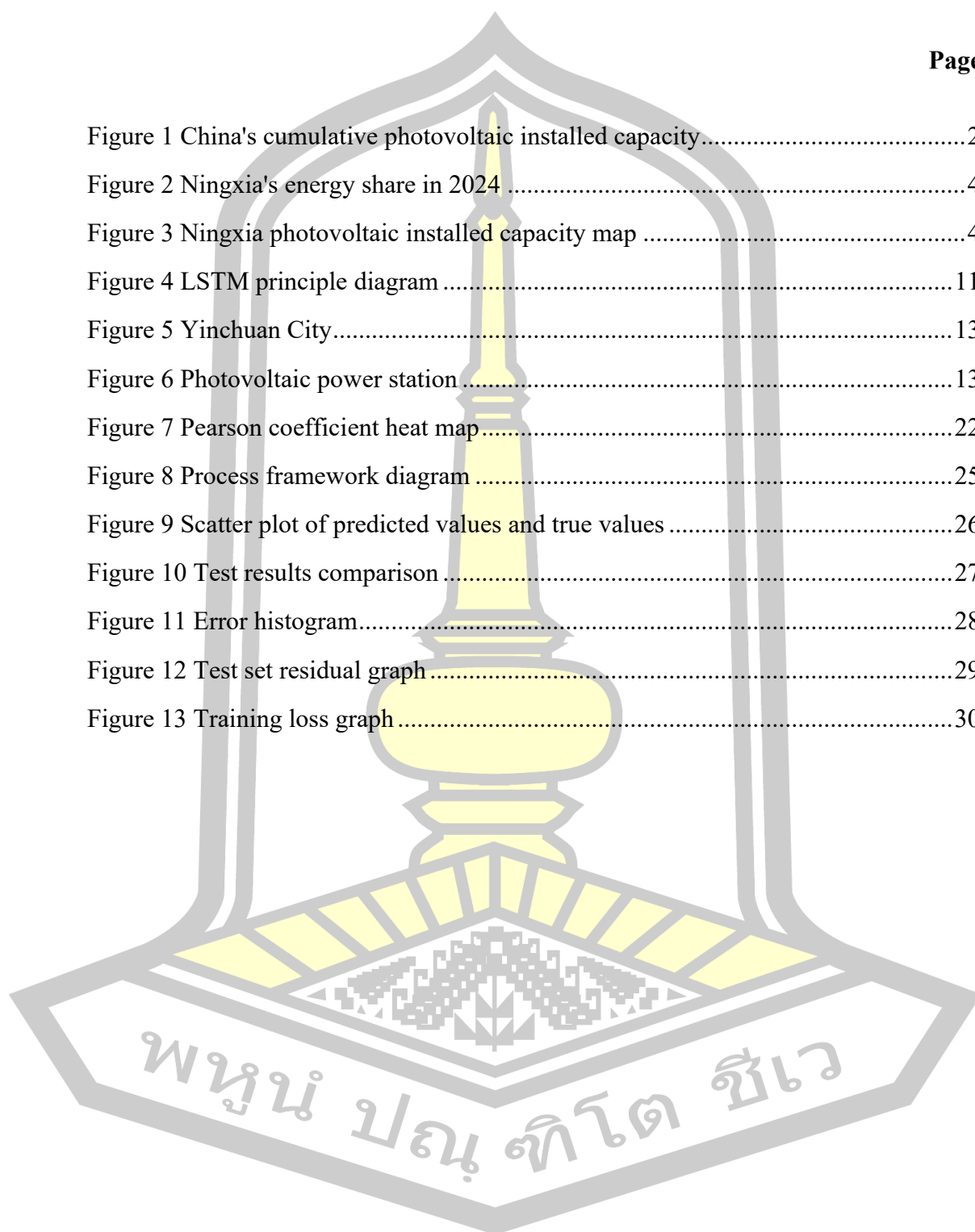
List of tables

	Page
Table 1 Select data for one day in January	14
Table 2 Unit table of influencing factors	18
Table 3 Hyperparameter configuration table	23
Table 4 Evaluation indicator table	31



List of figures

	Page
Figure 1 China's cumulative photovoltaic installed capacity.....	2
Figure 2 Ningxia's energy share in 2024	4
Figure 3 Ningxia photovoltaic installed capacity map	4
Figure 4 LSTM principle diagram	11
Figure 5 Yinchuan City.....	13
Figure 6 Photovoltaic power station.....	13
Figure 7 Pearson coefficient heat map	22
Figure 8 Process framework diagram	25
Figure 9 Scatter plot of predicted values and true values	26
Figure 10 Test results comparison	27
Figure 11 Error histogram.....	28
Figure 12 Test set residual graph	29
Figure 13 Training loss graph.....	30



Chapter 1: Introduction

1.1 Background and significance

In the past, fossil fuels such as oil, coal, and natural gas have always been the main energy sources to meet the growing world energy demand. Due to the large amount of polluting gases produced by fossil fuels during power generation, the greenhouse effect and environmental pollution problems have been further aggravated. In recent years, environmental protection, energy conservation and emission reduction have received increasing attention from countries around the world. Wind power and photovoltaic power have been vigorously promoted around the world due to their unique advantages such as clean and pollution-free, huge development potential, and recyclable use⁰. Photovoltaic power generation technology has gradually replaced traditional energy generation, which has not only greatly reduced greenhouse gas emissions, but also made full use of solar energy resources. In recent years, solar photovoltaic power generation has risen rapidly, and various countries have introduced measures to encourage the development of solar photovoltaic power generation technology, which has promoted the development of solar photovoltaic power generation [2]. From 2004 to 2024, the global new energy has developed rapidly. Photovoltaic power generation has increased by more than 400 times, occupying an important position in power supply. In the long run, solar photovoltaic power generation has transitioned from supplementary energy to alternative energy, and the proportion of photovoltaic power generation in total power generation is increasing day by day. Although the scale of photovoltaic power generation has grown rapidly, and the provincial and municipal governments in my country have strongly supported the construction of photovoltaic power stations, there is a lack of reasonable allocation and full utilization of photovoltaic power, resulting in a waste of construction funds and production equipment. The main reason is that photovoltaic power generation is subject to the influence of natural conditions and weather factors. It is volatile, random and intermittent . When a large-scale unstable photovoltaic system is connected to the power grid, it will bring challenges to the stable operation of the power grid system. In order to cope with the negative impact of photovoltaic grid connection and optimize

the operation of the power grid, it is necessary to establish a photovoltaic power prediction system[3].

The significance of accurately predicting photovoltaic power is as follows: (1) it can reduce the grid connection challenges caused by photovoltaic power fluctuations, assist dispatchers in formulating power dispatch plans, and improve the stability of power grid operation; (2) it is conducive to the grid to reasonably arrange the energy ratio of conventional energy and photovoltaic power generation, and enhance photovoltaic consumption; (3) it enables photovoltaic power generation companies to make full use of solar energy resources, avoid photovoltaic resource waste, and achieve greater economic benefits and social benefits of low-carbon green development. Therefore, the development of photovoltaic power prediction technology is of great significance for improving the utilization rate of photovoltaic power generation in daily power grids and increasing photovoltaic power generation.

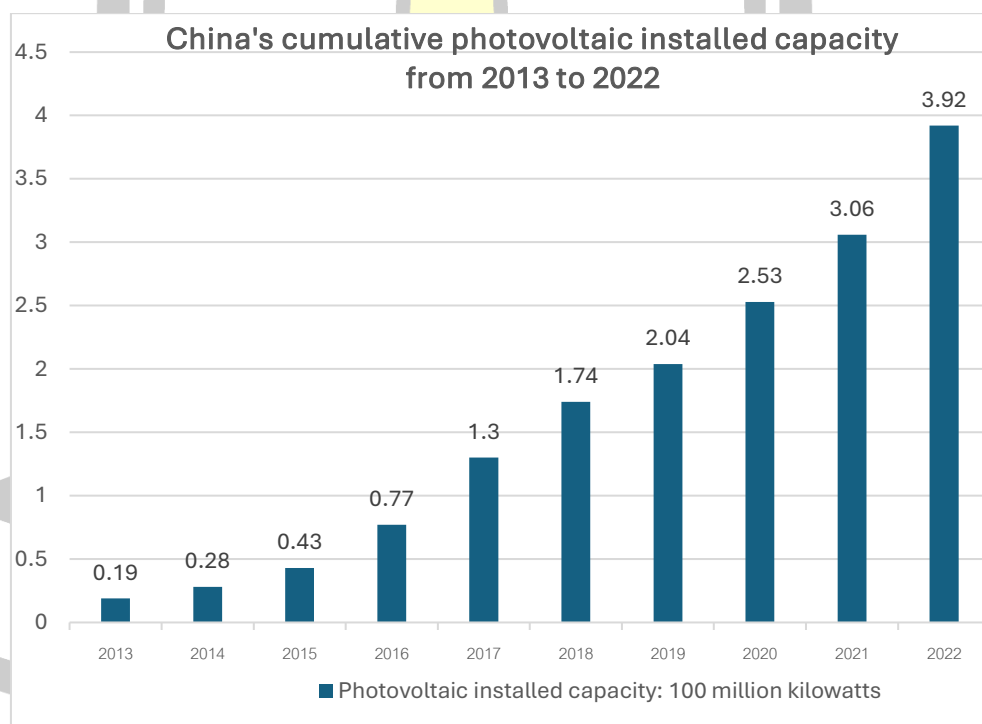
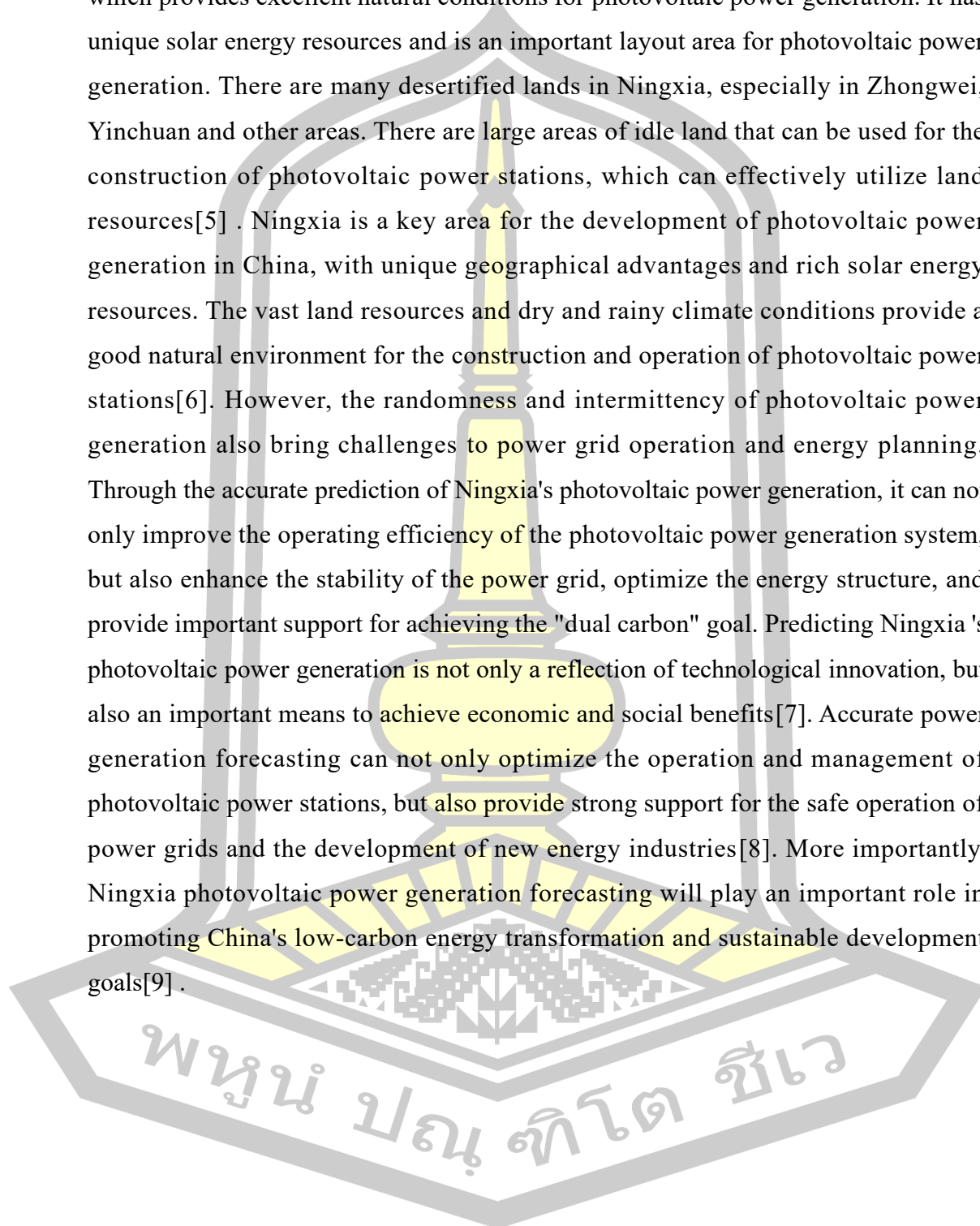


Figure 1 China's cumulative photovoltaic installed capacity

Ningxia is located in the northwest of China, with flat terrain and rich sunshine resources[4]. It is a typical solar energy resource category I region. The annual average

sunshine hours reach about 3,000 hours, and the solar radiation intensity is high, which provides excellent natural conditions for photovoltaic power generation. It has unique solar energy resources and is an important layout area for photovoltaic power generation. There are many desertified lands in Ningxia, especially in Zhongwei, Yinchuan and other areas. There are large areas of idle land that can be used for the construction of photovoltaic power stations, which can effectively utilize land resources[5]. Ningxia is a key area for the development of photovoltaic power generation in China, with unique geographical advantages and rich solar energy resources. The vast land resources and dry and rainy climate conditions provide a good natural environment for the construction and operation of photovoltaic power stations[6]. However, the randomness and intermittency of photovoltaic power generation also bring challenges to power grid operation and energy planning. Through the accurate prediction of Ningxia's photovoltaic power generation, it can not only improve the operating efficiency of the photovoltaic power generation system, but also enhance the stability of the power grid, optimize the energy structure, and provide important support for achieving the "dual carbon" goal. Predicting Ningxia's photovoltaic power generation is not only a reflection of technological innovation, but also an important means to achieve economic and social benefits[7]. Accurate power generation forecasting can not only optimize the operation and management of photovoltaic power stations, but also provide strong support for the safe operation of power grids and the development of new energy industries[8]. More importantly, Ningxia photovoltaic power generation forecasting will play an important role in promoting China's low-carbon energy transformation and sustainable development goals[9].



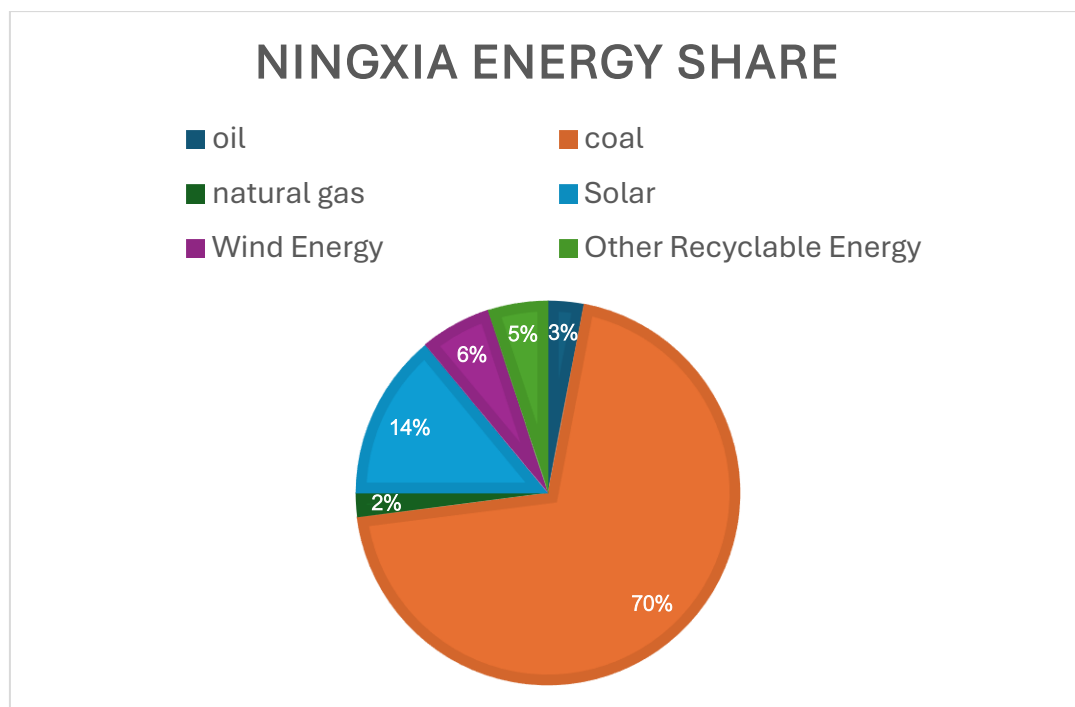


Figure 2 Ningxia's energy share in 2024

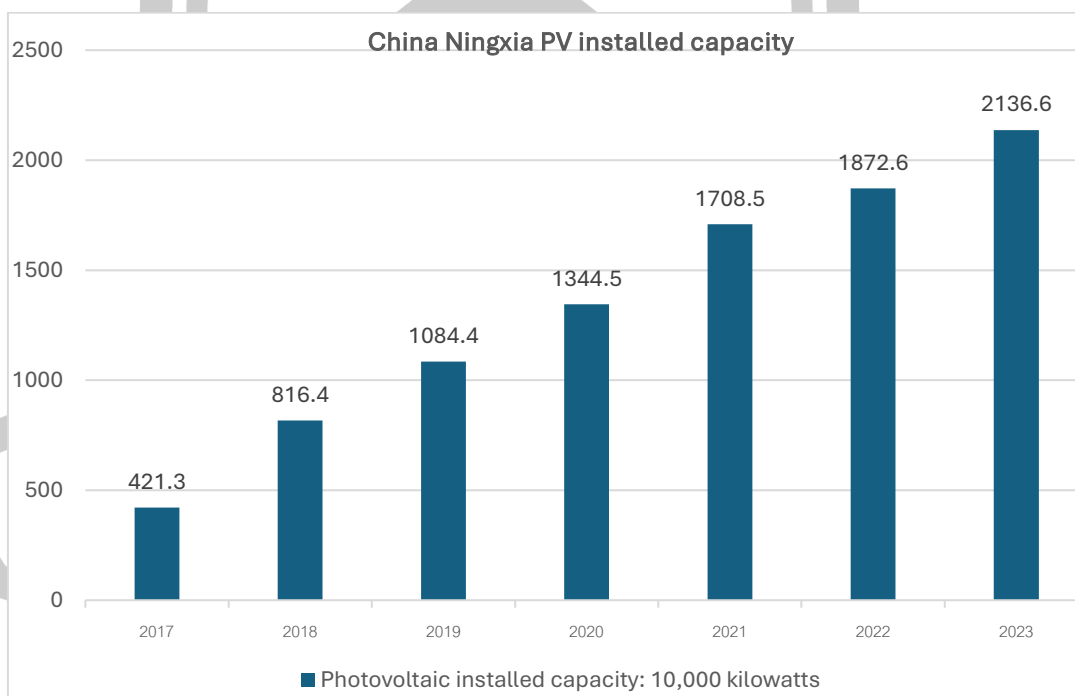


Figure 3 Ningxia photovoltaic installed capacity map

LSTM is a deep learning algorithm that specifically processes time series data. It can capture dependencies and trends in long time series and is particularly suitable for processing the complexity of photovoltaic power generation data. Through LSTM prediction, the prediction accuracy of short-term and ultra-short-term photovoltaic power generation can be significantly improved, providing a reliable basis for the

dispatch of power systems. Traditional methods (such as ARIMA, support vector machines, etc.) have limitations in dealing with the nonlinearity and randomness of

photovoltaic power generation, while the memory capacity and nonlinear modeling advantages of LSTM significantly improve the depth and breadth of research. Based on the LSTM algorithm, multi-source data (such as weather forecast data, historical power data, etc.) can be introduced to build a multi-input model to achieve more accurate predictions. Therefore, studying photovoltaic power generation based on the LSTM algorithm can not only improve the accuracy of power generation predictions and solve the shortcomings of traditional prediction methods, but also optimize power grid dispatch, promote the consumption of new energy, and promote the intelligent upgrade of energy systems. This is not only of great significance to the scientific research of photovoltaic power generation, but also has a far-reaching impact on achieving the "dual carbon" goal and building a green energy system.

1.2 Current research status at home and abroad

Research on LSTM in photovoltaic power generation prediction. Foreign scholars have earlier applied LSTM to short-term and ultra-short-term power prediction of photovoltaic power generation. For example, researchers in the United States and Germany used historical data and meteorological data of photovoltaic power stations to achieve high-precision prediction through LSTM models, especially showing advantages in dealing with long-term nonlinear problems. Some studies combined weather forecast data (such as cloud coverage, irradiance and temperature) and further improved the prediction accuracy by improving the LSTM network structure. For example, an Australian research team built a meteorological-assisted

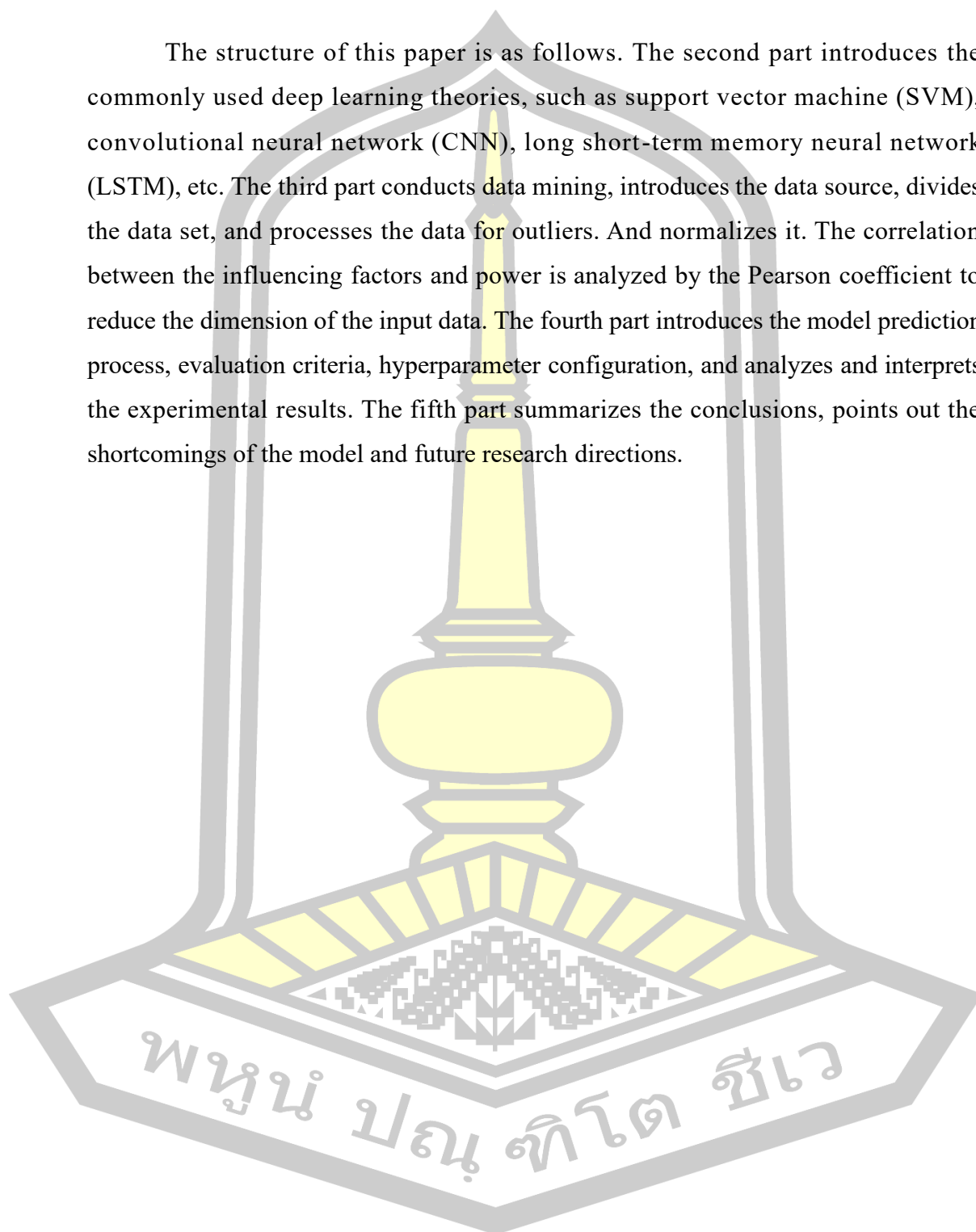
photovoltaic power generation prediction model based on LSTM, which significantly improved the power prediction effect under different weather conditions. Foreign researchers improved the traditional LSTM model and introduced bidirectional LSTM to better capture the contextual information in time series data. For example, some studies in the United Kingdom pointed out that BiLSTM performs better than unidirectional LSTM in photovoltaic power generation prediction.

Domestic scholars have widely applied LSTM networks to address the uncertainty and complexity of photovoltaic power generation prediction. For example, institutions such as the China Electric Power Research Institute and Tsinghua University have proposed a short-term photovoltaic power generation prediction method based on LSTM, which was verified by combining regional meteorological data to improve the prediction accuracy. Domestic research focuses more on photovoltaic power generation prediction in typical regions of China (such as Northwest China, North China and East China). For example, in high-sunshine areas such as Ningxia and Xinjiang, LSTM-based prediction models are studied to optimize regional power dispatch plans. In terms of optimizing hyperparameters, domestic researchers use genetic algorithms (GA) and particle swarm optimization (PSO) to optimize the hyperparameters of the LSTM model, thereby improving the model's convergence speed and prediction accuracy. Hybrid model: The combination of LSTM and other machine learning algorithms (such as CNN, support vector regression SVR, etc.) is a hot topic in domestic research. For example, Harbin Institute of Technology proposed a photovoltaic power prediction method based on a CNN-LSTM hybrid model, using CNN to extract spatial features and LSTM to process time series information.

Domestic and foreign research has shown that the LSTM algorithm has significant advantages in photovoltaic power generation forecasting. Foreign research is cutting-edge in model improvement and theoretical exploration, while domestic research focuses more on regional characteristics and practical engineering applications. In the future, with the development of new energy technologies, research based on LSTM will further promote the intelligent application of photovoltaic power generation in the energy Internet.

1.3 Chapter arrangement

The structure of this paper is as follows. The second part introduces the commonly used deep learning theories, such as support vector machine (SVM), convolutional neural network (CNN), long short-term memory neural network (LSTM), etc. The third part conducts data mining, introduces the data source, divides the data set, and processes the data for outliers. And normalizes it. The correlation between the influencing factors and power is analyzed by the Pearson coefficient to reduce the dimension of the input data. The fourth part introduces the model prediction process, evaluation criteria, hyperparameter configuration, and analyzes and interprets the experimental results. The fifth part summarizes the conclusions, points out the shortcomings of the model and future research directions.



Chapter 2 : Correlation Model Theory

2.1 Support Vector Machine

At present, the main directions of photovoltaic power prediction include cloud image recognition technology based on computer vision, probabilistic prediction and machine learning models. Cloud image recognition technology predicts photovoltaic power by analyzing sky cloud images, but the cost of cloud image recognition equipment is currently high. Probabilistic prediction takes into account the uncertainty of environmental factors and describes the prediction results through Monte Carlo simulation, Bayesian reasoning or probability distribution fitting. However, probabilistic prediction requires more computing resources for feature selection, feature engineering and dimensionality reduction, and has a great demand for computing resources. On the other hand, machine learning models provide the flexibility to flexibly select the complexity and number of sub-models according to specific applications, and have relatively low computing resource requirements. This flexibility has attracted great attention from researchers in recent years.

Support vector machine is a basic machine learning model and a supervised learning algorithm, which is widely used in classification and regression problems.[19] Its core idea is to find a hyperplane to maximize the interval between samples of different categories, thereby achieving data separation or fitting.

It has good performance for small samples and high-dimensional data. It can handle nonlinear problems, and with the help of kernel functions such as radial basis functions and Sigmoid, it can transform the training data set into a higher-dimensional space through nonlinear mapping[10]. With the advantages of not requiring a large amount of data and being stable and reliable, it is still popular in the field of time series prediction. However, SVM cannot automatically extract features, and the model performance is highly dependent on manually designed input features. If the input features are insufficient or inappropriate, the performance of SVM will drop significantly. Moreover, it is limited in high-dimensional complex scenarios, especially when the feature space of the problem is complex, the kernel function

selection may become difficult . Therefore, in the photovoltaic power prediction problem, the use of artificial neural network models may produce better results.

2.2 Convolutional Neural Networks

Deep learning provides a strong theoretical guidance for time series prediction tasks. Compared with machine learning models, deep learning can add more neural network unit layers and flexibly improve the network structure. These characteristics enable deep learning models to better solve complex tasks. As the amount of data increases, the prediction accuracy of deep learning models generally increases[12].

CNN is a type of deep learning model where the input layer is used to receive raw data. The convolutional layer equipped with convolutional filters performs a local spatial scan on this input data to extract basic local features. These collected features are then nonlinearly transformed using a nonlinear activation function to enhance the network's expressive power. The purpose of the pooling layer is to reduce the feature dimensionality and reduce the computational complexity of subsequent stages while retaining significant feature information. After several convolution-activation- pooling sequences , the architecture culminates in a fully connected layer that maps high-level features to the target output.[13] Due to its hierarchical and modular nature, CNN has also been successfully applied in different fields such as time series prediction.

The convolution layer is the core structure of the convolutional neural network. It performs weighted summation on the local feature matrix, that is, all elements in the local window are operated according to a certain rule, and then all local windows are traversed to perform the next layer of operation, and finally the local low-dimensional features are mapped to high-dimensional features. The convolution layer operation is generally accompanied by a pooling operation . Because adjacent features are usually similar, the pooling operation can prevent the model training from overfitting to a certain extent. After the pooling operation , the receptive field of the original input data will increase, and high-dimensional features can be extracted.[14]

The receptive field of CNN is limited, and it is necessary to increase the number of convolutional layers or pooling layers to expand the receptive field, which

may lead to gradient vanishing or gradient exploding problems . Therefore, CNN cannot process long-term time series data.

2.3 Long Short-Term Memory Neural Network

LSTM is a transformation based on the Recurrent Neural Network (RNN) structure. Its network structure has a gate mechanism and cell state. Therefore, LSTM has the function of saving and forgetting historical information. It has been widely used in the modeling of time series such as power load, stock prices, climate values and speech recognition[10]. LSTM is proposed to solve the dependencies between the previous and next time series, that is, the node information at the previous moment affects the node output at the next moment. Since the RNN node has been calculated for many stages, the characteristics of the previous time nodes have been covered[11]. LSTM adds memory gates, forget gates and cell states to the structure to adaptively save, forget and update hidden states, alleviate the problem of poor time series learning , and effectively alleviate the problem of gradient disappearance. It can capture long-term dependencies[11]. It has stronger modeling capabilities for nonlinear dynamic characteristics.

As shown in the figure, the internal mechanism of the LSTM unit adopts a gating mechanism. Specifically, the LSTM gating unit contains a forget gate f_t , an input gate i_t , and an output gate o_t . The forget gate f_t determines which information to discard or retain from the cell state. The input gate i_t controls the addition of new information, while the output gate o_t determines which part of the information based on the cell state should be passed to the hidden state. The cell state C_t acts as the "memory" of the network, retaining historical information. The associated computations can be found in Equation (1) through (3)

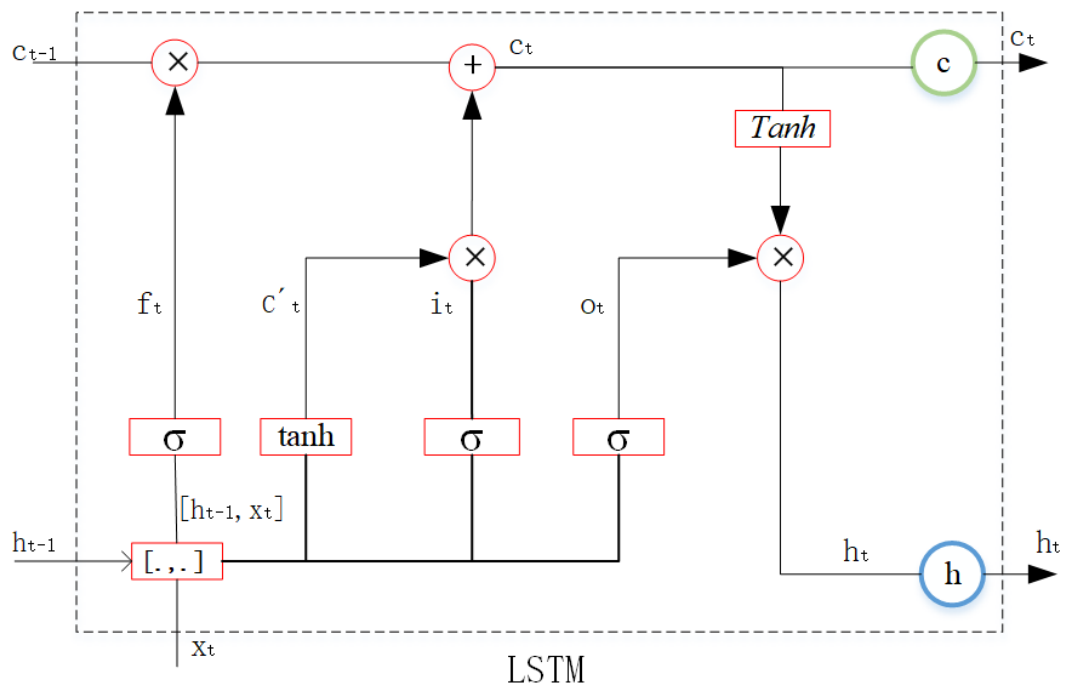
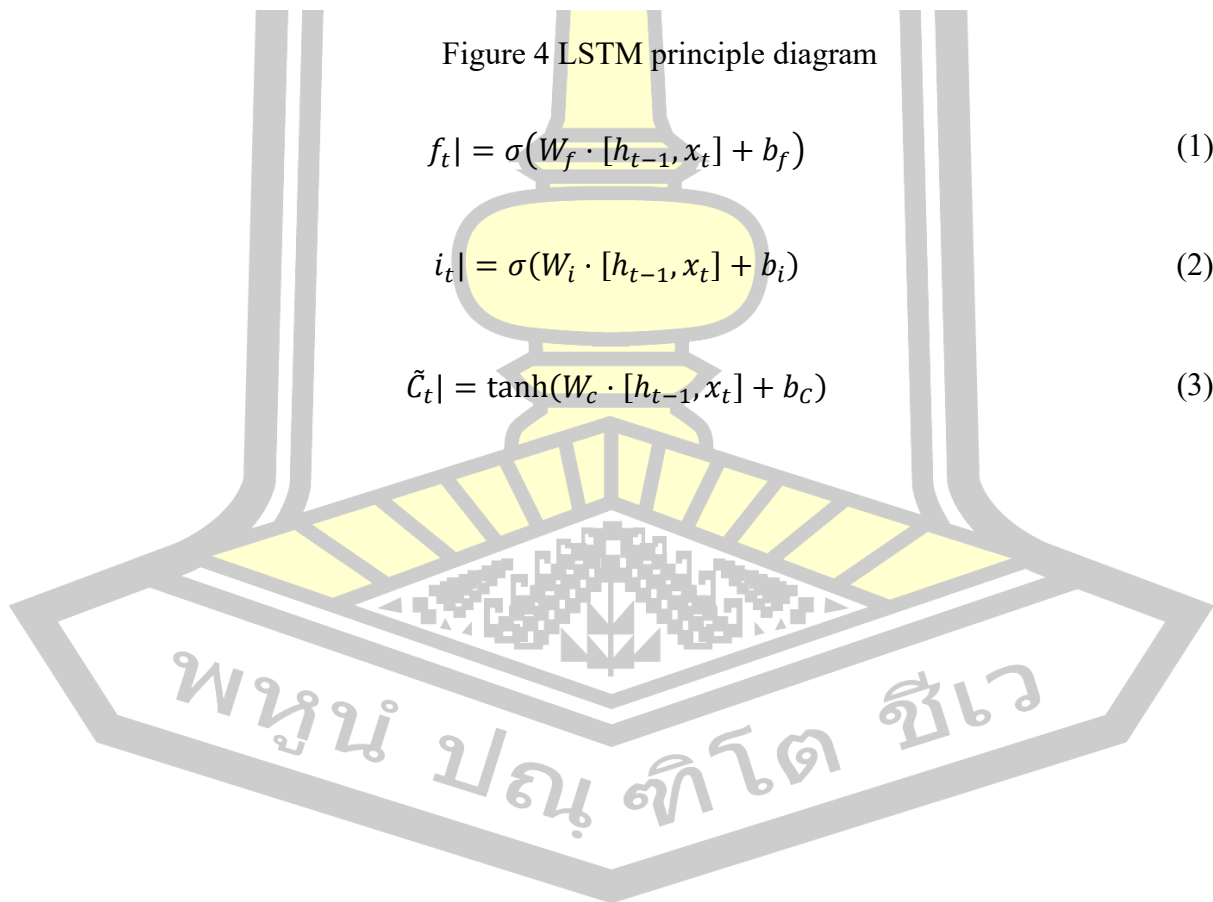


Figure 4 LSTM principle diagram

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$



Chapter 3: Data Analysis

The output power of photovoltaic power generation is determined by multiple factors, such as the influence of various conventional, accidental or sudden factors such as season, weather, solar irradiance, etc. The output power has certain uncertainty and periodic volatility, and belongs to a complex nonlinear time series. When predicting the output power, historical data of photovoltaic power generation and historical weather data are usually considered, and the randomness and volatility of the model itself are also considered. Therefore, this chapter mainly processes and analyzes the data features, and attempts to analyze the photovoltaic power generation data obtained from different angles to improve the prediction accuracy and stability of the model.

3.1 Data Source

All data in this article are derived from real data from a photovoltaic power plant in Ningxia, China. The data is from January 1 to December 31, 2023, with a time resolution of 15 minutes. The data features in this article include solar irradiance, temperature, relative humidity, cloud opacity, air pressure, wind speed, wind direction, and power generation. The power generation is taken as the output and other features are taken as input.

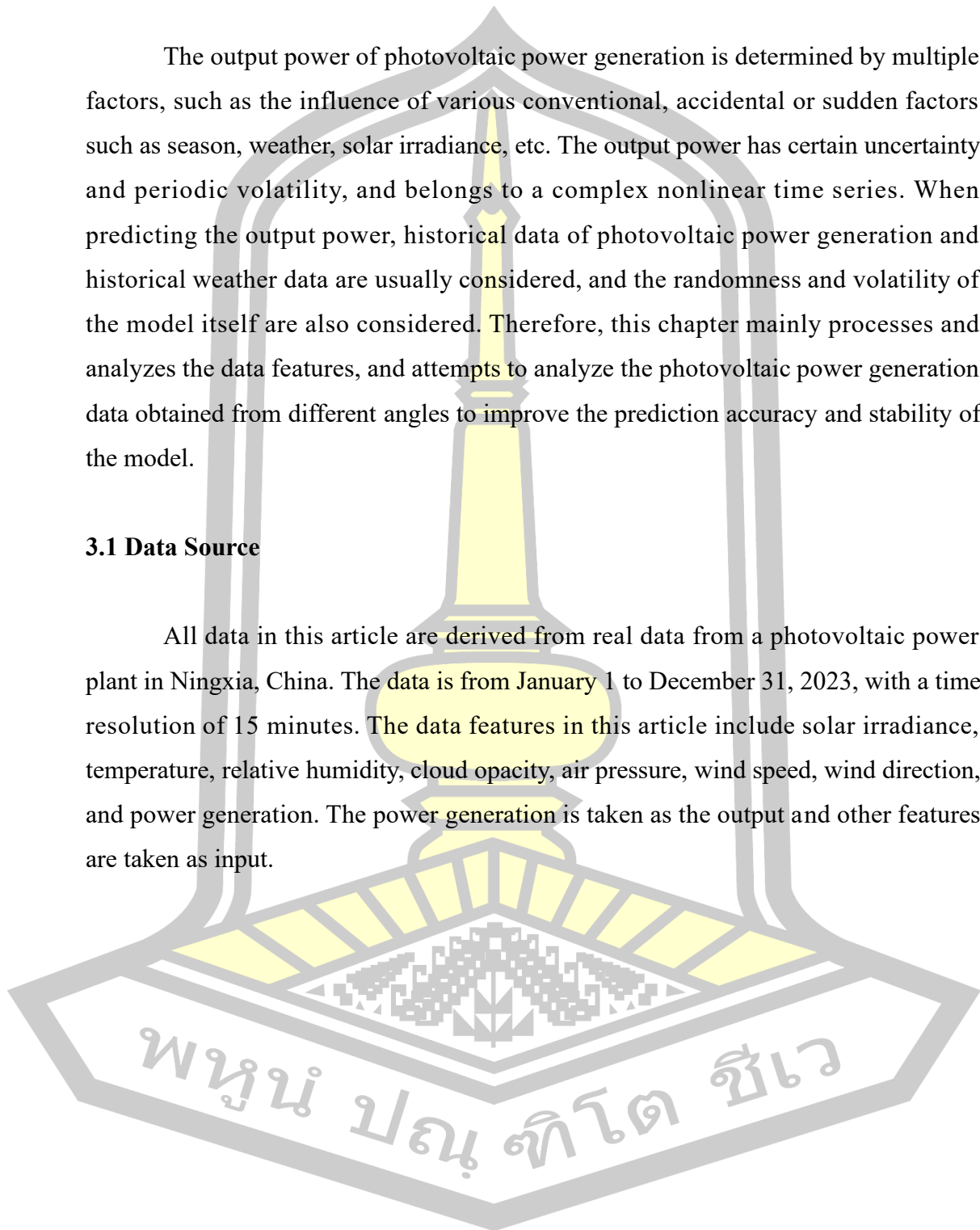




Figure 5 Yinchuan City



Figure 6 Photovoltaic power station

Table 1 Select data for one day in January

Temperature	Cloud Opacity	Irradiance	Relative Humidity	Humidity	Wind Direction	Wind Speed	Actual Power
-2.3	3.8	0	29.2	830.5	83	1.5	0
-2.6	15.8	0	29.7	830.5	90	1.5	0
-3	15.9	0	30.3	830.4	98	1.6	0
-3.3	15.9	0	30.9	830.4	106	1.6	0
-3.7	20.6	0	31.6	830.3	113	1.7	0
-4.1	35.2	0	32.2	830.2	119	1.8	0
-4.4	35.4	0	32.9	830.2	125	1.9	0
-4.8	35.4	0	33.6	830.1	130	2	0
-5.2	35.3	0	34.4	830	134	2.2	0
-5.5	35.2	0	35.1	830	138	2.3	0
-5.9	35.2	0	35.9	829.9	141	2.5	0
-6.3	35.2	0	36.7	829.9	144	2.7	0
-6.6	34.9	0	37.4	829.8	146	2.8	0
-6.9	35	0	38	829.7	147	3	0
-7.1	35	0	38.2	829.7	148	3	0
-7.3	35	0	38.5	829.6	148	3.1	0
-7.5	35	0	38.7	829.5	148	3.1	0
-7.7	35	0	39	829.4	148	3.2	0
-7.9	35.1	0	39.2	829.3	148	3.3	0
-8.2	35.6	0	39.5	829.3	149	3.3	0
-8.4	35.6	0	39.7	829.2	149	3.4	0
-8.6	35.6	0	40	829.1	149	3.5	0
-8.8	35.6	0	40.2	829	149	3.5	0
-9	35.6	0	40.5	829	149	3.6	0
-9.2	35.4	0	40.8	828.9	149	3.6	0

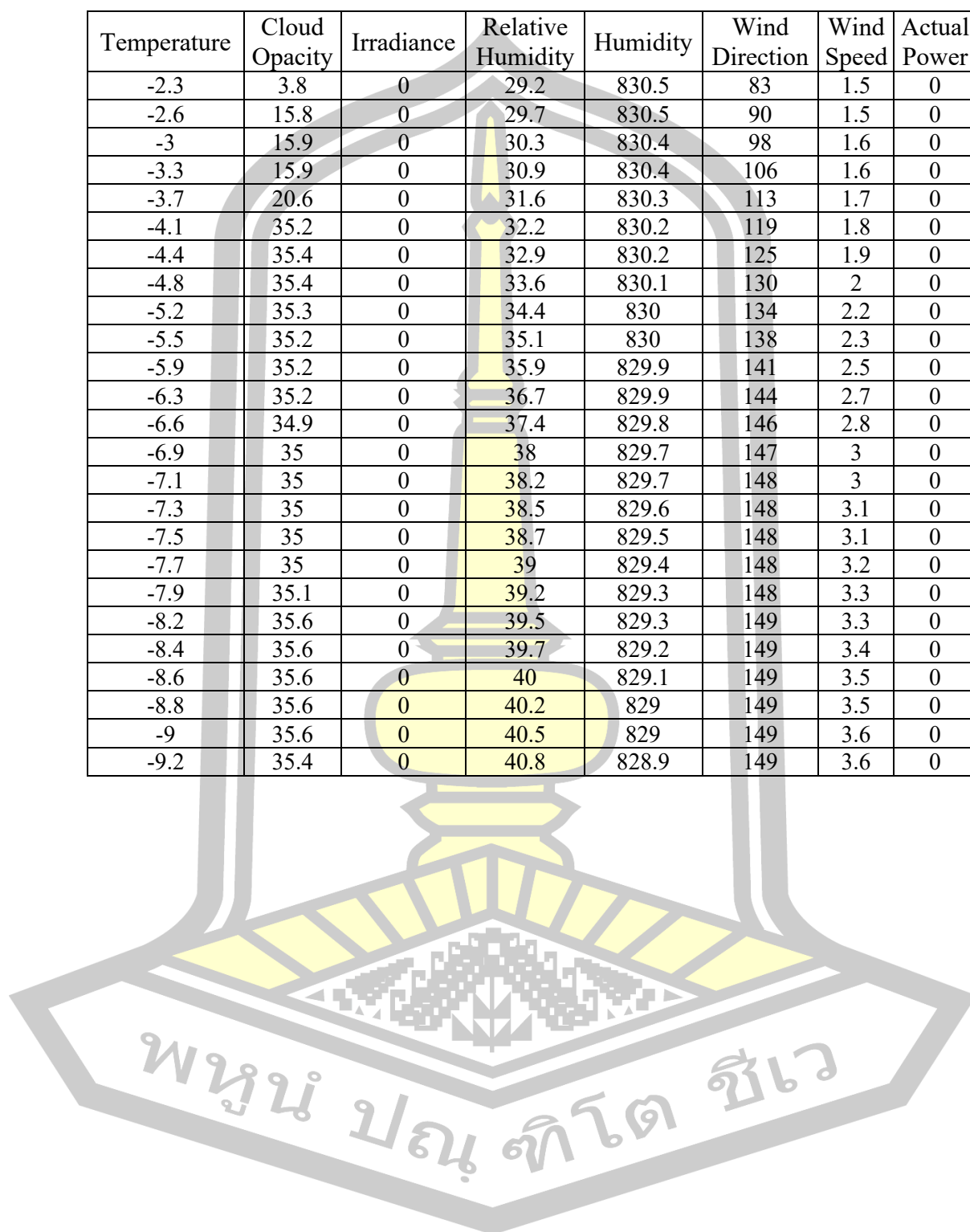


Table 1 Select data for one day in January (continue)

Temperature	Cloud Opacity	Irradiance	Relative Humidity	Humidity	Wind Direction	Wind Speed	Actual Power
-9.3	34.6	0	41	828.8	149	3.7	0
-9.3	34.6	0	41.3	828.8	149	3.7	0
-9.2	34.7	0	41.6	828.8	149	3.7	0
-9.2	34.8	0	41.9	828.8	148	3.7	0
-9.2	35.4	0	42.1	828.8	148	3.7	0
-9.2	35.8	0	42.4	828.8	148	3.7	0
-9.2	35.8	0	42.7	828.8	147	3.7	0
-9.1	35.8	0	43	828.8	147	3.7	0
-9.1	35.7	0	43.3	828.8	147	3.7	0
-9.1	36	0	43.5	828.8	146	3.7	0
-9.1	37.5	0	43.8	828.8	146	3.7	0
-9	41.5	0	44	828.8	146	3.7	0
-8.7	45.8	0	42.5	828.8	145	3.6	0
-7.9	46.9	1	39.5	828.9	144	3.5	0.12595
-7.2	44.7	13	36.7	828.9	143	3.3	1.63189
-6.5	40.4	35	34	829	142	3.2	4.37745
-5.7	37.3	64	31.6	829	141	3	8
-5	35.7	97	29.4	829.1	139	2.9	12.22103
-4.3	33.2	134	27.3	829.1	138	2.7	16.80092
-3.5	25.1	191	25.4	829.2	136	2.6	23.87882
-2.8	1.4	299	23.6	829.3	134	2.4	37.54543
-2.1	0	353	22	829.3	132	2.3	44.37916
-1.3	0	401	20.5	829.4	129	2.1	50.37362
-0.6	0	448	19.1	829.4	126	2	56.40768
0	0	491	18.2	829.4	121	1.9	61.62541



Table 1 Select data for one day in January (continue)

Temperature	Cloud Opacity	Irradiance	Relative Humidity	Humidity	Wind Direction	Wind Speed	Actual Power
0.5	0	532	17.8	829.3	114	1.8	66.64364
1	0	570	17.4	829.2	106	1.8	71.8086
1.5	0	605	17.1	829.1	98	1.7	75.94565
1.9	0	636	16.7	829	89	1.8	79.72896
2.4	0	663	16.4	828.9	82	1.8	83.3391
2.9	0	687	16	828.8	75	1.9	86.19789
3.4	0	706	15.7	828.7	68	2	88.29236
3.9	0	722	15.4	828.6	63	2.2	90.81316
4.4	0	733	15	828.5	58	2.3	91.8449
4.9	0	740	14.7	828.4	53	2.5	92.5222
5.3	0	743	14.4	828.3	50	2.7	93.41739
5.6	0	741	14.2	828.2	48	2.8	92.625
5.7	0	735	14	828.1	47	2.9	92.19105
5.8	0	725	13.8	827.9	46	2.9	91.06725
5.9	0	711	13.6	827.8	46	3	89.35848
6	0	692	13.4	827.7	45	3	87.17124
6.1	0	670	13.2	827.6	44	3.1	84.0984
6.2	0	644	13	827.5	44	3.1	80.73184
6.3	0	614	12.8	827.3	43	3.2	77.33944
6.4	0	580	12.6	827.2	42	3.2	72.9118
6.5	0	544	12.4	827.1	42	3.3	68.32096
6.6	1.7	495	12.2	827	41	3.3	62.04825
6.7	1.7	453	12.1	826.9	41	3.4	56.99646
6.6	4.1	398	12.1	826.8	41	3.4	49.84154
6.3	5.6	347	12.4	826.8	42	3.4	43.55891

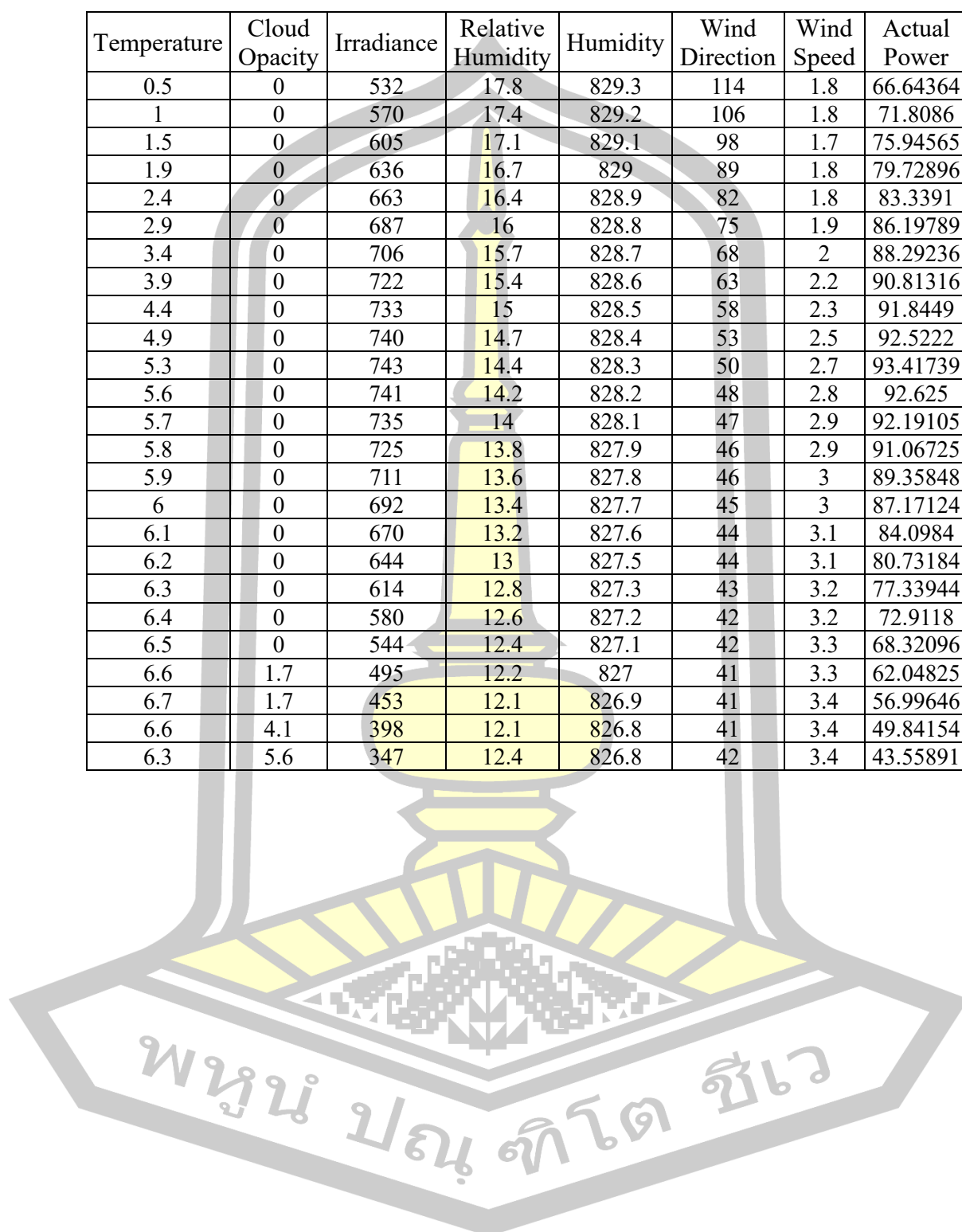


Table 1 Select data for one day in January (continue)

Temperature	Cloud Opacity	Irradiance	Relative Humidity	Humidity	Wind Direction	Wind Speed	Actual Power
6	1.4	314	12.7	826.8	44	3.4	39.44468
5.6	1.3	265	13	826.8	45	3.4	33.20185
5.3	10.4	195	13.3	826.8	46	3.4	24.41205
5	32.6	114	13.7	826.8	47	3.4	14.28192
4.7	42.3	66	14	826.9	49	3.3	8.25858
4.3	35.5	44	14.3	826.9	50	3.3	5.51408
4	5.9	28	14.7	826.9	51	3.3	3.52212
3.7	0	5	15	826.9	52	3.3	0.6282
3.4	0	0	15.4	826.9	54	3.3	0
3	0	0	15.8	826.9	55	3.3	0
2.6	0	0	16.3	826.9	58	3.3	0
2.2	0	0	16.9	826.8	62	3.3	0
1.7	0	0	17.6	826.7	67	3.3	0
1.2	0	0	18.2	826.6	72	3.3	0
0.8	0	0	19	826.6	76	3.3	0
0.3	0	0	19.7	826.5	80	3.4	0
-0.1	0	0	20.5	826.4	84	3.5	0
-0.6	0	0	21.3	826.3	88	3.6	0
-1.1	0	0	22.1	826.3	92	3.7	0
-1.5	0	0	23	826.2	96	3.8	0
-2	0	0	23.9	826.1	99	4	0
-2.3	3.8	0	29.2	830.5	83	1.5	0
-2.6	15.8	0	29.7	830.5	90	1.5	0
-3	15.9	0	30.3	830.4	98	1.6	0

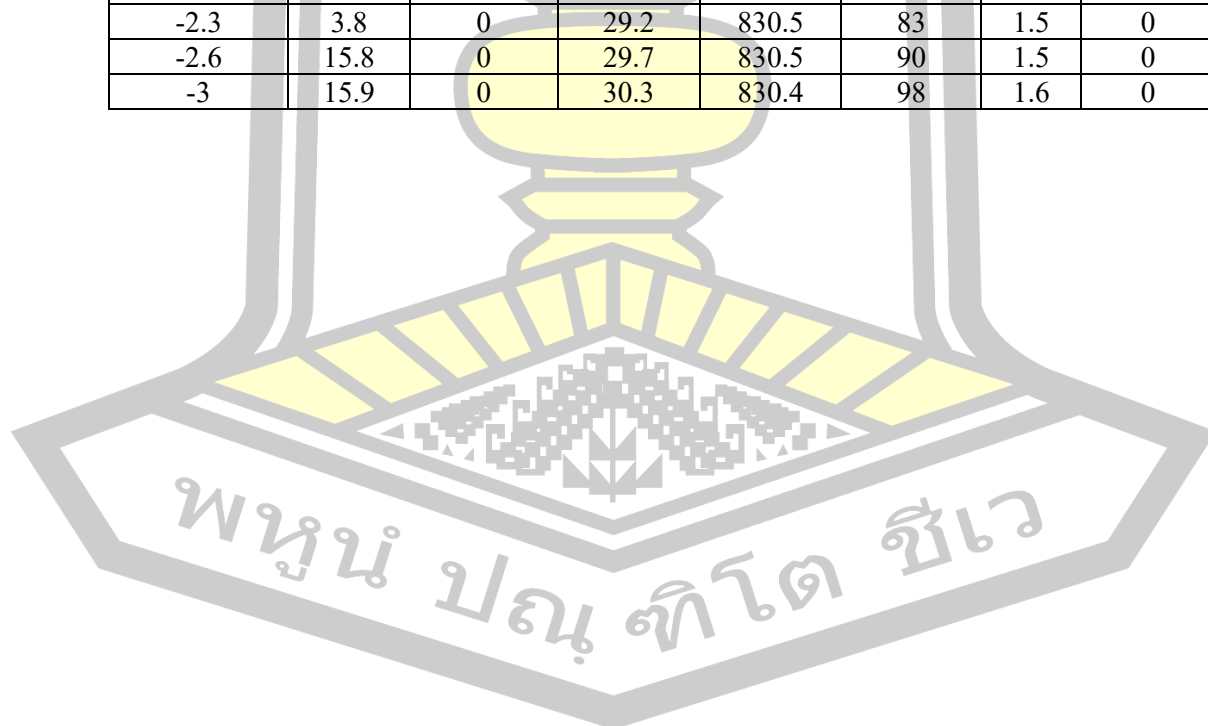


Table 2 Unit table of influencing factors

Variable Name	Unit
Temperature	°C
Cloud Opacity	%
Irradiance	W/m ²
Relative Humidity	%
Humidity	hPa
Wind Direction	°
Wind Speed	m/s
Actual Power	MW

3.2 Dataset Division

In photovoltaic power generation forecasting, data has significant seasonal characteristics. In summer, the irradiance is high and the power generation is large. In winter, the irradiance is low and the power generation is small. Therefore, the latter part of the data may correspond to winter or special periods with large weather changes. Whether the model can adapt to these changes is an important test goal. In addition, an important feature of time series data is that there is time dependence between data points, and subsequent data will be affected by previous data. Based on the above characteristics, this paper takes the first 80% of the data as the training set and the last 20% as the test set. This division method can improve the practicality of the model and ensure the robustness of the model to unknown data.

3.3 Data preprocessing

Photovoltaic data is the core of prediction and the cornerstone of ensuring prediction accuracy. Therefore, before conducting experiments, the original data must be preprocessed to ensure that the data is correct. Therefore, data outlier processing and data normalization operations are very important.

First, all photovoltaic data will be corrected for outliers. Outliers may be caused by sensor failure or manual recording, and usually deviate from the true value. In the field of photovoltaic power generation prediction, there are generally two types of outliers: data that violates logic. If outliers are not corrected or deleted, the results obtained by extreme value-sensitive algorithms or activation functions (mean absolute error) will be very poor. Therefore, outlier processing is very important. In this paper, the Z-score method (Z-Score) is used to detect outliers. The Z-score (the degree of deviation after standardization) of each data point is calculated. According to the properties of the normal distribution, most data points should fall near the mean. When the absolute value of the Z-score of a data point exceeds a certain threshold (usually 3), it is considered that the data point deviates too much from the mean and is an outlier.

$$Z = \frac{X - \mu}{\sigma} \quad (4)$$

Where x is the value of the data point. Where μ is the mean of the data. Where σ is The standard deviation of the data For the detected outliers.

After detecting the outliers, two methods were used to deal with them. Due to the large temperature difference between day and night in Ningxia, the irradiance and temperature outside 9:45-20:00 were 0, so they were directly deleted. The remaining outliers were uniformly replaced with the global mean.

Then the data is normalized. Through data normalization or standardization, the data of different units in the data set can be transformed dimensionlessly and mapped to the same scale for analysis and calculation. If the original photovoltaic power generation data is directly used for analysis, the data with larger dimensions may have a greater impact on the results, so it is necessary to balance the impact of different dimensions on the model performance. This article uses Min-Max normalization.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5)$$

Where X is the original feature; where X' is the regularized feature; where X_{max} is the maximum value of the input feature; where X_{min} is the minimum value of the input feature.

After normalization, the model is trained and the data obtained after training is also normalized data. The final predicted data needs to be compared with the actual data, which requires denormalization.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (6)$$

3.4 Data Correlation Analysis

The output power of photovoltaic power generation is affected by many external factors, such as solar irradiance, cloud cover, temperature, etc. However, the above characteristics have different effects on the output results of the model, so feature analysis is needed. In addition, although the small amount of data may cause the model to miss important information during the training process, too many input features will inevitably increase the complexity of the model, which will cause problems such as information redundancy. In order to make the model have better prediction performance and appropriately compress the input dimension of the input features, the photovoltaic data is analyzed for correlation. This paper uses the Pearson correlation coefficient to analyze the correlation between the input variables, solar radiation, temperature, cloud opacity, humidity, air pressure, wind speed, and wind direction, and the seven influencing factors and the power generation, and generates a heat map to intuitively display the correlation between the variables.

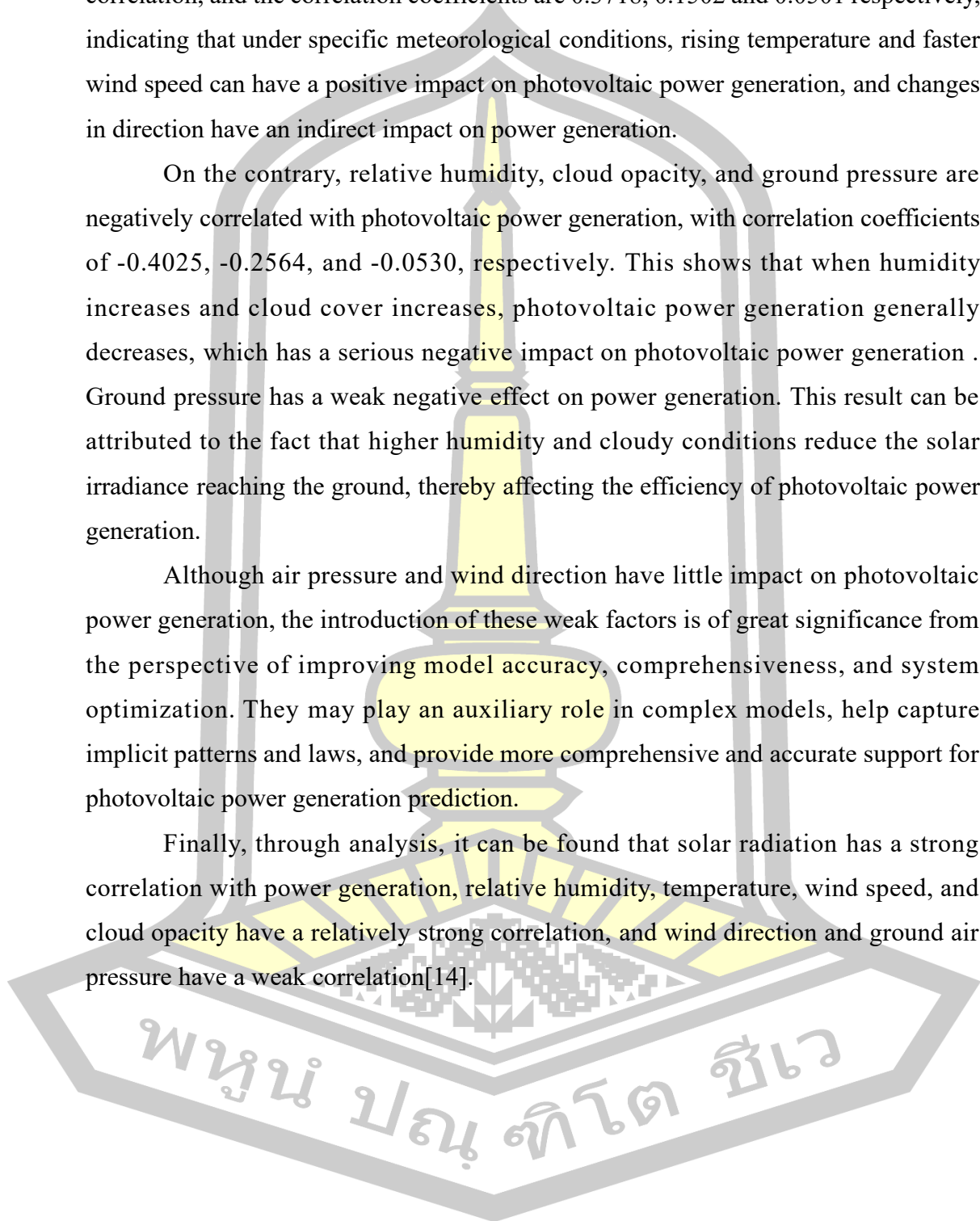
As shown in Figure 3, the correlation coefficients between various meteorological factors and photovoltaic power generation are shown through a heat map. From the data, it can be seen that solar radiation has a significant positive correlation with power generation, with coefficients of 0.8345. This highlights a concept: the increase in irradiance is conducive to a larger energy pool for photovoltaic cells to convert into electrical energy, thereby increasing power generation.

In addition, the temperature, wind speed and wind direction show a positive correlation, and the correlation coefficients are 0.3718, 0.1502 and 0.0501 respectively, indicating that under specific meteorological conditions, rising temperature and faster wind speed can have a positive impact on photovoltaic power generation, and changes in direction have an indirect impact on power generation.

On the contrary, relative humidity, cloud opacity, and ground pressure are negatively correlated with photovoltaic power generation, with correlation coefficients of -0.4025, -0.2564, and -0.0530, respectively. This shows that when humidity increases and cloud cover increases, photovoltaic power generation generally decreases, which has a serious negative impact on photovoltaic power generation. Ground pressure has a weak negative effect on power generation. This result can be attributed to the fact that higher humidity and cloudy conditions reduce the solar irradiance reaching the ground, thereby affecting the efficiency of photovoltaic power generation.

Although air pressure and wind direction have little impact on photovoltaic power generation, the introduction of these weak factors is of great significance from the perspective of improving model accuracy, comprehensiveness, and system optimization. They may play an auxiliary role in complex models, help capture implicit patterns and laws, and provide more comprehensive and accurate support for photovoltaic power generation prediction.

Finally, through analysis, it can be found that solar radiation has a strong correlation with power generation, relative humidity, temperature, wind speed, and cloud opacity have a relatively strong correlation, and wind direction and ground air pressure have a weak correlation[14].



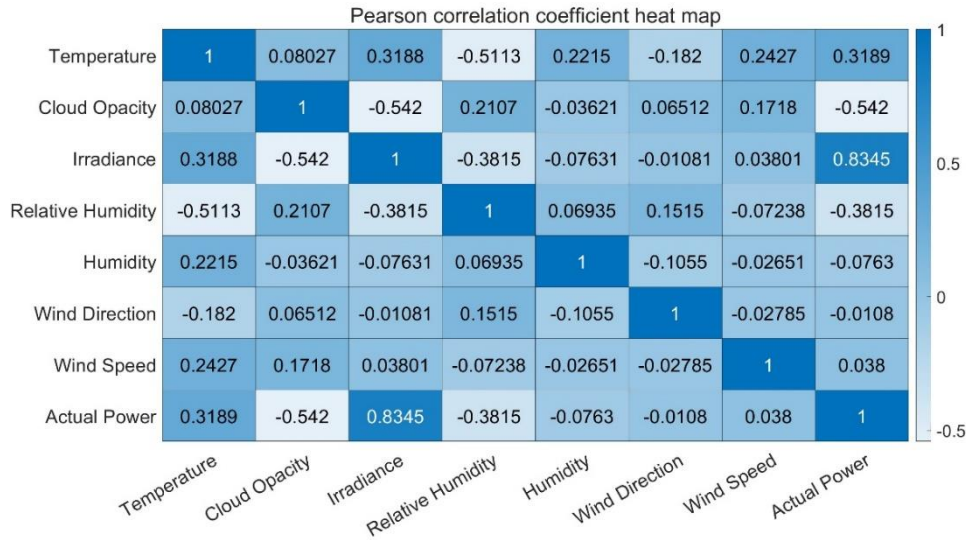


Figure 7 Pearson coefficient heat map

3.5 Hyperparameter Configuration

All program implementations are executed in MATLAB R2024b, and the main parameters of all models are listed in Table 1. Studies have shown that ReLU (Rectified Linear Unit) can enhance the nonlinearity of the network, promote model generalization and improve computational efficiency [20]. The Adam optimizer can adjust the learning rate of each parameter, which helps to smooth the optimization surface and speed up model training [23]. Therefore, ReLU is selected as the activation function and Adam as the optimizer.

Using fewer neurons can reduce model complexity and avoid model overfitting, so the number of neurons is set to 16. At the same time, the maximum number of iterations is set to 50 to prevent model overfitting [4]. In addition, all models are trained using mean square error as the loss function, and the batch size is selected as 320 to balance computational efficiency and hardware resources during training. Larger batches can more effectively utilize GPU to accelerate training. The gradient threshold is set to 0.1 to prevent large jumps in parameters due to large gradients during gradient updates. Therefore, the restriction on gradient changes is strong, which helps to maintain training stability in complex networks. The initial learning rate is set to 0.1 to quickly find the direction of loss reduction at the beginning

of model training. The learning rate is set to 0.0015 , and the step size of parameter updates is gradually reduced in the later stage of training to refine model parameters and improve model accuracy.

Table 3 Hyperparameter configuration table

Parameter	Value	Parameter	Value
MiniBatchSize	320	MaxEpochs	50
activation	ReLU	GradientThreshold	0.1
units	16	InitialLearnRate	0.01
Optimizer	Adam	LearningRate	0.0015

3.6 Evaluation Criteria

In order to more comprehensively evaluate the effectiveness and accuracy of the photovoltaic power generation prediction model, this study uses three main indicators: MAE (stability), RMSE (accuracy) and R2 (fit). These indicators provide accurate standards for the model's prediction.

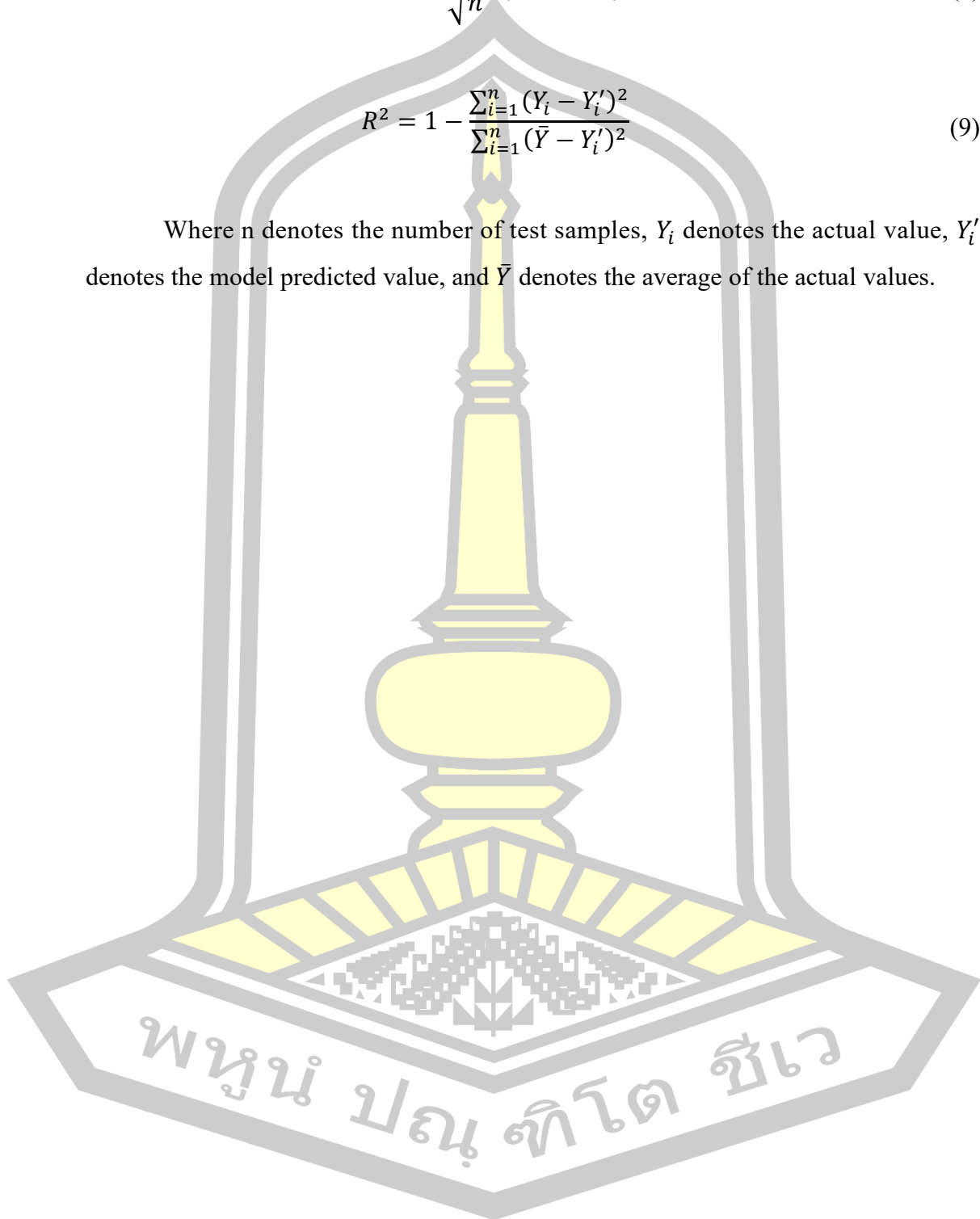
MAE is the average of the absolute errors between the predicted value and the true value . RMSE root mean square error is the square root of the square error between the predicted value and the true value . It reflects the actual difference between the predicted value and the observed value, and directly quantifies the prediction accuracy . The R2 coefficient indicates the degree of fit between the predicted value and the true value, and describes the ability of the prediction model to fit the actual data curve . The smaller the values of MAE and RMSE, the better the prediction performance. The larger the R2 value, the better the result, and the value range is [0,1]. The principles of the relevant indicator formulas are shown in (7)-(9)

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - Y_i'| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2} \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - Y'_i)^2}{\sum_{i=1}^n (\bar{Y} - Y'_i)^2} \quad (9)$$

Where n denotes the number of test samples, Y_i denotes the actual value, Y'_i denotes the model predicted value, and $\bar{Y}$ denotes the average of the actual values.



Chapter 4: Results and Discussion

4.1 Introduction

First, the historical power generation data and real-time meteorological data of photovoltaic power stations in Ningxia, China in 2019 were collected. Then the data were preprocessed and the data quality was ensured by cleaning outliers. Then the Pearson correlation coefficient method was used to analyze the correlation between the influencing factors and the power generation, and the data set was defined. The data set includes solar irradiance, air temperature, relative humidity, wind speed, cloud opacity, wind direction, ground degassing pressure, and photovoltaic power. Then the data were divided and normalized. Finally, a performance comparison experiment was carried out to verify the effectiveness of the proposed model.

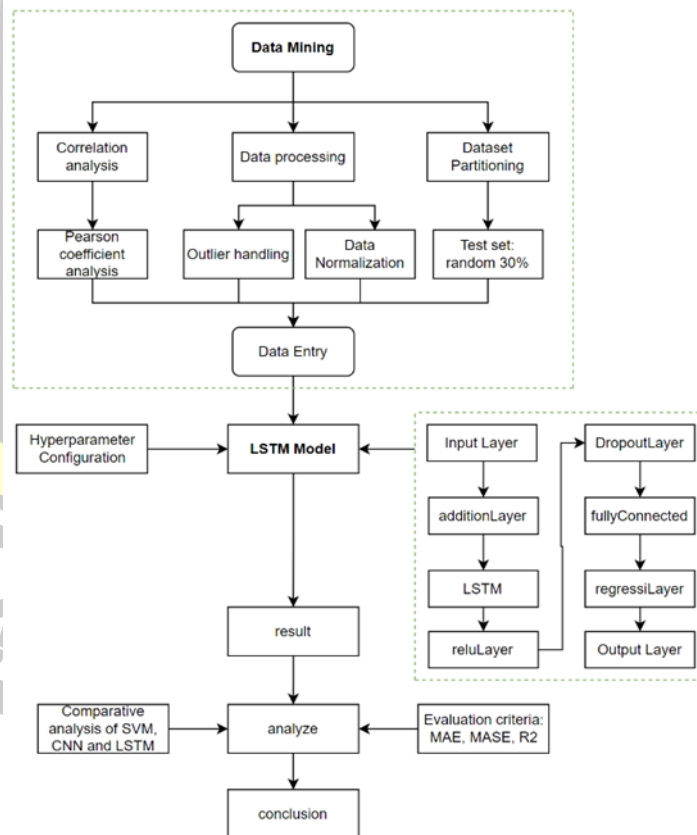


Figure 8 Process framework diagram

4.2 Results Analysis

The experimental results are visualized through MATLAB , showing the comparison chart of test value and true value, test value fitting scatter plot, loss curve chart, error bar chart, and test residual chart.

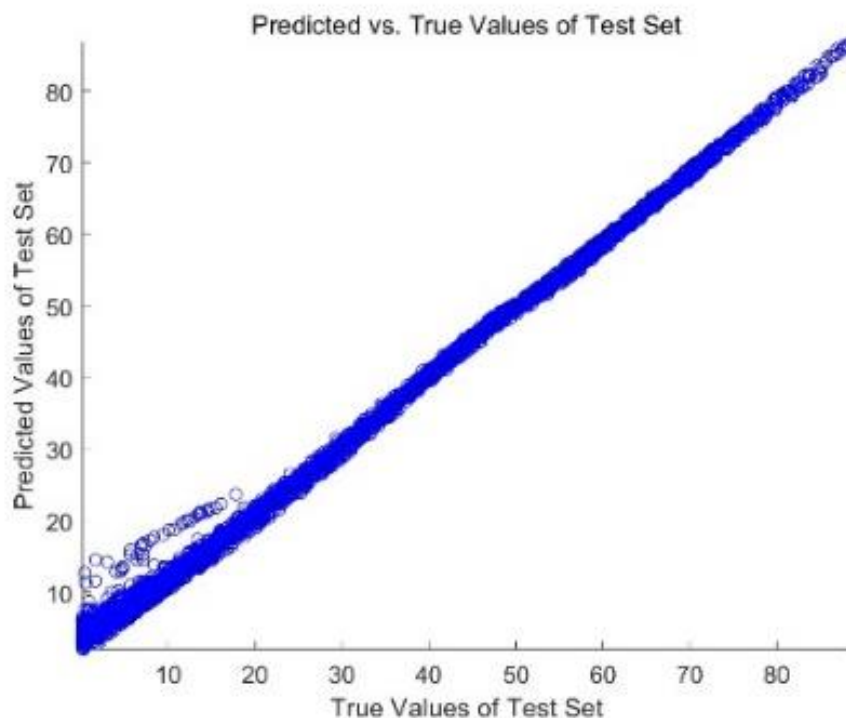


Figure 9 Scatter plot of predicted values and true values

This figure is a scatter plot of the predicted value and the true value . It can be seen from the observation that the predicted value is close to the diagonal line, indicating that the predicted value of the model is highly consistent with the true value. The scatter points are distributed very closely, indicating that the model has a good prediction effect in different true value ranges (such as low power and high power) without obvious systematic deviation. And the scatter points are symmetrically distributed on both sides of the dotted line, indicating that the predicted value is not systematically too high or too low. The scatter points in the low value area (such as

below 10) and the high value area (such as around 80) are still concentrated, indicating that the model's prediction ability at the boundary value is still relatively strong.

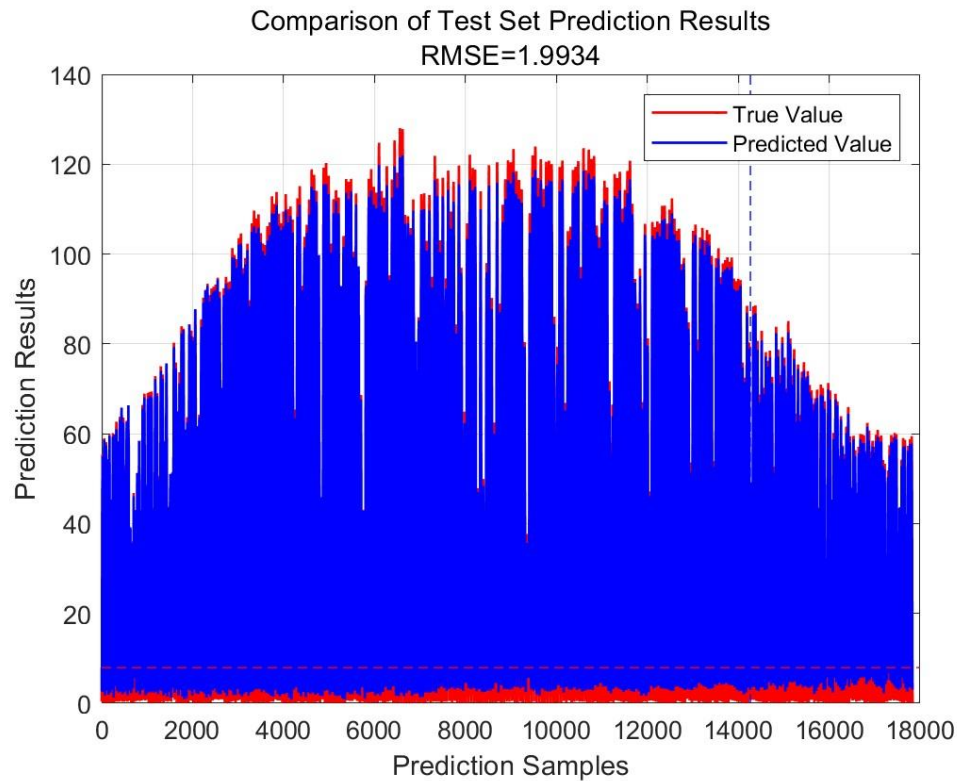


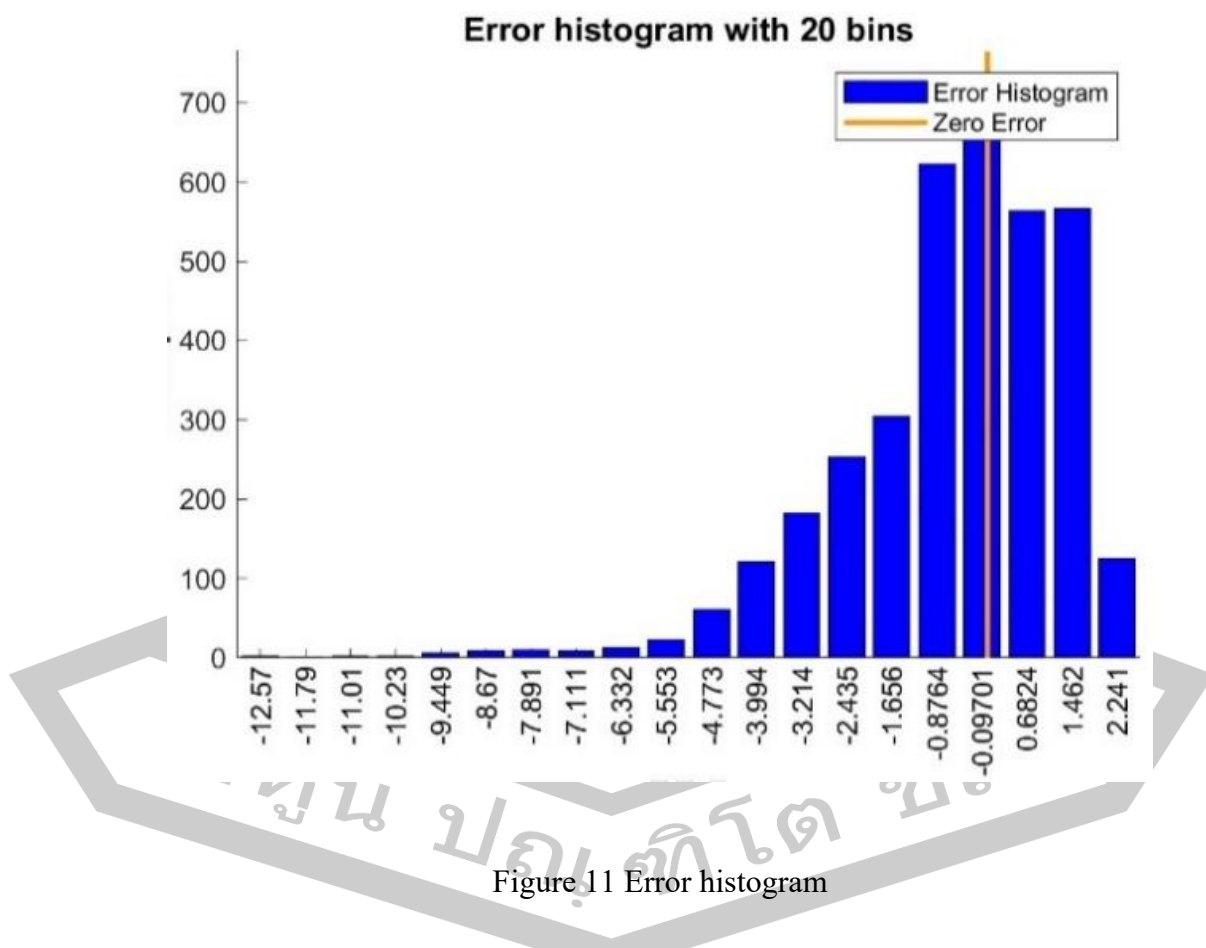
Figure 10 Test results comparison

This figure is a comparison between the test set and the true value. The red curve in the figure represents the true value of photovoltaic power generation, and the blue curve represents the model prediction value. From the overall trend, the model prediction results are highly consistent with the true value, and the main features are as follows:

Fitting effect : In the peak part of the data, the model captures the changing trend of the true value well, reflecting the adaptability of the model to complex nonlinear relationships. **Error analysis :** In the middle section where the data is densely distributed (sample range 4000 to 14000), the error between the predicted value and the true value is small. However, there is a certain degree of deviation in the low power generation interval at both ends of the sample, which may be related to the

uneven distribution of data or boundary effects. RMSE evaluation : The overall RMSE is 1.9934, indicating that the model has strong predictive ability, but there is still room for further optimization, especially in the prediction performance at the small power generation stage.

The experimental results show that the adopted deep learning model can effectively predict the actual output of photovoltaic power generation, and the predicted value is highly consistent with the true value . However, there is a certain error in the prediction in the low power generation range. In future research, it can be considered to further improve the robustness and accuracy of the model through data enhancement or the introduction of boundary correction technology.



This figure is an error histogram, which counts the prediction errors through 20 bins and shows the distribution characteristics of the errors. It can be seen from the error histogram that the prediction errors of the model are mainly distributed in

positions close to zero error, and at the same time show a certain normal distribution trend .

Most of the errors are concentrated in the range of $-2.2, 2.9-2.2$, $2.9 - 2.2$, 2.9 , especially reaching a peak near zero error (0.30.30.3), indicating that the overall prediction of the model is relatively accurate. The number of samples with large positive errors (> 3) and large negative errors (< -6) is small, indicating that the model has limited predictive ability for extreme cases and may need further optimization. The errors show a certain symmetry, but the positive errors are slightly higher than the negative errors, which may be related to the distribution of input features or the bias of the model. The number of samples in the bin near zero error (0.30.30.3) exceeds 600, reflecting the high prediction accuracy of the model for most samples.

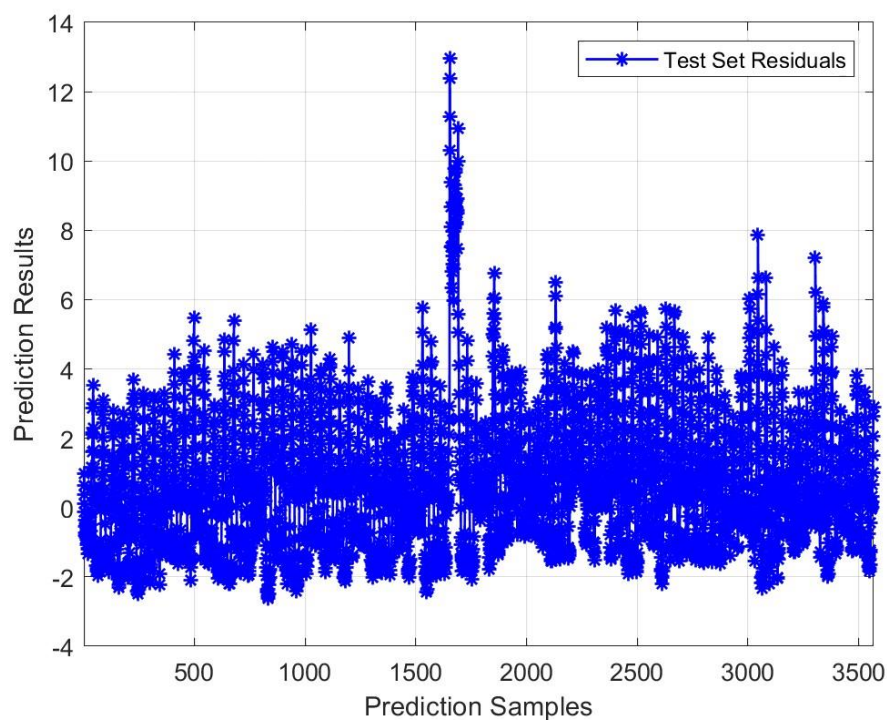


Figure 12 Test set residual graph

This figure is the residual graph of the test set . Through observation, it is found that most of the residual values are distributed between -2 and 2, indicating that the prediction error of the model is small and close to zero in most cases, and the

prediction results are generally accurate. The residual distribution is basically symmetrical around

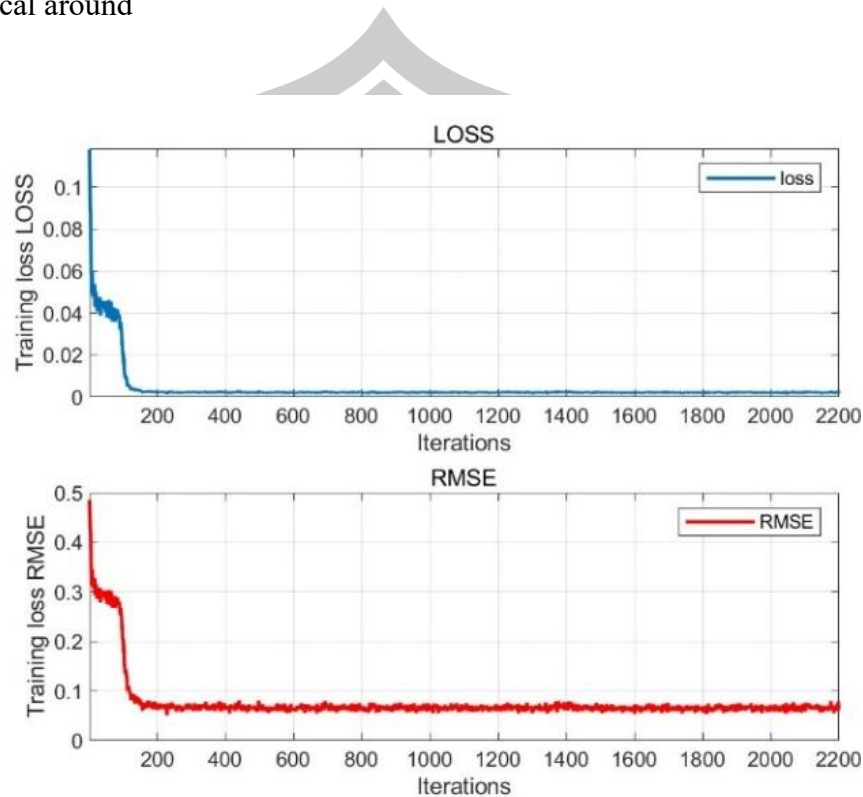


Figure 13 Training loss graph

0, indicating that the model does not have systematic over- or under-prediction problems. The overall residual fluctuation is relatively stable, indicating that the model's prediction error consistency on different samples is high, and there is no significant regularity with time or sample number changes. In some areas (such as around sample numbers 1500 and 2500), the residual fluctuation increases slightly, which may be because there are fewer samples in these areas in the training data, or the data features are more complex and difficult to predict.

4.3 Model performance comparison experiment

This experiment tests the SVM model, CNN model, and LSTM model respectively. These five graphs can clearly demonstrate the high-quality performance of the LSTM model. The evaluation indicators of LSTM are compared with those of

SVM and CNN models to comprehensively analyze the accuracy of the LSTM model in predicting photovoltaic power generation[15].

Table 4 Evaluation indicator table

	MAE (KW)	RMSE(KW)	R^2
SVM	4.325	4.935	0.9534
CNN	2.189	2.8782	0.9849
LSTM	1.4361	1.9934	0.9927

It can be seen from the evaluation index table that the overall effects of the three models are good, with MAE and RMSE both lower than 5 and R^2 higher than 0.95, indicating that the data preprocessing has achieved significant results and the predicted value fits the true value very well. However, through the comparative analysis of the three models, the MAE and RMSE of SVM are the highest and R^2 the lowest, which indicates that the error between the predicted value of SVM and the true value is too large. Although the SVM model can fit the photovoltaic power generation data to a certain extent, the prediction error is large due to the limited ability to model the nonlinear relationship between input features and output features. Moreover, the SVM model cannot capture the time dependence of photovoltaic power generation data, which limits the application of SVM in time series problems. Therefore, compared with deep learning, traditional machine learning, such as SVM, cannot process photovoltaic data with long-term dependence and complex dimensions, and is not suitable for photovoltaic power generation prediction.

Compared with SVM, the MAE of the CNN model dropped to 2.189kW, and the RMSE dropped to 2.8782kW, R^2 increasing to 0.9849. Deep learning has obvious advantages over machine learning when processing photovoltaic data. CNN extracts local features of data through convolution operations, effectively enhancing the model's modeling ability for complex input data, and can process and analyze multi-dimensional input features. Therefore, it performs better than SVM in capturing the

nonlinear relationship of photovoltaic power generation data. However, since CNN lacks the ability to model long time series and only relies on local feature extraction of input features, when the data has long time series features, it cannot memorize the long time features, which will cause gradient explosion. Therefore, it is still limited in processing data with long time series features in photovoltaic power generation.

Compared with CNN, the MAE of the LSTM model is reduced by 52.4%, RMSE is reduced by 44.3%. Unlike CNN, LSTM can smoothly capture the long time series characteristics in photovoltaic power data through internal memory units and gating mechanisms. This ability makes full use of historical data and extracts the dynamic change law between input data and power, thereby significantly improving the prediction accuracy. Compared with SVM, LSTM learns the complex nonlinear features between input and output through a multi-layer network structure, and quickly captures the complex relationship between power and input features. In comparison, SVM cannot automatically extract complex nonlinear features, and its feature capture ability is far inferior to the global dynamic modeling ability of LSTM. Among all models, the LSTM model has the lowest MAE and RMSE, with a MAE of 1.4361kW and a RMSE of 1.9934kW, which shows that LSTM performs best in prediction error and stability. In addition, it is worth mentioning that R^2 it reaches 0.9927, which is close to 1, indicating that LSTM can highly fit the real change trend of photovoltaic power generation data[16]. The data proves that LSTM has significant advantages in photovoltaic power generation prediction.

From the experimental results, it can be seen that the SVM model has the worst performance and is not suitable for photovoltaic prediction because it cannot handle the nonlinear characteristics of complex dimensional data. The CNN model enhances the extraction of nonlinear features of data through convolutional layers, but lacks long-term dependence on data[17]. In contrast, the LSTM model combines the advantages of nonlinear feature learning and long time series modeling, significantly reduces the prediction error, improves the stability and accuracy of the model, and is the best performing model in this experiment.

Chapter 5: Conclusion

Ningxia is one of the regions with the richest solar energy resources in China, with typical high radiation and large-scale photovoltaic power station characteristics. Based on the measured data of the region, this paper not only improves the accuracy of regional photovoltaic power prediction through data analysis and model optimization, but also provides a reference for prediction research in other similar regions[18]. The research results have an impact on the operation of regional photovoltaic power stations.

In view of the randomness, volatility and nonlinearity of photovoltaic power generation, this paper proposes a photovoltaic power prediction method based on LSTM (Long Short-Term Memory Network). Based on the historical data of photovoltaic power stations in Ningxia, China, combined with weather data, the study conducted comprehensive data preprocessing, feature correlation analysis and model comparison experiments. The performance comparison results with traditional machine learning methods (such as SVM) and other deep learning models (such as CNN) show that:

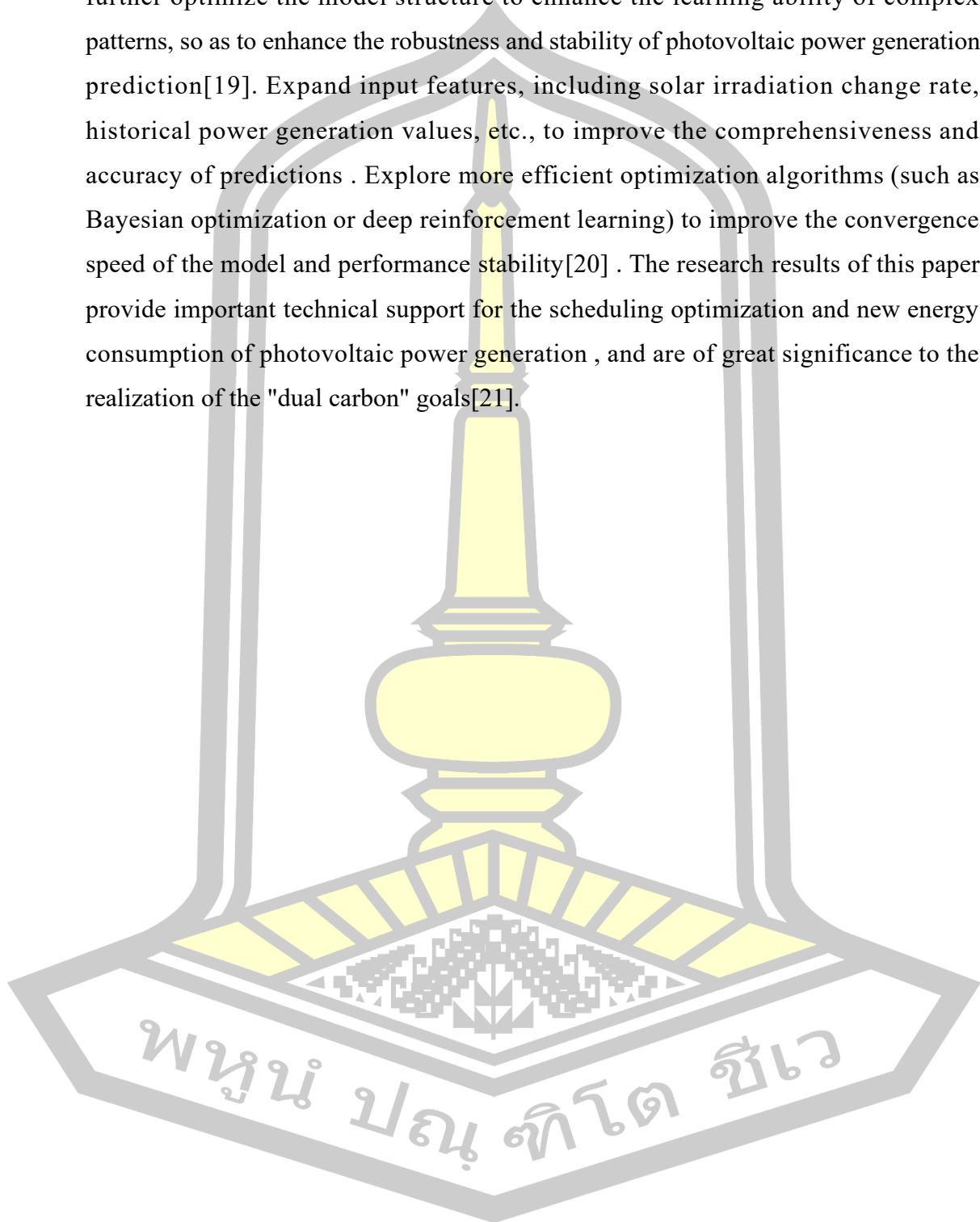
The LSTM model performed best in photovoltaic power generation forecasting due to its ability to remember time series data and nonlinear modeling capabilities, significantly reducing the prediction error (MAE is 1.4361kW, RMSE is 1.9934kW), and the fit degree reached 0.9927, which is better than SVM and CNN.

Data preprocessing (including outlier processing and normalization) and correlation analysis (based on the Pearson coefficient) effectively improved the model performance and verified the key impact of factors such as solar irradiance, temperature, and wind speed on power generation.

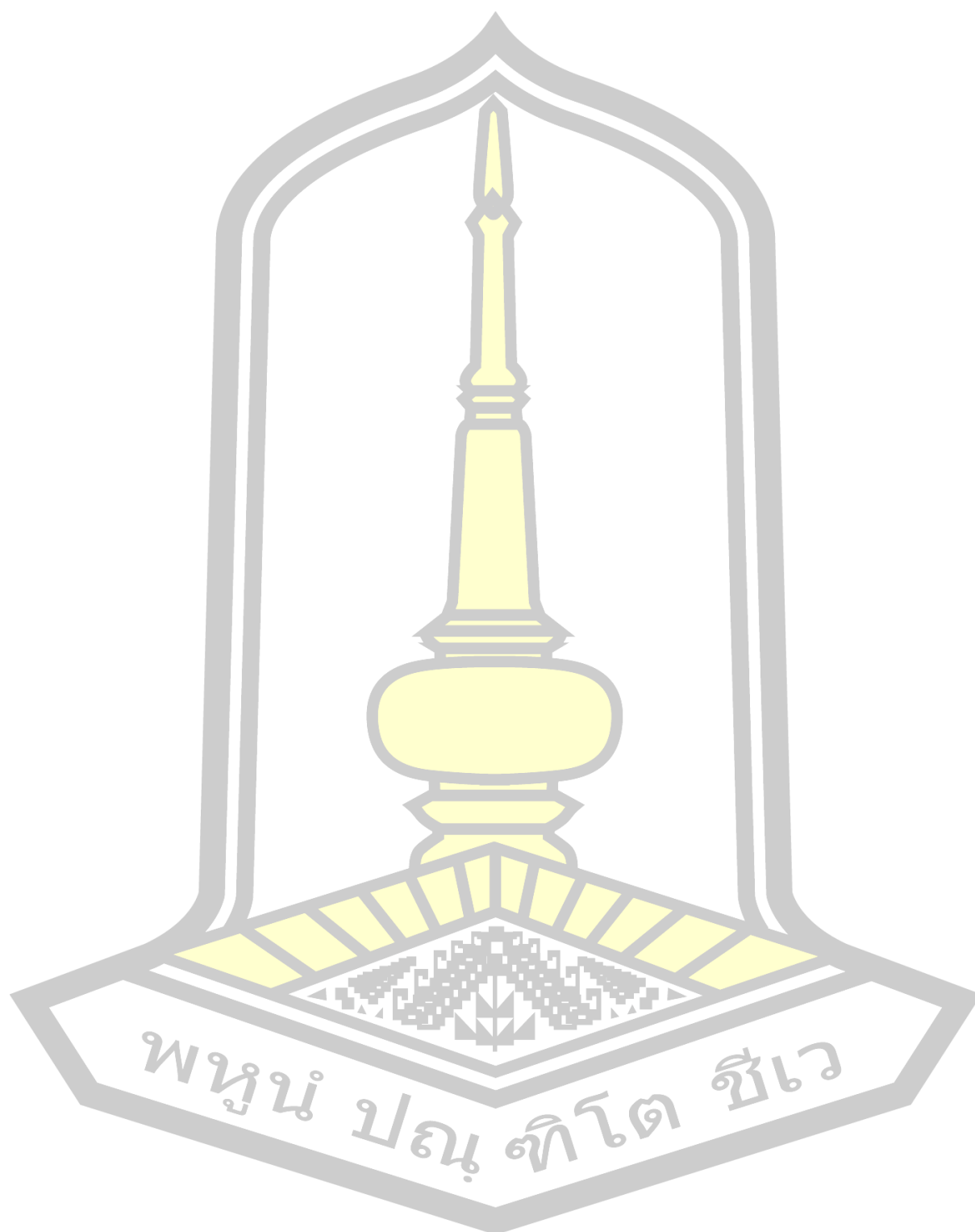
The LSTM model effectively captures the dynamic changes of photovoltaic power through a multi-layer network structure and gating mechanism, and has a strong modeling ability for complex nonlinear relationships, especially in processing long time series data.

Despite this, the study also found that the model has certain errors in the prediction of boundary values and extreme situations. Future work can further improve the prediction performance through data enhancement, feature expansion and model

optimization, and explore the potential of hybrid models (such as CNN-LSTM) to further optimize the model structure to enhance the learning ability of complex patterns, so as to enhance the robustness and stability of photovoltaic power generation prediction[19]. Expand input features, including solar irradiation change rate, historical power generation values, etc., to improve the comprehensiveness and accuracy of predictions . Explore more efficient optimization algorithms (such as Bayesian optimization or deep reinforcement learning) to improve the convergence speed of the model and performance stability[20] . The research results of this paper provide important technical support for the scheduling optimization and new energy consumption of photovoltaic power generation , and are of great significance to the realization of the "dual carbon" goals[21].

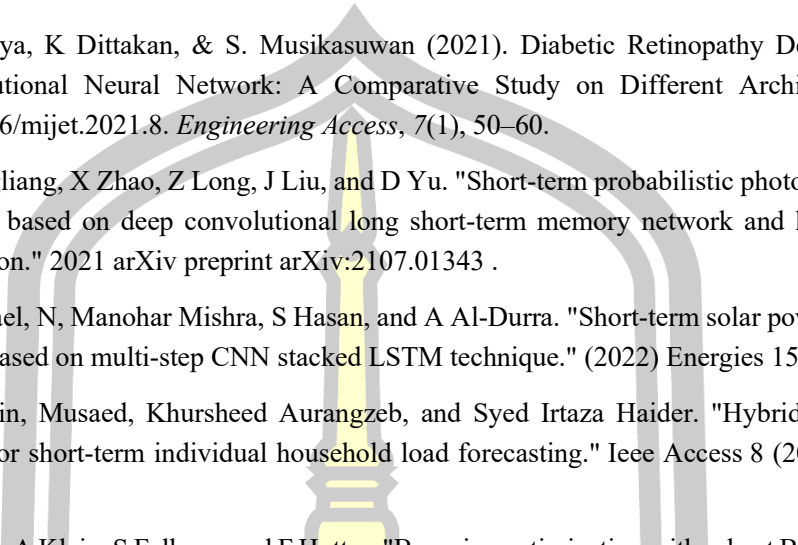


REFERENCES



REFERENCES

- [1] L Xingyun. Research on photovoltaic power generation time series prediction method[D]. Tianjin University of Science and Technology, 2023. DOI:10.27359/d.cnki.gtqgu.2023.000061.
- [2] C Yang. Photovoltaic power prediction based on deep learning[D]. Tianjin University of Science and Technology, 2023. DOI:10.27359/d.cnki.gtqgu.2023.000600
- [3] Z Manguo. Research on photovoltaic power generation prediction based on deep learning[D]. Jiangxi University of Science and Technology, 2022.DOI:10.27176/d.cnki.gnfyc.2022.000239.
- [4] F Lei, Z Jinsheng, L Haiyan, et al. Analysis of solar energy resource characteristics in Taiyangshan area of Ningxia based on field measurements [J]. *Acta Energiae Solaris Sinica*, 2020, 41(12): 146-153. DOI: 10.19912/j.0254-0096.2020.12.021.
- [5] S Yinchuan, B Yongqing, Z Hejiang. A preliminary study on the correction of localized WRF radiation forecast and photovoltaic power generation prediction method in Ningxia[J]. *Chinese Journal of Desertification*, 2012, 32(06):1738-1742.
- [6] W Lei, J Ning, Y Guangliang, et al. Research on the ability of Ningxia power system to accept new energy[J]. *Power System Technology*, 2010, 34(11): 176-181. DOI: 10.13335/j.1000-3673.pst.2010.11.039.
- [7] G Pengjiang, L Yi. Analysis of output characteristics of photovoltaic power stations in Ningxia[J]. *Electrical Automation*, 2016, 38(02): 33-36.
- [8] H Jiafeng, F Shuanglei, D Maosheng, et al. Wind-solar integrated power prediction system for Ningxia power grid [J]. *Ningxia Electric Power*, 2011
- [9] W Yingming. A brief analysis of the paths and measures for building solar photovoltaic integration in Ningxia[J]. *Architecture*, 2022, (17): 52-55.
- [10] C Chen, S Hua, and C Wang. "Assessment of different deep learning methods of power generation forecasting for solar PV system." *Applied Sciences* 12, no. 15 (2022): 7529.
- [11] S.Schmidhuber, J. Long short-term memory. *Neural Comput.* 1997, 9, 1735–1780.
- [12] F Gers. F Cummins. Learning to forget: Continual prediction with LSTM. *Neural Comput.* 2000, 12, 2451–2471.
- [13] S Singh, A Gupta, R Chandel, and S Tajjour. "Review of deep learning techniques for power generation prediction of industrial solar photovoltaic plants." *Solar Compass* 8 (2023): 100061.
- [14] L Zhiqin, D Jianqiang, N Bin, X Wangping, H Canyi, and L Huan. "A review of feature selection methods." *Computer Engineering and Applications*(2019).
- [15] Z, Xiangyu. "Mlp, cnn, lstm and hybrid svm for stock index forecasting task to indu and ftse100." (2020)Available at SSRN 3644034 .

- 
- [16] Z Jinsong, R Verschae, S Nobuhara. "Deep photovoltaic nowcasting." Solar Energy 176 (2018): 267-276.
- [17] T, Hattiya, K Dittakan, & S. Musikasuwana (2021). Diabetic Retinopathy Detection using Convolutional Neural Network: A Comparative Study on Different Architectures: doi: 10.14456/mijet.2021.8. *Engineering Access*, 7(1), 50–60.
- [18] B, Mingliang, X Zhao, Z Long, J Liu, and D Yu. "Short-term probabilistic photovoltaic power forecast based on deep convolutional long short-term memory network and kernel density estimation." 2021 arXiv preprint arXiv:2107.01343 .
- [19] E Michael, N, Manohar Mishra, S Hasan, and A Al-Durra. "Short-term solar power predicting model based on multi-step CNN stacked LSTM technique." (2022) *Energies* 15, no. 6: 2150.
- [20] Alhussein, Musaed, Khursheed Aurangzeb, and Syed Irtaza Haider. "Hybrid CNN-LSTM model for short-term individual household load forecasting." *Ieee Access* 8 (2020): 180544-180557.
- [21] J Tobias, A Klein, S Falkner, and F Hutter. "Bayesian optimization with robust Bayesian neural networks." (2016) *Advances in neural information processing systems* 29.

BIOGRAPHY

NAME	Mr.MingzeLei
DATE OF BIRTH	11 January 1999
PLACE OF BIRTH	china
ADDRESS	Li County, Changde City, Hunan Province
POSITION	teacher
PLACE OF WORK	Hunan Industrial and Trade School
EDUCATION	2014 High School Lixian No.2 Middle School 2017 College Hunan Mechanical and Electrical Vocational College 2020 Bachelor Hunan Institute of Technology 2024 Master Mahasarakham University College of Engineering

