

Article

A Hybrid Model for GRU Ultra-Short-Term Wind Speed Prediction Based on Tsfresh and Sparse PCA

Yaqi Wang  and Renzhou Gui *

School of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

* Correspondence: rzgui_eie@163.com

Abstract: Wind power is a popular renewable energy source, and the accurate prediction of wind speed plays an important role in improving the power generation efficiency of wind turbines and ensuring the normal operation of wind power equipment. Due to the instability and randomness of wind speed, it is difficult to achieve accurate prediction by traditional prediction methods. To improve the power generation efficiency of wind turbines and realize the predictability of wind speed, a hybrid wind speed prediction model based on GRUs (gated recurrent units) was constructed in this paper based on a deep neural network and feature extraction method. The hybrid model feature extraction module was implemented based on a combination of Tsfresh (a python package for time series feature extraction) and sparse PCA (sparse principal component analysis), and the network structure and other hyperparameters of the GRU module were determined through experiments. The model was validated using actual wind measurement data from a wind farm on the west coast of the United States. The results showed that the proposed model had less computational time and higher computational accuracy than the SARIMAX (seasonal auto-regressive integrated moving average with exogenous factors) and LSTM (long short-term memory) models.

Keywords: wind speed prediction; gate recurrent unit; deep learning; Tsfresh; sparse principal component analysis



Citation: Wang, Y.; Gui, R. A Hybrid Model for GRU Ultra-Short-Term Wind Speed Prediction Based on Tsfresh and Sparse PCA. *Energies* **2022**, *15*, 7567. <https://doi.org/10.3390/en15207567>

Academic Editor: Francesco Castellani

Received: 19 August 2022

Accepted: 10 October 2022

Published: 13 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Wind energy is an important form of renewable energy and the most attractive alternative to fossil energy. The wind strikes the turbine blades during motion, driving the blades to rotate, and the turbine converts the rotational energy of the blades into electrical energy [1]. In the process of wind power generation, unstable wind speed conditions can lead to degradation of power quality and even destabilize the power generation system [2,3]. If the wind speed conditions can be grasped in advance and the turbines and other equipment can be adjusted in time according to the wind speed conditions [4], the power generation efficiency of the turbines can be greatly improved and the hardware equipment losses can be reduced. Based on the wind speed data of the wind field over time and the current wind speed data for wind speed prediction, we can make point predictions or interval predictions for the future wind speed. Wind speed prediction is an important technical support module in the production decision of wind power.

In recent years, a lot of work has been carried out on wind speed prediction. The current wind speed prediction content is divided into four types: long-term prediction, medium-term prediction, short-term prediction, and ultra-short-term prediction [5]. The application of different prediction contents is shown in Table 1.

Table 1. Application of different time horizon wind speed forecasting models [6].

Forecast Content	Time Horizon	Application
ultra-short-term	≤ 1 h	Real-time grid operation, equipment monitoring, etc.
short-term	≤ 8 h	Load dispatch planning, load rationalization decision and power market operation guarantee, etc.
medium-term	≤ 24 h	Reserve demand decision, etc.
long-term	> 24 h	Wind farm design, optimization of operation cost, etc.

According to the characteristics of random, complex, and many influencing factors of wind speed, the current prediction model includes four types: the physical model, statistical model, neural network model, and hybrid method model [7]. Chen et al. [8]. developed an adaptive ARMA (auto-regressive moving-average) model for short-term load forecasting of a power system. Tyass et al. [9,10] realized short-term wind speed prediction based on the SARIMA (seasonal auto-regressive integrated moving average). Kritharas et al. [11] made use of the long-term historical wind speed record data from the ground station and used the SARIMAX model to achieve the monthly wind speed prediction. However, because of the solidified regression fitting method, the statistical model did not have an advantage in dealing with the time-series wind speed with nonlinear characteristics [12].

With the continuous development of artificial intelligence, model construction and prediction accuracy have changed dramatically [13]. Zhu Q et al. [14,15] combined the advantages of LSTM and CNNs (convolutional neural networks), put forward a spatiotemporal correlation prediction model, and realized wind speed prediction by extracting spatial features from CNNs and temporal features from LSTM. Liu M et al. [16] established a prediction system based on the GRU model, which used relevant historical data of the wind power system to predict results, and the prediction system could provide support for network structure analysis and the control strategy of a multi-energy system. Li C et al. [17] proposed a novel hybrid model based on gated recurrent unit neural network and variational mode decomposition for wind speed interval prediction, which had higher prediction interval coverage probability and narrower prediction interval width. With the deepening of research, some scholars found that the single network structure of most machine learning algorithms was not suitable for complex and variable wind speed conditions. To further optimize the accuracy of model prediction, some scholars have put forward hybrid models with multiple advantages by combining different research methods. Zhou B et al. [18,19] proposed a hybrid method for ultra-short-term wind power prediction considering meteorological characteristics (wind direction, wind speed, temperature, pressure, humidity, etc.) and seasonal information, using ensemble empirical modal decomposition to decompose wind energy data into smooth subseries and principal component analysis to reduce redundant meteorological characteristics and improve the model accuracy. Jung J et al. [20] used several historical meteorological variables, such as wind speed, temperature, humidity, pressure, dew point, and solar radiation, as indirect predictive characteristic quantities to achieve high accuracy wind speed prediction. The hybrid model improved the accuracy of wind speed prediction, but the feature expansion of wind speed data was less considered, thus requiring a large amount of feature data (e.g., meteorological features such as temperature) and high computational cost in the accuracy prediction.

Based on the above analysis, a hybrid short-term wind speed prediction model combining Tsfresh, Sparse PCA, and GRU is proposed in this paper. The model includes data pre-processing, feature extraction, feature selection, and data prediction parts. The model first converted the non-linear and non-smooth wind speed data into a data format suitable for the deep learning model. Then, the data of each subseries was feature transformed by the feature extraction and selection module to output the subseries with the most distinct data features. Finally, the subseries were predicted hierarchically using the prediction module, and the final prediction results were obtained after the reconstruction of different component predictions.

The contributions of this paper are as follows:

1. A new short-term wind speed hybrid prediction model was proposed;
2. A detailed comparison of the performance and computation time of the proposed model with neural network models and statistical models was made;
3. It was verified that the feature expansion approach was superior to the approach using data features related to wind speed in the point prediction of univariate wind speed time series data.

2. Materials and Methods

2.1. Tsfresh-Time Series Feature Extraction

Feature engineering is an important step used for many machine learning tasks, including time series prediction [21]. Feature processing generally includes feature extraction and feature selection [22]. Tsfresh is a python package that can automatically calculate a large number of time series features [23]. The principle is as follows:

Suppose a set of time series $D = \{X_1, X_2, \dots, X_N\}$ ($i = 1, 2, \dots, N$) were given as input, where X_i needed to be mapped to a feature space with a particular dimension M . There existed eigenvectors x_i :

$$x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,M}) \quad (1)$$

The feature vector x_i is constructed by applying the characterization method of time series with the corresponding time series X_i to obtain the corresponding feature vector.

$$f_j : X_i \rightarrow x_{i,j} \quad (2)$$

$$x_i = (f_1(X_i), f_2(X_i), \dots, f_M(X_i)) \quad (3)$$

M represents the number of features of the time series characterization method.

According to the above method, Tsfresh achieves feature extraction by parallel computation. Its computational method consists of three steps: feature extraction; p -value calculation; multiple computation, and testing. Tsfresh has 63 methods of characterizing time series and can calculate 794 features.

2.2. GRU

LSTM achieved “long-term memory” of wind speed data by controlling the transmission state through gating states, but too many gating states led to an increase in training parameters, which enhanced the training difficulty [24]. To solve this problem, GRU combines input gates and forgetting gates to make a network with fewer parameters [25], which can capture long-term dependencies more easily and improve the training effect while ensuring memory. GRU unit architecture is shown in Figure 1.

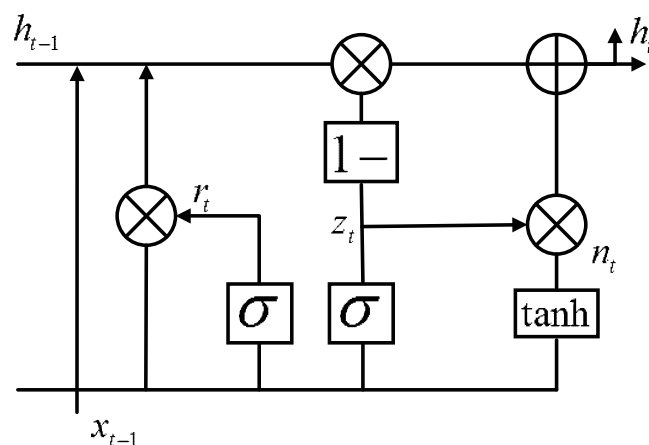


Figure 1. Structure of GRU Network.

The GRU network calculation process is as follows:

1. Firstly, we defined W_{ir} , W_{hr} , W_{iz} , W_{hz} , W_{in} , and W_{hn} as the weight matrices for the different phases, and b_{ir} , b_{hr} , b_{iz} , b_{hz} , b_{in} , and b_{hn} were the bias terms.
2. The GRU reset gate r_t was introduced through the reset gate calculated using the equation:

$$r_t = \sigma(W_{ir}x_t + b_{ir} + W_{hr}h_{t-1} + b_{hr}) \quad (4)$$

when r_t was 0, n_t contained the information of the current number signal input status.

3. The update gate z_t controlled the amount of information forgotten by the hidden layer at the previous moment:

$$z_t = \sigma(W_{iz}x_t + b_{iz} + W_{hz}h_{t-1} + b_{hz}) \quad (5)$$

4. The hidden state h_t carried the information of the current cell and passed it down to the network, and the last output value was chosen among n_t and h_t :

$$n_t = \tanh(W_{in}x_t + b_{in} + r_t \times (W_{hn}x_t + b_{hn})) \quad (6)$$

$$h_t = (1 - z_t) \times n_t + z_t \times h_{t-1} \quad (7)$$

2.3. Sparse PCA

Sparse PCA solved the problem that PCA did not have a sparse constraint effect on the load in the main data direction. Cai T. et al. [26] proposed the sparse PCA method by improving the interpretability of PCA through the method of sparse main direction load. Sparse PCA is calculated as follows:

1. A starts from the first k ordinary principal components $V [1: p, 1:k] (p \rightarrow \infty)$;
2. Given A as:

$$A = [\alpha_1, \alpha_2, \dots, \alpha_k] \quad (8)$$

By solving the elasticity problem:

$$\beta_j = \operatorname{argmin}(\alpha_j - \beta)^T X^T X (\alpha_j - \beta) + \lambda \|\beta\|^2 + \lambda_{1,j} \|\beta\|_1 \quad (9)$$

where $j = 1, 2, \dots, k$;

3. Given B as:

$$B = [\beta_1, \beta_2, \dots, \beta_k] \quad (10)$$

Calculate the singular value:

$$X^T X B = U D V^T \quad (11)$$

Finally, update A :

$$A = U V^T \quad (12)$$

4. Repeat steps 2 and 3 above until convergence.
5. Finally, standardization:

$$V_j = \frac{\beta_j}{\|\beta_j\|} \quad (13)$$

where $j = 1, 2, \dots, k$.

3. Realization of Prediction Model

Complex weather conditions such as wind speed and direction aggravated the difficulty of wind power generation. As shown in Figure 2, the wind turbine control system timely adjusted the wind turbine according to the short-term wind speed prediction results.

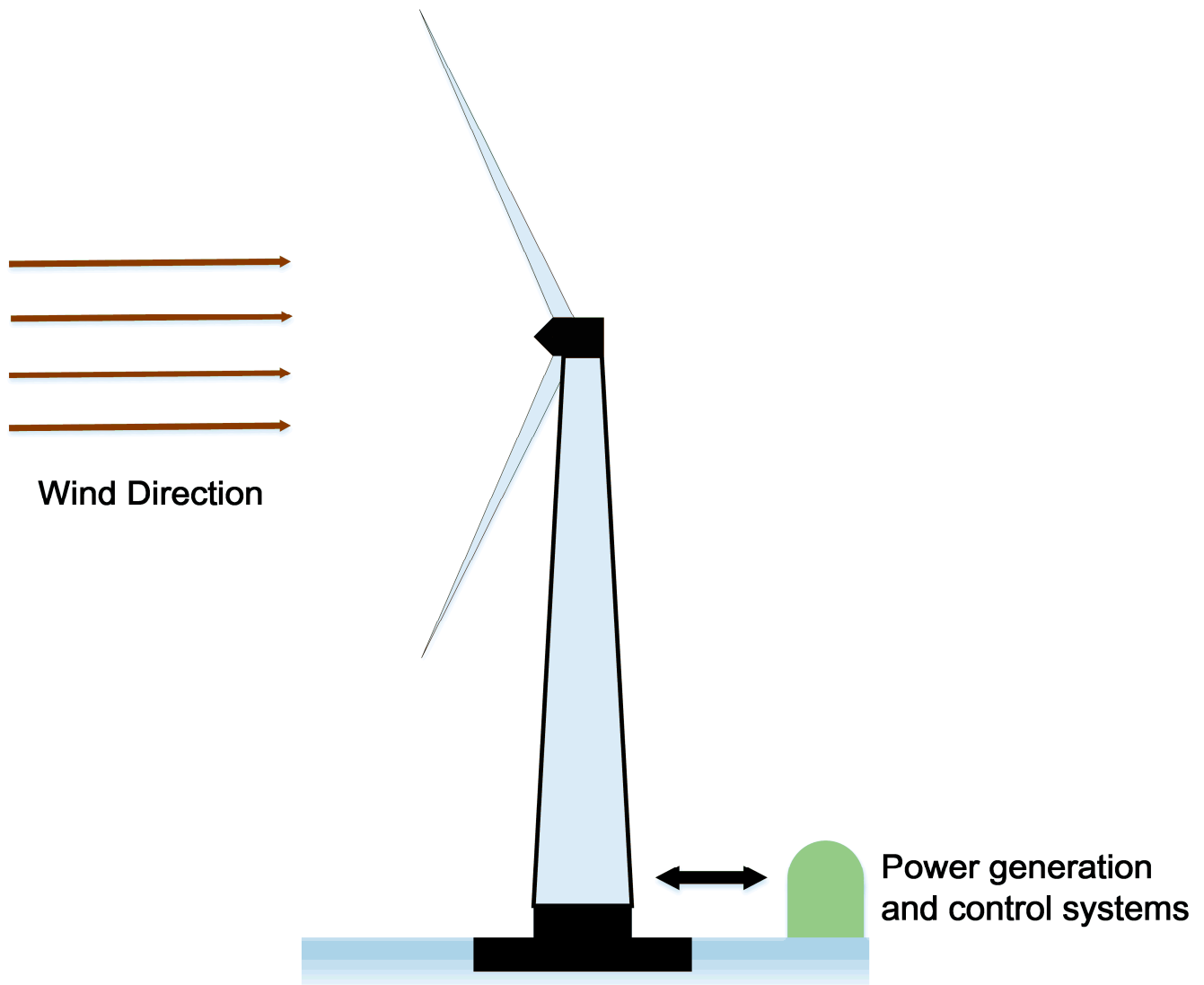


Figure 2. Schematic diagram of wind speed prediction and control. The control system made a forecast of the next moment wind speed according to the current wind speed conditions and adjusted the wind turbine in time.

The core of the proposed hybrid model was to create feature data similar to other meteorological features, select appropriate features, and predict with a GRU network. The block diagram of the prediction method is shown in Figure 3.

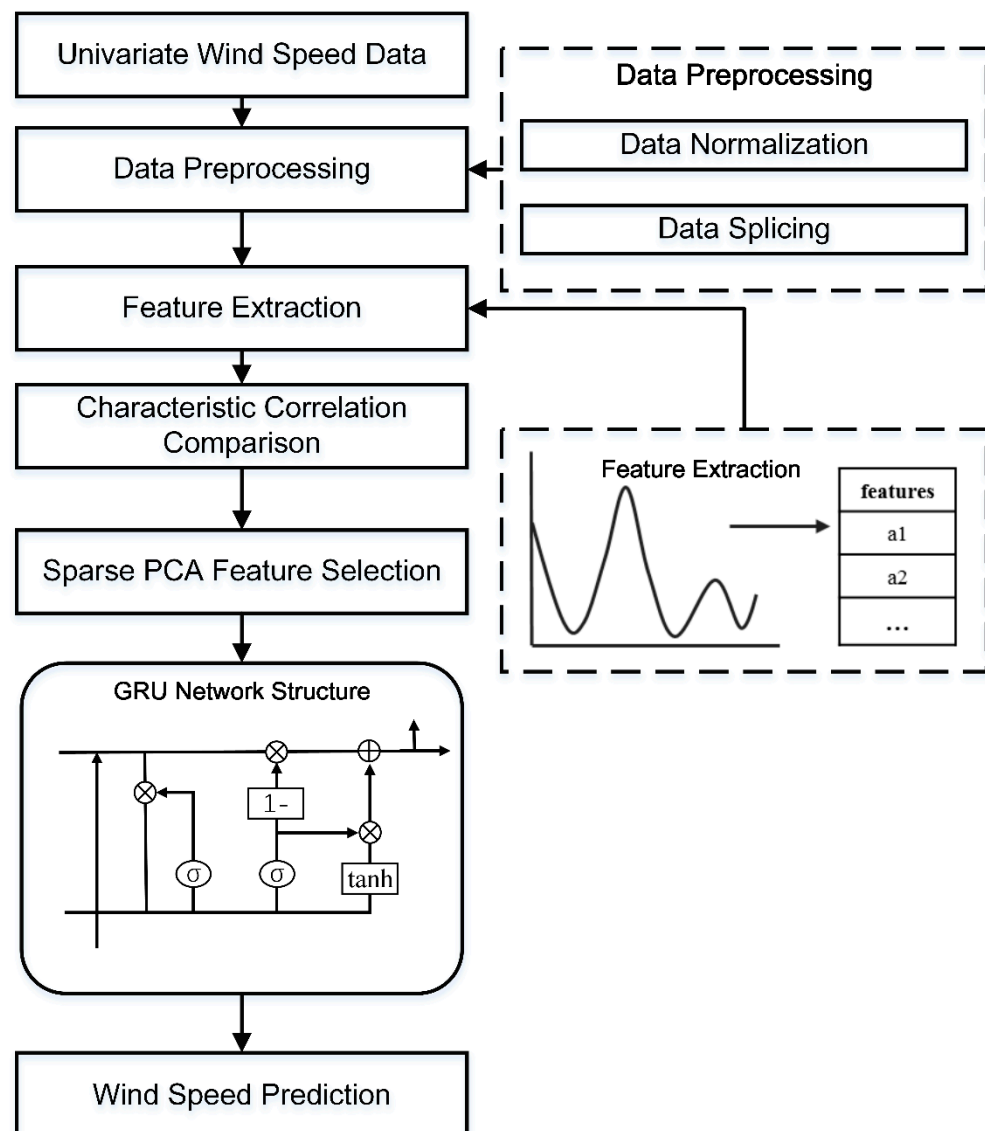


Figure 3. Flow chart of the prediction method framework. The hybrid model preprocessed the original univariate wind speed time series data, and the processing included data normalization and data binning. Feature processing was performed on the data, including feature extraction and feature selection. Finally, all feature data were predicted using a GRU network.

The hybrid model steps were divided into the following steps:

1. All data were sliced and normalized.
2. Data feature processing included feature extraction and feature selection, with Tsfresh in charge of feature extraction and Sparse PCA in charge of feature selection.
3. GRU accepted the input data for data fitting to obtain the predicted values and calculated the error based on the predicted and true values.
4. The weights of the layers of the GRU network were updated based on the BPTT algorithm [27].
5. After setting the number of training steps, steps 3 and 4 were repeated, and the Adam optimization algorithm was used to calculate the update steps until the loss function was minimized.
6. After the network was trained, the wind speed was predicted for each minute of the next hour.
7. Error analysis was carried out to evaluate the model performance.

During data normalization, outliers in the original data needed to be removed and the null values were filled using the average of the first seven historical data. All data were normalized using the Min-Max normalization method, as shown in Equation (14):

$$X_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (14)$$

x_i represents each value of the time series, $x_{\max}$ denotes the maximum value, $x_{\min}$ denotes the minimum value, and X_i represents the normalized value.

The normalized time series data needed to be sliced before feature extraction. The sliced data had more importance for discovering the detailed information of the time series data. The data slicing in the data preprocessing module was as shown in Figure 4. Suppose there was a set of time series. Its length is marked as “ $length(TimeSeriesData)$ ”. The preprocessing module decomposed all the time series data into n copies, l_t denoted the length of the subsequence, λ denoted the length of time each segment of the sequence needed to be predicted in advance, and the subsequence sliding window was δ .

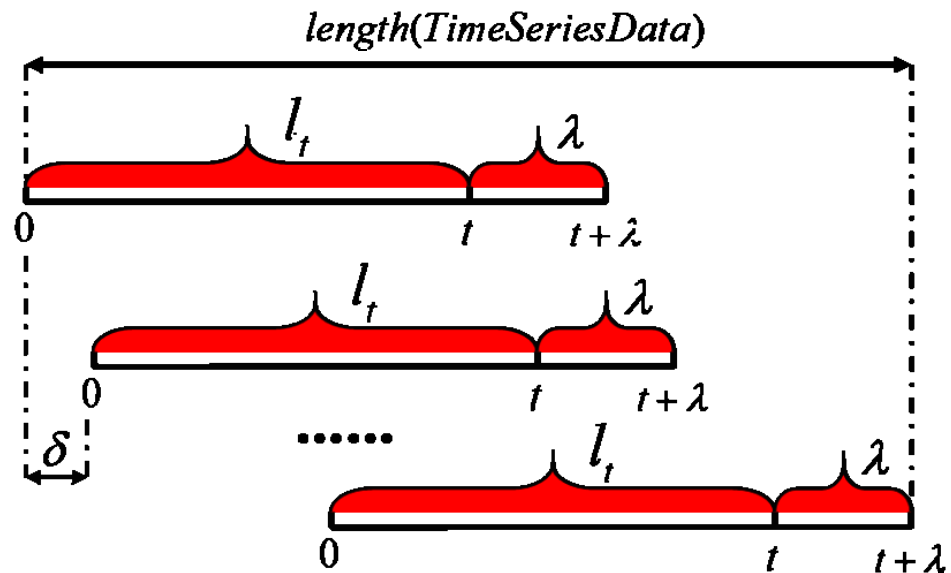


Figure 4. Data slicing in the data pre-processing module.

The evaluation indexes for wind speed regression prediction included MAE , MSE , $RMSE$, and $MAPE$, which were calculated as shown in Equations (15)–(18):

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (15)$$

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (16)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (17)$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (18)$$

y_i denoted the true value and $\hat{y}_i$ denoted the predicted value.

4. Case Analysis

4.1. Experimental Data

The dataset was obtained from the National Renewable Energy Laboratory (NREL). Data were readings from instruments mounted on a 270-foot weather tower with coordinates of 39°54′38.34″ N 105°14′5.28″ W. Short-term wind speed measurements were averaged from readings taken every two seconds and within one minute [28].

The data set for the research work in this paper was the data measured from the meteorological tower at 2 m height. The data were divided into four datasets based on the seasons, which were continuous data for ten random days in each season. The datasets had 57,600 data items, each containing wind speed data and associated with 12 other features, such as temperature, humidity, and barometric pressure. The last 60 min of data from each dataset were used for the test set, and the remaining data were divided into training and validation sets with a division ratio of 0.75:0.25. The wind speed time series of the four data sets are shown in Figure 5. The mathematical statistics of the wind speed data of each data set are shown in Table 2.

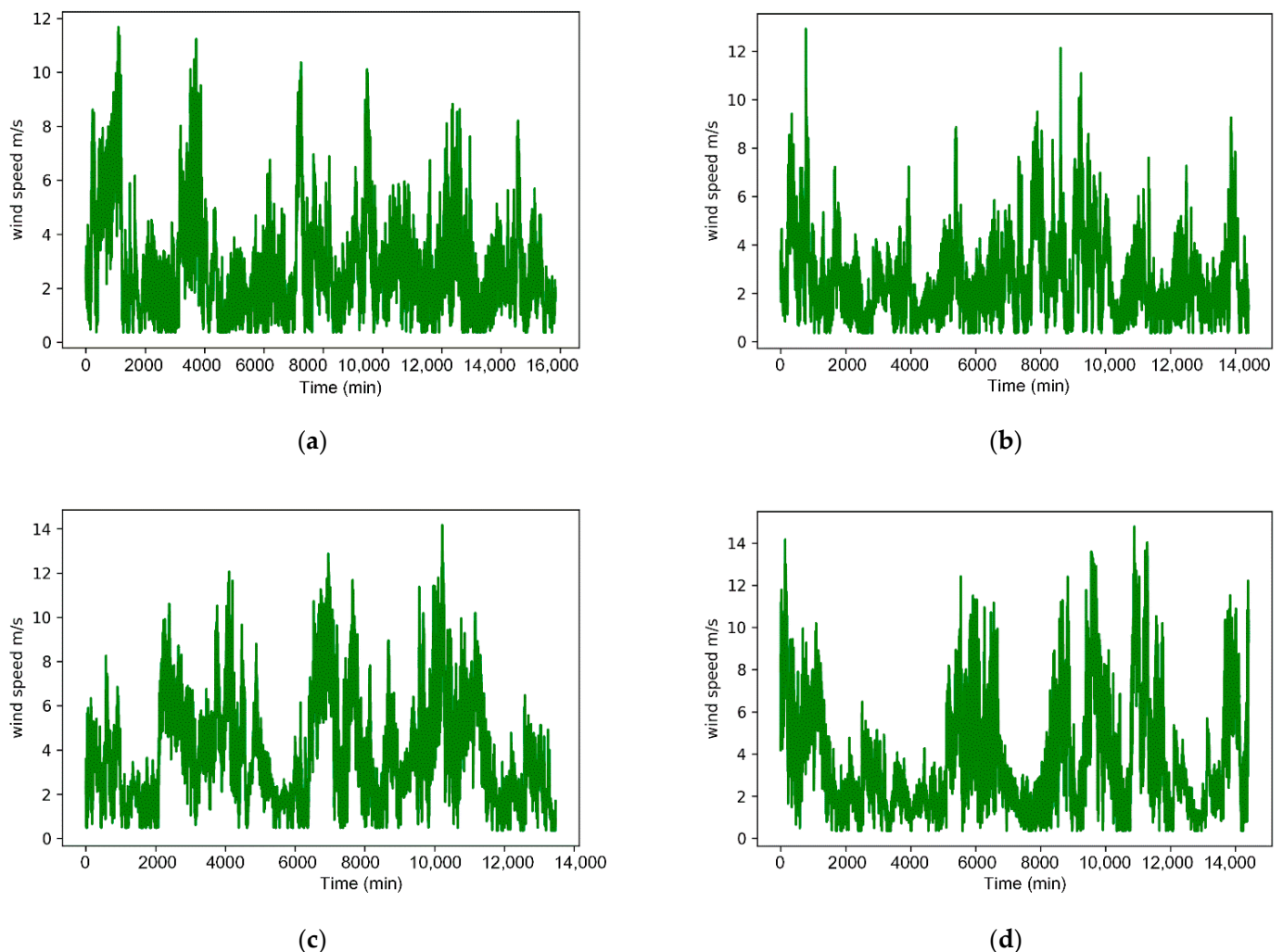


Figure 5. Time series representation of wind speed for the four datasets. (a) Time series representation of spring wind speed dataset; (b) time series representation of summer wind speed dataset; (c) time series representation of autumn wind speed dataset; (d) time series representation of winter wind speed dataset.

Table 2. Mathematical statistics of the four types of data.

Dataset	Samples	Length	Mathematical Statistical Indicators						
			Max	Min	Mean	Std	Median	Skew	Kurtosis
spring	all	15,840	11.68	0.36	2.76	1.78	2.32	1.29	1.67
	train	11,880	11.68	0.36	2.86	1.87	2.39	1.22	1.33
	validation	3240	8.83	0.36	2.45	1.4	2.2	1.17	1.69
	test	720	2.3	0.71	1.45	0.44	1.4	0.11	−1.3
summer	all	14,400	12.92	0.36	2.61	1.65	2.21	1.33	2
	train	10,800	12.92	0.36	2.7	1.74	2.28	1.25	1.59
	validation	3540	9.26	0.36	2.36	1.34	2.08	1.48	3.02
	test	60	3.21	0.37	1.66	0.67	1.64	0.16	−0.7
autumn	all	13,476	14.17	0.36	3.86	2.35	3.36	0.8	0.07
	train	10,107	12.87	0.48	3.93	2.33	3.49	0.76	−0.05
	validation	3309	14.17	0.36	3.7	2.42	3.08	0.95	0.46
	test	60	1.69	0.36	0.67	0.34	0.57	0.99	0.35
winter	all	14,400	14.78	0.36	3.64	2.59	2.81	1.13	0.72
	train	10,800	14.17	0.36	3.5	2.43	2.71	1.11	0.72
	validation	3540	14.78	0.36	4.04	2.96	3.1	1.03	2.53
	test	60	12.21	0.77	5.77	3.09	5.66	0.23	−1.14

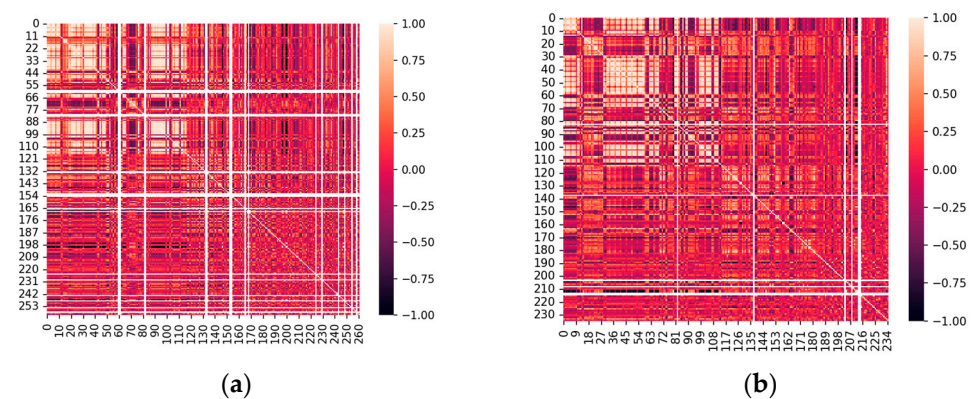
4.2. Feature Processing Result

The feature processing work was performed only for the wind speed data of the dataset. Four datasets were subjected to feature extraction using the Tsfresh framework and feature selection was performed uniformly by Sparse PCA. Table 3 showed the number of features after feature extraction and selection for each dataset.

Table 3. Number of feature-processed features for each data set.

Datasets	Number_Tsfresh	Number Sparse PCA
spring	789	253
summer	789	234
autumn	789	250
winter	789	286

The correlation representation between the extracted features after feature processing is shown in Figure 6. The darker the color, the stronger the correlation. The figure shows that the correlation of the selected features was not high.

**Figure 6.** Cont.

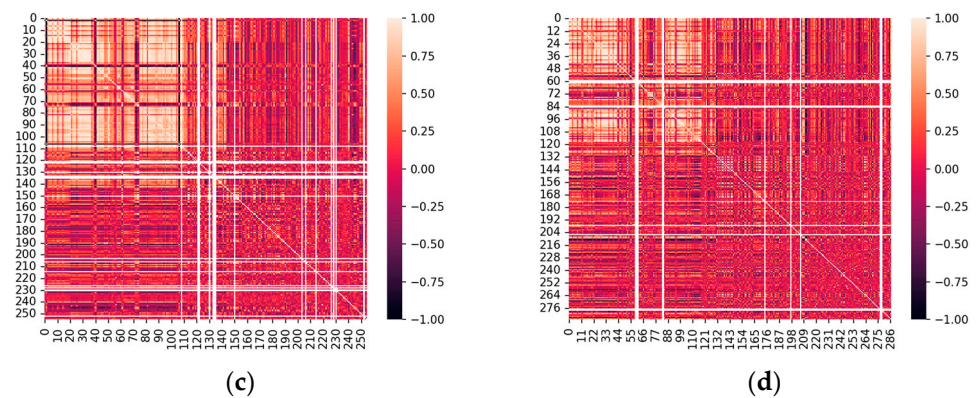


Figure 6. Correlation of features selected for the four datasets. (a) Feature correlation representation of spring dataset; (b) feature correlation representation of summer dataset; (c) feature correlation representation of autumn dataset; (d) feature correlation representation of winter dataset.

4.3. Analysis of Hybrid Model Prediction Results

In this part, the predicted results for the four datasets are analyzed and discussed in detail.

The wind speed prediction results of the hybrid model in four seasons are shown in Figure 7. The black line represents the real data and the red line represents the prediction results. The results showed that the wind speed prediction model had a good fitting trend and tracking ability, and the individual abrupt change results could not be fitted perfectly.

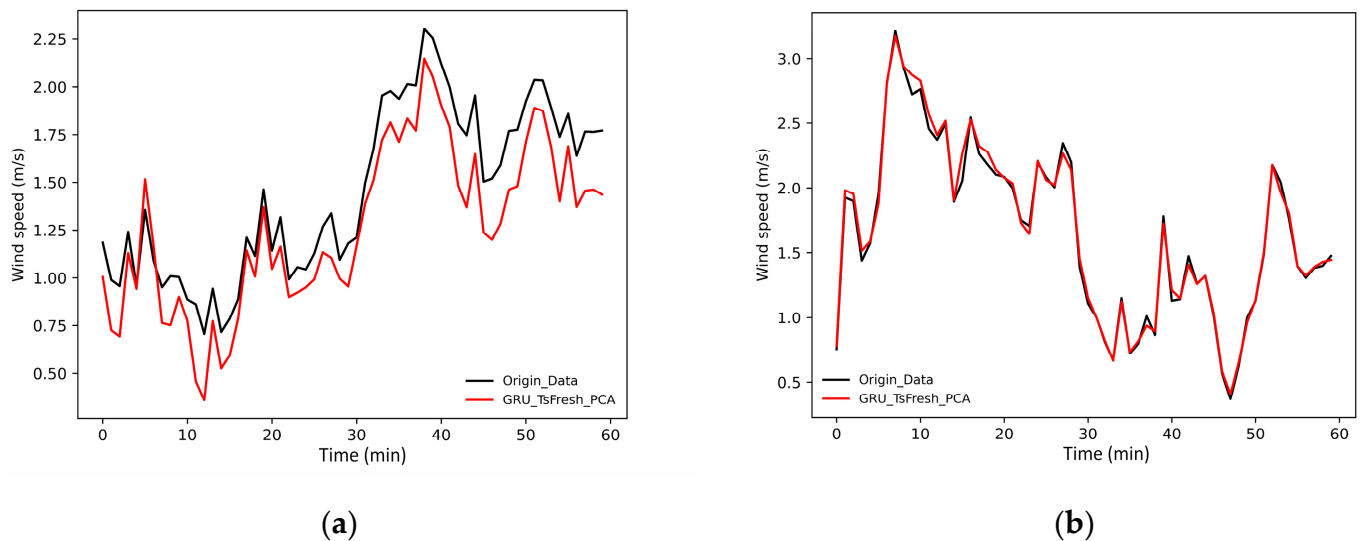
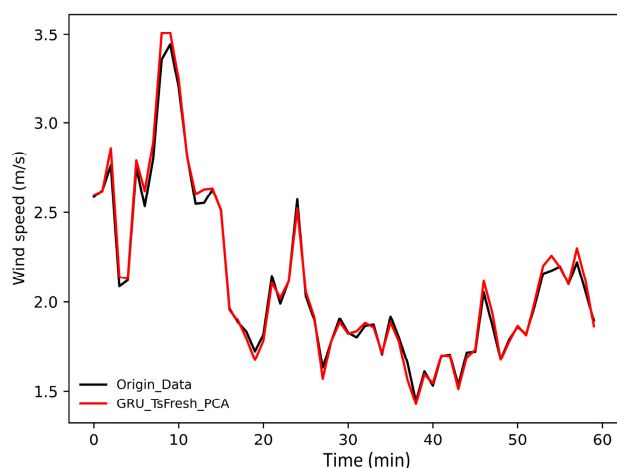
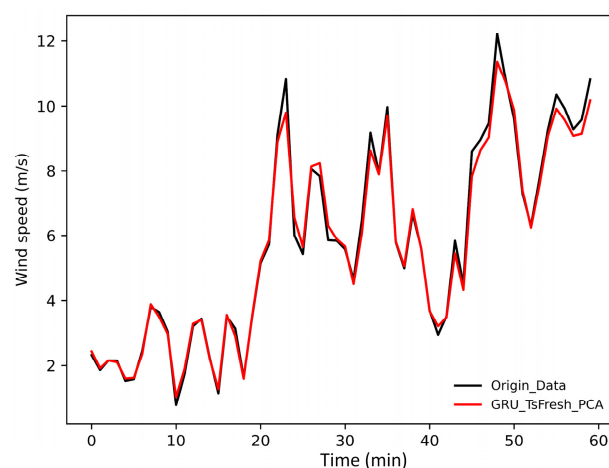


Figure 7. Cont.



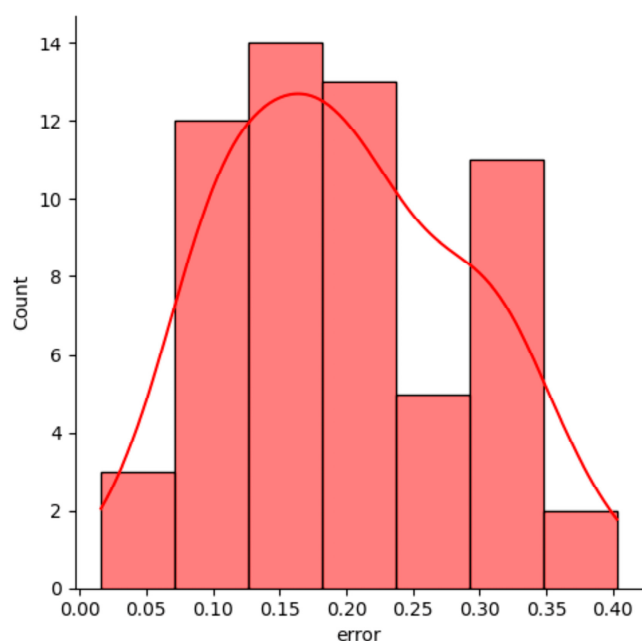
(c)



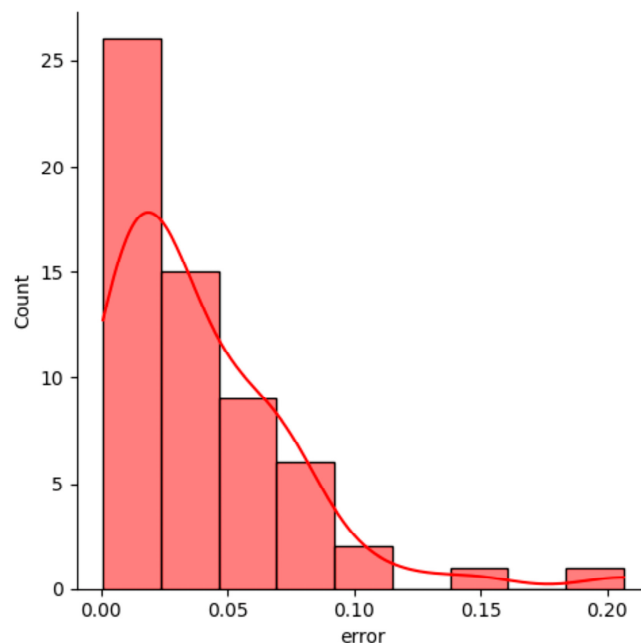
(d)

Figure 7. Comparison of predicted and actual values for the four datasets. (a) Comparison of predicted and actual values for the spring dataset; (b) comparison of predicted and actual values for the summer dataset; (c) comparison of predicted and actual values for the autumn dataset; (d) comparison of predicted and actual values for the winter dataset.

The prediction error histograms for the four seasonal data sets are shown in Figure 8. It could be seen that the maximum error was less than 1 m/s. According to the prediction evaluation index, the prediction accuracy of the hybrid model met the requirements.



(a)



(b)

Figure 8. Cont.

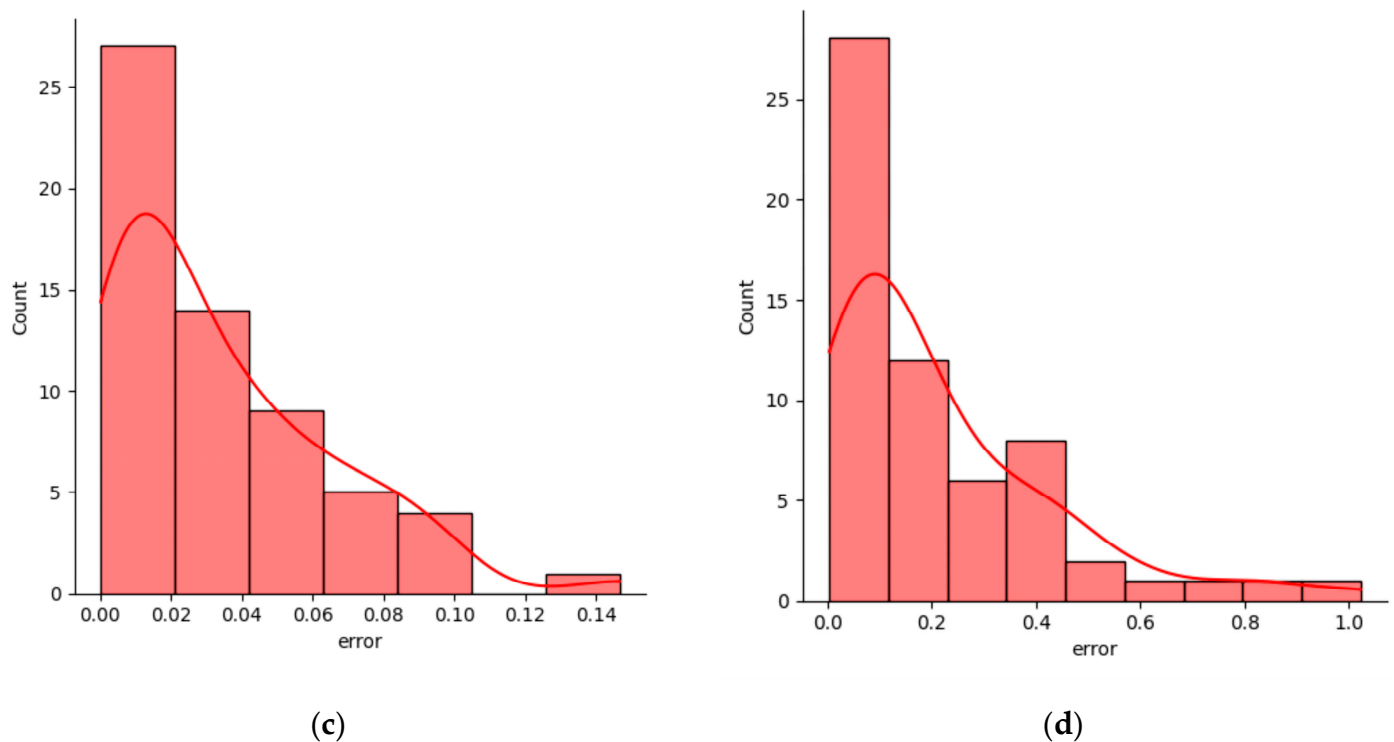


Figure 8. Histogram of the error distribution of predicted and actual values for the four datasets. (a) Histogram of the error distribution between predicted and actual values for the spring dataset; (b) histogram of the error distribution between predicted and actual values for the summer dataset; (c) histogram of the error distribution between predicted and actual values for the autumn dataset; (d) histogram of the error distribution between predicted and actual values for the winter dataset.

4.4. Comparison of Prediction Results of Different Models

4.4.1. Description of the Control Experiment

In order to measure the accuracy and feasibility of the proposed hybrid model, the SARIMAX model and the LSTM model were used as a comparison group for outcome prediction while using the same data set. The control flow is shown in Figure 9, and the module composition of the different prediction models is shown in Table 4. Each data set was divided into two data types. The first one contained only wind speed data for feature expansion, and was labeled as class A. The second one included wind speed data and 12 other relevant meteorological featured data, and was labeled as class B.

Table 4. The module composition of the different prediction models.

Model Name	Module Composition	Data Type
SARIMAX Model	SARIMAX	Class A
LSTM_Tsfresh_PCA Model	Tsfresh, Sparse PCA, LSTM Network	Class A & Class B
GRU_Tsfresh_PCA Model	Tsfresh, Sparse PCA, GRU Network	Class A & Class B
LSTM_Mul Model	LSTM Network	Class A & Class B
GRU_Mul Model	GRU Network	Class A & Class B

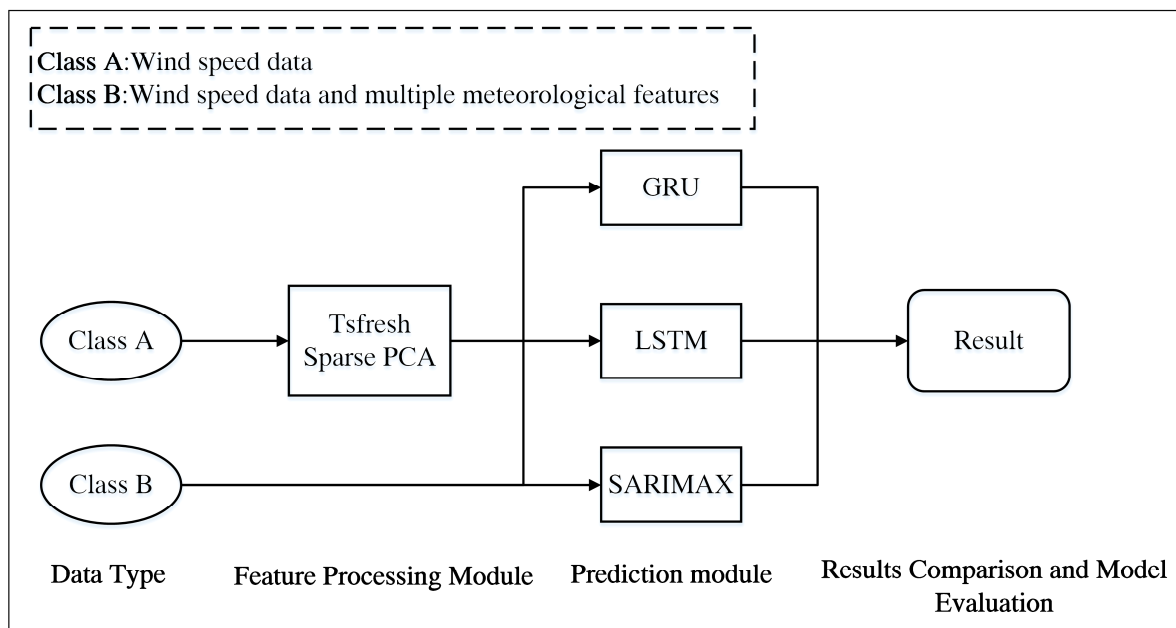


Figure 9. Experimental flow chart between dataset and model.

In terms of SARIMAX model parameter configuration, the parameters of the SARIMAX model were found by grid search to find the optimal hyperparameters. The LSTM network used a four-layer network structure, with the second and third layers being the hidden layers. The GRU network used a three-layer network structure, with the second layer being the hidden layer. GRU and LSTM networks used the same initial parameters, which included epoch, learning_rate, dropout, and activation function, etc. The hyperparameter optimization framework was used to find the optimal hyperparameters during the training phase.

4.4.2. Controlled Experimental Evaluation

In this part, the different forecasting results of all forecasting models on the four data sets are analyzed and discussed in detail. The *MSE*, *MAE*, *RMSE*, and *MAPE* metrics for all the aforementioned prediction models are shown in Table 5. The decreasing rates of the evaluation indicators obtained from the proposed model and the relevant comparison model are shown in Table 6.

Table 5. Evaluation of the prediction results of the five models.

Datasets	Data Type	Model Name	MSE	MAE	RMSE	MAPE
spring	Class A	SARIMAX	0.18415	0.34181	0.42913	21.5437
	Class B	LSTM_Tsfresh_PCA	0.00799	0.07536	0.08941	11.1132
	Class B	GRU_Tsfresh_PCA	0.01497	0.11156	0.12236	18.0979
	Class A	GRU_Mul	0.25500	0.49914	0.50498	36.8452
	Class A	LSTM_Mul	0.08446	0.26586	0.29063	17.4014
summer	Class A	SARIMAX	0.4618	0.57048	0.67962	46.122
	Class B	LSTM_Tsfresh_PCA	0.00359	0.04922	0.05996	13.6168
	Class B	GRU_Tsfresh_PCA	0.00108	0.02384	0.0329	9.8534
	Class A	GRU_Mul	0.22479	0.22479	0.47412	27.2866
	Class A	LSTM_Mul	0.21244	0.42376	0.46091	25.7521
autumn	Class A	SARIMAX	0.23448	0.34644	0.48424	14.6314
	Class B	LSTM_Tsfresh_PCA	0.00199	0.03317	0.04467	11.3925
	Class B	GRU_Tsfresh_PCA	0.00184	0.03148	0.04294	12.1422
	Class A	GRU_Mul	0.65398	0.80724	0.80869	39.6676
	Class A	LSTM_Mul	0.30338	0.54856	0.55079	26.7687

Table 5. Cont.

Datasets	Data Type	Model Name	MSE	MAE	RMSE	MAPE
winter	Class A	SARIMAX	0.39196	0.50293	0.62606	17.906
	Class B	LSTM_Tsfresh_PCA	0.01699	0.11946	0.13033	8.7932
	Class B	GRU_Tsfresh_PCA	0.01403	0.08312	0.11846	18.4467
	Class A	GRU_Mul	0.39577	0.48542	1.9894	36.0289
	Class A	LSTM_Mul	0.20425	0.61302	0.61429	23.9961

Table 6. Decreasing rates of evaluation metrics of other models and the proposed model.

Datasets	Model Name	P _{MSE} (%)	P _{MAE} (%)	P _{RMSE} (%)	P _{MAPE} (%)
spring	SARIMAX	0.918708	0.67362	0.714865	0.159945
	LSTM_Tsfresh_PCA	−0.87359	−0.48036	−0.36853	−0.6285
	GRU_Mul	0.941294	0.776496	0.757693	0.508813
	LSTM_Mul	0.822756	0.580381	0.578984	−0.04003
summer	SARIMAX	0.997661	0.958211	0.951591	0.786362
	LSTM_Tsfresh_PCA	0.699164	0.515644	0.451301	0.276379
	GRU_Mul	0.995196	0.893945	0.930608	0.638892
	LSTM_Mul	0.994916	0.943742	0.928619	0.617375
autumn	SARIMAX	0.992153	0.909133	0.911325	0.909133
	LSTM_Tsfresh_PCA	0.075377	0.05095	0.05095	0.05095
	GRU_Mul	0.997186	0.961003	0.961003	0.961003
	LSTM_Mul	0.993935	0.942613	0.942613	0.942613
winter	SARIMAX	0.964206	0.834728	0.834728	0.834728
	LSTM_Tsfresh_PCA	0.17422	0.304202	0.304202	0.304202
	GRU_Mul	0.96455	0.828767	0.828767	0.828767
	LSTM_Mul	0.93131	0.864409	0.864409	0.864409

Through comprehensive analysis of Tables 5 and 6, the following conclusions can be drawn:

- (1) Comparing the values of the single model and neural network model on the metrics *MSE*, *MAE*, *RMSE*, and *MAPE*, it was observed that the neural network model achieved superior performance on the four metrics, which indicated that the neural network could contribute to improving the model prediction. Taking the summer dataset as an example, the *MAE* of the SARIMAX model was 0.57048, and the *MAE* of the four neural network models were 0.4237, 0.2248, 0.0492, and 0.0238, respectively. The same situation occurred for *MSE*, *RMSE*, and *MAPE*. Similarly, the law continued to be maintained for the other three data sets.
- (2) In the neural network model, the neural network with the feature extraction module performed better than the neural network model using only 12 types of climate data as feature data. Taking the summer dataset as an example, the *MAE* of the SARIMAX model was 0.57048, and the *MAE* of the four neural network models were 0.4237, 0.2248, 0.0492, and 0.0238, respectively. The same situation was observed for *MSE*, *RMSE*, and *MAPE*. Similarly, the law continued to be maintained for the other three data sets.
- (3) Among the proposed hybrid models, the GRU_Tsfresh_PCA model had better predictive power than the LSTM_Tsfresh_PCA model in the summer, autumn, and winter datasets. In the spring dataset, the prediction accuracy of the two models was the same. Taking the winter dataset as an example, the GRU_Tsfresh_PCA model decreased 17.42% in *MSE*, 30.42% in *MAE*, 9.1% in *RMSE*, and 1.8% in *MAPE* compared with the LSTM_Tsfresh_PCA model.

4.4.3. Visual Analysis of the Prediction Results

In this part, the visualization of the experimental prediction results for each of the four data sets is discussed. The comparison of prediction curves for all experimental models is shown in Figure 10. The scatter plot of the degree of fit of each model is shown in Figure 11. It can be clearly seen from Figure 10 that the fitted curves of the individual models deviated from the actual wind speed values as the prediction length kept increasing, while the prediction curves of the hybrid model based on feature extraction were relatively stable. Moreover, based on the scatter plots of different data sets, it was clearly seen that the actual wind speed prediction values of the proposed hybrid model were most evenly distributed on the regression line. Compared with other comparative models, the prediction accuracy of the proposed model did not change significantly with the change of the data set. Therefore, the predictive ability of the hybrid model was well validated.

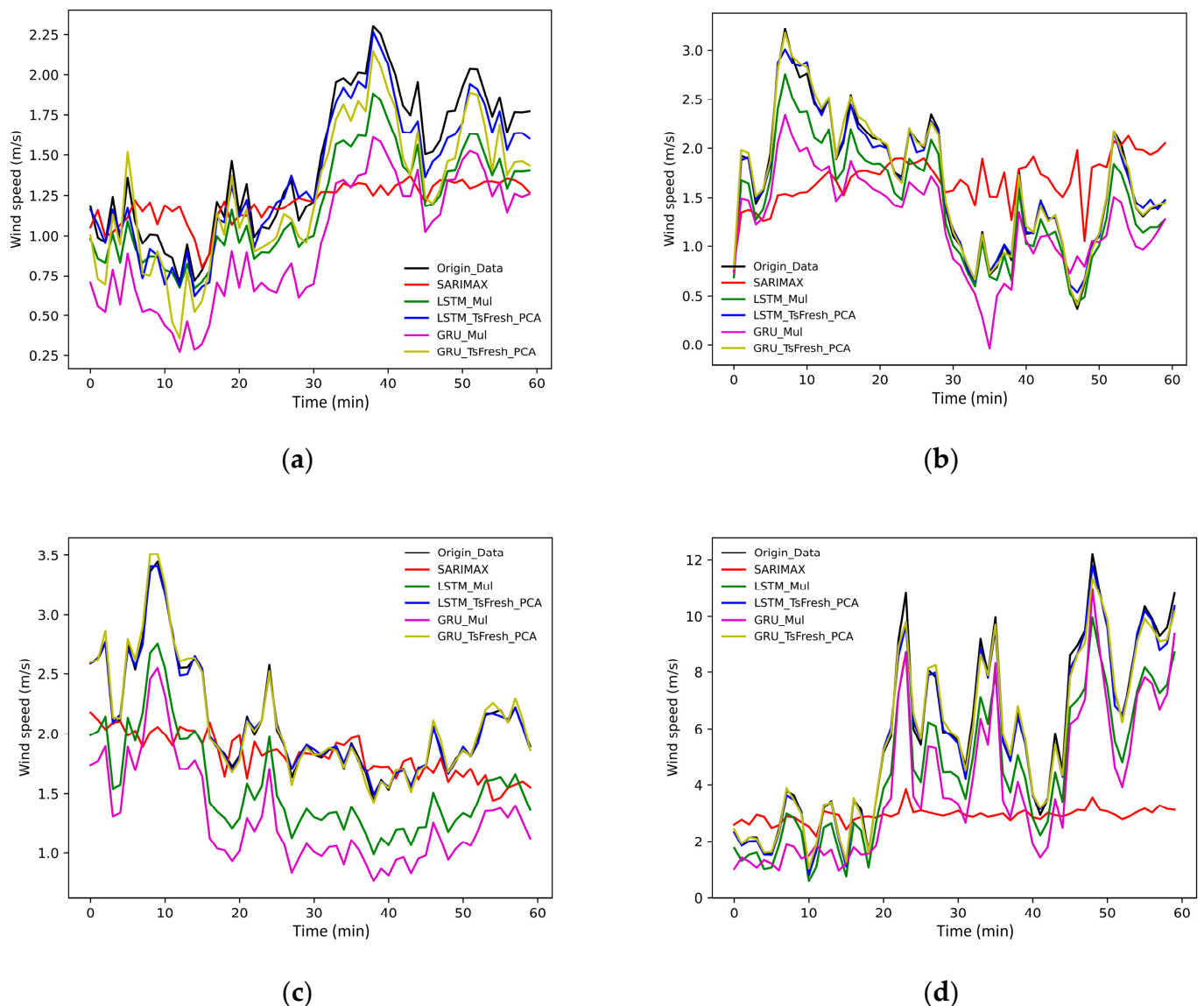


Figure 10. Prediction results of five models. (a) Five model predictions for the spring dataset; (b) five model predictions for the summer dataset; (c) five model predictions for the autumn dataset; (d) five model predictions for the winter dataset.

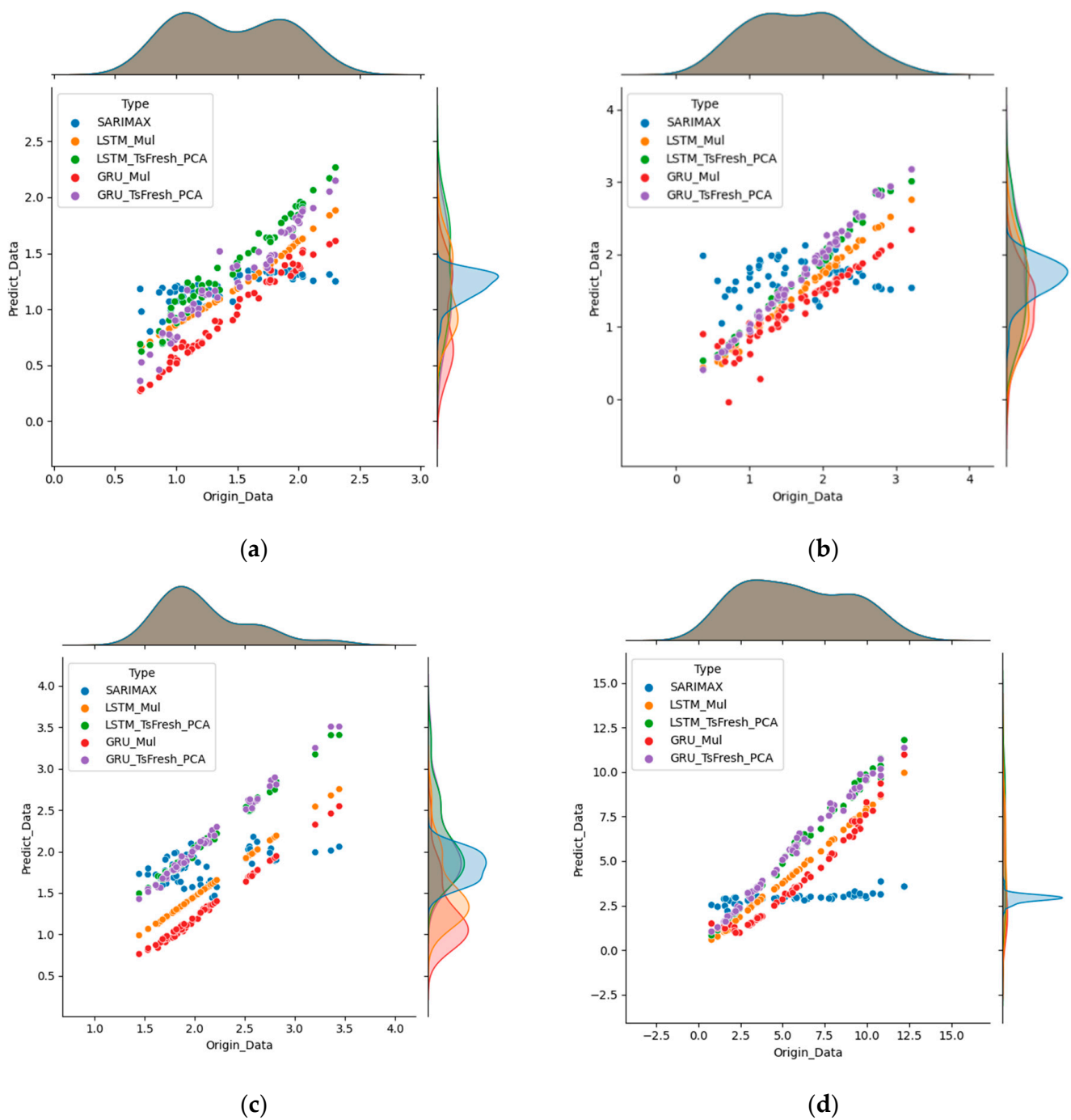


Figure 11. Scatter plots of different model fit. (a) Scatterplot of model fit for the spring dataset; (b) scatterplot of model fit for the summer dataset; (c) scatterplot of model fit for the autumn dataset; (d) scatterplot of model fit for the winter dataset.

In addition, the evaluation metrics shown in Table 5 are visualized in Figure 12 in the form of histograms. In the same subplot, each metric for different models in different datasets was shown sequentially. Specifically, it was observed from all the prediction results that the four metrics representing the prediction performance of GRU_TsFresh_PCA in the hybrid model were the smallest.

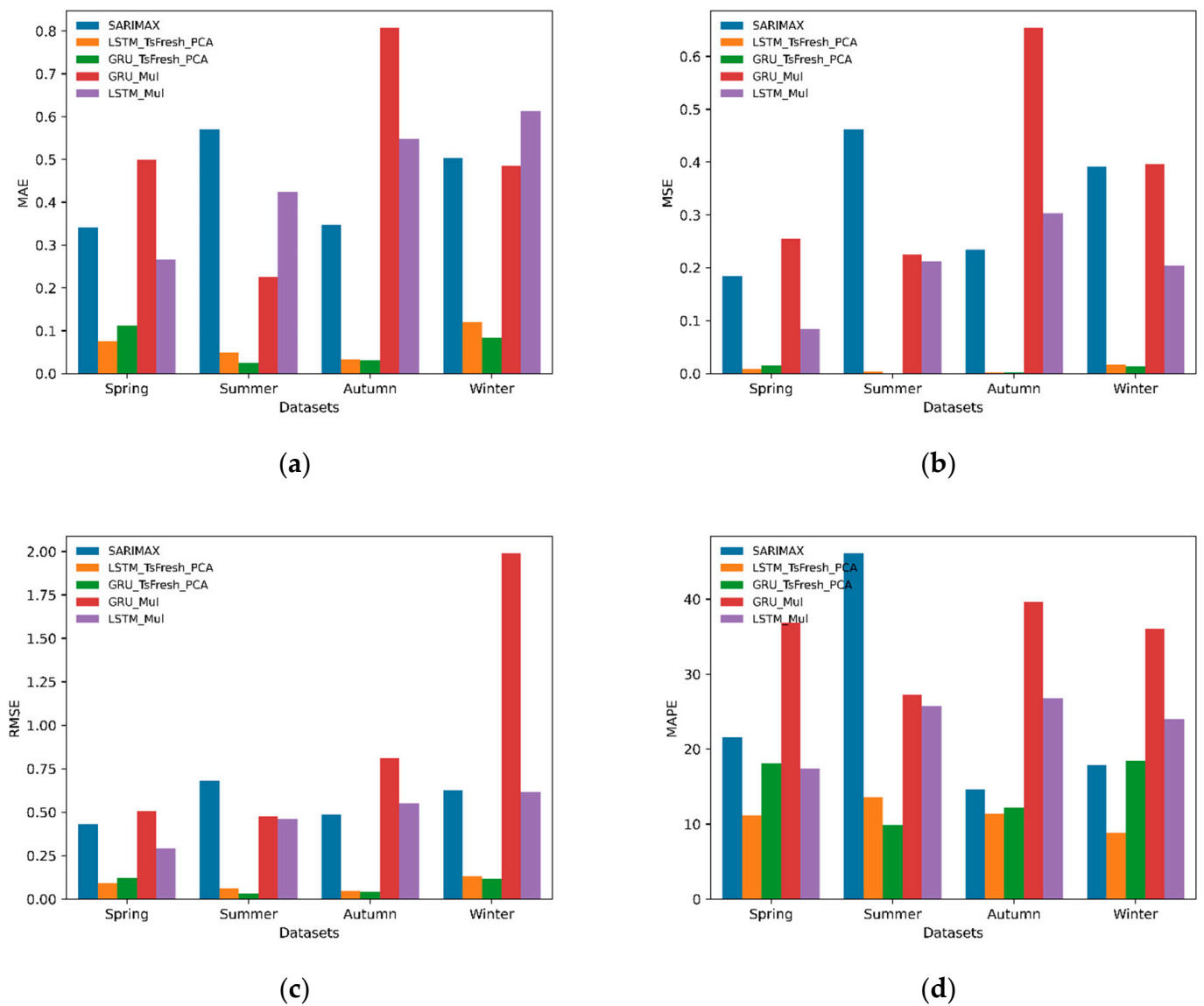


Figure 12. Comparison of the evaluation of different models. (a) MAE comparison of different models; (b) MSE comparison of different models; (c) RMSE comparison of different models; (d) MAPE comparison of different models.

To further evaluate the performance of the LSTM_TsFresh_PCA and GRU_TsFresh_PCA models, the model computation times were compared under the same computational conditions, as shown in Figure 13. In addition, the starting and ending moments of time were identical. From Figure 13, it was found that the GRU_TsFresh_PCA computation time was always lower than that of the LSTM_TsFresh_PCA model. This meant that GRU_TsFresh_PCA outperformed the LSTM_TsFresh_PCA model under the same conditions.

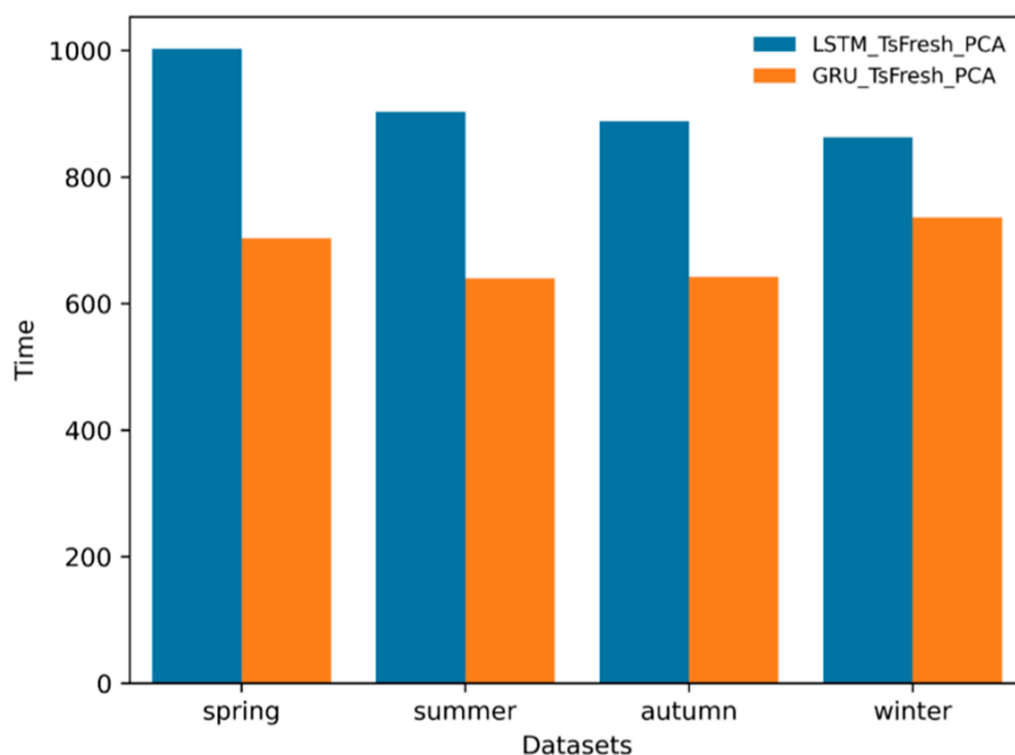


Figure 13. Comparison of prediction time of LSTM_TsFresh_PCA model and GRU_TsFresh_PCA model.

5. Conclusions

For the ultra-short-term wind speed prediction problem, a hybrid prediction method based on the GRU network, sparse PCA, and Tsfresh was proposed in this paper. The model first processed the univariate wind speed data for feature extraction and feature selection. Finally, the wind speed was predicted by the GRU network. To verify the predictive ability of the proposed model, this paper performed a comparative validation using different datasets and different models, respectively. The experimental results showed that:

- (1) The proposed model GRU_TsFresh_PCA had the best prediction results.
- (2) Experimentally, it was demonstrated that the GRU network outperformed the LSTM network under the same computational conditions after the wind speed data were processed by features.
- (3) The experimental results showed that the feature expansion could not only fit the existing meteorological features, but also fit the feature data more suitable for wind speed prediction than using the meteorological feature data related to the original wind speed data directly. These data were not directly obtained in reality.
- (4) In summary, the proposed method in this paper could be expected to have good application prospects in the future because of its low computational cost and good prediction performance.

However, there were still some limitations in this study, such as the quality of the raw wind speed data. In future research work, we will further optimize the model based on the prediction effect of the actual wind speed data to further improve the generalization ability of the model.

Author Contributions: Software, Y.W.; validation, Y.W. and R.G.; investigation, Y.W. and R.G.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W.; visualization, Y.W.; supervision, R.G.; project administration, R.G.; funding acquisition, R.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: For original wind speed data address, please visit <https://midcdmz.nrel.gov/apps/sitehome.pl?site=NWTC>, accessed on 5 July 2022. The processed data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Balat, M. A review of modern wind turbine technology. *Energy Sources Part A* **2009**, *31*, 1561–1572. [\[CrossRef\]](#)
- Bhadane, K.V.; Ballal, M.S.; Moharil, R.M. Investigation for causes of poor power quality in grid connected wind energy-A review. In Proceedings of the 2012 Asia-Pacific Power and Energy Engineering Conference, Shanghai, China, 27–29 March 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–6.
- Khadem, S.K.; Basu, M.; Conlon, M. Power quality in grid connected renewable energy systems: Role of custom power devices. *Renew. Energy Power Qual. J.* **2010**, *1*, 878–881. [\[CrossRef\]](#)
- Li, N.; He, F.; Ma, W. Wind power prediction based on extreme learning machine with kernel mean p-power error loss. *Energies* **2019**, *12*, 673. [\[CrossRef\]](#)
- Giebel, G.; Kariniotakis, G. Best practice in short-term forecasting—A users guide. In Proceedings of the CD-Rom Proceedings European Wind Energy Conference, London, UK, 7 May 2007.
- Aburiyana, G.; El-Hawary, M.E. An overview of forecasting techniques for load, wind and solar powers. In Proceedings of the 2017 IEEE Electrical Power and Energy Conference (EPEC), Saskatoon, SK, Canada, 22–25 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–7.
- Mosaico, G.; Saviozzi, M. A hybrid methodology for the day-ahead PV forecasting exploiting a Clear Sky Model or Artificial Neural Networks. In Proceedings of the IEEE EUROCON 2019—18th International Conference on Smart Technologies, Novi Sad, Serbia, 1–4 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
- Chen, J.F.; Wang, W.M.; Huang, C.M. Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting. *Electr. Power Syst. Res.* **1995**, *34*, 187–196. [\[CrossRef\]](#)
- Jeong, K.; Koo, C.; Hong, T. An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network). *Energy* **2014**, *71*, 71–79. [\[CrossRef\]](#)
- Tyass, I.; Bellat, A.; Raihani, A.; Mansouri, K.; Khalili, T. Wind Speed Prediction Based on Seasonal ARIMA model. *E3S Web Conf. EDP Sci.* **2022**, *336*, 00034. [\[CrossRef\]](#)
- Kritharas, P. Developing a SARIMAX Model for Monthly Wind Speed Forecasting in the UK. Ph.D. Thesis, Loughborough University, Loughborough, UK, 2014.
- Lydia, M.; Kumar, S.S.; Selvakumar, A.I.; Kumar, G.E.P. Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Convers. Manag.* **2016**, *112*, 115–124. [\[CrossRef\]](#)
- Duan, J.; Zuo, H.; Bai, Y.; Duan, J.; Chang, M.; Chen, B. Short-term wind speed forecasting using recurrent neural networks with error correction. *Energy* **2021**, *217*, 119397. [\[CrossRef\]](#)
- Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
- Zhu, Q.; Chen, J.; Shi, D.; Zhu, L.; Bai, X.; Duan, X.; Liu, Y. Learning temporal and spatial correlations jointly: A unified framework for wind speed prediction. *IEEE Trans. Sustain. Energy* **2019**, *11*, 509–523. [\[CrossRef\]](#)
- Liu, M.; Qiu, P.; Wei, K. Research on wind speed prediction of wind power system based on GRU deep learning. In Proceedings of the 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2), Changsha, China, 8–10 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1699–1703.
- Li, C.; Tang, G.; Xue, X.; Saeed, A.; Hu, X. Short-term wind speed interval prediction based on ensemble GRU model. *IEEE Trans. Sustain. Energy* **2019**, *11*, 1370–1380. [\[CrossRef\]](#)
- Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [\[CrossRef\]](#)
- Zhou, B.; Liu, C.; Li, J.; Sun, B.; Yang, J.; Arpino, F. A hybrid method for ultrashort-term wind power prediction considering meteorological features and seasonal information. *Math. Probl. Eng.* **2020**, *2020*, 1795486. [\[CrossRef\]](#)
- Jung, J.; Broadwater, R.P. Current status and future advances for wind speed and power forecasting. *Renew. Sustain. Energy Rev.* **2014**, *31*, 762–777. [\[CrossRef\]](#)
- Tang, J.; Alelyani, S.; Liu, H. Feature selection for classification: A review. In *Data Classification: Algorithms and Applications*; Chapman and Hall: New York, NY, USA, 2014; p. 37. [\[CrossRef\]](#)
- Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [\[CrossRef\]](#)
- Christ, M.; Braun, N.; Neuffer, J.; Kempa-Liehr, A.W. Time series feature extraction on basis of scalable hypothesis tests (Tsfresh—A python package). *Neurocomputing* **2018**, *307*, 72–77. [\[CrossRef\]](#)
- Zhou, G.B.; Wu, J.; Zhang, C.L.; Zhou, Z.H. Minimal gated unit for recurrent neural networks. *Int. J. Autom. Comput.* **2016**, *13*, 226–234. [\[CrossRef\]](#)

-
25. Rana, R. Gated recurrent unit (GRU) for emotion classification from noisy speech. *arXiv* **2016**, arXiv:1612.07778.
 26. Cai, T.T.; Ma, Z.; Wu, Y. Sparse PCA: Optimal rates and adaptive estimation. *Ann. Stat.* **2013**, *41*, 3074–3110. [[CrossRef](#)]
 27. Guo, J. *Backpropagation through Time*, Harbin Institute of Technology: Harbin, China, 2013; Volume 40, 1–6, *unpublished work*.
 28. Jager, D.; Andreas, A. *NREL National Wind Technology Center (NWTC): M2 Tower; Boulder, Colorado (Data)*; National Renewable Energy Lab. (NREL): Golden, CO, USA, 1996.