

## Article

# Forecasting the 10.7-cm Solar Radio Flux Using Deep CNN-LSTM Neural Networks

Junqi Luo <sup>1,2</sup> , Liucun Zhu <sup>1,\*</sup>, Kunlun Zhang <sup>1</sup>, Chenglong Zhao <sup>1</sup> and Zeqi Liu <sup>3</sup>

<sup>1</sup> Advanced Science and Technology Research Institute, Beibu Gulf University, Qinzhou 535000, China; luo.junqi@bbgu.edu.cn (J.L.); zhangkunlun1834@bbgu.edu.cn (K.Z.); zhaochenglong@bbgu.edu.cn (C.Z.)

<sup>2</sup> School of Electric and Information Engineering, Beibu Gulf University, Qinzhou 535000, China

<sup>3</sup> Fangchenggang Meteorological Bureau, Fangchenggang 538000, China; 2011401005@st.gxu.edu.cn

\* Correspondence: lczechu@bbgu.edu.cn

**Abstract:** Predicting the time series of 10.7-cm solar radio flux is a challenging task because of its daily variability. This paper proposed a non-linear method, a convolutional and recurrent neural network combined model to achieve end-to-end F10.7 forecasts. The network consists of a one-dimensional convolutional neural network and a long short-term memory network. The CNN network extracted features from F10.7 original data, then trained the feature signals in the long short-term memory network, and outputted the predicted values. The F10.7 daily data during 2003–2014 are used for the testing set. The mean absolute percentage error values of approximately 2.04%, 2.78%, and 4.66% for 1-day, 3-day, and 7-day forecasts, respectively. The statistical results of evaluating the root mean square error, spearman correlation coefficient shows a superior effect as a whole for the 1–27 days forecast, compared with the ordinary single neural network and combination models.

**Keywords:** solar radio flux; time series forecast; convolutional neural network; long short-term memory



**Citation:** Luo, J.; Zhu, L.; Zhang, K.; Zhao, C.; Liu, Z. Forecasting the 10.7-cm Solar Radio Flux Using Deep CNN-LSTM Neural Networks. *Processes* **2022**, *10*, 262. <https://doi.org/10.3390/pr10020262>

Academic Editor: Chien-Chih Wang

Received: 4 November 2021

Accepted: 26 January 2022

Published: 28 January 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

The 10.7 cm solar radio flux is the solar radio emission intensity in a 100-MHz-wide band centered at 2800 MHz (i.e., 10.7-cm wavelength). It is measured in units defined as solar flux units (sfu), where  $1 \text{ sfu} = 10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$ . The systematic F10.7 record started in 1947, measured by the Canada National Research Institute [1]. The F10.7 is primarily raised by the upper chromosphere and the base of the corona. It has a close relationship with the active regions of the sun. In addition, the formation of F10.7 includes thermal free-free emission and thermal gyroscopic resonance processes [2,3]. The thermal free-free emission takes place in the chromosphere and corona, as well as in plasma concentrations in the magnetic fields that support the active regions of the chromosphere and corona. On another hand, gyroresonance emissions occurs near sunspots with sufficiently strong magnetic fields. F10.7 is more measurable and independent of the weather on Earth than other metrics, such as sunspot number, extreme ultraviolet (EUV), ultraviolet (UV) and X-rays [4]. It is often used as an input parameter to predict and reconstruct atmospheric density in low-earth orbit (Ahluwalia 2016) and the density distribution of ionospheric electrons [5]. The prediction of F10.7 is of great significance in the aerospace field [6,7]. However, due to its daily variability, the high-precision forecasting in the short- and mid-term are often challenging to accomplish.

The traditional prediction method for F10.7 is to construct an autoregressive model, which usually uses linear mathematical expressions to express the functional relationship between the forecast value and the observed value. In Ref. [8], a Fourier series method, based on the short-period oscillation of solar radiation, was proposed to evaluate the effect of the F10.7 prediction. In Ref. [9], a 54-order autoregression model was used to forecast F10.7 for the subsequent 27 days, and the results showed that the average relative error of the forecast from 2003 to 2007 was 9.52%, which was similar to the accuracy of the prediction

model used by the US Air Force. Ref. [10] adopted a recursive linear autoregressive model, which predicted the future one day in the first  $n$  days. Then, the predicted value would be incorporated into the input value of the model and used for the next prediction. Ref. [4] proposed a linear model, which used the observations of the first 81 days to predict F10.7 for the next 1–45 days. The results showed that the correlation coefficient prediction correlation, with this method was 0.99 and 0.85 for 1-day and 40-day forecasting, respectively.

Furthermore, specific external parameters from solar activities have predicted F10.7 by physical approaches, such as an EUV image of the solar disk, the sunspot number, the ionospheric parameter  $F_o$  (critical frequency of the ionospheric layer) and magnetic flux. Ref. [11] used the index  $P_{SR}$  defined by the intensity values of solar EUV images to establish an empirical model for the F10.7 prediction. Refs. [3,12] have also successfully achieved short- and mid-term prediction of F10.7 by constructing a magnetic flux transport model, and its Spearman correlation index is generally higher than 0.95 for 3-day forecasts.

Most studies in F10.7 forecasting have only focused on constructing linear models, which usually have a relatively stable effect in mid- and long-term prediction. A significant drawback of the above methods is their lack of the high-quality short- and mid-term forecasting abilities. There had been substantial research on machine learning undertaken to alleviate this dilemma. In Ref. [13], a support vector machine (SVR) was used to make a short-term prediction of F10.7. This method reduced the computational complexity in the training process through a learning-based algorithm and achieves an accuracy close to that of traditional feed-forward neural network with fewer data. However, limited by the sample size, the literature does not test the case for moderate solar activity years. In addition, the lack of reference models makes it impossible to know its level of prediction. Ref. [14] adopted the backpropagation (BP) neural network algorithm to build the F10.7 prediction model, using the first 54 days as the input of the neural network to predict the future 1–3 days. The results show that this method is superior to the SVR method in short-term prediction. In order to further improve the prediction effect, some scholars tried to extract the signal characteristics of the original data by the signal processing method and then predict them by machine learning method. Common ones include empirical mode decomposition (EMD) [15,16], Fourier transform [17] and wavelet analysis [18,19]. Relevant literature above have exemplified the effectiveness of machine learning methods, signal processing technique and a combination of the two, but still faces the following limitations: (1) Traditional machine learning methods are usually shallow neural networks. With the increase of network layers, gradient disappearance or gradient explosion will occur, which will decrease detection accuracy. (2) Limited by the network depth, the shallow neural network algorithm extracts mostly surface signal's features, which is not sufficient to explain the model. In many cases, it still needs to extract sample features manually as the input samples. (3) The signal processing method may cause the loss or distortion of some signal features. In addition, signal features depend on manual experience, which easily causes the instability of feature extraction. These factors will limit the prediction effect of (2).

We proposed a time series prediction model for the short- and mid-term F10.7 forecasting to address the above deficiencies. The model is comprised two independent networks: the feature extraction network and the predicting network, respectively devised by a one-dimension convolutional neural network (CNN) and a long short-term memory network (LSTM). In the F10.7 prediction task, CNN first adaptively extracted the signal features from raw time series. These feature samples were then trained and predicted by the LSTM. The main contributions were summarized as follows: (1) Design a one-dimensional CNN to use in sequences, realizing the end-end forecasting model for the F10.7 data. (2) Design the comparative experiments to verify the strong robustness of the proposed model under a variant year and term conditions. (3) Use the combination model of CNN and LSTM, which benefit from the strength of both. As a result, the CNN-LSTM model provides a better prediction effect than other models.

The rest of this paper has been divided into five parts. Section 2 introduces the theoretical background of CNN and LSTM. Section 3 presents the framework of the proposed model. Section 4 deals with the results of the case studies. Section 5 gives the conclusion.

## 2. Theoretical Network

### 2.1. Convolutional Neural Network

CNN was proposed by Yann Lecun in 1998 [20] and has been primarily used in computer image processing since its proposal. The feature extraction network and classification network are contained in the CNN model. The feature extraction network consists of a convolutional and pooling layer, which extracts and compresses the input signals by a convolution operation and a down-sampling operation. The classification network consists of a fully connected layer. The parameters and weights of the two network parts are updated by forwarding propagation and error back-propagation mechanisms. The convolutional layer is a representative layer of the CNN network. In essence, it is a series of digital filters from which features extraction is possible. A convolution kernel used the input signal to extract the corresponding features by a local convolution operation. In addition, the local connectivity and weight sharing nature in the convolutional layer have been able to significantly reduce the number of network parameters to accelerate training and avoid over-fitting. The convolution operation formula is shown in Formula (1):

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * \omega_{ij}^l + b_j^l\right). \quad (1)$$

where  $l$  represents the layer number,  $i$  and  $j$  denote the serial number of the feature map, and  $M_j$  is the feature map of the  $(l-1)^{th}$  layer connected with the  $j^{th}$  feature map of the  $l^{th}$  layer.  $\omega_{ij}^l$  represents the convolution kernel weight of the  $i^{th}$  feature graph of the  $(l-1)^{th}$  layer that connected with the  $l^{th}$  layer's  $j^{th}$  feature graph.  $b_j^l$  represents the bias of the  $j^{th}$  feature graph at the  $l^{th}$  layer; The symbol  $*$  is the convolution operation.

The pooling layer is also known as the down-sampling layer, which is generally located between two convolution layers and aims to reduce the dimension of the feature graph while retaining important feature information. Pooling operations generally include maximum pooling, average pooling, and summation pooling. The maximum pooling is used more often, and is mathematically expressed as follows:

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l). \quad (2)$$

where  $\text{down}()$  is the down-sampling function;  $x_j^{l-1}$  represents the  $j^{th}$  feature map of the  $(l-1)^{th}$  layer;  $\beta_j^l$  and  $b_j^l$  represent the weight and bias of the  $j^{th}$  feature graph in the  $l^{th}$  layer, respectively.

The fully connected layer is located at the end of the CNN, usually using a traditional multilayer perceptron. The high dimensional feature signals are usually reduced to one-dimensional feature vectors before being input into the fully connected layer. The input layer and the output layer are fully connected. The activation function is used in the output layer to achieve multi-classification output. The forward propagation formula of full connection is as follows:

$$z_j^{l+1} = \sum_{i=1}^n \omega_{ij}^l a_i^l + b_j^l. \quad (3)$$

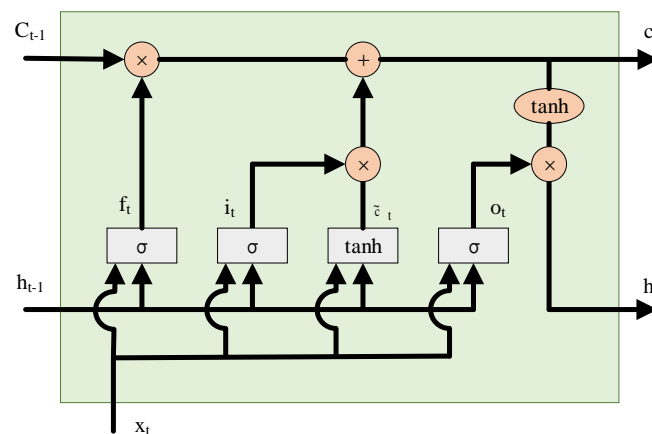
where  $\omega_{ij}^l$  represents the weight between the  $i^{th}$  neuron of the  $l^{th}$  layer and the  $j^{th}$  neuron of the  $(l+1)^{th}$  layer;  $b_j^l$  is the bias of the  $j^{th}$  neuron of the  $l^{th}$  layer;  $a_i^l$  is the activation value of the  $i^{th}$  neuron of the  $l^{th}$  layer;  $z_j^{l+1}$  represents the output value of the  $j^{th}$  neuron at the  $(l+1)^{th}$  layer.

## 2.2. Implementation Scheme of CNN for F10.7 Prediction

CNN was initially used for 2D image processing. Due to its excellent depth feature extraction capability, more and more scholars have explored its application in the field of time series prediction in recent years [21–23]. The most significant difference between CNN in time series prediction and image recognition tasks is the type of input signal. The original signal is two-dimensional data in the image recognition task, which can be used directly for training. In contrast, the input signal is usually one-dimensional in the time series prediction task, such as vibration, voltage, and sound signals. In order to be adopted by CNN, it is necessary to make corresponding changes. Such as the convolution kernel and pooling layer need to be reset to one-dimensional dimensions, or the time-series signal reconstructed for two-dimensional vector. Compared with the latter scheme, using the original one-dimensional signal as the network's input will be beneficial in preserving the original information and reducing the computational effort. It takes full advantage of the characteristics of the CNN network integrating automatic signal feature extraction, signal dimensionality reduction, feature selection and pattern classification to achieve “end-to-end” prediction. Prior studies have adopted the above scheme and verified the algorithm's validity [24,25].

## 2.3. Long Short-Term Memory Network

LSTM model is an improved RNN architecture proposed by Hochreiter [26]. Its special hidden units endow the capacity to store long-term information. Recently, a substantial body of literature has demonstrated that the LSTM models yield impressive results in a variety of time series tasks, including language modelling [27], traffic speed prediction [28] and travel time prediction [29]. The internal structure of the LSTM unit is shown in Figure 1. The training process of LSTM includes forward calculation and error backpropagation, in which forward calculation is mainly realized by forget gate, input gate and output gate. The specific formula of forwarding calculation is as follows:



**Figure 1.** Structure of cell in LSTM. Adapted from Le [30].

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f). \quad (4)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i). \quad (5)$$

$$\tilde{c}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c). \quad (6)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t. \quad (7)$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o). \quad (8)$$

$$h_t = o_t \circ \tanh(c_t). \quad (9)$$

where  $w_i$  and  $w_c$  are the weight matrix of the input gate.  $w_f$  and  $w_o$  are the weight matrix of forget gate and output gate, respectively.  $b$  denotes the bias of correspondent gates.

$[h_{t-1}, x_t]$  stands for concatenating  $h_{t-1}$  and  $x_t$ .  $\sigma$  and  $\tanh$  are the activation functions, shown in Formulas (10) and (11). LSTM has three inputs at time  $t$ : the network input value  $x_t$  at the current time, the LSTM output value  $h_{t-1}$  at the last time, and the unit state  $c_{t-1}$  at the last time.

$$\sigma = \frac{1}{1 + \exp(-x)}. \quad (10)$$

$$\tanh = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}. \quad (11)$$

The forgetting gate determines the proportion of unit states inherited from the preceding moment. The input gate affects how much of the current input is available to be saved to the cell state. The output gate controls part of the cell state as the current output value of the LSTM. The backpropagation of LSTM is represented as the following three steps:

Step 1. The output value of each neuron is calculated forward by Formulas (4)–(9);

Step 2. Reversely calculate the error of each neuron. Like RNN, the backpropagation of LSTM error also includes two directions: one is the backpropagation along time, that is, the error term at each moment is calculated from the current moment  $t$ ; The other is to propagate the error backwards;

Step 3. Calculate the gradient of each weight according to the corresponding error.

### 3. The Framework of the Proposed Method

#### 3.1. The Overall Framework

Some literature on the time series problems [31,32] refined the signal characteristics before the machine-learning processes, aiming to improve the predicted accuracy further. Nonetheless, such practice usually needs to be done manually, which relies on expert experience and is time-consuming. Furthermore, some signal feature extracting [33,34] may lose or distort some raw information. That would be an adverse effect on forecasting. The CNN-LSTM model in this paper implemented unattended raw signal depth feature extraction by end-to-end signal processing. It consisted of two functional networks in series: the feature extraction network (i.e., the CNN model) and the sequence prediction network (i.e., the LSTM model), and their structure is schematically shown in Figure 2, with the following core procedure:

Step 1. Normalize the F10.7 data set and divide it into the training set and test set. Take the consecutive 27-day sequence as a sample. The training samples and test samples were made using the sliding window method (sliding window set to 1). These training samples were randomly fed into the network for training.

Step 2. CNN performs feature extraction on F10.7 and outputs feature signals.

Step 3. LSTM trains the feature signals and gives the prediction results. Denormalization of the prediction results to obtain the final predicted value.

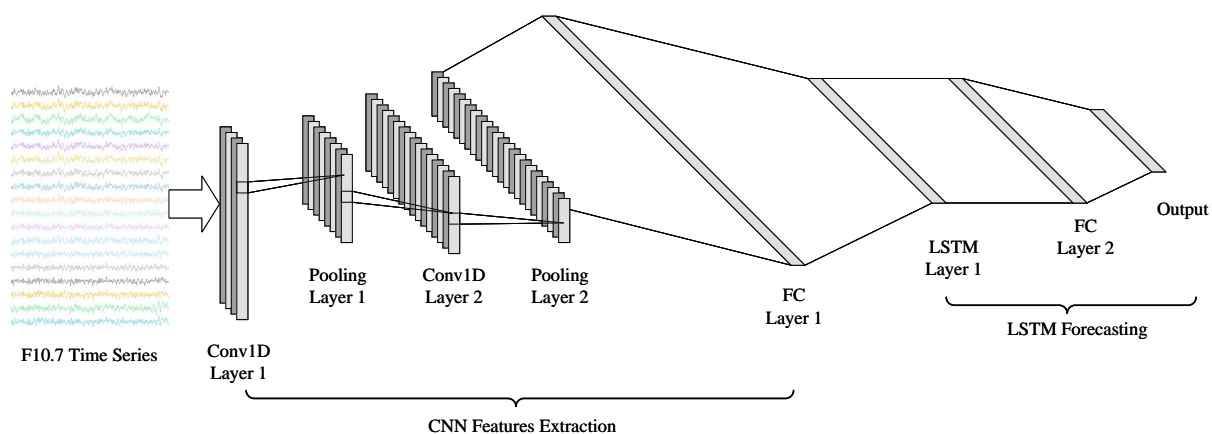


Figure 2. Structure diagram of CNN-LSTM.

### 3.2. The Networks Architectures and Parameters Settings

The CNN-LSTM is cascade connection, by CNN and LSTM, which are the role of features extraction and sequences prediction, respectively. Their detailed frame and parameters setting were shown in Table 1. The CNN are composed of convolutional layers, pooling layers and flattened layers. The CNN's input is the raw F10.7 time series. Its two convolutional layers use ReLU as the activation function and perform batch normalization. The pooling layer remains the main features and reduces the parameters and computation, prevents overfitting, and improves generalization. The flatten layer allows output the one-dimension signals to the LSTM networks. The sequences prediction network consists of a three-layer LSTM network. It uses the ADAM as the optimizer, the learning rate is set to be 0.001 and the training iterations are set to 500.

**Table 1.** Network structure and parameters Settings of CNN-LSTM.

Layer Number	Layer Type	Parameters
1	convolutional layer	filters = 16, kernel size = (1, 14), strides = (1, 2)
2	maximum pooling layer	pool size = (1, 4)
3	convolutional layer	filters = 32, kernel size = (1, 4), strides = (1, 2)
4	maximum pooling layer	pool size = (1, 2)
5	fully connected layer	flatten
6	LSTM layer	neurons number = 64
7	LSTM layer	neurons number = 32
8	fully connected layer	output layer

## 4. Experimental Results Analysis

### 4.1. Dataset Introduction

The dataset used in this paper was acquired from the Center for Space Standards and Innovation (<http://www.celestrak.com/SpaceData/SW-All.txt> accessed on 3 February 2021). Furthermore, the training and testing data selected the F10.7\_ADJ parameter, which means the observed 10.7-cm solar radio flux, adjusted to one astronomical unit. We selected the F10.7 data from 1980–2002, which included 8401 samples as the training sets. The data from 2003–2014 contained 4383 samples and was set to be the testing set. The proportion between the training set and testing set was around 2:1.

### 4.2. Evaluation Metric

In the experiment, multiple evaluation indexes are used to evaluate the prediction effect of comparative algorithms, as shown in Formulas (12)–(15). Mean Absolute Percentage Error (MAPE) is calculated by dividing the difference between the actual value and the estimated value by the actual value. The absolute value in this ratio is summed for every forecasted point in time and divided by the sample number  $n$ . Root-mean-square error (RMSE) reflects the overall stability of the forecast. The correlation coefficient ( $R$ ) and Spearman correlation coefficient ( $\rho$ ) measure the estimation and observation correlation. A high correlation coefficient indicates that the estimated value and the observed value increase or decrease synchronously, and the expected result is satisfactory.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (13)$$

$$R = \frac{\sum_{i=1}^n y_i \hat{y}_i}{\sqrt{\sum_{i=1}^n y_i^2} \sqrt{\sum_{i=1}^n \hat{y}_i^2}}. \quad (14)$$

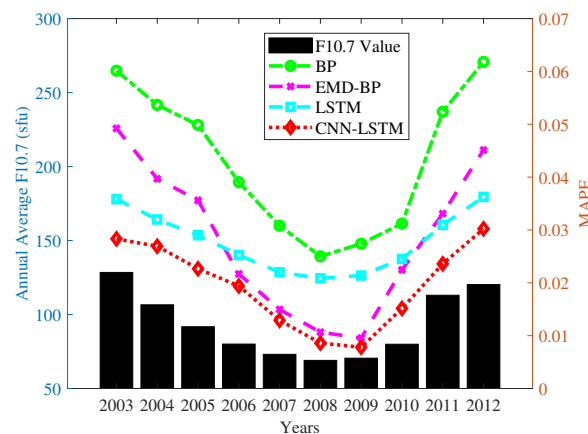


$$\rho = 1 - \frac{6 \sum_{i=1}^n (y_i - \hat{y}_i)^2}{n(n^2 - 1)}. \quad (15)$$

where  $n$  is the sample number,  $Y_i$  and  $\hat{y}_i$  denote the  $i^{th}$  observed and estimated values, correspondingly.

#### 4.3. The Annual Accuracy Analysis of F10.7 Prediction

A positive correlation between MAPE and solar activity level has been reported in the literature [4,10]. This view was also supported in this paper. Figure 3 showed the annual average MAPE results of comparative algorithms from 2003–2012 on 3-day forecasts. The comparison models include the BP model [14], a three-layer feed-forward network, the EMD-BP model [15] and a stacked model of EMD and BP. In addition, there is the traditional three-layer LSTM model. The MAPE is positive correlation with the level of annual average F10.7 for all algorithms. For each algorithm, the MAPE values in periods of moderate and high solar activity (i.e.,  $F10.7 > 100$  sfu) are higher than that of periods of low solar activity ( $F10.7 < 100$  sfu). Furthermore, the MAPE values of the BP algorithm are inferior to the other three algorithms in all solar activity years. The LSTM algorithm has a relatively stable MAPE, on the whole. The EMD-BP has a diminutive MAPE value under low solar activity but sharp deterioration under medium and high solar activity. The CNN-LSTM algorithm has the optimal performance in all testing years. Its average MAPE value in low-solar-activity years is 1.52%, lower than 3.39%, 1.91% and 2.39% of BP, EMD-BP and LSTM, respectively. The average MAPE is 2.85% in moderate- and high-solar-activity years, which is lower than 4.95%, 4.18% and 3.38% of BP, EMD-BP and LSTM, respectively.



**Figure 3.** Distribution of Yearly MAPE in different years.

#### 4.4. Short- and Mid-Term Prediction Effect Analysis

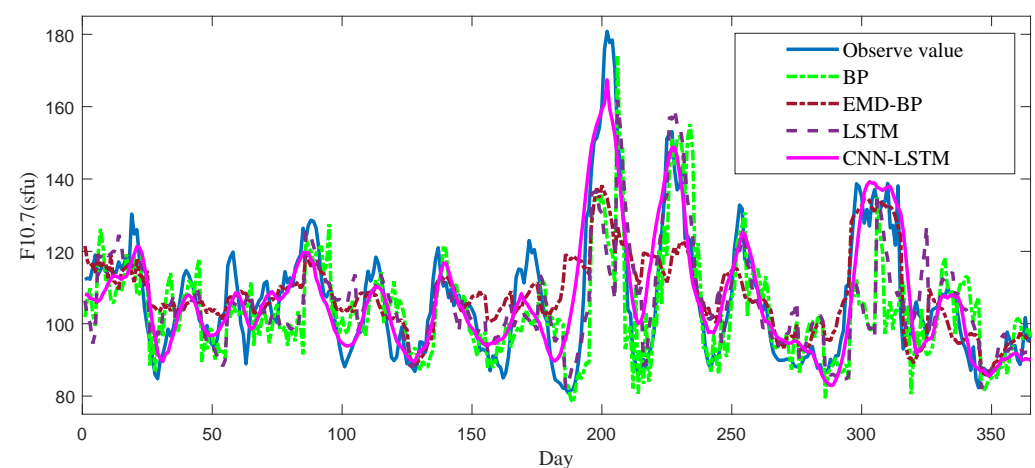
Table 2 shows the prediction results of CNN-LSTM, BP, EMD-BP and LSTM models for prediction lengths of 1–27 days. From a general view, the MAPE and RMSE values rise with the increase of forecast days. Contrarily, the Spearman correlation coefficient has a dropping trend. Our method has an average MAPE of 3.41% in the short-term forecast (i.e., 1–7 forecasting days), which has a lead of 32.7%, 21.2% and 12.3% over BP, LSTM and EMD-BP, respectively. In the index RMSE, our method has a lead of 33.9%, 25.6% and 9.0% over with the other three. In the mid-term forecast (i.e., 8–27 forecasting days), the MAPE and RMSE have a dramatic rise; our algorithm is ahead by 30.7%, 16.3% and 8.1% in MAPE and 28.5%, 20.5% and 7.4% in RMSE. As for Spearman correlation coefficients, all the algorithms are higher than 0.9 over the 1–27 days forecast, and our method also performed better.

**Table 2.** Comparison of performance between BP, LSTM, EMD-BP and CNN-LSTM by multi-step forecast.

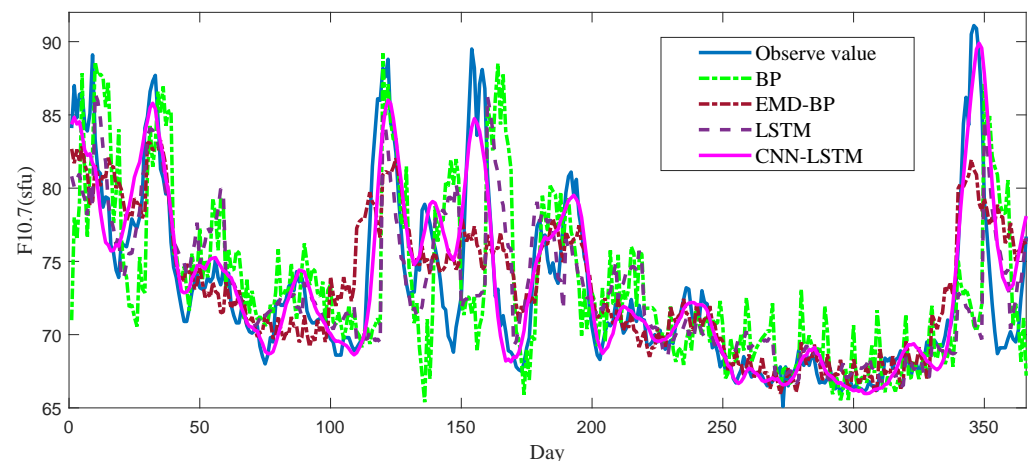
Forecast Day	MAPE (%)				RMSE				$\rho$			
	BP	LSTM	EMD-BP	CNN-LSTM	BP	LSTM	EMD-BP	CNN-LSTM	BP	LSTM	EMD-BP	CNN-LSTM
1	2.94	2.58	2.33	<b>2.04</b>	5.48	4.94	4.26	3.91	0.99	0.99	0.99	0.99
2	3.70	3.26	2.49	2.28	6.95	5.99	4.37	4.17	0.98	0.98	0.99	0.99
3	4.28	4.10	3.07	<b>2.78</b>	7.73	7.27	5.48	5.03	0.97	0.98	0.98	0.98
4	5.50	4.41	4.29	3.55	9.25	8.21	7.66	6.48	0.97	0.97	0.98	0.98
5	5.94	4.89	4.71	4.00	10.88	9.33	7.32	6.77	0.96	0.97	0.97	0.98
6	6.73	5.38	5.31	4.56	11.76	10.10	9.00	7.92	0.95	0.96	0.97	0.98
7	6.39	5.69	5.04	<b>4.66</b>	11.94	11.05	8.39	8.03	0.95	0.96	0.97	0.97
8	6.92	6.05	5.36	4.86	12.87	11.66	9.47	8.72	0.95	0.95	0.97	0.97
9	7.78	6.41	5.68	5.17	13.17	12.21	9.95	9.19	0.94	0.95	0.96	0.97
10	7.95	6.65	6.41	5.65	14.46	12.49	11.72	10.2	0.92	0.94	0.95	0.97
11	7.83	6.93	6.42	5.71	13.79	13.06	11.70	10.31	0.93	0.94	0.95	0.96
12	9.06	7.06	6.25	5.65	14.71	13.49	11.02	10.07	0.90	0.93	0.95	0.96
13	8.61	7.29	6.29	5.80	15.09	13.61	11.22	10.32	0.92	0.93	0.95	0.96
14	8.88	7.46	6.16	<b>5.85</b>	15.01	13.89	11.15	10.42	0.91	0.93	0.95	0.96
15	9.16	7.49	6.52	6.08	15.56	14.10	11.88	10.84	0.90	0.93	0.95	0.96
16	8.96	7.64	6.79	6.19	16.20	14.18	11.98	10.93	0.90	0.93	0.95	0.95
17	8.95	7.75	6.68	6.26	15.78	14.38	12.48	11.42	0.91	0.92	0.95	0.95
18	9.81	7.79	6.89	6.34	16.87	14.47	12.37	11.5	0.89	0.92	0.94	0.95
19	9.28	7.83	7.64	6.82	15.61	14.59	12.64	11.57	0.91	0.92	0.94	0.95
20	10.00	8.04	7.80	7.02	15.41	14.70	12.96	11.94	0.91	0.92	0.94	0.95
21	9.68	8.06	7.72	6.97	16.50	14.86	12.98	12.03	0.90	0.92	0.94	0.95
22	9.87	8.18	7.11	6.74	16.51	14.83	12.47	11.88	0.90	0.92	0.94	0.94
23	9.88	8.10	7.61	6.97	16.30	14.98	12.90	12.13	0.91	0.92	0.94	0.94
24	9.98	8.32	7.68	7.14	16.84	15.03	13.28	12.51	0.90	0.91	0.94	0.94
25	10.33	8.25	7.48	7.08	17.61	15.15	12.25	12.05	0.89	0.91	0.94	0.94
26	10.17	8.22	7.81	7.24	17.97	15.06	13.24	12.54	0.89	0.91	0.93	0.94
27	10.19	8.39	7.92	<b>7.40</b>	17.16	15.14	13.44	12.78	0.89	0.91	0.93	0.94

#### 4.5. Fitting Effect Analysis of the Mid-Term Forecast

Figure 4 shows the fitting curves of predicted values and observed values among comparative algorithms for 10-day forecasts in 2003 (i.e., the year of high solar activity). All the predictive curves are fitted well with the trend of the observed curve. We observe some relatively significant deviations between the predicted curves and the observed curve in local extremum cycles of the fitting curves of BP, EMD-BP and LSTM. The fitting curve of CNN-LSTM fits the curve well most of the time, but there is still a relatively large deviation in some periods (i.e., around the 90th, 170th and 200th days). Figure 5 shows the fitting effect among comparative algorithms for 10-day forecasts in 2007 (i.e., a year of low solar activity). From the curves, we observe that the CNN-LSTM model still shows the optimal fitting effect. The fitting curve of 10-day forecasts represents a mid-term forecast, which reveals the ability of trend prediction. Compared with the fitting curves of the BP, EMD-BP and LSTM models, the CNN-LSTM model demonstrates a smoother and smaller deviation, especially at high-solar-activity days.

**Figure 4.** The fitting curves of 10-day forecasts in 2003.





**Figure 5.** The fitting curves of 10-day forecasts in 2007.

#### 4.6. Comparison with Physical Methods in Short-Term Prediction

This part comprehensively evaluates the performance of the short-term forecast. The CH-MF model [12] mixed the heteroscedasticity linear model and Bayesian optimization method, and a solar magnetic flux transport model namely Y model [12] was compared with the CNN-LSTM model. The results are shown in Table 3. All the three methods have good and comparable effects in the metrics of Spearman coefficients, Pearson coefficients and correlation coefficients. Comparing the Y model and CH-MF model, our method has a lead of 39.9% and 42% in MAPE for 1-day forecasts and has a lead of 35.1% and 2.4% in RMSE. In 3-day forecasts, it leads by 28.6% and 51.8% in MAPE and 16.5% and 29.9% in RMSE.

**Table 3.** Comparison between the observed and forecast values using the CH-MF and Y models.

Forecast Days	Model	Spearman Coefficients	MAPE (%)	RMSE	Pearson Coefficients	R
1	Y	0.98	3.38	5.56	0.98	No provide
	CH-MF	No provide	3.50 <sup>a</sup>	3.70 <sup>a</sup>	No provide	0.99
	CNN-LSTM	0.99	2.03	3.61	0.99	0.99
3	Y	0.98	4.12	6.13	0.97	No provide
	CH-MF	No provide	6.10 <sup>a</sup>	7.30 <sup>a</sup>	No provide	0.99
	CNN-LSTM	0.99	2.94	5.12	0.99	0.99

<sup>a</sup> means an approximate value, acquired from the figures of corresponding reference literature.

## 5. Conclusions

We employed the combination model CNN-LSTM, which stacked the specified CNN and LSTM network in series for the short- and mid-term F10.7 forecasting. The CNN network was used to extract the features of the F10.7 daily sequence adaptively. The LSTM network then performed the predicted task. We compared the CNN-LSTM model with other F10.7 forecasting models under different solar activity years and forecast days. The results showed that the CNN-LSTM model had MAPE values of 2.04% (1-day ahead), 2.78% (3-day ahead) and 4.66% (7-day ahead) in short-term forecasts. Additionally, mid-term forecasts had MAPE values of 5.85% (14-day ahead) and 7.40% (27-day ahead). For the Spearman correlation coefficients and RMSE, the CNN-LSTM model also outperformed the BP, EMD-BP, and LSTM model. Further, the CNN-LSTM model was found to be effective in improving the prediction accuracy of high-solar-activity years. The CNN-LSTM was also compared with the Y model and the CH-MF models, and showed distinct advantages in a variety of evaluation metrics. We summarized the reasons that our method had impressive performance in short and medium-term prediction accuracy of F10.7: (1) Compared with the common shallow neural network, the LSTM model was more suitable for F10.7 prediction because of its memory nature. (2) The feature extractor of the convolutional neural network achieved adaptive extraction, which avoided the instability of manually extracted

features and facilitated further prediction of time series by LSTM. (3) Compared with other machine learning algorithms, the CNN-LSTM model provided a deeper neural network architecture, extracting depth features from the raw signal.

Nevertheless, we realized that the prediction error increased near the peak points, which might be due to the low proportion of samples in the vicinity of the extreme value points relative to the whole sample, making the model less effective in prediction in the face of this unbalanced sample problem. Future work will explore the use of down-sampling techniques and generative adversarial networks to improve this problem.

**Author Contributions:** Conceptualization, J.L. and L.Z.; methodology, J.L.; validation, K.Z., C.Z. and Z.L.; writing—original draft preparation, review and editing, J.L.; visualization, supervision, project administration, funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by Special Fund for Bagui Scholars of the Guangxi Zhuang Autonomous Region, under Grand 2019A08; The Basic Ability Enhancement Program for Young and Middle-aged Teachers of Guangxi, under Grand 2020KY10018.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Svalgaard, L. Reconstruction of solar extreme ultraviolet flux 1740–2015. *Sol. Phys.* **2016**, *291*, 2981–3010. [CrossRef]
2. Tapping, K. The 10.7 cm solar radio flux (F10.7). *Space Weather* **2013**, *11*, 394–406. [CrossRef]
3. Henney, C.; Toussaint, W.; White, S.; Arge, C. *Forecasting F10.7 with Solar Magnetic Flux Transport Modeling (Postprint)*; Technical Report; Air Force Research Lab Kirtland Afb Nm Space Vehicles Directorate: Albuquerque, NM, USA, 2012.
4. Lampropoulos, G.; Mavromichalaki, H.; Tritakis, V. Possible estimation of the solar cycle characteristic parameters by the 10.7 cm solar radio flux. *Sol. Phys.* **2016**, *291*, 989–1002. [CrossRef]
5. Krivova, N.; Solanki, S.; Wenzler, T.; Podlipnik, B. Reconstruction of solar UV irradiance since 1974. *J. Geophys. Res. Atmos.* **2009**, *114*. [CrossRef]
6. Warren, H.P.; Emmert, J.T.; Crump, N.A. Linear forecasting of the F 10.7 proxy for solar activity. *Space Weather* **2017**, *15*, 1039–1051. [CrossRef]
7. Zhang, W.; Zhao, X.; Feng, X.; Liu, C.; Xiang, N.; Li, Z.; Lu, W. Predicting the Daily 10.7-cm Solar Radio Flux Using the Long Short-Term Memory Method. *Universe* **2022**, *8*, 30. [CrossRef]
8. Hong-bo, W.; Jian-ning, X.; Chang-yin, Z. The mid-term forecast method of solar radiation index. *Chin. Astron. Astrophys.* **2015**, *39*, 198–211. [CrossRef]
9. Liu, S.Q.; Zhong, Q.Z.; Wen, J.; Dou, X.K. Modeling research of the 27-day forecast of 10.7 cm solar radio flux (I). *Chin. Astron. Astrophys.* **2010**, *34*, 305–315.
10. Lean, J.; Picone, J.; Emmert, J. Quantitative forecasting of near-term solar activity and upper atmospheric density. *J. Geophys. Res. Space Phys.* **2009**, *114*. [CrossRef]
11. Lei, L.; Zhong, Q.; Wang, J.; Shi, L.; Liu, S. The Mid-Term Forecast Method of F10.7 Based on Extreme Ultraviolet Images. *Adv. Astron.* **2019**, *2019*, 1–14. [CrossRef]
12. Liu, C.A.; Zhao, X.H.; Chen, T.; Li, H.C. Predicting short-term  $F_{10.7}$  with transport models. *Astrophys. Space Sci.* **2018**, *363*, 266. [CrossRef]
13. Huang, C.; Liu, D.D.; Wang, J.S. Forecast daily indices of solar activity, F10.7, using support vector regression method. *Res. Astron. Astrophys.* **2009**, *9*, 694. [CrossRef]
14. Xiao, C.; Cheng, G.; Zhang, H.; Rong, Z.; Shen, C.; Zhang, B.; Hu, H. Using Back Propagation Neural Network Method to Forecast Daily Indices of Solar Activity F10.7. *Chin. J. Space Sci.* **2017**, *37*, 1–7.
15. Luo, J.; Zhu, H.; Jiang, Y.; Yang, J.; Huang, Y. The 10.7-cm radio flux multistep forecasting based on empirical mode decomposition and back propagation neural network. *IEEE Trans. Electr. Electron. Eng.* **2020**, *15*, 584–592. [CrossRef]
16. Lee, T. EMD and LSTM hybrid deep learning model for predicting sunspot number time series with a cyclic pattern. *Sol. Phys.* **2020**, *295*, 1–23. [CrossRef]
17. Roy, S.; Prasad, A.; Panja, S.C.; Ghosh, K.; Patra, S.N. A Search for Periodicities in F 10.7 Solar Radio Flux Data. *Sol. Syst. Res.* **2019**, *53*, 224–232. [CrossRef]
18. Kasde, S.; Sondhiya, D. Study of phase relationship of Sunspot Numbers with F 10.7 cm Solar Radio-Flux and Coronal Index using Wavelet-Transform technique. *Editor. Off.* **2021**, *25*, 59–67.

19. Deng, L.; Li, B.; Zheng, Y.; Cheng, X. Relative phase analyses of 10.7 cm solar radio flux with sunspot numbers. *New Astron.* **2013**, *23*, 1–5. [\[CrossRef\]](#)
20. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
21. Wang, K.; Li, K.; Zhou, L.; Hu, Y.; Cheng, Z.; Liu, J.; Chen, C. Multiple convolutional neural networks for multivariate time series prediction. *Neurocomputing* **2019**, *360*, 107–119. [\[CrossRef\]](#)
22. Liu, P.; Liu, J.; Wu, K. CNN-FCM: System modeling promotes stability of deep learning in time series prediction. *Knowl.-Based Syst.* **2020**, *203*, 106081. [\[CrossRef\]](#)
23. Jain, A.K.; Grumber, C.; Gelhausen, P.; Häring, I.; Stolz, A. A toy model study for long-term terror event time series prediction with CNN. *Eur. J. Secur. Res.* **2020**, *5*, 289–309. [\[CrossRef\]](#)
24. Yao, D.; Li, B.; Liu, H.; Yang, J.; Jia, L. Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit. *Measurement* **2021**, *175*, 109166. [\[CrossRef\]](#)
25. Cavalli, S.; Amoretti, M. CNN-based multivariate data analysis for bitcoin trend prediction. *Appl. Soft Comput.* **2021**, *101*, 107065. [\[CrossRef\]](#)
26. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
27. Aufa, B.Z.; Suyanto, S.; Arifianto, A. Hyperparameter Setting of LSTM-based Language Model using Grey Wolf Optimizer. In Proceedings of the 2020 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 5–6 August 2020; pp. 1–5.
28. Zhao, J.; Gao, Y.; Bai, Z.; Wang, H.; Lu, S. Traffic speed prediction under non-recurrent congestion: Based on LSTM method and BeiDou navigation satellite system data. *IEEE Intell. Transp. Syst. Mag.* **2019**, *11*, 70–81. [\[CrossRef\]](#)
29. Petersen, N.C.; Rodrigues, F.; Pereira, F.C. Multi-output bus travel time prediction with convolutional LSTM neural network. *Expert Syst. Appl.* **2019**, *120*, 426–435. [\[CrossRef\]](#)
30. Le, X.H.; Ho, H.V.; Lee, G.; Jung, S. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water* **2019**, *11*, 1387. [\[CrossRef\]](#)
31. Karthikeyan, L.; Kumar, D.N. Predictability of nonstationary time series using wavelet and EMD based ARMA models. *J. Hydrol.* **2013**, *502*, 103–119. [\[CrossRef\]](#)
32. Rhif, M.; Ben Abbes, A.; Farah, I.R.; Martínez, B.; Sang, Y. Wavelet transform application for/in non-stationary time-series analysis: A review. *Appl. Sci.* **2019**, *9*, 1345. [\[CrossRef\]](#)
33. Lei, Y.; He, Z.; Zi, Y. EEMD method and WNN for fault diagnosis of locomotive roller bearings. *Expert Syst. Appl.* **2011**, *38*, 7334–7341. [\[CrossRef\]](#)
34. Yu, C.; Li, Y.; Zhang, M. An improved wavelet transform using singular spectrum analysis for wind speed forecasting based on elman neural network. *Energy Convers. Manag.* **2017**, *148*, 895–904. [\[CrossRef\]](#)