Forecasting tropical cyclone tracks in the Northwest Pacific based on a deep-learning model

Liang Wang[1,2], Bingcheng Wan[3], Shaohui Zhou[3], Haofei Sun[1,2], Zhiqiu Gao[1,3]*

[1] State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China

[2] University of Chinese Academy of Sciences, Beijing, 100049, China

[3] School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, 210044, China

Correspondence to: Dr. Zhiqiu Gao(zgao@mail.iap.ac.cn)

Abstract. Tropical cyclones (TCs) are one of the most severe meteorological disasters, making rapid and accurate track forecasts crucial for disaster prevention and mitigation. Because TC tracks are affected by various factors (the steering flow, thermal structure of the underlying surface, and atmospheric circulation), their trajectories present highly complex nonlinear behavior. Deep learning has many advantages in simulating nonlinear systems. In this paper, we explore the movement of TCs in the Northwest Pacific from 1979 to 2021 based on deep-learning technology, divided into training (1979–2014), validation (2015–2018), and test sets (2019–2021), and create 6–72 h TC track forecasts. Only historical trajectory data are used as input for evaluating the forecasts of the three recurrent neural networks utilized: recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU) models. The GRU approach performed best; to further improve forecast accuracy, a model combining GRU and a convolutional neural network (CNN) called GRU_CNN is proposed to capture the characteristics varying with time. By adding reanalysis data of the steering flow, sea-surface temperatures, and geopotential height around the cyclone, we can extract sufficient information on the historical trajectory features and three-dimensional spatial features. The results show that GRU_CNN outperforms other deep-learning models without CNN layers. Furthermore, by analyzing three additional environmental factors through control experiments, it can be concluded that the historical steering flow of TCs plays a key role, especially for short-term predictions within 24 h, while sea-surface temperatures and geopotential height can gradually improve the 24–72-h forecast.

28 The average distance errors at 6 h and 12 h are 17.22 km and 43.90 km, respectively. Compared with

29 the 6-h and 12-h forecast results (27.57km and 59.09km) of the Central Meteorological Observatory,

30 the model proposed herein is suitable for short-term forecasting of TC tracks.

31 **1 Introduction**

32 The Northwest Pacific is the most active basin for tropical cyclones (TCs) in the world, generating over

33 one-third of the total number of TCs (Gray, 1968). China, located on the western side of the Pacific

34 Ocean with a coastline longer than 18,000 km, is one of the countries most severely influenced by TCs.

35 These storm systems are accompanied by strong winds, heavy precipitation, and storm surges, resulting

36 in severe disasters that affect human lives and economic growth (Goldenberg et al., 2001). Studies have

37 shown that global warming will progressively intensify TCs over time (Emanuel, 2017; Schulthess et

38 al., 2019). Since disasters caused by TCs are unavoidable and potentially destructive, accurately

39 predicting the movement of TCs can provide sufficient preparation time for people in affected areas to

40 implement disaster mitigation strategies.

41 Given the uncertainty of TC movements, the complexity and nonlinearity inherent in the

42 atmospheric system, and the scarcity of ocean-based observational data, accurately predicting the

43 center positions and intensities of TCs is a challenge. Currently, forecasting methods for TCs are

44 mainly divided into two categories, with the primary method being numerical weather prediction

45 (NWP). NWP calculates the approximate solution of partial differential equations involving

46 atmospheric state variables when the initial conditions and boundary conditions of the atmosphere are

47 known. In this way, some elements, such as the tracks and intensities of the TCs, can be solved

48 iteratively; GRAPES-TYM (CMA), GFS (NCEP), and IFS (ECWMF) are the main NWP models.

49 Although these model forecasts can provide accurate results, there are limitations in methods relying on

50 high-performance computers and requiring precise initial conditions. At the same time, ensemble

51 forecast methods (GRAPES-GEFS, ECMWF-EPS, NCEP-GEFS) have been used to reduce the

52 influence of various uncertainties on the numerical prediction results (Goerss, 2000). The other

53 forecasting method is a statistical model, which generally utilizes multiple regression. The statistical

54 model is mainly based on the relationship between the movement of the TC and its specific historical

55 characteristics, but it usually does not consider any physical processes. The National Hurricane Center

56   has successively adopted statistical models such as NHC64 (taking observational data and historical

57   12h movements as factors), NHC67 (increasing factors based on NHC64), CLIPER (climate

58   persistence factors) (Neumann and Hope, 1972), and NHC72 (a combination of NHC67 and CLIPER).

59   Most traditional TC statistical models adopt a linear regression model, and it is difficult for this

60   approach to address the nonlinear problems in TC track forecasting (Roy and Kovordányi, 2012). At

61   the same time, manual feature selection is unable to produce accurate predictions. CLP5 had the largest

62   Mean absolute error(MAE) of all models for TCs occurring from the Eastern Pacific and North Atlantic

63   (Boussioux et al., 2022). Li-Min et al. (2009) used the BP neural network to predict that the average

64   distance error of the 6h movement track of six typhoons in 2005 improved by 36.9km, compared with

65   CLIPER.

66   Deep learning is an emerging application of supercomputing that is continuously being developed;

67   many researchers have tried to adopt this technology to forecast weather and meteorological elements,

68   including visibility (Ortega et al., 2022), wind speeds (Liu et al., 2018), radar echoes (Klein et al.,

69   2015), and precipitation nowcasting (Shi et al., 2015). Deep learning is a statistical model that solves

70   nonlinear and complex relationships from historical sample data based on neural network algorithms.

71   The weight factor between network nodes is automatically adjusted through repeated training; thus,

72   neural network algorithms have the advantages of strong adaptability and fault tolerance. TCs have

73   complex dynamic mechanisms and are easily affected by many factors, including environmental

74   steering flow, Beta effects, underlying surface conditions, the asymmetric structure of the inner core,

75   and mesoscale circulations (Chan and Kepert, 2010). Artificial neural networks (ANNs) have been

76   applied to predict TC tracks due to their strong learning ability and advantages in simulating nonlinear

77   systems. Until the 2010s, ANN and back propagation (BP) networks were the mainstream neural

78   network methods for forecasting TC tracks (Ali et al., 2007; Li-Min et al., 2009; Wang et al., 2011).

79   Since the mid-2010s, more new methods have been introduced into TC prediction due to the

80   development of deep-learning technology. Recurrent neural networks (RNNs) are suitable for TC track

81   forecasting owing to their ability to handle time series data of arbitrary lengths. Moradi Kordmahalleh

82   et al. (2015) applied a sparse RNN to Atlantic hurricane trajectory prediction using the dynamic time

83   warping (DTW) method to measure the hurricane most similar to the target hurricane for training. Gao

84   et al. (2018) used long short-term memory (LSTM) to predict typhoon tracks in the Northwest Pacific

85   Ocean; the ratio of the cyclone training set and test set was set at 8:2, and the 24-h prediction error

86   could reach 105 km. Alemany et al. (2018) proposed an RNN based on a grid system to predict

87   hurricanes in the Atlantic, potentially improving the 6-h prediction accuracy with a root mean square

88   error (RMSE) of 0.11 for the test set. Kim et al. (2018) performed a TC identification task based on

89   ConvLSTM to train WRF-simulated data, and the results show the average precision of the forecast

90   was improved by 78.99% than those of a convolutional neural network (CNN). These CNNs have

91   attracted attention given their suitability for processing 2D image data; they maintain spatial

92   correlations by implementing convolution layers and then pooling layers for feature extraction.

93   Giffard-Roisin et al. (2020) combined historical trajectory data with wind field reanalysis data as input

94   to a CNN and predicted Atlantic hurricane tracks since 1979, with an average error of 32.9 km for 6-h

95   predictions.

96   Making full use of different types of data is essential for deep learning. TC-related data are mainly

97   divided into the following three categories: observational trajectory data, remote sensing data, and

98   meteorological reanalysis field data. A multi-modal approach enables more accurate predictions than

99   an approach using a single data source does. Zhang et al. (2018) developed a matrix neural network

100   (MNN) model that preserves the spatial information of the TC tracks, and it has demonstrated the

101   ability to provide more accurate results compared with other models (GRU, LSTM, Multi-Layer

102   Perceptron, and RNN). Ruttgers et al. (2019) built generation adversarial networks (GANs) adding

103   satellite images to predict the coordinates of the typhoon center and generate cloud maps of future

104   typhoons. Liu et al. (2022) proposed a new deep learning–based model, DBFNet, to effectively fuse the

105   inherent features of cyclones and reanalyze 2D pressure field data. The above studies have shown that

106   deep-learning models that incorporate multiple data types can improve the track forecast of TCs to a

107   certain extent. Still, most of them have neglected to describe and analyze the meteorological factors

108   that affect the movement of TCs, ignoring valuable features. The 6-hour average distance error

109   between predicted and real location by the fusion network (wind+track) is 32.9 km, while the network

110   prediction results without adding wind variables are 35km (Giffard-Roisin et al., 2018), which

111   indicates that the addition of meteorological field variables can effectively improve the prediction

112   accuracy.

113   This paper attempts to propose a new method for TC track prediction based on a combination of

114   CNN and GRU models that incorporate data regarding the trajectory, steering airflow, sea-surface

115   temperatures, and geopotential height as input features, aiming to improve the accuracy of TC track

116   forecasts by leveraging big data. The main contents of this paper are as follows: Section 2 introduces

117   the necessary data and data preprocessing. Section 3 describes the experimental design and the

118   framework of the fusion model (GRU_CNN) proposed in this paper. Section 4 presents the

119   experimental results and comparative analysis, and Section 5 provides a summary and discussion of

120   shortcomings and directions for future work.

121   **2 Data and data preprocessing**

122   **2.1 Data**

123   The data used in this paper are trajectory data and reanalysis environmental data. The TC track data

124   come from the International Best Track Archive for Climate Stewardship (IBTrACS), which

125   encompasses all TCs globally. For each TC, the latitude, longitude, central pressure, maximum wind

126   speed, direction, moving speed, and other data are recorded at 3-h intervals. The IBTrACS dataset

127   contains data from different basins where cyclones show different characteristics; thus, this paper only

128   selects TCs that occur in the Northwest Pacific Ocean. To better mine the hidden information, 19

129   movement characteristics were obtained, including the past 24-h longitude, latitude, central

130   atmospheric pressure, maximum wind speed, meridional moving speed, zonal moving speed, moving

131   direction/speed, the difference between those values and those at the current time, and the angle, zonal

132   distance, and meridional distance formed between the data over the past 24 h and in the present

133   moment. Because they are influenced by the earth's rotation, TCs will be biased to the northwest

134   (Kitade, 1981). The Coriolis parameters corresponding to the latitude of the TCs in the past 24 hours

135   are also included.

136   Both observational and theoretical studies have shown that TC movement is closely related to

137   large-scale airflow fields (Holland, 1983), and TC movement is mainly affected by the steering flow

138   (Brand et al., 1981; Chan, 1984). Interactions among weather systems, the subtropical anticyclone,

139   Westerlies, and the Tibetan High will also affect the movement of cyclones (George and Gray, 1976;

140 Chan et al., 1980). The geopotential heights of 300 hpa, 500 hpa, and 700 hpa are selected as the

141 locations for the high, middle, and low-level circulation data, respectively. In addition, the underlying

142 surface conditions must be considered, and, in the case of a weak guidance environment, TCs tend to

143 move toward warmer sea-surface temperatures (Sun et al., 2017; Katsube and Inatsu, 2016).

144 Meteorological environmental data are obtained by downloading high-resolution ERA5 reanalysis data

145 from the European Centre for Medium-Range Forecasting (ECMWF). Holland (1984) noted that the

146 deep mean circulation from 850 hpa to 300 hpa can better represent the direction of a TC. Therefore,

147 the environmental data for the preceding 24 h were extracted as follows:

148 (1) The *u*- and *v*-component data of the wind field on the four isobaric surfaces (300 hpa, 500 hpa, 700

149 hpa, 850 hpa): Centered the TC, extend 10 degrees outward in the zonal and meridian direction

150 respectively. Since the resolution of the selected reanalysis data is $1° × 1°$, a $21 × 21$ grid can be

151 formed.

152 (2) The sea-surface temperature (SST): 10 degrees outward in the zonal and meridian direction is again

153 extended from the TC center to form a $21 × 21$ grid.

154 (3) The geopotential heights of 300 hpa, 500 hpa, and 700 hpa: A grid is extended $+35°$ to the north,

155 $−10°$ to the south, $−40°$ to the west, and $+40°$ to the east from the center of the TC, forming a $46 × 81$

156 grid.

157 **2.2 Data preprocessing**

158 **2.2.1 Devortexing**

159 Because the actual weather circulation is very complex and includes information about the TC itself,

160 the surrounding airflow, and the interaction between the two, it is necessary to separate the cyclone

161 vortex from the surrounding airflow to obtain the steering flow. The most commonly used method

162 (Lownam, 2001; Galarneau and Davis, 2013) corrects the vorticity and divergence by solving the

163 change in the velocity stream and potential functions, respectively, and then calculates the modified

164 velocity field. The modified flow field can be interpreted as having a non-rotating wind and

165 non-diverging wind. There must be potential velocity in the irrotational motion and a stream function

166 in the non-divergent motion. The relationship between them can be expressed as follows:

167
$$\nabla^2 \psi = \zeta \tag{1}$$

168
$$v_\psi = \hat{k} \times \nabla \psi \tag{2}$$

169 Where ψ is the stream function without divergence, ζ is the relative vorticity, and v_ψ is the

170 non-divergent wind (rotating wind). To define the rotating wind, the vorticity outside the vortex radius

171 is set to zero, and $\psi = 0$ is specified on the horizontal boundary. The iterative relaxation method is used

172 to solve the stream function of Eq. (1) at all layers and then to calculate v_ψ using Eq. (2). In the case of

173 divergence, Eqs. (1) and (2) are replaced by:

174
$$\nabla^2 \chi = \delta \tag{3}$$

175
$$v_\chi = \nabla \chi \tag{4}$$

176 Where χ is the potential velocity, δ is the divergence, and v_χ is the non-vorticity wind. The divergence

177 outside the vortex radius is set to zero and the potential function $\chi = 0$ on the boundary of the region.

178 The velocity potential can be solved in the same manner to calculate v_χ. The ambient wind field with

179 the vortex removed can be obtained by subtracting the rotating wind and divergent wind from the

180 original wind field, V:

181
$$v_{env}(x, y, p) = V(x, y, p) - v_\chi(x, y, p) - v_\psi(x, y, p) \tag{5}$$

182 **2.2.2 Random forest**

183 By sorting features based on importance, random forest selects the best feature combination and

184 reduces the input feature dimensions that efficiently direct variables for machine learning models

185 (Díaz-Uriarte and Alvarez De Andrés, 2006; Genuer et al., 2010). The random forest contains N

186 decision trees, and N is generally set to 100. Since bootstrapping (random sampling with replacement)

187 is used to generate the random decision tree, all samples are not in the generation process of a tree, and

188 the unused samples are called "out-of-bag" (OOB) samples. Through OOB samples, the accuracy of

189 this tree can be evaluated.

190 Before model training, it is necessary to determine whether the 19 trajectory features all have an

191 impact on the prediction results. Figure 1(a) shows the long name corresponding to the short name of

192 19 input features and Fig 1(b) shows the 19 features' order of importance calculated using the random

193 forest method. For forecasting the difference in longitude and latitude within the following 72 hours,

characteristics like the historic longitude or the angle formed by the historical moment and the current moment are significant. The decision about whether to exclude some less important features, however, requires further consideration. The OOB scores under different input feature dimensions are computed, with variables input in the order of importance, as shown in Fig. 1(c). In the case in which the first 11 features are sorted by importance, the OOB score is the highest, and the features added later will no longer affect the result; in other words, the best combination is that of the first 11 features.

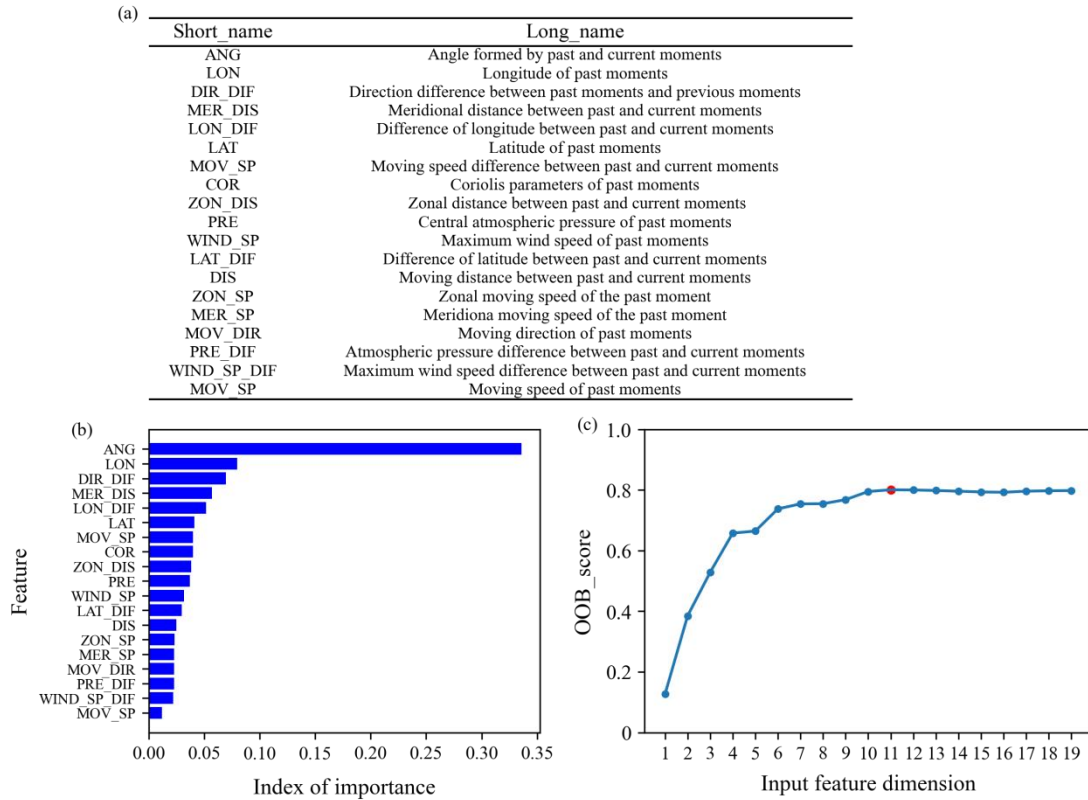



Figure 1: (a) Table displaying the short and long names of features, (b) the importance index of features, and (c) the OOB_score of different feature combinations based on the random forest (red dot indicates the maximum value).

3 Experiment

3.1 Experimental design

Our goal is to predict the TC movements for the following 6–72 h using the trajectory data and surrounding environmental field from the previous 24 h. We explore TC movement in the Northwest Pacific from 1979 to 2021 and consider the longitudinal and latitudinal changes in the following 6–72 h

209   as the quantitative prediction variables, with the center of the TC at the current time as the reference

210   point. Since the maximum forecast hour is 72 and the input sequence time length is 24 h, TCs that

211   persist for longer than 96 hours are removed. All samples obtained based on the sliding window of the

212   input-prediction sequence length are divided into three groups in chronological order: training set

213   (1979–2014), validation set (2015–2018), and test set (2019–2021). There are 36473 samples, of which

214   90% are trained, and the remaining 10% are validated; 49 TCs from 2019 to 2021 are used for testing,

215   and the number of test samples is 2095.

216   **3.2 Model framework**

217   **3.2.1 Recurrent Neural Network**

218   RNNs can process sequences of any length using neurons with self-feedback, characterized by

219   architectural features intentionally designed to preserve historical information, showing a remarkable

220   ability to process sequential data (Graves et al., 2013; Bathla, 2020; Wang and Fu, 2020). However,

221   simple RNNs have difficulty in dealing with the long-term dependence of the sequence; when the

222   sequence length exceeds a certain threshold, the information may disappear during the transmission

223   process, resulting in large deviations in prediction accuracy. The LSTM network proposed by

224   Hochreiter and Schmidhuber (1997) can avoid the gradient disappearance and explosion phenomena

225   that occur in the standard RNN. While GRU (Cho et al., 2014) is an improved and optimized neural

226   network based on LSTM, it has a faster convergence speed and maintains accuracy levels close to those

227   of LSTM.

228   **3.2.2 Convolutional Neural Network**

229   CNNs can extract features automatically by processing the input patterns and translating the same

230   convolution kernel from top to bottom and from left to right. The spatial relationship is fixed with the

231   distribution of neurons, and the local connection and weight sharing of neurons reduce the training

232   complexity by reducing the number of parameters. Lecun et al. (1998) first used CNN for handwritten

233   character recognition with average pooling and the tanh activation function. Krizhevsky et al. (2012)

234   proposed the AlexNet model in the ImageNet competition, using the ReLU function instead of the

235   traditional tanh function to introduce nonlinearity and solve the gradient disappearance problem of the

236 activation function when the network was relatively deep, employing maximum pooling to avoid the

237 blurring effect of average pooling. Ioffe and Szegedy (2015) applied batch normalization to image

238 classification models, which significantly accelerated the training of deep networks, and batch

239 normalization helped alleviate the problem of gradient exploding or vanishing.

240 **3.2.3 GRU_CNN**

241 Due to differences in the data sources, a new model must be developed to integrate the four

242 information sources into the neural network using the Keras deep-learning framework. The specific

243 model structure is shown in Fig. 2. For the 3D meteorological data, the data are superimposed on the

244 geopotential height (pressure level), so the input data for the CNN consist of multiple two-dimensional

245 matrices. the gray filling arrow represents the TimeDistributed layer that is applied to a series of

246 tensors in the processing of the time dimension. In addition, the CNN adopts a typical architecture with

247 alternating convolution layers (Conv layers) and maximum pooling layers (Maxpool layers), The black

248 solid arrow means the Flatten layer converting two-dimensional data into one-dimensional vectors at

249 the end of the CNN network and the dashed black arrow represents the fully connected layer in the

250 network framework. All hidden layers are equipped with batch normalization, and this paper employs

251 ReLU as the activation function.

252 For the two-dimensional trajectory data of the TCs, x_i^j represents the input value of the ith feature

253 at the jth timestamp, $i \in (1, n)$, $j \in (1, t)$, and they are input into GRU. The model is based on the Adam

254 optimizer and trained with the RMSE between the forecast and the actual value as a loss function. Due

255 to the different properties among the wind field, pressure field, SSTs, and past trajectory data, different

256 learning rates are required for the neural network. Therefore, the parameters of each branch in the

257 model can be trained with the same task, and then the branches can be fused into one network (Concat

258 layer), that is, the dashed red arrow means the merging of multiple vectors into one vector. It is

259 eventually stitched with output with a fully connected layer; thereafter, the parameters can be adjusted

260 slightly. Table2 lists the input and output size of each layer in the network framework, including

261 convolution kernel size, stride, and channel number.

262

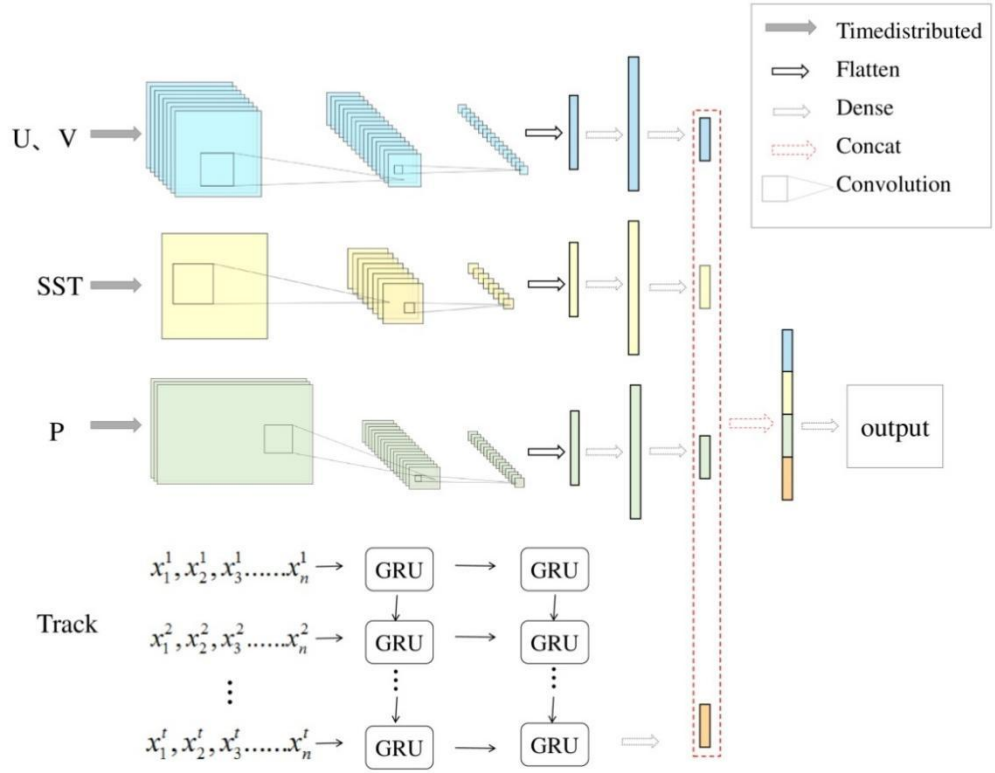



Figure 2: The model framework and network structure of GRU_CNN.

Table 1 Each layer architecture of the GRU_CNN

Layers	Kernel Size	Stride	Channel	Input Size	Output Size
Conv_uv	7×7	2	8	21×21	8×8
MaxPool_uv	4×4	4	16	8×8	2×2
Flatten_uv	-	-	16	2×2	64
Dense_uv_1	-	-	-	64	128
Dense_uv_2	-	-	-	128	32
Conv_sst	7×7	2	1	21×21	8×8
MaxPool_sst	4×4	4	8	8×8	2×2
Flatten_sst	-	-	8	2×2	32
Dense_sst_1	-	-	-	32	128
Dense_sst_2	-	-	-	128	32
Conv_p	14×25	4	3	46×81	9×15
MaxPool_p	5x11	4	16	9×15	2×2
Flatten_p	-	-	16	2×2	64
Dense_p_1	-	-	-	64	128
Dense_p_2	-	-	-	128	32
GRU_1	-	-	-	8×11	8×128

263

264

					8×128	128
GRU_2	-	-	-	8×128	128
Dense_GRU	-	-	-	128	32
Concat_layer	-	-	-	-	128

4 Results

Three types of recurrent neural networks (RNN, LSTM, GRU) are used to train samples with eight timestamps and 11 features selected by the random forest method, according to their importance; the results of analyzing 49 TCs in 2019–2021 are then evaluated. We set the value of the batch size to 64 and the epoch to 100 and found that the model performed best when the number of neurons in the hidden layer is set to 128; this was determined via experiments using different numbers of neurons in the hidden layer. Early stopping is used to prevent overfitting. When the performance of the model in the validation set begins to decline, training is stopped to avoid overfitting due to continued training.

Table 2 Model performance evaluation (RMSE) for RNN, LSTM, and GRU

		Lat						Lon					
Forecast Hour		6 h	12 h	18 h	24 h	48 h	72 h	6 h	12 h	18 h	24 h	48 h	72 h
Training Sets	RNN	0.146	0.368	0.624	0.912	2.232	4.069	0.153	0.412	0.735	1.118	3.181	6.236
	LSTM	0.126	0.335	0.584	0.867	2.34	3.986	0.149	0.391	0.703	1.077	3.232	6.172
	GRU	**0.112**	**0.312**	**0.555**	**0.83**	**2.282**	**3.883**	**0.134**	**0.376**	**0.681**	**1.041**	**3.152**	**6.031**
Validation Sets	RNN	0.171	0.441	0.76	1.116	2.723	4.493	0.182	0.483	0.865	1.337	3.739	6.715
	LSTM	0.157	0.428	0.751	1.103	**2.675**	**4.349**	0.183	**0.480**	**0.855**	**1.305**	3.69	6.761
	GRU	**0.157**	**0.42**	**0.735**	**1.086**	2.699	4.434	**0.179**	0.484	0.868	1.33	**3.632**	**6.608**
Test Sets	RNN	0.166	0.411	0.685	0.988	2.481	4.207	0.176	0.461	0.797	1.178	3.534	6.157
	LSTM	0.149	0.389	0.661	0.965	2.473	4.154	0.169	**0.456**	0.812	1.215	3.346	5.989
	GRU	**0.149**	**0.387**	**0.653**	**0.951**	**2.446**	**4.143**	**0.167**	0.457	**0.8**	**1.185**	**3.325**	**5.969**

The performance evaluation of the three RNN models is displayed in Table 2 by calculating the RMSE values between the predicted longitude (latitude) and the actual longitude (latitude), including training, validation, and test sets; the best results are highlighted in bold font. It is clear that the GRU-based and LSTM-based models significantly outperformed the RNN-based model, which suggests that the RNN is inferior in handling the problem of long-term dependence. GRU is a variant of LSTM that combines the forget and input gates in LSTM into an update gate and also merges the cell and hidden states. Hence, the parameter amounts of GRU are less than those of LSTM, which results in the overall training speed of GRU being faster than that of LSTM. GRU is theoretically similar to LSTM and can achieve the same accuracy as LSTM (or even better), so the results of GRU

284 and LSTM are close and their RMSE values are much lower than that of RNN. GRU achieves the best

285 performance in all forecast hours, with the smallest RMSE in the test set. Therefore we use GRU as a

286 part of the fusion network model called GRU_CNN, adding meteorological environment data

287 processed with CNN.

288 **Table 3 Comparison of the average absolute distance errors (km) predicted by multiple deep-learning**

289 **models**

	6 h	12 h	18 h	24 h	48 h	72 h
CLIPER(Demaraia,1992)	—	—	—	213	442	659
BP	23.86	59.58	101.01	146.91	377.64	634.42
RNN	21.43	55.46	94.59	138.09	373.12	625.17
LSTM	19.65	52.38	91.76	136.05	360.32	614.76
GRU	19.51	52.6	91.21	134.73	357.25	607.44
NMSTN(Huang,2022)	27.52	59.09	—	139.18	336.16	544.16
GRU_CNN	**17.22**	**43.9**	**72.74**	**106.16**	**281.52**	**502.71**

290

291 Table 3 compares the results between GRU_CNN and various deep-learning models, showing the

292 forecast results in the form of the mean absolute distance error. It is evident that GRU_CNN presents

293 an absolute advantage in long-term forecasting. Both LSTM and GRU retain important features

294 through various gate functions, which ensures that they will not be lost during long-term propagation.

295 They can better predict the medium and long-term tracks of the TCs, compared with standard RNNs

296 and two traditional methods named CLIPER and BP. The GRU_CNN is more accurate than the models

297 without CNN. The average distance errors at 6 h, 24 h, 48 h, and 72 h are 17.22 km, 106.16 km, 281.52

298 km, and 502.71 km, respectively. The error is also reduced compared with the NMSTN method

299 proposed by Huang et al. (2022). In addition, although there is a big difference between the long-term

300 forecast and the numerical prediction results, the average distance prediction results are better than the

301 results provided by the Central Meteorological Observatory (CMO) in the short-term forecasts,

302 including the 6 h (27.57 km) and 12 h (59.09 km) forecasts.

303

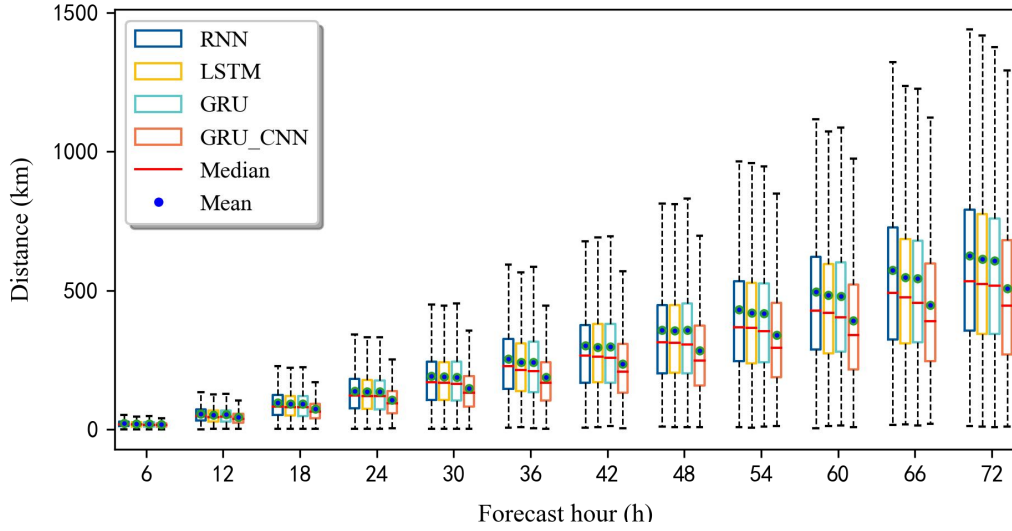



Figure 3: The absolute average distance boxplot of the three kinds of recurrent neural networks (RNN,

LSTM, GRU) and the method in this paper (GRU_CNN) creating 6–72 h forecasts (interval 6 h).

As shown in Fig. 3, the maximum distance errors predicted by the three RNNs at 48 h and 72 h

are over 500 km and 1000 km, respectively. Only considering the trajectory characteristics of the TCs

in the RNN while ignoring the external atmospheric environmental characteristics will cause instability

in the prediction of the TC tracks. The errors of the maximum and average values predicted by the

GRU_CNN model are both significantly reduced. To illustrate GRU_CNN more comprehensively and

intuitively, Fig. 4 shows a scatter plot of the predicted and actual values. The distance between the data

points and the diagonal line represents the prediction error. The higher the wind speed, the stronger the

intensity of the TCs, and the closer the predicted value is to the actual value. In addition, with the

increase in the forecast time, in high latitude and longitude forecasts when the TC is moving towards

the northwest, the predicted value is often lower than the actual value.

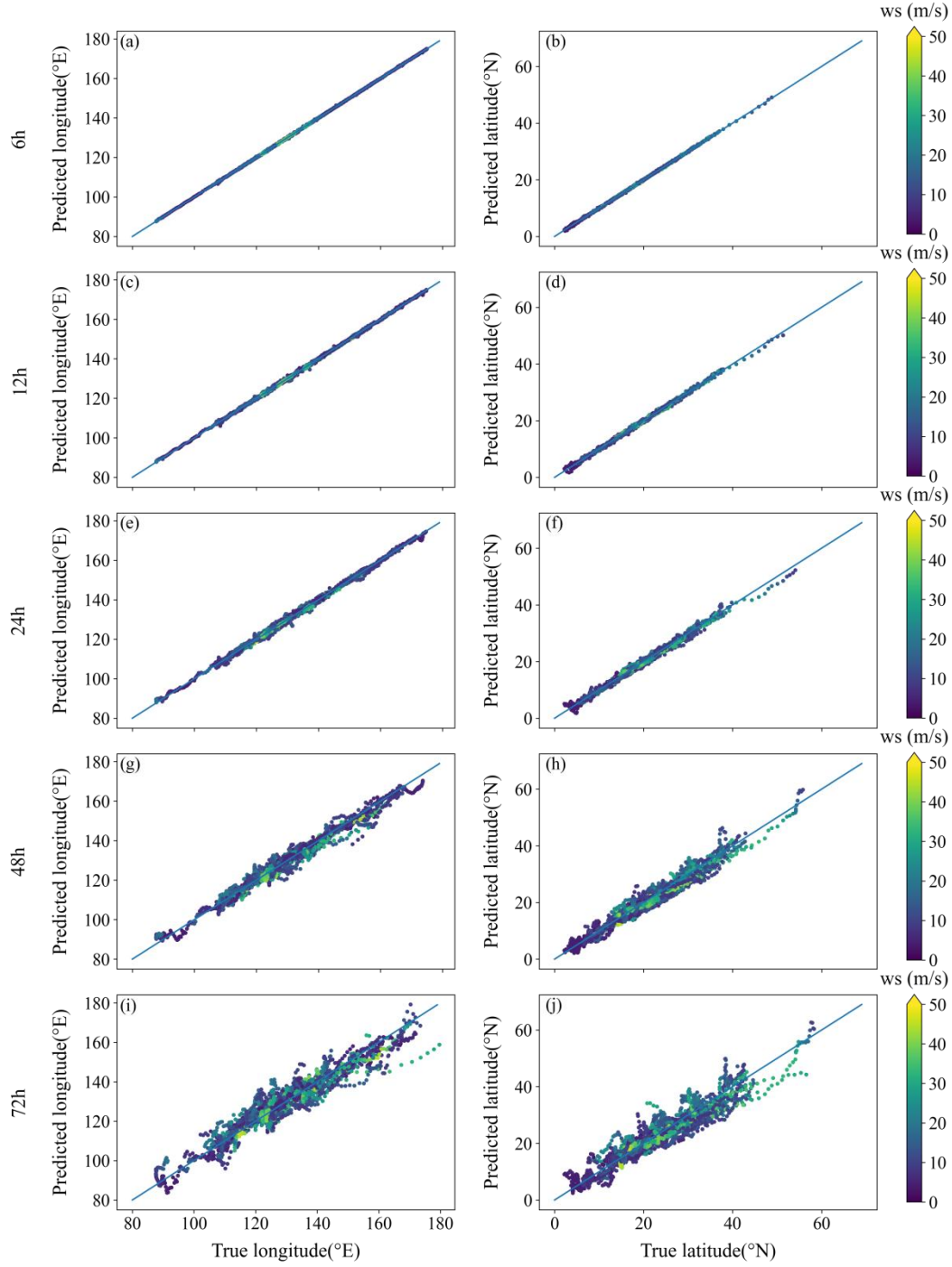



Figure 4: Scatter plot distributions of latitude predictions. The color bar represents the maximum wind speed, including the longitude and latitude forecasts at (a–b) 6 h, (c–d) 12 h, (e–f) 24 h, (g–h) 48 h, and (i–j) 72 h.

Data from three environmental fields are used in this paper: SST, geopotential height (pressure), and wind field (*u*- and *v*-component) data. Different environmental input variables show different

322   effects in the model (Table 4). GRU+SST (pressure, UV) represents only the combination of the

323   trajectory characteristics and SST (geopotential height, wind field), while GRU+CNN is the result of

324   the fusion of the three. The results in Table 4 indicate that GRU+UV performed best, followed by

325   GRU+pressure and then GRU+SST, indicating that the steering flow plays a dominant role in TC

326   forecasting, especially in the short-term < 24-h forecast. The forecasting results of adding only the

327   steering flow are close to those of GRU_CNN, while the results at 48 h and 72 h illustrate that the

328   influence of the SST and geopotential height on the long-term TC forecast track gradually increases.

329   **Table 4 Comparison of trajectory data combining different environmental features. RMSE is the root mean**

330   **square error of latitude and longitude, and the distance is the average absolute distance error (km).**

	GRU+SST		GRU+Pressure		GRU+UV		GRU+CNN	
	RMSE	Distance (km)	RMSE	Distance (km)	RMSE	Distance (km)	RMSE	Distance (km)
6 h	0.154	19.35	0.132	16.15	0.137	16.94	0.138	17.22
12 h	0.419	52.37	0.352	44.22	0.347	43.74	0.35	43.9
18 h	0.739	92.25	0.598	76.16	0.575	72.81	0.575	72.74
24 h	1.103	137.78	0.883	112.93	0.841	106.63	0.837	106.16
48 h	2.858	358.25	2.462	306.03	2.379	302.76	2.248	281.52
72 h	4.913	588.08	4.52	557.86	4.385	524.88	4.146	502.71

331

332   To better show the model forecast of GRU_CNN, Figures 5–7 present the observed and forecast

333   tracks at 6 h and 24 h of TCs FAXAI, MITAG, and IN-FA, respectively, and the forecast tracks of

334   other TCs in the test set are presented in Supplementary Fig. S1-51. The blue lines represent the

335   observed tracks, while the red and yellow lines indicate the 6-h and 24-h forecast tracks. In general, it

336   is particularly hard to forecast unexpected turns in the TC track. The three TCs shown all exhibit a

337   sudden northward or northwestern turn in the TC track. For the 6-h forecast, the predicted path is

338   approximately consistent with the actual track, while the 24-h forecast has some deviations. The

339   average distance predicted near the northwest turn of FAXAI is 91.35 km; the error for MITAG's first

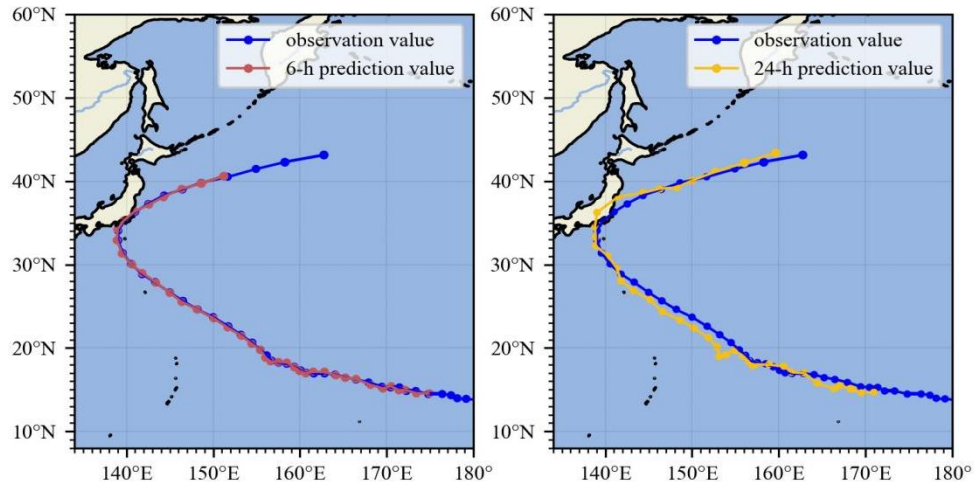
340   turn to the north is 127.02 km, and the error for the second turn to the northwest is 121.91 km. The two

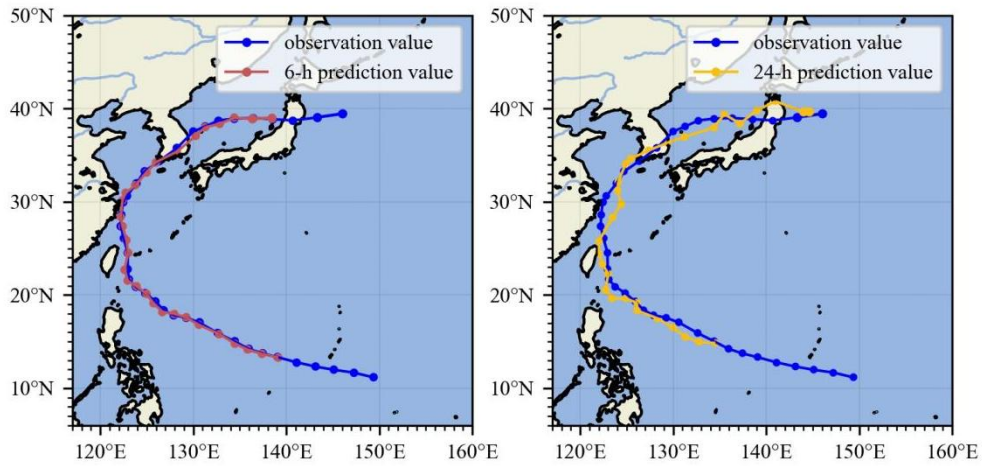
341   average errors in the track forecast for In-fa are 84.27 km and 82.37 km. It can be seen that there is no

342   significant deviation in the forecast around the steering point, but, for some abnormal track changes,

343   such as crossing back over the same location, samples with more significant errors will be generated,

344   reducing the overall average absolute distance error.

345

346 **Figure 5: Forecast tracks of Tropical Cyclone FAXAI (2019) (left: 6 h, right: 24 h).**

347

348 **Figure 6: Forecast tracks of Tropical Cyclone MITAG (2019) (left: 6 h, right: 24 h).**

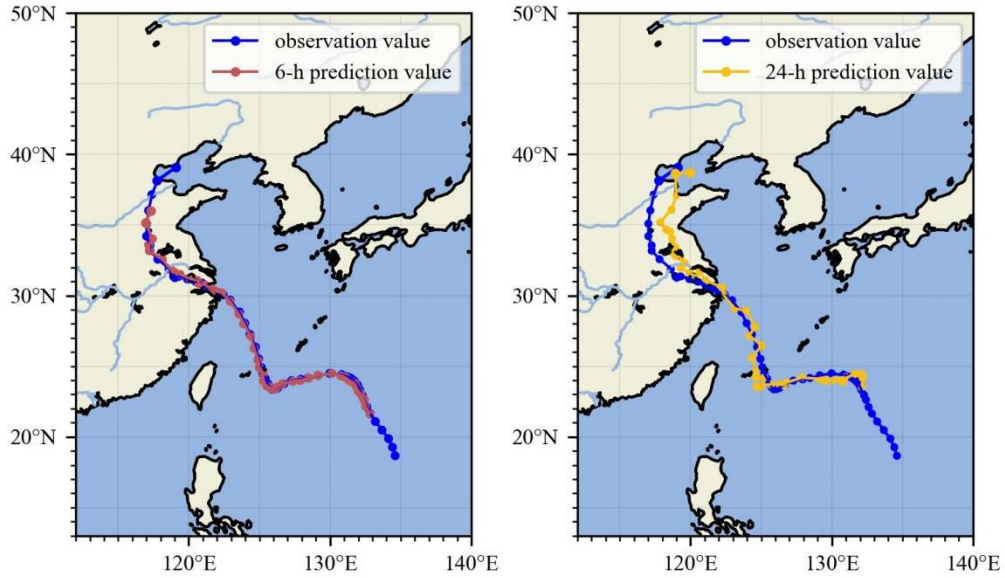



Figure 7: Forecast tracks of Tropical Cyclone IN-FA (2021) (left: 6 h, right: 24 h).

5 Conclusion

The past 24-h TC trajectory and meteorological field data have been used to forecast TC tracks in the Northwest Pacific from hours 6–72 using deep learning methods. First, in order to eliminate data redundancy and reduce the complexity of the prediction model, the random forest algorithm was used for feature extraction of the two-dimensional movement data. Second, three kinds of recurrent neural networks (RNN, LSTM, GRU) were used to evaluate and compare the models based on the input of trajectory features, and it was concluded that GRU performed relatively better in predicting TC tracks. Eventually, we combined GRU with CNN by adding the pre-processed meteorological environmental data around the cyclones (removing the vortex to obtain the steering flow); the CNN models the selected meteorological variables and extracts features, while GRU processes trajectory sequences. GRU_CNN has better prediction results than traditional single deep-learning methods do.

When a new TC generates in the ocean, the GRU_CNN model can quickly provide the forecast track within seconds. Short-term predictions within 12 h of initialization can provide better results than CMO can, and the average distance errors of the forecasts at 6 h and 12 h are 17.22 km and 43.9 km. When the forecast goes beyond 24 h, the model's accuracy declines. The historical steering flow of cyclones has a significant effect on improving the accuracy of short-term forecasting, while, in long-term forecasting, the SST and geopotential height will have a particular impact, which is regarded

368   as a crucial way to expand and improve the application of deep-learning models in TC track forecasting,

369   In addition, the model can accurately predict TCs that suddenly turn to the north or northwest, but there

370   will be a considerable distance error for abnormal trajectories, possibly due to a lack of synoptic

371   analysis in our study.

372   Cyclone prediction has been a challenge in weather forecasting for a long time. With future

373   scientific and technological advances, it is becoming increasingly convenient to obtain meteorological

374   data, and the database has gradually expanded. At the same time, deep-learning models are flexible and

375   can easily be expanded upon. In the future, more data can be integrated, and more valuable features can

376   be extracted to improve the prediction accuracy of the deep-learning model. In addition, model

377   predictor variables will be considered in future work, the inclusion of which can predict more useful

378   information, such as cyclone intensity, rainfall, and wind speed.

379   *Code availability.* The code and model are available as a free access repository on GitHub at .

380   https://zenodo.org/record/7454324#.Y58Dwv3P2UK.

381   *Data availability.* IBTrACS that we used in this study is publicly available. It can be down-loaded at

382   https://www.ncei.noaa.gov/data/international-best-track-archive-for-climate-stewardship-ibtracs/v04r0

383   0/access/netcdf/IBTrACS.WP.v04r00.nc. ERA5 data can be obtained from Copernicus Climate Data

384   Store (https://cds.climate.copernicus.eu).

385   *Author contributions.* Liang Wang wrote the paper and conducted most of the code implementation and

386   data analysis. Bincheng Wan designed the research framework. Shaohui Zhou provided the code and

387   revised the paper. Haofei Sun was involved in data collation and Zhiqiu Gao was responsible for

388   supervision.

389   *Competing interests.* The authors declare that they have no known competing financial interests or

390   personal relationships that could have appeared to influence the work reported in this paper.

391 *Acknowledgments.* This study was supported by the National Key Research and Development Program

392 of the Ministry of Science and Technology of China (2018YFC1506405), and by the National Natural

393 Science Foundation of China (Grants 42175082 and 42222503). We are very grateful to anonymous

394 reviewers for their careful review and valuable comments, which led to substantial improvement of this

395 manuscript.

396 **References**

397 Alemany, S., Beltran, J., Perez, A., and Ganzfried, S.: Predicting Hurricane Trajectories Using a

398   Recurrent Neural Network, Proceedings of the AAAI Conference on Artificial Intelligence, 33,

399   https://doi.org/10.1609/aaai.v33i01.3301468, 2018.

400 Ali, M. M., Kishtawal, C. M., and Jain, S.: Predicting cyclone tracks in the north Indian Ocean: An

401   artificial neural network approach, Geophysical Research Letters, 34,

402   https://doi.org/10.1029/2006gl028353, 2007.

403 Bathla, G.: Stock Price prediction using LSTM and SVR, 2020 Sixth International Conference on

404   Parallel, Distributed and Grid Computing (PDGC), 6-8 Nov. 2020, 211-214,

405   https://doi.org/10.1109/PDGC50313.2020.9315800,

406 Boussioux, L., Zeng, C., Guenais, T., and Bertsimas, D.: Hurricane Forecasting: A Novel Multimodal

407   Machine Learning Framework, Weather and Forecasting, https://arxiv.org/abs/2011.06125, 2022.

408 Brand, S., Buenafe, C. A., and Hamilton, H. D.: Comparison of Tropical Cyclone Motion and

409   Environmental Steering, Monthly Weather Review, 109, 908-909,

410   https://doi.org/10.1175/1520-0493(1981)109<0908:cotcma>2.0.co;2, 1981.

411 Chan, J. and Kepert, J.: Global Perspectives on Tropical Cyclones: From Science to Mitigation,

412   https://doi.org/10.1142/7597, 2010.

413 Chan, J. C.-L.: An Observational Study of the Physical Processes Responsible for Tropical Cyclone

414   Motion, Journal of Atmospheric Sciences, 41, 1036-1048,

415   https://doi.org/10.1175/1520-0469(1984)041<1036:aosotp>2.0.co;2, 1984.

416 Chan, J. C. L., Gray, W. M., and Kidder, S. Q.: Forecasting Tropical Cyclone Turning Motion from

417 Surrounding Wind and Temperature Fields, Monthly Weather Review, 108, 778-792,

418 https://doi.org/10.1175/1520-0493(1980)108<0778:FTCTMF>2.0.CO;2, 1980.

419 Cho, K., Merriënboer, B. v., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.:

420 Learning Phrase Representations using RNN Encoder--Decoder for Statistical Machine

421 Translation, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language

422 Processing (EMNLP), https://doi.org/10.3115/v1/D14-1179, 2014.

423 Díaz-Uriarte, R. and Alvarez de Andrés, S.: Gene selection and classification of microarray data using

424 random forest, BMC bioinformatics, 7, 3, https://doi.org/10.1186/1471-2105-7-3, 2006.

425 Emanuel, K.: Will Global Warming Make Hurricane Forecasting More Difficult?, Bulletin of the

426 American Meteorological Society, 98, 495-501, https://doi.org/10.1175/bams-d-16-0134.1, 2017.

427 Galarneau, T. J. and Davis, C. A.: Diagnosing Forecast Errors in Tropical Cyclone Motion, Monthly

428 Weather Review, 141, 405-430, https://doi.org/10.1175/mwr-d-12-00071.1, 2013.

429 Gao, S., Zhao, P., Pan, B., Li, Y., Zhou, M., Xu, J., Zhong, S., and Shi, Z.: A nowcasting model for the

430 prediction of typhoon tracks based on a long short term memory neural network, Acta

431 Oceanologica Sinica, 37, 8-12, https://doi.org/10.1007/s13131-018-1219-z, 2018.

432 Genuer, R., Poggi, J.-M., and Tuleau-Malot, C.: Variable selection using random forests, Pattern

433 Recognition Letters, 31, 2225-2236, https://doi.org/10.1016/j.patrec.2010.03.014, 2010.

434 George, J. E. and Gray, W. M.: Tropical Cyclone Motion and Surrounding Parameter Relationships,

435 Journal of Applied Meteorology and Climatology, 15, 1252-1264,

436 https://doi.org/10.1175/1520-0450(1976)015<1252:TCMASP>2.0.CO;2, 1976.

437 Giffard-Roisin, S., Yang, M., Charpiat, G., Kégl, B., and Monteleoni, C.: Fused Deep Learning for

438 Hurricane Track Forecast from Reanalysis Data, Climate Informatics Workshop Proceedings 2018,

439 Boulder, United States, 2018-09-19, https://hal.science/hal-01851001, 2018.

440 Giffard-Roisin, S., Yang, M., Charpiat, G., Kumler Bonfanti, C., Kégl, B., and Monteleoni, C.:

441 Tropical Cyclone Track Forecasting Using Fused Deep Learning From Aligned Reanalysis Data,

442 Front Big Data, 3, 1, https://doi.org/10.3389/fdata.2020.00001, 2020.

443   Goerss, J. S.: Tropical Cyclone Track Forecasts Using an Ensemble of Dynamical Models, Monthly

444   Weather Review, 128, 1187-1193,

445   https://doi.org/10.1175/1520-0493(2000)128<1187:tctfua>2.0.co;2, 2000.

446   Goldenberg, S. B., Landsea, C. W., Mestas-Nuñez, A. M., and Gray, W. M.: The Recent Increase in

447   Atlantic Hurricane Activity: Causes and Implications, Science, 293, 474-479,

448   https://doi.org/doi:10.1126/science.1060040, 2001.

449   Graves, A., Jaitly, N., and Mohamed, A.: Hybrid speech recognition with Deep Bidirectional LSTM,

450   2013 IEEE Workshop on Automatic Speech Recognition and Understanding, 8-12 Dec. 2013,

451   273-278, https://doi.org/10.1109/ASRU.2013.6707742,

452   Gray, W. M.: Global View of the Origin of Tropical Disturbances and Storms, Monthly Weather

453   Review, 96, 669-700, http://dx.doi.org/10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2,

454   1968.

455   Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural computation, 9, 1735-1780,

456   https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

457   Holland, G. J.: Tropical Cyclone Motion: Environmental Interaction Plus a Beta Effect, Journal of the

458   Atmospheric Sciences, 40, 328-342,

459   https://doi.org/10.1175/1520-0469(1983)040<0328:tcmeip>2.0.co;2, 1983.

460   Holland, G. J.: Tropical Cyclone Motion. A Comparison of Theory and Observation, Journal of

461   Atmospheric Sciences, 41, 68-75,

462   https://doi.org/10.1175/1520-0469(1984)041<0068:tcmaco>2.0.co;2, 1984.

463   Huang, C., Bai, C., Chan, S., and Zhang, J.: MMSTN: A Multi-Modal Spatial-Temporal Network for

464   Tropical Cyclone Short - Term Prediction, Geophysical Research Letters, 49,

465   https://doi.org/10.1029/2021gl096898, 2022.

466   Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing

467   Internal Covariate Shift, Proceedings of the 32nd International Conference on International

468   Conference on Machine Learning, 37, 448-456, https://doi.org/10.48550/arXiv.1502.03167, 2015.

469   Katsube, K. and Inatsu, M.: Response of Tropical Cyclone Tracks to Sea Surface Temperature in the

470   Western North Pacific, Journal of Climate, 29, 1955-1975,

471   https://doi.org/10.1175/jcli-d-15-0198.1, 2016.

472   Kim, S., Kang, J.-S., Lee, M., and Song, S.-k.: DeepTC: ConvLSTM Network for Trajectory Prediction

473   of Tropical Cyclone using Spatiotemporal Atmospheric Simulation Data, In Spatiotemporal

474   Workshop at 31st Conference on Neural Information Processing Systems, 2018.

475   Kitade, T.: A Numerical Study of the Vortex Motion with Barotropic Models, Journal of the

476   Meteorological Society of Japan. Ser. II, 59, 801-807, https://doi.org/10.2151/jmsj1965.59.6_801,

477   1981.

478   Klein, B., Wolf, L., and Afek, Y.: A Dynamic Convolutional Layer for short rangeweather prediction,

479   2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7-12 June 2015,

480   4840-4848,   https://doi.org/10.1109/CVPR.2015.7299117,

481   Krizhevsky, A., Sutskever, I., and Hinton, G.: ImageNet Classification with Deep Convolutional

482   Neural Networks, Neural Information Processing Systems, 25, https://doi.org/10.1145/3065386,

483   2012.

484   Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-Based Learning Applied to Document

485   Recognition, Proceedings of the IEEE, 86, 2278-2324, https://doi.org/10.1109/5.726791, 1998.

486   Li-min, S., Gang, F. U., Xiang-chun, C., and Jian, Z.: Application of BP neural network to forecasting

487   typhoon tracks, Journal of Natural Disasters, 18, 104-111,

488   https://doi.org/0.3969/j.issn.1004-4574.2009.06.018, 2009.

489   Liu, H., Mi, X., and Li, Y.: Smart multi-step deep learning model for wind speed forecasting based on

490   variational mode decomposition, singular spectrum analysis, LSTM network and ELM, Energy

491   Conversion and Management, 159, 54-64, https://doi.org/10.1016/j.enconman.2018.01.010, 2018.

492   Liu, Z., Hao, K., Geng, X., Zou, Z., and Shi, Z.: Dual-Branched Spatio-Temporal Fusion Network for

493   Multihorizon Tropical Cyclone Track Forecast, IEEE Journal of Selected Topics in Applied Earth

494   Observations and Remote Sensing, 15, 3842-3852, https://doi.org/10.1109/JSTARS.2022.3170299,

495   2022.

496   Lownam, C.: The NCAR-AFWA tropical cyclone bogussing scheme. A report prepared for the Air

497   Force Weather Agency (AFWA), National Center for Atmospheric Research, 2001.

498   Moradi Kordmahalleh, M., Gorji Sefidmazgi, M., Homaifar, A., and Liess, S.: Hurricane Trajectory

499   Prediction Via a Sparse Recurrent Neural Network, 2015.

NEUMANN, C. J. and HOPE, J. R.: Performance Analysis of the HURRAN Tropical Cyclone Forecast System, Monthly Weather Review, 100, 245-255, https://doi.org/10.1175/1520-0493(1972)100<0245:paotht>2.3.co;2, 1972.

Ortega, L. C., Otero, L. D., Solomon, M., Otero, C. E., and Fabregas, A.: Deep learning models for visibility forecasting using climatological data, International Journal of Forecasting, https://doi.org/10.1016/j.ijforecast.2022.03.009, 2022.

Roy, C. and Kovordányi, R.: Tropical cyclone track forecasting techniques ― A review, Atmospheric Research, 104-105, 40-69, https://doi.org/10.1016/j.atmosres.2011.09.012, 2012.

Ruttgers, M., Lee, S., Jeon, S., and You, D.: Prediction of a typhoon track using a generative adversarial network and satellite images, Sci Rep, 9, 6057, https://doi.org/10.1038/s41598-019-42339-y, 2019.

Schulthess, T. C., Bauer, P., Wedi, N., Fuhrer, O., Hoefler, T., and Schär, C.: Reflecting on the Goal and Baseline for Exascale Computing: A Roadmap Based on Weather and Climate Simulations, Computing in Science & Engineering, 21, 30-41, https://doi.org/10.1109/MCSE.2018.2888788, 2019.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W. K., and Woo, W.-c.: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, Proceedings of the 28th International Conference on Neural Information Processing Systems, 1, 802-810, https://doi.org/10.48550/arXiv.1506.04214, 2015.

Sun, Y., Zhong, Z., Li, T., Yi, L., Camargo, S. J., Hu, Y., Liu, K., Chen, H., Liao, Q., and Shi, J.: Impact of ocean warming on tropical cyclone track over the western north pacific: A numerical investigation based on two case studies, Journal of Geophysical Research: Atmospheres, 122, 8617-8630, https://doi.org/10.1002/2017JD026959, 2017.

Wang, C. and Fu, Y.: Ship Trajectory Prediction Based on Attention in Bidirectional Recurrent Neural Networks, 529-533, https://doi.org/10.1109/isctt51595.2020.00100, 2020.

Wang, Y., Zhang, W., and Fu, W.: Back Propogation(BP)-neural network for tropical cyclone track forecast, Proceedings - 2011 19th International Conference on Geoinformatics, Geoinformatics 2011, 1-4, https://doi.org/10.1109/GeoInformatics.2011.5981095, 2011.

528 Zhang, Y., Chandra, R., and Gao, J.: Cyclone Track Prediction with Matrix Neural Networks, 2018

529 International   Joint   Conference   on   Neural   Networks   (IJCNN),   1-8,

530 https://doi.org/10.1109/IJCNN.2018.8489077, 2018.

531