The Application Research of FCN Algorithm in Different Severe Convection Short-Time Nowcasting Technology in China, Gansu Province

Wubin Huang ¹, Jing Fu ¹*, Xin Xin Feng ², Runxia Guo ¹, Junxia Zhang ¹ and Yu Lei ³

¹ Lanzhou Central Meteorological Observatory, Lanzhou 730020, China; hwb_0707@outlook.com (W.H.); guorunxia25@126.com (R.G.); zjx01182@163.com (J.Z.)
² Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong 999077, China; xfengba@connect.ust.hk
³ Meteorological Bureau of Lanzhou, Lanzhou 730020, China; ylei08@163.com
* Correspondence: fujing_1228@163.com

Abstract: This study explores the application of the fully convolutional network (FCN) algorithm to the field of meteorology, specifically for the short-term nowcasting of severe convective weather events such as hail, convective wind gust (CG), thunderstorms, and short-term heavy rain (STHR) in Gansu. The training data come from the European Center for Medium-Range Weather Forecasts (ECMWF) and real-time ground observations. The performance of the proposed FCN model, based on 2017 to 2021 training datasets, demonstrated a high prediction accuracy, with an overall error rate of 16.6%. Furthermore, the model exhibited an error rate of 18.6% across both severe and non-severe weather conditions when tested against the 2022 dataset. Operational deployment in 2023 yielded an average critical success index (CSI) of 24.3%, a probability of detection (POD) of 62.6%, and a false alarm ratio (FAR) of 71.2% for these convective events. It is noteworthy that the predicting performance for STHR was particularly effective with the highest POD and CSI, as well as the lowest FAR. CG and hail predictions had comparable CSI and FAR scores, although the POD for CG surpassed that for hail. The FCN model’s optimal performances in terms of hail prediction occurred at the 4th, 8th, and 10th forecast hours, while for CG, the 6th hour was most accurate, and for STHR, the 2nd and 4th hours were most effective. These findings underscore the FCN model’s ideal suitability for short-term forecasting of severe convective weather, presenting extensive prospects for the automation of meteorological operations in the future.

Keywords: severe convective weather classification; FCN algorithm; short-term nowcasting; Qinghai–Tibet Plateau; Gansu

1. Introduction

Severe convective weather (SCW) typically refers to a range of disastrous weather phenomena including thunderstorms and/or lightning, hail, convective gusts (CG), short-duration heavy rainfall (STHR), and tornadoes, all of which are generated by deep moist convection (DMC) [1]. This type of weather has a widespread distribution across the globe, with China being one of the regions with a high incidence of SCW. The rapid changes, short life spans, and strong local characteristics of SCW pose significant challenges to daily life, agriculture, the aviation industry, etc. [2,3]. In operational weather forecasting, SCW is defined as any weather event that includes hail with a diameter of 5 mm or more, tornadoes of any intensity, CG of more than 17 m/s, and STHR exceeding 20 mm/h [3].

China’s Gansu province, a confluent area of westerlies and monsoons in the summer, is located on the northeastern side of the Qinghai–Tibet Plateau [4,5]. Due to the interactive atmospheric circulation characteristics and complex topographic effects, SCW can induce hail, CG, and STHR, causing serious casualties and property losses [6]. Consequently, it
is of utmost importance to improve the accuracy of forecasts for such weather events to mitigate related disasters [7].

Given the meso- and micro-scale nature of severe convective systems, forecasts focus on monitoring and short-term nowcasting [8]. Sun Jisong [9] and Yu Xiaoding [10] have both provided clear and in-depth discussions regarding forecasting severe convective weather. In China, the forecasting service typically refers to the 0–2 h range as nowcasting and 2–12 h as short-term forecasting [11]. Nowcasting primarily relies on empirical forecasting and the extrapolation of radar echoes and cloud images [7,12,13], while traditional short-term forecasting of severe convection mainly falls into two categories: pattern recognition [14,15] and ingredient-based forecasting [16,17]. With the development of high-resolution numerical weather prediction models, the predictive capabilities of forecasters have been challenged. Forecasters should extract useful dynamic data as well as moisture-, energy-, and topographical-related information that is conducive to severe convective weather from a huge amount of data to form a reliable forecast [4]. However, this approach has its limitations. For example, the significant climatic differences across regions make it difficult to accurately forecast severe convection in different areas using a unified threshold [18–22].

With the rapid development of numerical weather prediction and meteorological observation networks, the amount of available meteorological information has proliferated enormously, making it challenging for forecasters to detect and synthesize valuable data from the vast array.

In recent years, machine learning algorithms have been widely applied in meteorological detection and forecasting and have achieved certain accomplishments [23–27]. For instance, Chen Jinpeng [28] demonstrated that the hourly precipitation forecast corrected by the convolutional neural network significantly outperforms the frequency-matching method, and Zhou Kanghui [29] established a short-term forecasting method for lightning strike areas using the LightningNet-NWP algorithm, which yielded better results than using multi-source observation data and high-resolution numerical model forecast data alone. Zhang Yuchen [30] proposed a deep learning model for nowcasting extreme precipitation events, integrating physical evolution laws and conditional learning methods into a neural network framework, which enhanced the forecasting ability for 0–3 h extreme precipitation events. In general, deep neural networks have unmatched advantages in automatic high-level feature extraction and image classification [31,32]. Applying deep learning technology can effectively extract data features and improve classification accuracy [33]. With the development of deep learning, many neural network algorithms for semantic segmentation have been proposed, such as fully convolutional networks (FCNs) [34], pyramid scene parsing network (PSPNets) [35], SegNet [36], etc. Many scholars have introduced these algorithms into the meteorological field [37], such as the deep semantic segmentation model that extracts multi-source observation data from satellites, radar, and lightning detectors introduced by Zhou Kanghui [38]. This model effectively achieves the integration of multi-source observation data, with good forecasting results for 0–1 h lightning events. Li Mengya [39] established a satellite detection method for nighttime sea areas/low clouds over the Northwest Pacific based on the FCN-CRF algorithm, the performance of which surpassed the dual-channel interpolative method.

Classifying severe convection is essentially a semantic segmentation task. We explored how deep learning can effectively utilize multi-source observation data and high-resolution numerical model prediction data, thereby enhancing the accuracy of short-term forecasting for classified severe convection, particularly within the region of Gansu, where severe convection samples are scarce. Furthermore, sample classes are imbalanced, which requires further elucidation through research findings. Therefore, based on the occurrence and development characteristics of severe convective systems, this study aims to establish a deep-learning semantic segmentation model suitable for Gansu via the utilization of multi-source observation data and high-resolution numerical model forecasts. The objective of this paper is to extract effective convective forecast information to achieve more accurate
short-term forecasting of classified severe convection, providing a reference for better utilizing these forecast products to enhance operational forecasting capabilities.

2. FCN Algorithm and Model Construction

Semantic segmentation is a technology that partitions a scene image into several meaningful image regions, and then assigns a class label to each pixel [40]. This technique can be applied to areas such as the classification of severe convective weather, transforming multi-source meteorological data into image product-like forms to capture meteorological characteristics of different severe convective weather events and the corresponding correlation analysis. In 2015, Long et al. first introduced the FCN [34], which is capable of pixel-level semantic segmentation for images of any size. The advantages of the FCN model include accelerated training speed, reduced operational time, and enhanced classification accuracy [41], as illustrated in Figure 1. Below is a brief introduction to the main components of the FCN algorithm:

![Figure 1. Model structure diagram of FCN model.](image)

Convolutionalization: Traditional convolutional neural networks (CNNs) typically use fully connected layers for classification at the end. The FCN, however, replaces this fully connected layer with a convolutional layer to enable direct pixel-level segmentation. Through five rounds of convolution and pooling, the image size is progressively reduced by factors of 2, 4, 8, 16, and 32, extracting the primary information and roughly determining the category of each region.

Upsampling: In the network process, multiple downsampling operations are conducted to reduce the size of the feature layer. To obtain a prediction layer that matches the size of the original image, a 32-fold upsampling operation is applied to the last layer. This ensures that the resolution of the prediction results matches that of the original image, similar to a deconvolution operation.

Skip connections: Classification predictions are made using multi-layer information by interconnecting different levels of information. This maintains information from other layers while extracting the main information, which contributes to the enhancement of the model's performance.

The innovation of the FCN model lies in the utilization of traditional CNNs for pixel-level semantic segmentation tasks while implementing this goal through convolutionalization, upsampling, and skip connections. These technical improvements not only enhance training speed and classification accuracy, but also enable the model to process images of any size, which is of great significance in segmentation tasks. This technological advancement is of vital importance for applications such as severe convective weather forecasting.

Model training typically includes five key steps (Figure 2): data collection, data processing, model training, cross-validation, and model evaluation. Data collection (Step 1) is a crucial step to ensure that the model has sufficient information for training and prediction, involving the gathering of classified severe convective reality data (dependent variables), ground reality observations, and numerical model products during severe convective
episodes (independent variables). Data processing (Step 2) involves selection and cleaning, i.e., removing samples with incomplete feature variables or containing outliers, to ensure data quality. It also includes normalizing sample data to ensure consistent scales between different features, aiding in the model’s stable training. The model training step (Step 3) uses the processed dataset to establish and train the FCN model. Training typically involves multiple iterations, where the model continuously tries to fit the data and adjusts weights and parameters through optimization algorithms. During cross-validation (Step 4), the ratio of model training samples to the cross-validation set is 8:2. The model is trained on the training samples and then evaluated on the validation set. This helps to verify whether the model is overfitting or underfitting and provides guidance for hyperparameter tuning. Once the model is optimized, it is assessed based on the cross-validation set and test set using common model evaluation algorithms (Step 5) to evaluate accuracy and generalizability.

Figure 2. Flow chart of severe convection algorithm of FCN model.

3. Data Used and Model Validation Methods
3.1. Introduction of Independent Variables for Modeling
Extensive research indicates that different types of severe convective weather events exhibit different environmental conditions and threshold characteristics of physical quantity characteristics [26,42,43]. Drawing upon the distinctive characteristics of various severe convective weather events observed in Gansu Province, our study integrates the findings of Huang Yuxia [44] and Taszarek et al. [45], as well as additional research that leverages the ERA5 reanalysis data developed by the European Center for Medium-Range Weather Forecasts (ECMWF). Through a meticulous statistical analysis, we explored the correlation between multifaceted atmospheric variables at different altitudes and the occurrence of severe convective weather events. Our rigorous selection process identified variables that exhibit a strong correlation with CG, Hail, and STHR events. Key variables such as convective available potential energy (CAPE), potential vorticity (PVORT), temperature (TMP), geopotential height (HGT), divergence (DIV), horizontal wind (UV), and specific humidity (Q) at the 700 hPa and 500 hPa levels were scrutinized. Furthermore, we incorporated vertical wind (W), saturated specific humidity (SPFH), total precipitable water (PW), and 2 m dew point temperature (DPT-2M) into our analysis. In conclusion, 24 feature quantities were included in the analysis (Table 1).
Table 1. Characteristic quantity of FCN model.

<table>
<thead>
<tr>
<th>Factors</th>
<th>PVORT</th>
<th>TMP</th>
<th>HGT</th>
<th>DIV</th>
<th>UV</th>
<th>Q</th>
<th>W</th>
<th>SPFH</th>
<th>CAPE</th>
<th>PW</th>
<th>DPT-2M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level (hPa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
</tr>
<tr>
<td>surface</td>
<td>surface</td>
<td>all layers</td>
<td>surface</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ECMWF’s numerical forecast products have a time resolution of 3 h, with a horizontal resolution of 0.125° × 0.125° for the surface layer (CAPE, DPT-2M, PW) and 0.25° × 0.25° for the pressure layers. Temporal and spatial downscaling of the 24 feature quantities was necessary to align with the spatiotemporal resolution of the intelligent grid forecasting service. In the spatial domain, the technique of bilinear interpolation was employed, while in the temporal realm, we applied non-decreasing cubic spline interpolation [46] with a spatial resolution of 5 km and a time resolution of 1 h.

Prior to the model input, standardization or normalization of the image data was required. The ECMWF feature quantities, which comprise various dimensions, also underwent similar data preprocessing [28]. Unlike images, the ECMWF feature quantities had clear upper and lower bounds, and their numerical spatial distribution and relative magnitudes were crucial for the model’s output. To preserve the spatial and interval information regarding the values, normalization was employed as follows (1):

\[
x' = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]

where \(x_{\text{max}}\) and \(x_{\text{min}}\) represent the maximum and minimum values of the feature quantity data \(x\), respectively.

3.2. Introduction of Dependent Variables and Samples for Modeling

Data belonging to STHR, hail, and CG—three types of severe convective weather in Gansu Province from 2017 to 2023—were collected and organized. STHR is defined as hourly rainfall exceeding 20 mm, hail is characterized by the presence of hailstones with a diameter equal to or exceeding 5 mm, and CG is determined in accordance with the prevailing operational standards proposed by the China Meteorological Administration. Hourly accumulated lightning stroke data from the National Lightning Detection Network were accumulated onto a 0.5° × 0.5° (multiples of 0.5° in latitude and longitude) grid using the nearest neighbor method, forming an hourly lightning count grid data, which was then interpolated to the station locations with the application of the nearest neighbor method. A station was considered to have encountered CG if the hourly cumulative number of lightning strokes was \(\geq 1\) and the hourly gust wind speed was \(\geq 17\) m/s within an hour.

Based on the longitude and latitude of the weather stations where severe convective weather was recorded, along with the corresponding times of occurrence, data from the model products within 3 h before and after the events were extracted as samples. For instance, if STHR was observed at a location from 18:00 to 22:00, the sample extraction time would be from 15:00 to 01:00 on the subsequent day. Meanwhile, the 18:00 to 22:00 period would be selected as the severe convective weather sample, and the 15:00 to 17:00 and 23:00 to 01:00 periods would be chosen as non-severe convective weather samples. Moreover, due to the infrequency of severe convective weather in Gansu Province, samples from days without severe convective weather were also collected to better establish the identification of non-severe convective weather, ensuring a sample ratio of approximately 10:1 between days characterized by non-severe and severe convective weather, respectively.

With the methods described above, a dataset of 26,747 severe convective weather samples and 269,746 non-severe convective weather samples for the years 2017–2023 was established. Given that the FCN model requires a large training set for modeling,
186,095 samples from 2017–2021 were used as the training set, with a training-to-cross-validation-set ratio of 8:2. The 2022 samples served as an independent test set, comprising 4093 severe convective weather samples and 51,024 non-severe convective weather samples. The 2023 samples were used as an independent validation set to verify the effectiveness in operational application, including 3667 severe convective weather samples and 51,614 non-severe convective weather samples. To facilitate the recognition of machine language, this study set labels for four types: hail, CG, STHR, and no severe convection. Within the sample set, the configuration of the labels were as follows: the designation of label 1 denoted the absence of severe convection, label 2 represented hail, label 3 represented CG, and label 4 represented STHR. The distribution of training and validation period sets and label classification are shown in Table 2.

Table 2. Sample numbers and label values of each type of severe convective weather in the training set (2017–2021), testing set (2022), and validation set (2023) of the model.

<table>
<thead>
<tr>
<th></th>
<th>Non-SCW</th>
<th>Hail</th>
<th>CG</th>
<th>STHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>167,108</td>
<td>1125</td>
<td>11,972</td>
<td>5890</td>
</tr>
<tr>
<td>Testing set</td>
<td>51,024</td>
<td>168</td>
<td>2258</td>
<td>1667</td>
</tr>
<tr>
<td>Validation set</td>
<td>51,614</td>
<td>167</td>
<td>2331</td>
<td>1169</td>
</tr>
<tr>
<td>Label value</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

3.3. Model Validation Indicators

In the meteorological field, the percentage of doom (POD), false alarm rate (FAR), and critical success index (critical success index) (CSI) can accurately evaluate the recognition and prediction effect of short-term heavy precipitation [47], which facilitates the assessment of the identification and forecasting effectiveness of severe convective weather. The formulas are as follows (2)–(5):

\[ \text{FIR} = \frac{\text{false identification number}}{\text{total sample number}} \times 100\% \quad (2) \]

\[ \text{POD} = \frac{\text{correct identification number}}{\text{total sample number}} \times 100\% \quad (3) \]

\[ \text{CSI} = \frac{\text{correct identification number}}{\text{total sample number} + \text{false identification number}} \times 100\% \quad (4) \]

\[ \text{FAR} = \frac{\text{false identification number}}{\text{correct identification number} + \text{false identification number}} \times 100\% \quad (5) \]

The criterion used in this paper is consistent with the current forecasting business, that is, the point-to-point comparison and verification of the prediction results of all time models in the duration period of severe convective weather observed on the ground. The time range of discrimination is the occurrence period of the observational reality, and the space range is a radius of 40 km. As illustrated in Figure 3, the evaluation methodology for the FCN models’ performance involves a spatial criterion centered on observed CG occurrences. Specifically, if CG is detected at a given time, the predictive accuracy of all FCN models is assessed within a 40 km radius around the CG point. Predictions within this radius are deemed accurate (labeled as a “hit”) if the model correctly identifies the occurrence as a short-duration, severe weather event. Conversely, if the model’s prediction within this designated area either categorizes the event as a different type of severe convective weather or fails to recognize it as a severe weather event altogether, the prediction is classified as a misjudgment.
Based on the above modeling steps and testing methods, we established a classified severe convection weather forecast model for the northeast Tibetan Plateau based on the FCN algorithm. The model output the deterministic forecast results of CG, Hail, STHR, and non-SCW weather types with a spatial resolution of 5 km and a temporal resolution of 1 h.

4. Results Analysis

4.1. Training Set Effectiveness Evaluation

Using the 2017–2021 severe convective weather observational data as the training set, the FCN model was trained and cross-validated. The evaluation of the training set, as presented in Table 3, scrutinized the rates of misjudgment for three types of severe convective weather: hail, CG, and STHR. STHR exhibited the lowest misjudgment rate, which was only 21.8%. The model mainly misclassified STHR as CG. Conversely, hail showed the highest misjudgment rate, reaching 33.2%, which was primarily due to the misclassification of hail as CG. As for CG, the misjudgment rate settled at 24.9%, with the primary misclassifications including STHR and non-severe convection. The misjudgment rate for non-severe convective weather was relatively low, at 15.7%. By aggregating the total misjudgments across all categories and dividing this value by the total number of training set samples, the overall misjudgment rate for the FCN model for severe convective weather was only 16.6%. Considering these metrics, the FCN model demonstrated high accuracy in categorizing severe convective weather.

Table 3. Severe convection weather effect evaluation of FCN model in training period cross-validation set.

<table>
<thead>
<tr>
<th>Obs.</th>
<th>Classification of the FCN Model</th>
<th>Overall FIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hail</td>
<td>150 46 16 13</td>
<td>33.2%</td>
</tr>
<tr>
<td>CG</td>
<td>36 1798 320 241</td>
<td>24.9%</td>
</tr>
<tr>
<td>STHR</td>
<td>22 194 921 41</td>
<td>21.8%</td>
</tr>
<tr>
<td>Non-SCW</td>
<td>366 4619 262 28,174</td>
<td>15.7%</td>
</tr>
</tbody>
</table>

Figure 4 displays the cross-validation scores for the model during the training period. CSI, POD, and FAR were used to evaluate the three types of severe convective weather: hail, CG, and STHR. STHR performed the best, with CSI and POD reaching 51.9% and 78.2%, respectively. In comparison, hail and CG had similar CSI scores, but were notably lower than STHR. The POD for CG was slightly lower than for STHR, but significantly higher than for hail, which was 66.8%. STHR exhibited the lowest FAR score, the value of which was 39.4%. Hail and CG had similar FAR values. The model’s average CSI was 33.3%, accompanied by an average POD of 73.4% and an average FAR of 62.1%. These results indicate that, after sufficient training, the FCN model is capable of classifying severe convective weather with high accuracy and efficiency.
4.2. Independent Test Set Effectiveness Evaluation

The FCN model, constructed based on the 2017–2021 training set, was applied to an independent 2022 dataset containing 55,117 samples. The results of the independent sample assessment, as shown in Table 4 and Figure 5, revealed misjudgment rates of 48.2%, 29.0%, and 27.2% for hail, CG, and STHR, respectively. In contrast, the misjudgment rate for non-severe convective weather was the lowest, at 14.5%. Considering all categories, the overall misjudgment rate was 18.6%. Similarly to the cross-validation results during the training period, hail was primarily misclassified as CG, and STHR was mainly misclassified as CG and non-severe convection. Although the test set’s performance was slightly lower than that of the training set, the average CSI still reached 25.8%, with an average POD of 65.2% and an average FAR of 70.0%. Similarly to the cross-validation set during the training period, the CSI and POD for STHR proved to be superior, accompanied by the lowest FAR. CSI and FAR scores for CG and hail were similar and significantly lower than for STHR. Moreover, the POD for CG closely approached the POD of STHR, while hail exhibited the lowest POD at 51.8%, which is consistent with the training set results. These results suggest that, although there was a slight decline in the independent test set, the model still demonstrated notable accuracy and stability in classifying severe convective weather.

Figure 4. Classification forecast score results of FCN model in the training period cross-validation set.

Figure 5. Classification forecast results score of FCN model in independent test set.
Table 4. Severe convection weather effect evaluation of FCN model in independent test set.

<table>
<thead>
<tr>
<th>Obs.</th>
<th>Classification of FCN Model</th>
<th>Hail</th>
<th>CG</th>
<th>STHR</th>
<th>Non-SCW</th>
<th>FIR</th>
<th>Overall FIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hail</td>
<td></td>
<td>87</td>
<td>44</td>
<td>18</td>
<td>19</td>
<td>48.2%</td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td></td>
<td>67</td>
<td>1603</td>
<td>128</td>
<td>460</td>
<td>29.0%</td>
<td>18.6%</td>
</tr>
<tr>
<td>STHR</td>
<td></td>
<td>44</td>
<td>281</td>
<td>1214</td>
<td>128</td>
<td>27.2%</td>
<td></td>
</tr>
<tr>
<td>Non-SCW</td>
<td></td>
<td>281</td>
<td>6117</td>
<td>980</td>
<td>43,646</td>
<td>14.5%</td>
<td></td>
</tr>
</tbody>
</table>

4.3. Independent Validation Set Effectiveness Evaluation

The well-trained FCN model was implemented in operational service and evaluated using 55,281 independent samples from 2023. The results of the evaluation conducted for the business application, as depicted in Table 5 and Figure 6, revealed misjudgment rates of 46.7%, 34.0%, and 31.5% for hail, CG, and STHR, respectively. Notably, the misjudgment rate for non-severe convective weather was lower than 13.3%, and the overall misjudgment rate was 18.3%.

Table 5. Severe convection weather effect evaluation of FCN model in independent validation set.

<table>
<thead>
<tr>
<th>Obs.</th>
<th>Classification of FCN Model</th>
<th>Hail</th>
<th>CG</th>
<th>STHR</th>
<th>Non-SCW</th>
<th>FIR</th>
<th>Overall FIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hail</td>
<td></td>
<td>89</td>
<td>58</td>
<td>12</td>
<td>8</td>
<td>46.7%</td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td></td>
<td>58</td>
<td>1539</td>
<td>340</td>
<td>394</td>
<td>34.0%</td>
<td></td>
</tr>
<tr>
<td>STHR</td>
<td></td>
<td>23</td>
<td>222</td>
<td>801</td>
<td>123</td>
<td>31.5%</td>
<td></td>
</tr>
<tr>
<td>Non-SCW</td>
<td></td>
<td>388</td>
<td>6053</td>
<td>421</td>
<td>44,752</td>
<td>13.3%</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. Classification forecast results score of FCN model in independent validation set.

Similarly to the training and test set results, hail was primarily misclassified as CG, and STHR mainly as CG and non-severe convection. Although the performance of the independent validation set was slightly lower than that of the training period cross-validation set and test set, the gap was not significant. The model’s average CSI was 24.3%, with an average POD maintained at 62.6% and an average FAR of 71.2%. In accordance with previous findings, the CSI and POD for STHR were the best, while the CSI and FAR for CG and hail exhibited resemblances and were notably lower than that of STHR. Furthermore, the POD for CG was close to that for STHR, surpassing hail by 10%, which aligned with the training period cross-validation set and test set results. These outcomes indicate that, despite a slight decrease in performance during operational implementation, the model still maintained a certain level of accuracy and stability in classifying severe convective weather, thereby supporting its suitability for operational application.
Across the entire cross-validation set, test set, and independent validation set (Figure 7) for the training period, the FCN model’s average CSI scores were: hail at 17.6%, CG at 20.3%, and STHR with the highest at 45.5%. In terms of average POD, hail was the lowest at 57.3%, while CG and STHR both exceeded 70%. As for the average FAR, hail and CG exhibited similarities, both of which were higher than 77%, while the average FAR for STHR was the lowest at 45.5%. These outcomes, in comparison with previous CNN-based models for severe convective weather forecasting [3], demonstrate that the FCN model exhibited superior performance in terms of CSI and POD for hail, CG, and STHR, also in addition to excelling in FAR. Overall, the FCN model-based classification of severe convective weather shows promising accuracy and has wide-ranging prospects in the automated identification of and early warning applications for severe convective weather.

![Figure 7](image_url)

**Figure 7.** Classification prediction results score of FCN model in the whole sample.

### 4.4. Hourly Effectiveness Analysis

To explore the efficacy of the FCN model in classifying forecasts, an evaluation of hourly results was conducted. In order to visually illustrate the variations among the three categories of severe convective weather across different validation indicators for each hour, anomalies in the validation results were graphed (Figure 8). Specifically, the score of each indicator for each type of severe convective weather at each hour in the different sample sets was calculated by subtracting the score of that indicator for the entire sample set (as shown in Figure 7). For instance, the anomaly for the hail CSI at the first hour in the cross-validation set (7.1%) is the hail CSI score in the cross-validation set (24.7%) minus the average hail CSI score for the entire sample set (17.6%).

Based on the CSI depicted in Figure 7, it is evident that, except for the second and eleventh hours in the cross-validation set, hail consistently scored above the average by more than 4.0%, reaching its peak anomaly of over 10.0% in the eighth hour. For the test set, apart from the fourth hour, the scores for most hours were lower than the average value, especially the first, sixth, and eleventh hours, which showed a deviation of more than 4.0% below the average. Within the independent validation set, the third hour exhibited the most exemplary performance, with a score value of 0.5%, whereas the second hour was the poorest, the deviation of which was −8.0%. Additionally, the first, fifth, seventh, and twelfth hours also scored more than 4.0% below average.

For CG in the cross-validation set, the eleventh hour obtained the lowest score, and the tenth hour achieved the highest score. Furthermore, the first, second, fourth, fifth, sixth, ninth, and tenth hours were all significantly higher than the average value by more than 5.0%. In the test set, the sixth hour emerged as the best outstanding performer with the highest score, while the fifth hour was the worst. Similarly to hail, most hours scored below average, a similar deviation as that observed for the hail results.
<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hail CSI</strong></td>
<td>7.1</td>
<td>1.3</td>
<td>4.3</td>
<td>7.4</td>
<td>5.3</td>
<td>4.2</td>
<td>5.4</td>
<td>10.2</td>
<td>8.9</td>
<td>5.8</td>
<td>2.6</td>
<td>5.5</td>
<td>7.0</td>
<td>6.0</td>
<td>3.3</td>
<td>5.3</td>
<td>5.5</td>
<td>5.1</td>
<td>4.0</td>
<td>2.5</td>
<td>5.9</td>
</tr>
<tr>
<td><strong>CG CSI</strong></td>
<td>-4.7</td>
<td>-1.7</td>
<td>-0.6</td>
<td>1.0</td>
<td>-3.4</td>
<td>-4.2</td>
<td>-2.7</td>
<td>-1.3</td>
<td>-1.9</td>
<td>-1.4</td>
<td>-4.1</td>
<td>-2.4</td>
<td>-4.3</td>
<td>-8.0</td>
<td>0.5</td>
<td>-2.9</td>
<td>-4.3</td>
<td>-1.6</td>
<td>-5.3</td>
<td>-0.5</td>
<td>-2.2</td>
</tr>
<tr>
<td><strong>STHR CSI</strong></td>
<td>3.6</td>
<td>8.8</td>
<td>8.5</td>
<td>7.3</td>
<td>3.8</td>
<td>7.0</td>
<td>5.0</td>
<td>3.9</td>
<td>4.3</td>
<td>7.1</td>
<td>9.2</td>
<td>4.0</td>
<td>0.7</td>
<td>-0.7</td>
<td>-2.8</td>
<td>-1.2</td>
<td>-1.4</td>
<td>-2.1</td>
<td>-4.0</td>
<td>-2.2</td>
<td>-1.4</td>
</tr>
<tr>
<td><strong>Hail POD</strong></td>
<td>15.6</td>
<td>3.0</td>
<td>4.2</td>
<td>12.2</td>
<td>15.5</td>
<td>8.4</td>
<td>8.3</td>
<td>16.1</td>
<td>14.4</td>
<td>11.8</td>
<td>4.2</td>
<td>6.8</td>
<td>-9.7</td>
<td>1.6</td>
<td>-1.4</td>
<td>-1.8</td>
<td>-9.4</td>
<td>-10.2</td>
<td>-3.8</td>
<td>-1.0</td>
<td>-6.8</td>
</tr>
<tr>
<td><strong>CG POD</strong></td>
<td>-4.5</td>
<td>-4.4</td>
<td>0.5</td>
<td>0.9</td>
<td>-8.2</td>
<td>-1.5</td>
<td>-5.2</td>
<td>3.4</td>
<td>2.7</td>
<td>2.3</td>
<td>2.0</td>
<td>-2.0</td>
<td>-8.9</td>
<td>-2.0</td>
<td>-4.2</td>
<td>-3.6</td>
<td>-4.0</td>
<td>-6.8</td>
<td>-1.8</td>
<td>-6.1</td>
<td>-4.3</td>
</tr>
<tr>
<td><strong>STHR POD</strong></td>
<td>5.7</td>
<td>5.9</td>
<td>-2.1</td>
<td>9.0</td>
<td>6.4</td>
<td>7.2</td>
<td>-2.5</td>
<td>3.8</td>
<td>8.2</td>
<td>9.9</td>
<td>1.0</td>
<td>0.8</td>
<td>-2.6</td>
<td>4.2</td>
<td>0.8</td>
<td>-0.8</td>
<td>-7.0</td>
<td>5.0</td>
<td>2.8</td>
<td>4.4</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Hail FAR</strong></td>
<td>2.6</td>
<td>9.2</td>
<td>8.4</td>
<td>7.3</td>
<td>1.8</td>
<td>8.1</td>
<td>-0.9</td>
<td>2.6</td>
<td>0.9</td>
<td>9.3</td>
<td>11.2</td>
<td>0.8</td>
<td>4.3</td>
<td>4.2</td>
<td>4.4</td>
<td>3.5</td>
<td>2.0</td>
<td>-2.6</td>
<td>4.1</td>
<td>2.8</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>CG FAR</strong></td>
<td>-2.6</td>
<td>-8.7</td>
<td>-4.4</td>
<td>-2.7</td>
<td>-1.3</td>
<td>-9.2</td>
<td>-0.1</td>
<td>-7.0</td>
<td>-3.5</td>
<td>-8.6</td>
<td>-0.1</td>
<td>-9.0</td>
<td>1.6</td>
<td>-3.4</td>
<td>-3.8</td>
<td>8.1</td>
<td>8.3</td>
<td>1.6</td>
<td>-0.6</td>
<td>4.4</td>
<td>-0.1</td>
</tr>
<tr>
<td><strong>STHR FAR</strong></td>
<td>12.8</td>
<td>1.2</td>
<td>0.3</td>
<td>10.3</td>
<td>1.6</td>
<td>2.8</td>
<td>3.2</td>
<td>11.3</td>
<td>12.3</td>
<td>10.8</td>
<td>0.2</td>
<td>1.2</td>
<td>9.8</td>
<td>2.6</td>
<td>-3.0</td>
<td>-0.7</td>
<td>11.2</td>
<td>-0.1</td>
<td>9.8</td>
<td>3.1</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>CG FAR</strong></td>
<td>9.9</td>
<td>9.9</td>
<td>-1.8</td>
<td>5.0</td>
<td>-1.4</td>
<td>0.1</td>
<td>5.7</td>
<td>3.3</td>
<td>5.1</td>
<td>7.1</td>
<td>5.8</td>
<td>9.9</td>
<td>-4.5</td>
<td>-9.7</td>
<td>-7.5</td>
<td>-11.0</td>
<td>-1.4</td>
<td>-13.9</td>
<td>-4.6</td>
<td>0.8</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>STHR FAR</strong></td>
<td>6.0</td>
<td>3.2</td>
<td>3.0</td>
<td>-1.9</td>
<td>-2.1</td>
<td>8.3</td>
<td>-3.3</td>
<td>7.7</td>
<td>9.9</td>
<td>7.4</td>
<td>-4.2</td>
<td>5.6</td>
<td>-1.0</td>
<td>-0.3</td>
<td>7.4</td>
<td>-2.3</td>
<td>-2.8</td>
<td>1.6</td>
<td>6.9</td>
<td>2.1</td>
<td>-3.0</td>
</tr>
</tbody>
</table>

**Figure 8.** The score anomaly of the FCN model in classification prediction results hour by hour, unit: % (a, b, c are the score results of cross-validation set, testing set, and independent validation set).

In the independent validation set, the performances in the first and second hours were higher than the average, which is slightly better than the test set in terms of hourly count. However, the overall classification performance was lower than the test set results, particularly with a noticeable drop after the sixth hour. For STTHR, the cross-validation set demonstrated a remarkable, above-average performance, particularly during the second,
third, fourth, sixth, seventh, tenth, and eleventh hours, where scores exceeded the average by more than 5.0%. The eleventh hour stood out as the best performance period, while the first hour ranked as the worst time period, with a deviation of 3.6% above the average. Within the test set, apart from the first hour, the remaining time intervals exhibited a tendency to score below the average, especially the seventh and twelfth hours, all of which were particularly lower than the average. However, the overall performance of classification for STHR was still slightly better than that for hail and CG. In the independent validation set, the scores of all time periods were lower than average, particularly the first, sixth, eighth, and tenth hours, which exhibited the most substantial deviations below the average.

From the POD results, hail in the cross-validation set was significantly higher than average, especially in the first, fourth, fifth, eighth, ninth, and tenth hours, which were more than 10.0% higher. Relative to the cross-validation set results, the score of the test set was noticeably lower, particularly in the first, fifth, sixth, and eleventh hours, which were around 10.0% below the score of the cross-validation set. In the independent validation set, the fifth, seventh, and twelfth hours scored more than 5.0% below average, while the third, fourth, eighth, and tenth hours exceeded average scores, with the eighth hour being the highest at 3.4%. CG in the cross-validation set had a different pattern from hail, scoring below average in the third and seventh hours. The tenth hour performed the best, but did not exceed 10.0%. In the test set, apart from the first, fourth, fifth, eleventh, and twelfth hours, the scores for the aforementioned time periods were below the average value. The remaining seven hours were above average. The highest score corresponded to the sixth hour, which was 5.0%. In the independent validation set, only the second hour was close to average, while the scores of all remaining hours were lower than the average value. In particular, scores corresponding to the fourth and seventh to twelfth hours were much lower than the average, with the twelfth hour being the lowest at −9.2%. Similar to CG, the score of STHR in the seventh hour in the cross-validation set was only slightly lower than that −0.9%, while the scores of all other hours were higher than the average. The maximum score corresponded to the eleventh hour. In the test set, only the third, seventh, eleventh, and twelfth hours scored below average, whereas the scores of all other times were above average. The independent validation set scored below average across all times, particularly during the second, sixth, eighth, tenth, and twelfth hours, which exhibited pronounced deviations below the average.

From the FAR results, it was found that for hail, most hours in the cross-validation set scored below the average, particularly at the first, fourth, fifth, eighth, ninth, and tenth hours, which were more than 10.0% below the average score. In the test set, the second, third, fourth, eighth, and tenth hours also scored below the average, while the remaining hours were above. In the independent validation set, only the third, fourth, and sixth hours were below average, while the fifth hour was notably 11.2% higher than the average. For CG, the third and twelfth hours were above average in the cross-validation set. In the test set, the first, fourth, and eleventh hours were particularly higher than the others, with the second hour being the lowest at −4.5%. In the independent validation set, apart from the third, fifth, and sixth hours, the remaining hours exhibited notable elevations in the performance, particularly the second and twelfth hours, which were significantly higher than other time periods.

As for STHR, all hours except the eighth, ninth, and twelfth were below the average in the cross-validation set, especially the fourth, sixth, and eleventh hours, all of which are more than 10.0% below. In the test set, the first, second, fourth, fifth, ninth, and tenth hours were lower than average, while the others were higher.

Within the independent validation set, it was revealed that scores corresponding to the fourth, fifth, seventh, and eleventh hours were lower than the average, while the scores of the remaining temporal intervals were much higher. The ninth hour, in particular, reached the maximum, with a value of 9.9%.\"
In summary, the hourly analysis revealed that the POD and FAR for hail were lower in the test set than in the independent validation set. Upon reviewing the hail weather processes, it was found that the main reason lay in the diminished frequency of hail events throughout 2023. However, it was observed that each event was accompanied by a multitude of occurrences. For CG and STHR, the independent validation set had a lower POD and a higher FAR value as compared to the test set. This disparity was primarily due to the recurrent CG events, characterized by a reduced number of incidents per event throughout 2023. Additionally, there was an overall decrease in the occurrences of STHR within the Gansu province, with fewer instances per event in comparison to an average year.

Considering the CSI, POD, and FAR indices together across the hourly cross-validation set, test set, and independent validation set, it is apparent that the FCN model had a superior performance in relation to hail occurrences during the fourth, eighth, and tenth hours. This was closely followed by the performance during the third and ninth hours, while the least favorable outcomes were observed during the fifth, eleventh, and twelfth hours. In comparison, CG had the best performance at the sixth hour, followed by the first, second, and fifth hours, and the worst at the eleventh and twelfth hours. STHR scored best in the second and fourth hours, followed by the third, fifth, sixth, tenth, and eleventh hours, with the worst performance in the eighth and twelfth hours.

This comprehensive hourly analysis suggests that the FCN model, despite exhibiting variations in its performance across diverse categories of severe weather and different time intervals, still holds considerable potential for accurately classifying severe convective weather events. Notably, it demonstrates a good performance during specific hours, which may offer valuable insights for targeted improvements in forecasting practices.

5. Conclusions and Discussions

This paper constructed an FCN model using ECMWF numerical model data, with the aim of evaluating its efficacy in predicting severe convective weather phenomena. The following main conclusions were drawn:

1. Within the 2017–2021 training set, the FCN model attained an overall misjudgment rate of 16.6% for severe convective weather. Specifically, the lowest misjudgment rate was for STHR at 21.8%, while the highest was for hail at 33.2%, and the misjudgment rate for non-severe convective weather was 15.7%. In terms of scoring, the FCN model’s average CSI for the three types of severe convective weather was 33.3%, with an average POD of 73.4% and an average FAR of 62.1%. Among these, STHR had the highest POD and CSI, accompanied by the lowest FAR values. On the other hand, CG and hail had similar CSI and FAR scores, but CG had a higher POD than hail.

2. The FCN model was tested using ground observation data of severe convective weather from 2022, indicating an overall misjudgment rate of 18.6% for the three types of severe convective weather as well as non-severe convective weather. The misjudgment rates for hail, CG, and STHR were 48.2%, 29.0%, and 27.2%, respectively. The misjudgment rate was 14.5% for non-severe convective weather. The scores obtained from the test set were lower than those in the training set, with an average CSI of 25.8%, an average POD of 65.2%, and an average FAR of 70.0%. Nevertheless, STHR was still the best in terms of forecast performance.

3. After putting the FCN model into operation, independent validation using ground observation data of severe convective weather from 2023 demonstrated an overall misjudgment rate of 18.3%. The misjudgment rates for hail, CG, and STHR were 46.7%, 34.0%, and 31.5%, respectively, with a rate of 13.3% for non-severe convective weather. The average CSI was 24.3%, the average POD remained at 62.6%, and the average FAR was 71.2%, with STHR continuing to have the best forecast performance. The performance of the independent validation set was slightly lower than that of the training period cross-validation set and the test set, indicating that, despite a slight
performance decline in operational implementation, the model still demonstrated a degree of accuracy and stability in classifying severe convective weather events.

(4) The hourly analysis illustrated the fluctuations in the CSI, POD, and FAR metrics across different time intervals. In a comprehensive assessment, it was found that the FCN model exhibited optimal performances in hail classification analysis during the fourth, eighth, and tenth hours. As for CG, its peak performance was observed at the sixth hour, whereas STHR demonstrated the best accurate performance during the second and fourth hours. Conversely, the least favorable performance was witnessed at the twelfth hour across all three categories of severe convective weather. Across the entire 2017–2023 sample, hail had a CSI of 17.6%, CG of 20.3%, and STHR of 45.5%. The best POD for STHR was 73.2%, followed by CG, with hail being the lowest at 57.3%. The FAR for hail and CG was comparable, with STHR having the lowest score, which was 45.5%. Thus, the FCN can be treated as a reliable short-term forecasting model which can provide an accurate classification forecast for severe convective weather events.

The FCN model constructed using ECMWF numerical model data and ground observation data can rapidly and automatically forecast severe convective weather with considerable accuracy, significantly enhancing the level of short-term forecasting for severe convective weather. Moreover, the FCN algorithm can integrate radar echo extrapolation products and satellite data to improve the prediction accuracy of severe convective weather forecasts within the 0–2 h time period, which will be one of the focal points in the future. However, the FCN algorithm constructed in this paper refrained from categorizing the intricate circulation patterns of various severe convective weathers, which will be another focus for improving the algorithm later.

Although there has been extensive research based on the classification and recognition of single types of severe convective weather, it is crucial to acknowledge that multiple types often occur simultaneously or in different regions. Consequently, the limitations of systems and technologies that can only discern a singular type of severe convective weather become evident, as they fail to adequately address this requirement. Therefore, through the utilization of the FCN model, this paper has successfully attained automated classification forecasting for the three main types of transient severe weather with commendable outcomes, which is a major innovation of this study with broad application prospects in future short-term severe weather forecasting. Another contribution of this paper lies in the recognition that, although a series of severe convective weather observation samples were observed and recorded in China’s central and western regions, these valuable data have not been fully utilized in conjunction with artificial intelligence technology in actual operational work. This paper can be treated as a benchmark for how to explore the application value of these data. Additionally, the advanced FCN algorithm is employed in this paper, thereby broadening the horizons of artificial intelligence’s application in meteorological operations. This work, in turn, could also bring a significant contribution to the automation of meteorological services.


Funding: Natural Science Foundation of Gansu (Grant 23JRRA1572), “Flying Clouds” youth top talent (Grant 2425rczx), Gansu Youth Science and Technology Foundation (Grant 23JRRA1332), Key project of meteorological research of Gansu Meteorological Bureau (Grant Zd20284-B-2), Drought meteorological science research Foundation project (Grant IAM202308).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
**Data Availability Statement:** Data are contained within the article.

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

**References**


17. Moller, A.R. *Severe Local Storms Forecasting*; *Doswell III C A. Severe Convective Storms*; American Meteorological Society: Boston, MA, USA, 2001; pp. 433–480. [CrossRef]

18. Huang, W.B.; Zhao, Y.; Sun, C.; Wang, H.M.; Wang, X. Climate modulation of summer rainstorm activity in eastern China based on the Tibetan Plateau spring heating. *Arab. J. Geosci.* 2020, 13, 126. [CrossRef]


27. Tian, Y.; Zhao, H.; Sun, S.W.; Loo, J.J.; Oh, S.G.; Wang, Y.J. A Deep-Learning Ensemble Method to Detect Atmospheric Rivers and Its Application to Projected Changes in Precipitation Regime. *Atmos. JGR* 2023, 13, e2022JD037041. [CrossRef]


Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.