Theoretical and Applied Climatology (2024) 155:5909–5923
https://doi.org/10.1007/s00704-024-04984-w

RESEARCH

PP-Loss: An imbalanced regression loss based on plotting position for improved precipitation nowcasting

Lei Xu[1] · Xuechun Li[1] · Hongchu Yu[2] · Wenying Du[1] · Zeqiang Chen[1] · Nengcheng Chen[1]

Received: 23 November 2023 / Accepted: 24 April 2024 / Published online: 2 May 2024

Abstract

Precipitation nowcasting refers to the accurate and timely forecasting of rainfall in a certain area for the near future, which plays a key role in severe convective weather warnings, water resources, and flood risk management. Rainfall data sample imbalance has always been an important issue that needs to be addressed in precipitation forecasting. It has been shown that an appropriate constraint can be introduced into the loss function of deep learning models to reduce the negative impact of imbalanced data on model training. In this work, we propose a sample weighting approach for imbalanced precipitation datasets called *Plotting Position Weight* (PP-weight), and a cost-sensitive learning approach based on weighting schemes, called *Plotting Position Loss* (PP-Loss). The developed PP-Loss is used to adjust the impact of each data point on the loss through the PP-weight calculated based on the ranking of the position of the data values. We explore the manipulation of imbalanced regression loss function by the convolutional long short-term memory (ConvLSTM) model on the ERA5-Land precipitation dataset. Experimental results suggest that the ConvLSTM model using PP-Loss is comparable to that using state-of-the-art loss functions (mean square error, Focal-R, Gumbel Loss, and label distribution smoothing), obtaining an R^2 of 0.886 and a root mean square error of 2.220 mm when using 3-h precipitation data as a learning sample to predict precipitation for the next hour. Therefore, the proposed approach may be expected to contribute to applying imbalanced regression methods in deep learning precipitation forecasting research.

1 Introduction

Precipitation is a vital variable in the regional hydrological cycle and is essential for water resource management and climate change monitoring (Xu et al. 2022; Fang et al. 2023). However, global warming causes drastic changes in precipitation, triggering droughts and floods that pose a serious threat to life safety and sustainable economic development (Xu et al. 2020a). The World Meteorological Organization (Kann 2018) refers to forecasts for the next 0–2 h as nowcasting. Precipitation nowcasting works by providing very short-range localized rainfall forecasts using radar echo, rain gauges, and numerical weather prediction models (Prudden et al. 2020). Precipitation nowcasting is of significant importance in early warning of severe convective weather, heavy rainfall warning, flood prevention and mitigation, and water resource management (Chen et al. 2020).

Existing mainstream precipitation prediction methods can be broadly categorized into two groups. One is numerical weather prediction (NWP) (Al-Yahyai et al. 2010), which is based on dynamic weather processes. NWP is based on the numerical solution of partial differential equations constrained by the laws of dynamics and thermodynamics and requires a large amount of computational power (Bauer et al. 2015). Moreover, NWP methods are inherently limited because the intrinsic mechanisms of atmospheric behavior are not yet fully understood. Another category of methods is data-driven methods, and artificial intelligence methods for precipitation prediction have received much attention in recent years. These methods learn feature representations of large-scale historical data to achieve accurate forecasts. Specifically, they utilize deep learning techniques to infer the spatial and temporal correlations and mechanisms of change in radar echoes or precipitation images to predict future radar echoes or precipitation distributions (Chen et al. 2021). This approach views precipitation nowcasting as a

✉ Wenying Du
duwenying@cug.edu.cn

[1] National Engineering Research Center for Geographic Information System, China University of Geosciences (Wuhan), Wuhan 430074, China

[2] School of Navigation, Wuhan University of Technology, Wuhan 430063, China

spatio-temporal sequence prediction problem by predicting the next consecutive frame sequence after the previous observation sequence(Shi et al. 2015).When it comes to deep learning, compared to the traditional recurrent neural network (RNN), long short-term memory (LSTM) (Graves and Graves 2012; Greff et al. 2016) introduces memory units and gating mechanisms to better capture long-term dependencies in time series. ConvLSTM (Shi et al. 2015) possesses both the temporal relations of LSTM and the spatial feature extraction capability of convolutional neural networks (CNN) by extending fully connected LSTM(FC-LSTM). To capture spatio-temporal features more effectively, the trajectory gated recurrent unit (TrajGRU) (Shi et al. 2017) model actively studies the trajectories of cyclically connected location-varying structures. predictive recurrent neural network (PredRNN) (Wang et al. 2017), with spatio-temporal LSTM (ST-LSTM) at its core, can extract and store temporal and spatial information at the same time. The spatial temporal long short-term Memory based on the self-attentive mechanism (ST-LSTM-SA) (Liu et al. 2022) model combines 3D convolution and self-attention modules for better temporal characterization. Weather Fusion Net (Kaparakis and Mehrkanoon 2023) is a weather fusion prediction model inspired by U-Net, which improves the performance for accurate prediction of precipitation during the next three hours by integrating variables such as wind speed and precipitation.

The imbalance of precipitation data samples has remained an important problem to be addressed in precipitation forecasting. Previous studies have shown that appropriate constraints can be applied to the loss function for deep learning models to minimize the negative impact of imbalanced data on model training (You et al. 2023). Data imbalance problems, also known as long-tail identification (Ali et al. 2019). A vast majority of previous work has been concentrated on the imbalanced classification problem. Proposals can be categorized into data-based solutions and model-based solutions, mainly for binary classification problems (Johnson and Khoshgoftaar 2019). Data-based solutions involve resampling the data, which mainly includes oversampling (Chawla et al. 2002; He et al. 2008) of the minority classes, undersampling (Babar and Ade 2015) of the majority classes, and mixed sampling. For example, a classical SMOTE (Chawla et al. 2002) algorithm generates synthetic samples of the minority class by linearly interpolating the samples in the same class. Model-based solutions, also called cost-sensitive learning methods (Abd Elrahman and Abraham 2013), aim to adjust the loss of different categories during training using a loss function. In addition to the methods mentioned, alternative learning paradigms can be leveraged to compensate for class imbalances (Yang et al. 2021). Representation learning focuses on learning a smaller representation bias (Cheng et al. 2020). Metric learning (or contrast learning) looks to better model the boundaries near fewer classes (Gautheron et al. 2019). Two-stage training (or decoupling) is done by normal sampling in the feature learning stage and balanced sampling in the classifier learning stage (Kang et al. 2019).

Current techniques for dealing with imbalanced data focus on targets with categorical indexes, i.e., dealing with imbalances between different classes. In many tasks, however, continuous objectives are involved, where there are no clear boundaries between classes. This situation is referred to as a regression problem, in which the objective values can be infinite. Due to the characteristics of regression problems, it may not be possible to directly apply former imbalanced classification methods to imbalanced regression problems. Compared to imbalanced classification, regression over imbalanced data is not as well explored (Gong et al. 2022). One of the early works applied algorithms for classification directly to regression. SMOTER (Torgo et al. 2013) modifies SMOTE (Chawla et al. 2002) to make it suitable for regression tasks where the goal is to predict rare extremes of the target variable. SMOGN (Branco et al. 2017) appends Gaussian noise enhancement to it. These works all attempted to resample the training set by synthesizing new samples for minority targets. Another is to adopt several approaches to imbalanced classification for regression scenarios. Inspired by Focal loss (Lin et al. 2017), one regression variant, Focal- R (Yang et al. 2021), reweights the loss values by replacing the scale factor with a continuous function. Another regression variant, Focal L1 Loss (Zhang et al. 2022), reduces the bounding box regression error by redesigning the loss function. In imbalanced regression, continuity in the label space can provide extra information about the relationship of data instances. To create meaningful connections between different labels, current state-of-the-art approaches recommend smoothing the target distribution using kernel density estimation. Label Distribution Smoothing (LDS) (Yang et al. 2021) and DenseLoss (Steininger et al. 2021) follow an analogous concept. They estimated the empirical label density by applying a Gaussian kernel function to obtain an effective label density distribution that accounts for label continuity. However, the disadvantage of kernel density estimation is related to bandwidth. Large bandwidths may lead to over-smoothness, masking the true structure of the data; small bandwidths produce spiky, multi-peaked density estimates (Chen 2017). Solutions for imbalanced regression tasks are relatively scarce and far less explored for cost-sensitive learning. Cost-sensitive learning has known strengths over sampling-based approaches in classification tasks (Krawczyk 2016; Cui et al. 2019).

In this paper, we explore the effectiveness of imbalanced regression methods for precipitation prediction and propose a cost-sensitive learning method for neural network imbalanced regression called Plotting Position Loss (PP-Loss). The key contributions of this work are listed as follows: (i) Plotting Position Weight (PP-Weight), a sample weighting

approach for regression on imbalanced data, is proposed. This study innovatively applies the plotting position formula (a form of empirical probability distribution function) to the computation of the density function of the objective value, which forms the basis of PP-weight. (ii) PP-Loss is proposed as a cost-sensitive method for neural network regression modeling of imbalance data by combining PP-weight with sample weighting of the loss function. PP-Loss assigns a weight to each data point in the training set based on PP-Weight, increasing the impact of rare values on the loss and gradient. (iii) This study applies PP-Loss on the ERA5-Land precipitation dataset and competes with different imbalanced regression loss functions, showing that it can improve model performance in practice.

2 Data

The precipitation data used in this work were downloaded from the Copernicus Climate Change Service Climate Data Store (CDS) of the ECMWF. ERA5-Land is the land component of the fifth generation of European Reanalysis (ERA5) producing an enhanced global dataset (Muñoz-Sabater et al. 2021), ERA5-Land contains elevation corrections for the thermodynamic near-surface state, which can provide more precise and higher spatial and temporal resolution data for relevant studies (Wu et al. 2021).

Experimental data were collected on total precipitation from January 1, 2011, to January 1, 2021 (10 years) with a temporal resolution of one hour. The study area consists of 20×20 grid points covering the city of Wuhan between 29.5–31.4°N and 113.4–115.3°E with a spatial resolution of 0.1°. Geomorphologically, Wuhan city is located in the center of China, surrounded by mountains in the north and south, whereas the central part of the city has a lower topography (Wu et al. 2021). Wuhan city is the only megacity in six provinces in central China, with the study area map shown in Fig. 1.

To facilitate computational convergence during the training process, the eigenvalues of the samples are normalized by max–min normalization so that they vary in the range of [0, 1]. Within these data, the last two years between 2019 and 2021 were treated as a test set. The experiments were carried out for six years from 2011 to 2017 for training and two years from

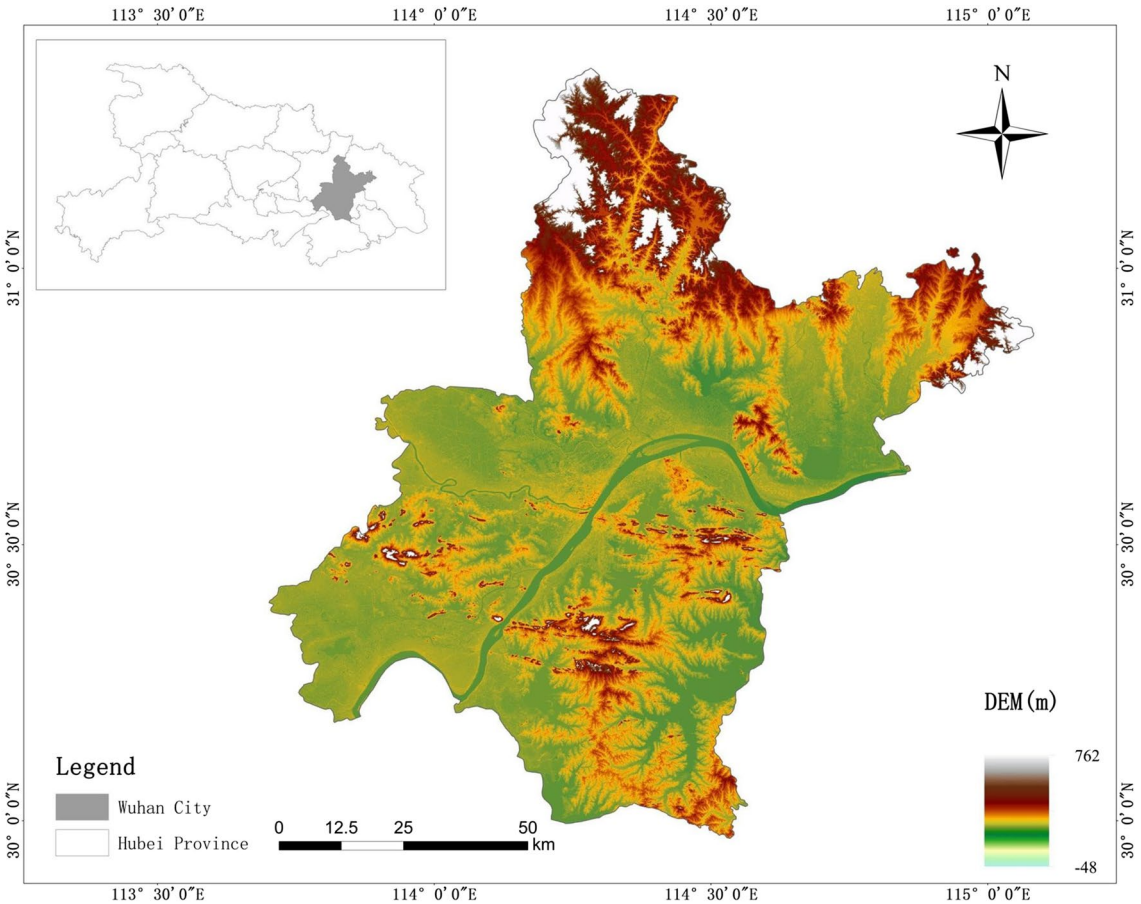

Fig. 1 The Digital Elevation Model (DEM) map of the study area

2017 to 2019 for validation to determine the optimal model architecture (Scheepens et al. 2023).

3 Methodology

3.1 The ConvLSTM model

ConvLSTM is a variant of LSTM introduced (Shi et al. 2015) to learn the spatial information in spatially sequential datasets. In the domain of spatio-temporal series prediction, complex spatio-temporal correlations are commonly existing within the data (Shi and Yeung 2018; Wang et al. 2020). The major drawback of FC-LSTM in dealing with spatio-temporal data is its lack of encoding spatial information. ConvLSTM replaces the full concatenations of input-to-state and state-to-state transitions with convolutional operations, which is the main difference between ConvLSTM and FC-LSTM. ConvLSTM networks can better capture spatio-temporal correlations and model spatio-temporal relationships. The structure of the ConvLSTM used in this study is shown in Fig. 2.

The equations for ConvLSTM are as follows.

$$i_t = \sigma\left(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i\right) \quad (1)$$

$$f_t = \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f) \quad (2)$$

$$C_t = f_t \circ \mathcal{C}_{t-1} + i_t \circ \tanh\left(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c\right) \quad (3)$$

$$O_t = \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \circ \mathcal{C}_t + b_o) \quad (4)$$

$$H_t = o_t \circ \tanh\left(\mathcal{C}_t\right) \quad (5)$$

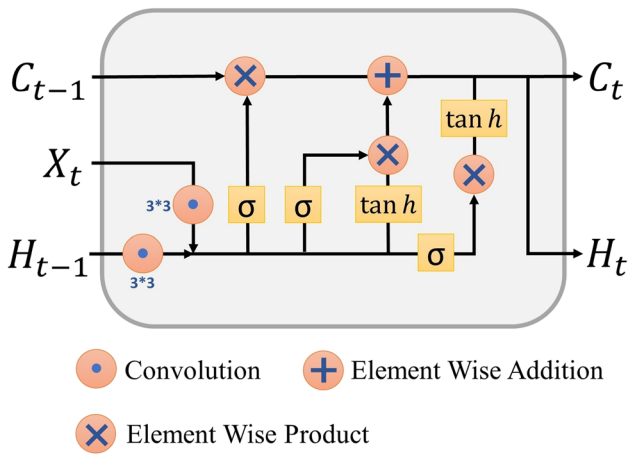

Fig. 2 The ConvLSTM structural cell has three inputs: the current input feature $\mathcal{X}_t$, the hidden output $\mathcal{H}_{t-1}$, and the cell state $\mathcal{C}_{t-1}$ of the previous time step

where σ stands for the sigmoid function, $*$ stands for the convolution operation, $\circ$ stands for the element-wise operation, tanh is the hyperbolic tangent function; t stands for the moment t; f, i, and O stand for the forgetting gate, the input gate, and the output gate; C and $\mathcal{H}$ stand for the memorized cell state and the hidden state, respectively; and W is the weight matrix and b is the bias value.

Our work employs a deep ConvLSTM model architecture for spatio-temporal sequence prediction, using an encoding-forecasting network structure that includes several stacked ConvLSTM layers. The model is implemented and trained using Pytorch (Paszke et al. 2019), which investigates the performance of different loss functions on ConvLSTM. The hyperparameters of the ConvLSTM model were adjusted according to the loss functions of the validation set. The flow architecture of the ConvLSTM model is shown in Fig. 3.

The model is implemented as a multi-frame prediction model with inputs of true precipitation data for three consecutive hours and outputs of forecasted precipitation data for the first hour of the future and the third hour of the future (Shi et al. 2017; Scheepens et al. 2023). As an input, the model incorporates a tensor of size $3\times20\times20$, consisting of precipitation for the previous 3 h on a 20×20 grid. Then the various hidden states are simultaneously encoded by the encoding network and decoded into a forecast tensor of the same size for the following 1 h by the decoding network.

For this paper, we focus on comparing the proposed PP-Loss with different loss functions to improve the prediction of precipitation by applying the imbalanced regression methods to the model. To evaluate the accuracy of the model, three metrics were chosen to test the model's prediction results: RMSE, R^2, and Correlation. R^2 measures the correlation between the predicted values and the true values, RMSE measures the variance, and Correlation measures the linear correlation. The correlation formula calculated for the evaluation metrics is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \quad (7)$$

$$Correlation = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(\hat{y}_i - \bar{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \quad (8)$$

where n is the number of pixels of the predicted value, y_i is the true value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the average of the true value, and $\bar{\hat{y}}$ is the average of the predicted value.

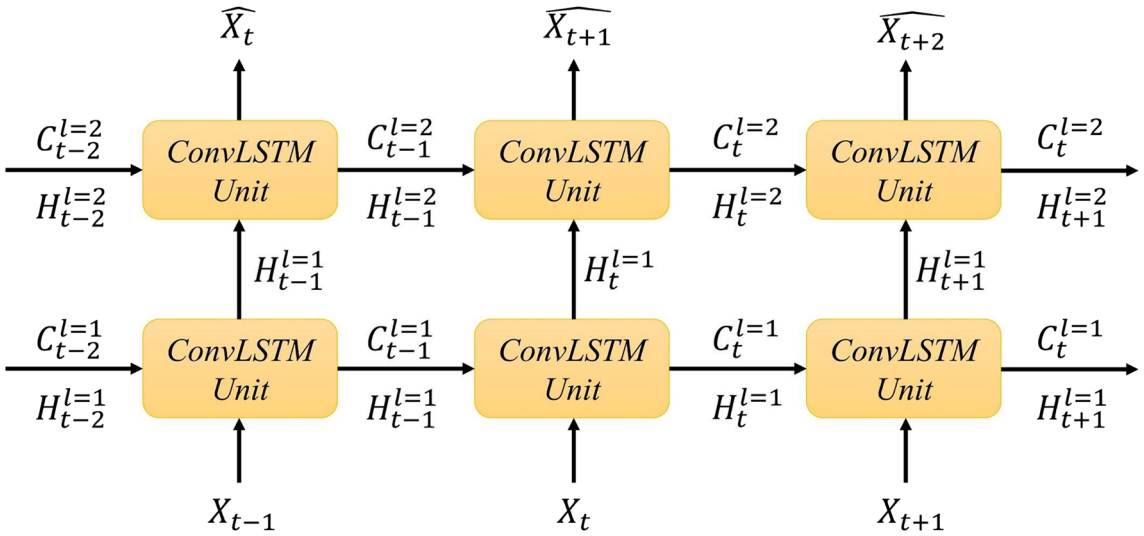

Fig. 3 Schematic diagram of ConvLSTM memory flow architecture. In this study, we used two ConvLSTM layers (each with 32 hidden states) and 3×3 kernels for stacking. During training, the model was trained using an Adam optimizer with a learning rate of $5*10^{-4}$. The number of training rounds for the model was set to 150 with a batch size of 256

3.2 Imbalanced regression method

3.2.1 Focal-R

Focal Loss is a widely used loss function that addresses category imbalance problems and a dynamically scaled cross-entropy loss based on binary cross-entropy (Lin et al. 2017). The introduction of Focal loss is primarily intended to resolve the problem of extreme imbalance between the number of positive and negative samples in one-stage target detection. With a dynamic scaling factor, one can dynamically reduce the weight of examples that are easy to classify correctly during training to focus rapidly on those examples that are hard to classify correctly. Ultimately, it achieves the effect of suppressing the imbalance between the number of easy to classify correctly and hard to classify correctly examples.

Inspired by the classification loss function Focal Loss originally used for classification error measurement, (Zhang et al. 2022) proposed a regression version of the focus loss called Focal L1 Loss. This loss function is intended to enhance the performance of bounding box regression (BBR) models for target detection, especially for optimizing the contribution of high-quality anchor frames with large Intersection over Unions (IOUs). (Yang et al. 2021) proposed a version of the regression called Focal-R. The Focal Loss used in this paper takes the regression form of the Focal Loss proposed by Yang, where the scale factor is replaced by a continuous function that maps the absolute error into [0, 1]. The Focal-R loss based on the L2 distance can be written in the following form:

$$\frac{1}{n}\sum_{i=1}^{n}\sigma(|\,\beta e_i\,|\,)^{\gamma}e_i \tag{9}$$

where e_i is the L2 error of the i-th sample, σ is the Sigmoid function, and β, γ are hyper-parameters. L1 Loss also known as Mean absolute error (MAE) and is defined as $MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$. L2 Loss is also known as Mean squared error (MSE) and is defined as $MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$. Here, N is the number of data samples, y_i is the true value, and $\hat{y}_i$ is the predicted value.

3.2.2 Gumbel loss

In the context of loss functions, (Ding et al. 2019) introduced Extreme Value Loss (EVL) to predict the future occurrence of extreme events. However, this loss function is based on a binary classification layer, making it difficult to implement for regression problems. To circumvent this, (Zhang et al. 2021) proposed a generalized EVL framework inspired by the kernel density estimator (KDE) and introduced the Gumbel Loss function (Xie and Mallick 2023). Gumbel loss is defined as the following equation:

$$\frac{1}{n}\sum_{i=1}^{n}(1-e^{-\delta^2})^{\gamma}\delta^2 \tag{10}$$

where $\delta = y_i - \hat{y}_i$, y_i is the true value, and $\hat{y}_i$ is the predicted value. γ is a hyperparameter. Notice that with $\gamma > 1$, it discounts smaller errors more aggressively than the MSE and is therefore more sensitive to outliers.

3.2.3 Label distribution smoothing

Label Distribution Smoothing (LDS) (Yang et al. 2021) is an approach to estimating the effective label density distribution in the case of continuous labeling. It draws from the idea of kernel density estimation in the field of statistical learning. Specifically, given a continuous empirical label density distribution, LDS employs a symmetric kernel function as the smoothing function. By convolving the empirical density distribution with the kernel function, a version of kernel smoothing called effective label density is obtained. Such a version can intuitively reflect the problem of information overlap between neighboring labeled samples. A symmetric kernel is any kernel that satisfies: $k\left(y, y^{'}\right) = k(y^{'}, y)$ and $\nabla_y \mathrm{k}(y, y^{'}) + \nabla_{y'} \mathrm{k}(y^{'}, y) = 0, \forall y, y' \in \mathcal{Y}$. Note that a Gaussian or a Laplacian kernel is a symmetric kernel, while $\mathrm{k}(y, y') = yy'$ is not. The symmetric kernel characterizes the similarity between target values y' and any y w.r.t. their distance in the target space. LDS calculates the effective label density distribution with the following equation:

$$\widetilde{p}(y') \triangleq \int_{y} \mathrm{k}(y, y') p(y) dy \tag{11}$$

where $p(y)$ is the number of times the label of y appears in the training data, $\mathrm{k}(y, y')$ is the symmetric kernel, and $\widetilde{p}(y')$ is the effective density of the label y'. Next the estimates of effective label density were weighted according to the available estimates using a cost-sensitive reweighting method. This is achieved by multiplying the loss function by the inverse of the LDS estimated label density for each target value.

3.3 The proposed plotting position loss

A general practice for cost-sensitive learning methods for imbalanced regression problems starts out weighting individual data points in terms of the rarity of the target value (Cui et al. 2019). We would like to compute a weight for each sample that is inversely proportional to the probability of occurrence of the target value. It is necessary to devise a weighting function with specific attributes that shall fulfill the following conditions:

1. The weights should be smaller if the sample is a common target value and larger if it is a rare target value.
2. All data points should have non-negative weights to ensure that no portion of the data is ignored by the model.
3. The weights should not be zero since the model is inclined to maximize the difference between the predicted and original values of any of these data.
4. The average weight of all data points should be 1 to simplify the gradient descent optimization process.

Weights that meet the aforementioned conditions may be applied to any machine learning or deep learning model that supports the weighting of samples (Steininger et al. 2021). Models would be better suited for the estimation of rare cases. In this work, we make use of the weighting function to realize the cost-sensitive imbalanced regression method.

Typically, hydrological data can be analyzed by using statistical methods based on frequency analysis such as statistical distributions (Samat and Othman 2023). Plotting Position is a common technique to convert data values to percentile ranks for presenting data in frequency distribution tables or probability plots (Cunnane 1978). Plotting position can provide an estimate of exceedance probability for observed events. We can use them to develop empirical frequency curves or provide a visual comparison with a computed analytical frequency curve, such as a flow frequency curve. The traditional plotting position rules $\frac{i}{n}$ or $\frac{i-1}{n}$ are problematic in terms of treating extreme values and symmetry (Leon Harter 1984). Normally the following calculation methods are recommended (Subramanya 2008).

Weibull formula as equation:

$$P = \frac{m}{(N+1)} \tag{12}$$

Hazen formula as equation:

$$P = \frac{m - 0.5}{N} \tag{13}$$

Gringorten formula as equation:

$$P = \frac{m - 0.44}{(N + 0.12)} \tag{14}$$

where m is the rank of the data and N is the number of samples.

Depending on the plotting position formula mentioned above, we suggest a sample weighting approach for imbalanced regression datasets named PP-Weight. Moreover, based on it, we propose a cost-sensitive learning method for the imbalanced regression of neural networks named PP-Loss. PP-Loss incorporates the cost-sensitive learning method through the ranking function along with the plotting position formula. PP-Loss allocates weight to each data point in the training set based on PP-Weight, which gives more weight to rare data values than to common data values. In this way, PP-Loss increases the impact of rare data points on loss and gradient during model training and overcomes the problem of data imbalance in a better way.

With reference to LDS, PP-Loss divides the target space into finite intervals through smoothing. Initially, we smoothed and sorted the precipitation data. In other words, when the data value is 0, the zero value is taken; when the data value is a data value greater than 0, it is rounded upwards. Let r_k denotes the ranking function, a denotes the precipitation dataset, then $r_k(a)$ is the sequence

of permutations containing the ordering of each element in a, n denotes the number of elements in a, and s_i denotes the number of elements in the interval where the elements of a_i are contained. The i-th element in $r_k(a)$ is given by:

$$r_k(a_i) = 1 + \left| \left\{ \sum s_j : s_j < s_i \right\} \right| \tag{15}$$

Now it is necessary to determine the weight function f_w, which allows weighting the data points according to the rarity of the target value. During model training, there is no significant difference in the final accuracy of PP-Loss using different plotting position formulas. The Weibull formula is chosen to determine the density function PP of the target data:

$$PP(a_i) = \frac{r_k(a_i)}{n+1} \tag{16}$$

The normalization leads to the PP-Weight weight function f_w:

$$f_w(a_i) = \frac{PP(a_i)}{\frac{1}{n}\sum_{i=1}^{n} PP(a_i)} \tag{17}$$

Neural networks are typically optimized using a gradient descent optimization algorithm. This algorithm aims to obtain a minimization metric M that can be incorporated into a loss function L to which sample weighting is then applied. Given a model estimate $\widehat{Y} = \{\widehat{a_1}, \widehat{a_2}, \ldots, \widehat{a_n}\}$, the formula for PP-Loss is obtained as follows:

$$L_{PP-Loss} = \frac{1}{n}\sum_{i=1}^{n} f_w(a_i) \bullet M(\widehat{a_i}, a_i) \tag{18}$$

Compared to models trained with balanced sample weights, models trained with PP-Loss take into account the imbalance of samples, allowing the model to better deal with rare samples, improving the prediction capacity and robustness of the model. For PP-Loss, Algorithm 1 describes it more formally (Usharani 2023).

Algorithm 1 Plotting position loss algorithm

INPUT: ERA5-Land precipitation data

OUTPUT: Reduced loss value of the output f(x) data

1. Training set X.
2. Data distribution smoothing and sorting.
3. Initialize weights w. Randomly assigns weights and bias.
4. Repeat
   a) Calculate new I = (I_1, I_2, ... I_n).
   b) Apply plotting position method to each value. Return weight value corresponding to each position.
   c) Minimize weights w.
   d) Observe true values y and calculate the predict values $\hat{y}$.
5. Sum the weights using the function f to compute the loss with specified hyperparameters.
6. Update weights and bias using the values in the dataset. Go to step 4.
7. Repeat until the error converges.

4 Result

4.1 The forecasting skill of the developed model

In this paper, we have constructed a forecasting model by employing the proposed cost-sensitive learning method PP-Loss based on a weighting scheme for imbalanced regression using the ConvLSTM model. Experiments are conducted on the ERA5-Land precipitation dataset, using the 3-h precipitation data as the learning sample to predict the next 1-h precipitation data up to the 3-h precipitation data. From the evaluation indexes of the model, it is evident that PP-Loss is more efficient at forecasting the precipitation within the next 1 ~ 3 h in the target area. Among them, the R^2 of the model reaches 0.886, the Correlation reaches 0.942, and the RMSE is 2.220 mm, all outperforming the state-of-the-art methods. PP-Loss we proposed shows higher R^2 and Correlation, lower RMSE, and a more concentrated distribution of results. Likewise, it can be seen that the forecasting skill of the method decreases with the increase in the forecasting duration. Specific results can be found in Table 1.

To verify the applicability of PP-Loss on other precipitation datasets, GPM data was selected for this study for supplementary experiments. The data come from NASA's Global Precipitation Measurement (GPM) website's Level 3 fused precipitation product integrated multi-satellite retrievals for GPM (IMERG), which is a multi-satellite, multi-sensor, and multi-algorithmic integrated inversion product for GPM. Since most evaluations show that Final Run products usually outperform Early Run, and Late Run products and better capture the temporal and spatial distribution of precipitation (Zhang et al. 2023). We selected GPM satellite data retrieved from version 6 of the IMERG Final product over Wuhan (IMERG V06 Final) between 2011 and 2020. This product provides a relatively fine spatial resolution of $0.1° \times 0.1°$ and a high temporal resolution of half an hour. In order to maintain the same temporal resolution as ERA-5 Land's one-hour resolution, we processed the GPM data to a temporal resolution of one hour. Experiments are conducted on the GPM precipitation dataset to predict the future 1 h to 3 h precipitation data using the 3 h precipitation data as a learning sample. It can be seen that our proposed PP-Loss has an R^2 of 0.681 in predicting the future one-hour precipitation, which obtains the highest R^2 and correlation and the lowest RMSE compared to other loss functions. The specific experimental results are shown in Table 2.

The weaker results of the GPM data compared to the ERA-5 dataset may be due to poorer data quality. Satellite remote sensing technology has some limitations in capturing precipitation information, and malfunctioning or inaccurate calibration of satellite sensors may affect the quality of the data. Also, as the precipitation intensity increases or the time scale decreases, the accuracy of the GPM precipitation products in capturing precipitation events decreases subsequently. The ERA-5 Land precipitation dataset is a reanalyzed data that has undergone a rigorous validation and verification process by the ECMWF. It employs a sophisticated data assimilation methodology that combines multiple observational sources and model outputs. Despite the overall overestimation, the ERA-5 Land data is reliable and widely accepted, and used in research and applications (Xin et al. 2022). Therefore, we choose ERA-5 Land data as the main dataset in our experiments, and the following analytical discussions are all based on ERA-5 Land data.

Table 1 Comparison of the overall accuracy of each model on the ERA5-Land dataset

Loss Function	ConvLSTM(1 h)			ConvLSTM(3 h)		
	R^2	RMSE(mm)	Correlation	R^2	RMSE(mm)	Correlation
MSE Loss	0.846	2.582	0.921	0.648	3.732	0.823
Focal-R	0.863	2.430	0.929	0.686	3.671	0.840
Gumbel Loss	0.804	2.909	0.930	0.542	4.430	0.820
LDS	0.868	2.383	0.933	0.689	3.655	0.830
PP-Loss	**0.886**	**2.219**	**0.942**	**0.708**	**3.539**	**0.845**

Table 2 Comparison of the overall accuracy of each model on the GPM dataset

Loss Function	ConvLSTM(1 h)			ConvLSTM(3 h)		
	R^2	RMSE(mm)	Correlation	R^2	RMSE(mm)	Correlation
MSE Loss	0.615	1.465	0.785	0.237	2.056	0.496
Focal-R	0.613	1.468	0.784	0.246	2.045	0.502
Gumbel Loss	0.580	1.531	0.782	0.002	2.352	0.438
LDS	0.624	1.448	0.790	0.254	2.034	0.517
PP-Loss	**0.681**	**1.334**	**0.826**	**0.290**	**1.984**	**0.539**

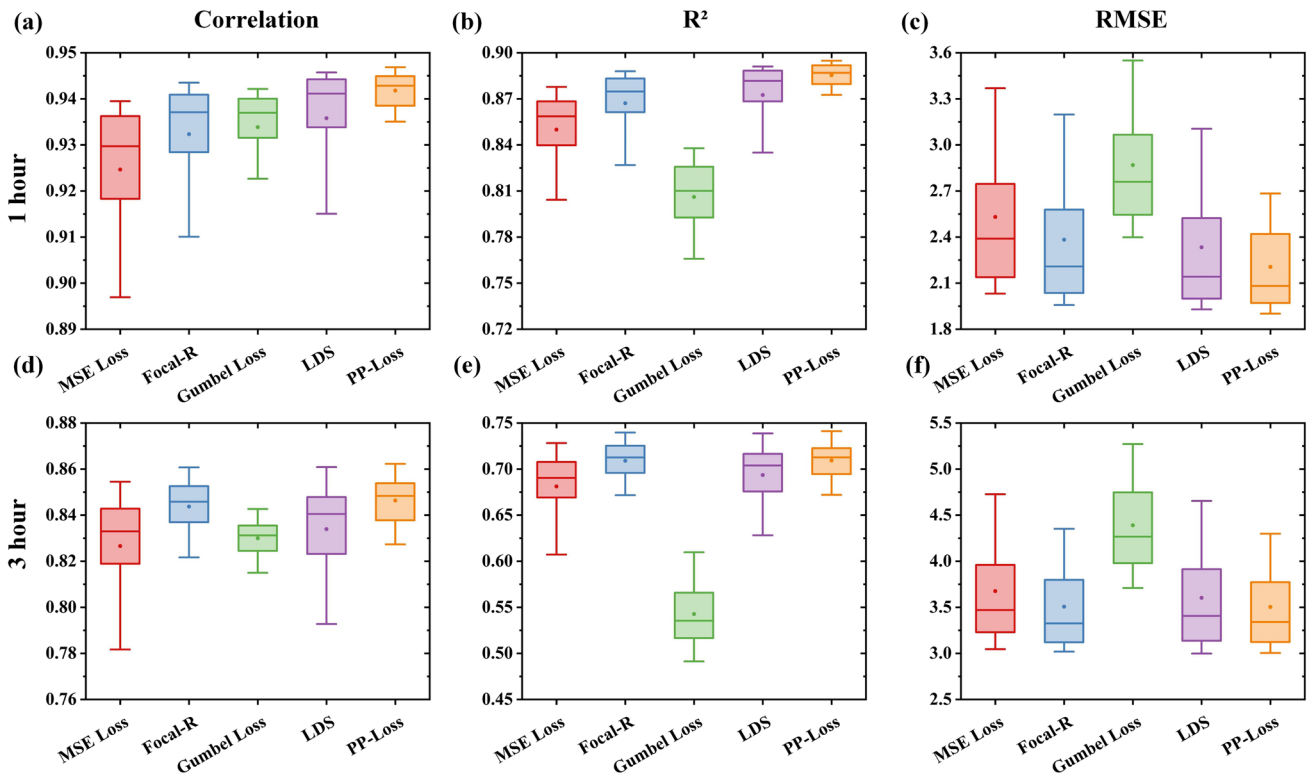

Fig. 4 Box plots of the prediction accuracies for each model. Confidence intervals are single-point accuracy values from 10 to 90%, the box plots range from upper to lower quartiles, the horizontal line in the center of the box plot indicates the median, and the circle indicates the mean

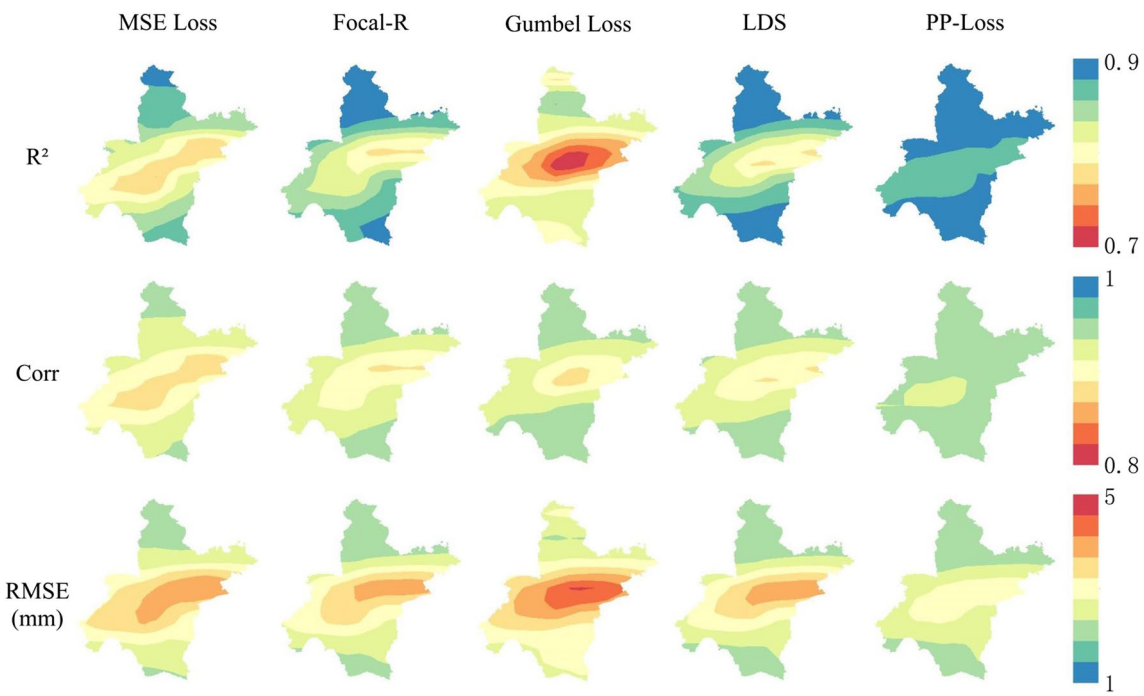

Fig. 5 The ConvLSTM model employs different loss functions to predict the spatial distribution of experimental results for the next 1-h precipitation data. Comparing the prediction results of alternative methods, it can be observed that the PP-Loss forecasting model achieves comparatively excellent performance in the south and north of Wuhan City

Regarding the ConvLSTM model, Mean Square Error Loss (MSE Loss) is adopted as a baseline to compare the PP-Loss proposed in this paper with the current imbalanced regression loss functions Focal-R, Gumbel Loss, and LDS. We may observe that the spatial patterns of the forecasting skills across the imbalanced loss functions are comparable to each other. However, variations in performance may exist in localized areas and prediction skills, which indicates the reliability of the PP-Loss proposed in this paper. In Fig. 4, it is possible to observe that the model using PP-Loss to predict the precipitation data for the next hour has marked strengths in terms of median, median and maximum and minimum values of the evaluated metrics, all of which are superior to the current state-of-the-art methods. As can be observed in Fig. 5, the model using PP-Loss obtains more accurate results in a wider range of regions. We also find that the broadly used MSE loss function may not be efficient in imbalanced regressions.

4.2 Spatial accuracy evaluation of the developed model

As seen from Fig. 5, spatially, the higher predictability is mainly distributed in the southern and northern parts of Wuhan. Different lead times show comparable forecasting ability and the forecasting accuracy tends to decrease with increasing lead time. The model works out better in predicting the following 1-h precipitation data. According to the spatial distribution characteristics of precipitation, each approach can be seen to attain a relatively favorable prediction result. However, the PP-Loss we propose obtains more accurate precipitation values over a more precise precipitation range. To demonstrate this, we conducted experiments on precipitation forecasting models using ConvLSTM models trained with different loss functions. Such loss functions include MSE Loss, Focal-R, Gumbel Loss, LDS, and the proposed PP-Loss. For the next step, we randomly selected three precipitation cases for comparison. Figure 6 shows the predicted values obtained using different methodologies in comparison with the corresponding measured precipitation.

In Fig. 6, it is seen that both forecasts and observations characterize the spatial distribution of precipitation. However, there exists a certain degree of estimation error in the forecast of precipitation. To further explore the spatial distribution of the error, the difference between the precipitation predictions and observations is calculated, as shown in Fig. 7 Regarding the forecasts of precipitation, the errors of PP-Loss primarily occur in the southern part of Wuhan City. Negative errors indicate a certain underestimation of precipitation, but PP-Loss has the smallest overall error. Gumbel Loss shows the largest overall error, and precipitation is overestimated. The spatial overestimation of MSE Loss, Focal-R and LDS mainly occurs in the western and northern part of Wuhan, and the prediction deviations are large. For the precipitation prediction in the precipitation event center, all other loss functions except PP-Loss can be seen to exhibit significant overestimation.

Let's take the most frequently used MSE Loss in deep learning as an example. MSE values are represented as

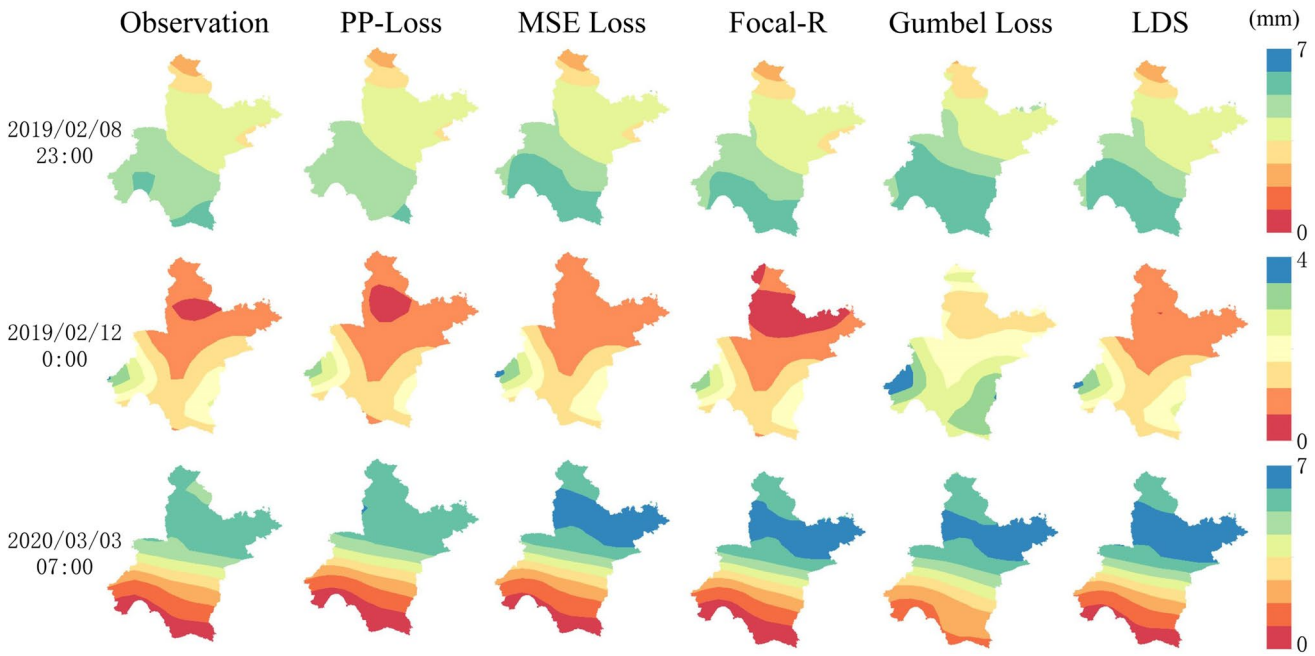

Fig. 6 Comparison of precipitation predictions and observations at different times

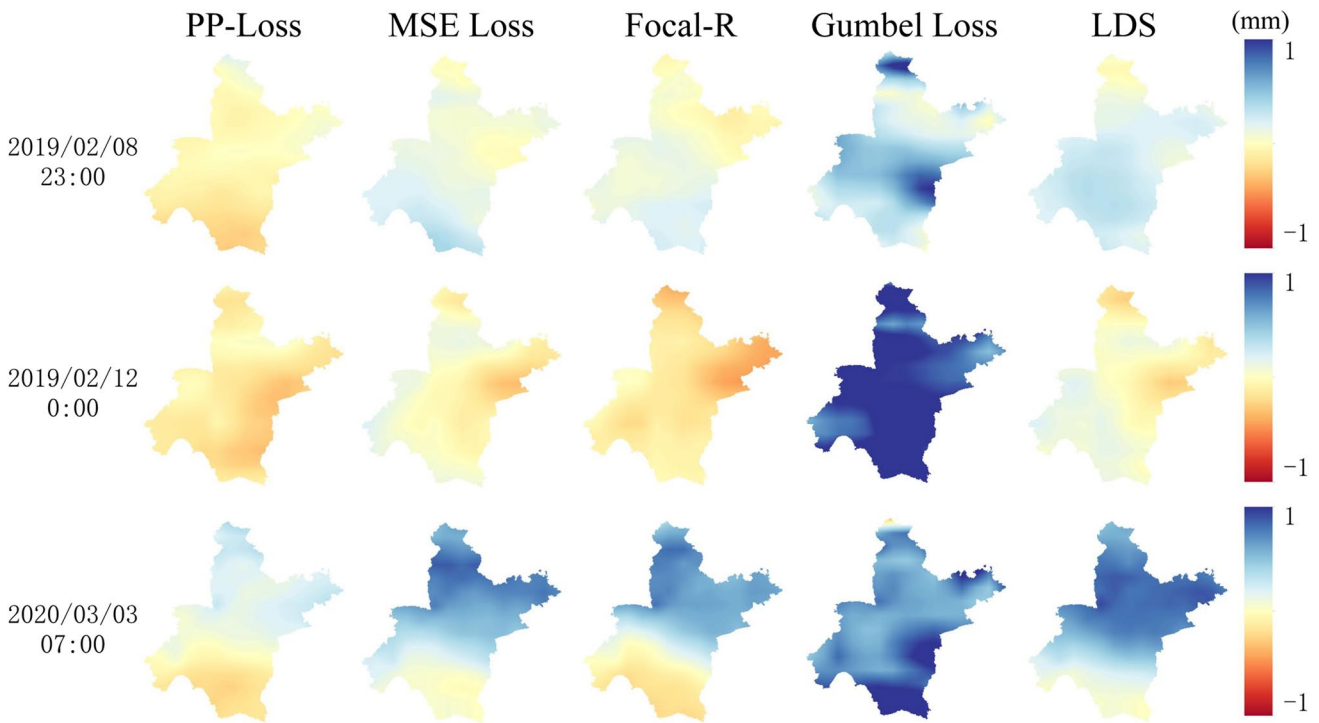

Fig. 7 Differences between precipitation predictions and observations at different times

quadratic equations, which help to penalize the model in case of outliers (Jadon et al. 2022). Focal Loss together with its variants is susceptible to noisy data and thus may mislead the direction of model learning. Hyperparameter combinations in the formulas require constant experimentation to reach their optimization, which can be very time-consuming. Attention paid to hard examples may cause the model to ignore other samples that have been properly categorized, leading to overfitting (Lin et al. 2017; Hammad et al. 2021). Gumbel Loss takes into account the prediction of extreme events and enhances the ability of a representative time series predictor to model extreme events. The poor performance of the results may be due to the fact that Gumbel Loss is not applicable to spatio-temporal forecasting tasks. LDS (Yang et al. 2021) and DenseLoss (Steininger et al. 2021) apply a Gaussian kernel to the empirical label density for estimating an effective label density distribution considering label continuity. However, they both only consider nearby label values but not the global relationship between labels and distances in the feature space (Gong et al. 2022). Moreover, bandwidth is correlated with kernel density estimation, which can be too large or too small to affect the estimation results.

We innovatively apply the plotting position formula (a form of empirical probability distribution function) to the design of the loss function. PP-Loss adjusts the impact of each data point on the loss calculation by weighting the data points according to their positional ranking of the target value. Eventually, better results are attained in most areas. The experiments indicate that the method proposed herein achieves relatively better prediction results for the southern and northern parts of Wuhan city. The poorer results in the middle region of Wuhan may be explained by the fact that precipitation will be more unpredictable under future warming scenarios (Seino et al. 2018; Xu et al. 2020b). Urbanization also makes it more challenging to predict precipitation in cities (Ajaaj et al. 2018).

With the ongoing development of cities, the number of buildings, roads, and other man-made structures has increased. This has led to significant changes in the surface characteristics of urban areas (Ke et al. 2013). Such changes affect atmospheric circulation and humidity distribution, which in turn have an impact on precipitation patterns, leading to significant changes in the spatial distribution of precipitation (Dai et al. 2023). As a result, urbanization has made it harder to predict urban precipitation correctly. More research and improved prediction models are needed to meet this challenge.

4.3 Prediction accuracy of developed models for different precipitation thresholds

When it comes to precipitation nowcasting, one of the most significant applications is the forecasting of extreme rainfall events. This is due to the fact that extreme rainfall events

Table 3 Rain rate statistics in the HKO-7 benchmark

Rain Rate (mm/h)	Proportion (%)	Rainfall Level
$0 \le x < 0.5$	68.6%	No/ Hardly noticeable
$0.5 \le x < 2$	11.6%	Light
$2 \le x < 5$	8.2%	Light to moderate
$5 \le x < 10$	5.3%	Moderate
$10 \le x < 30$	5.2%	Moderate to heavy
$30 \le x$	1.1%	Rainstorm

may lead to disasters such as flash floods and severe thunderstorms (Franch et al. 2020). Therefore, in this section, we evaluate the forecast accuracy of the proposed PP-Loss as well as other loss functions on ERA5-Land precipitation data for different precipitation thresholds, especially for performance of heavy rain in Wuhan. For the hourly scale rainfall levels, reference is made to the classification criteria of the HKO-7 benchmark(Shi et al. 2017). Table 3 shows the percentage distribution of different precipitation levels in ERA5-Land precipitation dataset. As can be observed, the precipitation distribution is highly imbalanced.

In this section, we comprehensively analyze the R^2 and RMSE scores for different precipitation thresholds to evaluate the model's nowcasting performance. Specifically, we selected precipitation levels larger than 0.5, 2, 5, 10, and 30 mm/h (You et al. 2023) and yielded the results of the model's breakdown for five different precipitation thresholds, as shown in Table 4.

Based on the results in Table 4, we observe that the model using PP-Loss outperforms Focal-R, Gumbel Loss, LDS, and MSE Loss in predicting the five precipitation levels in terms of both R^2 and RMSE. The imbalanced regression loss functions PP-Loss, Focal-R, and LDS were all improved to a certain extent compared to MSE Loss at the thresholds of 0.5 mm, 2 mm, and 5 mm/h. However, MSE Loss achieves good results as well, with the lowest R^2 of the model being 0.850. At the thresholds of 10 mm/h and 30 mm/h, we can see that the accuracy of MSE Loss is significantly reduced compared to 0.5 mm/h and 2 mm/h. This is due to the fact that the widely used MSE Loss is not applicable to tackle imbalanced data. On the contrary, PP-Loss, Focal-R, and LDS all show significant improvement compared to MSE Loss, which also indicates that the imbalanced regression loss function plays an important role in precipitation nowcasting. Gumbel Loss performs poorly for different rainfall thresholds, especially for the threshold of 30 mm/h. The poor performance may be due to the fact that Gumbel Loss is suitable for temporal prediction tasks rather than spatio-temporal prediction tasks. However, at a threshold of 30 mm/h, i.e., for predicting heavy rainfall, PP-Loss is much more effective than the other loss functions. Compared to the best of the other loss functions, LDS, PP-Loss improves the R^2 by 0.225 and reduces the RMSE by 5.196. This is because PP-Loss takes into account information about position of the target values and their ranking, and weights the loss function using cost-sensitive learning. Thus, the impact of each data point on the loss function can be better adjusted, which reduces the negative impact of sample imbalance on model training. This also shows that PP-Loss not only achieves good results in overall precipitation forecasting, but also shows outstanding advantages in heavy precipitation nowcasting.

5 Discussion

Precipitation forecasting is a challenging problem in which forecasting uncertainty stems mainly from data uncertainty and modeling uncertainty (Xu et al. 2021). For this reason, researchers are committed to developing new modeling techniques to improve the accuracy of precipitation forecasting. In this context, machine learning and deep learning techniques have been extensively used to correct precipitation forecasting models. Among them, the loss function serves a crucial role. It guides the direction of model training and affects the forecasting capability of deep learning models.

The imbalance in precipitation data samples remains a challenge, especially when dealing with relatively scarce samples of intense precipitation. Such imbalance may result in poor performance of precipitation forecasting models in predicting heavy precipitation (You et al. 2023). In contrast to previous studies, we innovatively apply a plotting location formulation of the empirical probability distribution function to the design of the loss function. In this study, a new

Table 4 Comparison of the accuracy of different rainfall intensities

Loss Function	$r \ge 0.5$		$r \ge 2$		$r \ge 5$		$r \ge 10$		$r \ge 30$	
	R^2	RMSE (mm)	R^2	RMSE (mm)	R^2	RMSE (mm)	R^2	RMSE (mm)	R^2	RMSE (mm)
MSE Loss	0.850	3.992	0.857	4.441	0.855	5.085	0.814	6.519	0.480	15.117
Focal-R	0.870	3.712	0.879	4.087	0.880	4.628	0.849	5.875	0.589	13.437
Gumbel Loss	0.830	4.249	0.823	4.942	0.792	6.087	0.717	8.035	0.279	17.805
LDS	0.880	3.576	0.891	3.869	0.897	4.278	0.875	5.344	0.665	12.143
PP-Loss	**0.902**	**3.226**	**0.919**	**3.336**	**0.937**	**3.345**	**0.941**	**3.679**	**0.890**	**6.947**

technique PP-Loss is proposed. By ranking and weighting the positions of the target values, this approach adjusts the impact of each data point on the loss function. In this way, the sample imbalance problem could be effectively dealt with and the accuracy of precipitation forecasts could be improved. We evaluate and interpret the usefulness of PP-Loss in suppressing sample imbalance while exploring its application in precipitation forecasting. This approach provides a new way to address the challenges in precipitation forecasting and is expected to improve the performance of forecasting models and enhance the accuracy of precipitation forecasts.

Although we achieved improved results in this study compared to the current state-of-the-art imbalanced regression practices, there remain many directions for further research that need to be explored. To begin with, a limitation of our approach is that it neglects the spatial correlation of precipitation predictions and the numerical similarity of precipitation amounts, which also weakens the model's robustness and generalization ability in prediction skills. Secondly, accurate prediction of extreme precipitation is crucial for disaster prevention and mitigation as well as water resource management, which we failed to achieve. Future work could concentrate on exploring forecasting approaches for extreme precipitation, including feature extraction and modeling of extreme events. By continuously improving the loss function based on imbalanced regression and the prediction model, more effective solutions can be provided to deal with extreme precipitation events.

6 Conclusion

In this work, we propose a sample weighting approach PP-Weight for imbalanced regression targeting imbalanced precipitation data. Based on that weighting scheme, we propose a cost-sensitive learning approach PP-Loss. We trained a multilayer ConvLSTM network using a variety of imbalanced regression loss functions, including Focal-R, Gumbel Loss and LDS, as well as the traditional loss function MSE Loss, to train a multilayer ConvLSTM network. These models were trained and tested on the ERA5-Land precipitation data from ECMWF, using the 3-h precipitation data as the learning sample to predict the next 1-h precipitation data up to the 3-h precipitation data. The experimental results show that the PP-Loss method, a weighting scheme based on position ordering proposed in this paper, exhibits better performance under most circumstances. With the benefit of its unique weighting scheme, our method applies to the spatio-temporal imbalance regression task. Future work includes further improving the capability of the PP-Loss for extreme precipitation prediction and combining the approach with other training models.

Among the various imbalanced regression loss functions studied in this paper, we argue that the proposed PP-Loss method performs best for ConvLSTM in the task of precipitation nowcasting with imbalanced regression. With these results, it is hoped that this work will make a valuable contribution to the field of deep learning in spatio-temporal imbalanced regression and its application in precipitation forecasting research.

Acknowledgements This research was supported by the National Natural Science Foundation of China (42201509, 42101429), the National Key Research and Development Program for Young Scientist (2021YFF0704400) and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (162301212687).

Author contribution Lei Xu contributed to the study's conception and design. Material preparation, data collection and analysis were performed by Xuechun Li. Wenying Du was involved in data curation, results analysis and supervision. The first draft of the manuscript was written by Xuechun Li. Lei Xu, Hongchu Yu, Wenying Du, Zeqiang Chen and Nengcheng Chen commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Code availability Code will be available on request to the corresponding author.

Declarations

Ethics approval Not applicable.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication All authors agreed to let the paper published when considered for publication.

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abd Elrahman SM, Abraham A (2013) A review of class imbalance problem. J Netw Innov Comput 1:332–340

Ajaaj AA, Mishra AK, Khan AA (2018) Urban and peri-urban precipitation and air temperature trends in mega cities of the world using multiple trend analysis methods. Theoret Appl Climatol 132:403–418

Ali H, Salleh MNM, Saedudin R et al (2019) Imbalance class problems in data mining: A review. Indonesian J Electric Eng Comput Sci 14:1560–1571

Al-Yahyai S, Charabi Y, Gastli A (2010) Review of the use of numerical weather prediction (NWP) models for wind energy assessment. Renew Sustain Energy Rev 14:3192–3198

Babar VS, Ade R (2015) A review on imbalanced learning methods. Int J Comput Appl 975:23–27

Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. Nature 525:47–55

Branco P, Torgo L, Ribeiro RP (2017) SMOGN: a pre-processing approach for imbalanced regression. In: First international workshop on learning with imbalanced domains: Theory and applications. PMLR, pp 36–50

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357

Chen L, Cao Y, Ma L, Zhang J (2020) A deep learning-based methodology for precipitation nowcasting with radar. Earth Space Sci 7:e2019EA000812

Chen Y-C (2017) A tutorial on kernel density estimation and recent advances. Biostat Epidemiol 1:161–187

Chen Z, Zeng Y, Shen G et al (2021) Spatiotemporal characteristics and estimates of extreme precipitation in the Yangtze River Basin using GLDAS data. Int J Climatol 41:E1812–E1830

Cheng L, Guo R, Candan KS, Liu H (2020) Representation learning for imbalanced cross-domain classification. In: Proceedings of the 2020 SIAM international conference on data mining. SIAM, pp 478–486

Cui Y, Jia M, Lin T-Y, et al (2019) Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 9268–9277

Cunnane C (1978) Unbiased plotting positions—a review. J Hydrol 37:205–222

Dai X, Wang L, Li X et al (2023) Characteristics of the extreme precipitation and its impacts on ecosystem services in the Wuhan Urban Agglomeration. Sci Total Environ 864:161045

Ding D, Zhang M, Pan X, et al (2019) Modeling extreme events in time series prediction. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp 1114–1122

Fang W, Qin H, Liu G et al (2023) A Method for Spatiotemporally Merging Multi-Source Precipitation Based on Deep Learning. Remote Sensing 15:4160

Franch G, Nerini D, Pendesini M et al (2020) Precipitation nowcasting with orographic enhanced stacked generalization: Improving deep learning predictions on extreme events. Atmosphere 11:267

Gautheron L, Habrard A, Morvant E, Sebban M (2019) Metric learning from imbalanced data. In: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, pp 923–930

Gong Y, Mori G, Tung F (2022) RankSim: Ranking similarity regularization for deep imbalanced regression. arXiv preprint arXiv:220515236

Graves A, Graves A (2012) Long short-term memory. Supervised sequence labelling with recurrent neural networks 37–45

Greff K, Srivastava RK, Koutník J et al (2016) LSTM: A search space odyssey. IEEE Trans Neural Netw Learn Syst 28:2222–2232

Hammad M, Alkinani MH, Gupta BB, Abd El-Latif AA (2021) Myocardial infarction detection based on deep neural network on imbalanced data. Multimedia Systems 1–13

He H, Bai Y, Garcia EA, Li S (2008) ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). Ieee, pp 1322–1328

Jadon A, Patil A, Jadon S (2022) A Comprehensive Survey of Regression Based Loss Functions for Time Series Forecasting. arXiv preprint arXiv:221102989

Johnson JM, Khoshgoftaar TM (2019) Survey on deep learning with class imbalance. J Big Data 6:1–54

Kang B, Xie S, Rohrbach M, et al (2019) Decoupling representation and classifier for long-tailed recognition. arXiv preprint arXiv:191009217

Kann A (2018) Statement of guidance for nowcasting and very short range forecasting (VSRF)

Kaparakis C, Mehrkanoon S (2023) WF-UNet: Weather Fusion UNet for Precipitation Nowcasting. arXiv preprint arXiv:230204102

Ke X, Wu F, Ma C (2013) Scenario analysis on climate change impacts of urban land expansion under different urbanization patterns: a case study of Wuhan metropolitan. Adv Meteorol 2013:1–12

Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 5:221–232

Leon Harter H (1984) Another look at plotting positions. Commun Stat Theory Meth 13:1613–1633

Lin T-Y, Goyal P, Girshick R, et al (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp 2980–2988

Liu J, Xu L, Chen N (2022) A spatiotemporal deep learning model ST-LSTM-SA for hourly rainfall forecasting using radar echo images. J Hydrol 609:127748

Muñoz-Sabater J, Dutra E, Agustí-Panareda A et al (2021) ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth Syst Sci Data 13:4349–4383

Paszke A, Gross S, Massa F, et al (2019) Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems 32

Prudden R, Adams S, Kangin D, et al (2020) A review of radar-based nowcasting of precipitation and applicable machine learning techniques. arXiv preprint arXiv:200504988

Samat SR, Othman N (2023) Plotting Position for Low Flow Frequency Analysis at Jempol River Streamflow Station. In: IOP Conference Series: Earth and Environmental Science. IOP Publishing, p 012021

Scheepens DR, Schicker I, Hlaváčková-Schindler K, Plant C (2023) Adapting a deep convolutional RNN model with imbalanced regression loss for improved spatio-temporal forecasting of extreme wind speed events in the short to medium range. Geosci Model Dev 16:251–270

Seino N, Aoyagi T, Tsuguti H (2018) Numerical simulation of urban impact on precipitation in Tokyo: How does urban temperature rise affect precipitation? Urban Climate 23:8–35

Shi X, Chen Z, Wang H, et al (2015) Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems 28

Shi X, Gao Z, Lausen L, et al (2017) Deep learning for precipitation nowcasting: A benchmark and a new model. Advances in neural information processing systems 30

Shi X, Yeung D-Y (2018) Machine learning for spatiotemporal sequence forecasting: A survey. arXiv preprint arXiv:180806865

Steininger M, Kobs K, Davidson P et al (2021) Density-based weighting for imbalanced regression. Mach Learn 110:2187–2211

Subramanya K (2008) Engineering hydrology. McGraw-Hill

Torgo L, Ribeiro RP, Pfahringer B, Branco P (2013) Smote for regression. In: Portuguese conference on artificial intelligence. Springer, pp 378–389

Usharani B (2023) ILF-LSTM: Enhanced loss function in LSTM to predict the sea surface temperature. Soft Comput 27:13129–13141

Wang S, Cao J, Philip SY (2020) Deep learning for spatio-temporal data mining: A survey. IEEE Trans Knowl Data Eng 34:3681–3700

Wang Y, Long M, Wang J, et al (2017) Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. Advances in neural information processing systems 30

Wu Z, Feng H, He H et al (2021) Evaluation of soil moisture climatology and anomaly components derived from ERA5-land and GLDAS-2.1 in China. Water Resour Manage 35:629–643

Xie Y, Mallick T (2023) A Comparative Study of Loss Functions: Traffic Predictions in Regular and Congestion Scenarios. arXiv preprint arXiv:230815464

Xin Y, Yang Y, Chen X et al (2022) Evaluation of IMERG and ERA5 precipitation products over the Mongolian Plateau. Sci Rep 12:21776

Xu L, Chen N, Chen Z et al (2021) Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions. Earth Sci Rev 222:103828

Xu L, Chen N, Moradkhani H, et al (2020a) Improving global monthly and daily precipitation estimation by fusing gauge observations, remote sensing, and reanalysis data sets. Water Resour Res 56:e2019WR026444

Xu L, Chen N, Yang C et al (2022) Quantifying the uncertainty of precipitation forecasting using probabilistic deep learning. Hydrol Earth Syst Sci 26:2923–2938

Xu L, Zhang C, Chen N, et al (2020b) Potential precipitation predictability decreases under future warming. Geophys Res Lett 47:e2020GL090798

Yang Y, Zha K, Chen Y, et al (2021) Delving into deep imbalanced regression. In: International Conference on Machine Learning. PMLR, pp 11842–11851

You X, Liang Z, Wang Y, Zhang H (2023) A study on loss function against data imbalance in deep learning correction of precipitation forecasts. Atmos Res 281:106500

Zhang M, Ding D, Pan X, Yang M (2021) Enhancing time series predictors with generalized extreme value loss. IEEE Trans Knowl Data Eng 35:1473–1487

Zhang Y, Zheng X, Li X, et al (2023) Evaluation of the GPM-IMERG V06 Final Run products for monthly/annual precipitation under the complex climatic and topographic conditions of China. J Appl Meteorol Climatol

Zhang Y-F, Ren W, Zhang Z et al (2022) Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing 506:146–157

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com