Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Improved freezing rain forecast using machine learning

Qiuzi Han Wen [a,c], Dingyu Wan [c], Quan Dong [b,*], Yan Yan [d], Pingwen Zhang [c,e]

[a] *Huaneng Clean Energy Research Institute, Beijing, 102209, PR China*
[b] *National Meteorological Centre, Beijing, 100081, PR China*
[c] *Peking University, Beijing, 100871, PR China*
[d] *Beijing Aviation Meteorological Institute, Beijing, 100085, PR China*
[e] *Wuhan University, Wuhan, Hubei, 430072, PR China*

ARTICLE INFO

ABSTRACT

Freezing rain is one of the most damaging weather phenomena in winter or early spring in many parts of the world, affecting traffic, power lines and agriculture. Thus, reliable and computationally efficient prediction of its occurrence is urgently needed in weather forecast operations. However, there are different thermodynamic processes that can lead to freezing rain, resulting in unsatisfactory forecasting performance of the state-of-the-art Numerical Weather Prediction (NWP) models. Here a data-driven forecasting method for freezing rain using machine learning technologies is proposed. Observations of weather phenomenon collected from 2 515 national weather stations of China for winter of 2016–2019 and the corresponding atmospheric predictors derived from ERA5 reanalysis are used. The prediction function is constructed based on the classification and regression tree, and the predicting variables include temporal and vertical profiles of fundamental thermodynamic and kinematic parameters from 500 hPa to 1000 hPa, with a total dimension of 2 304. The LightGBM (Light Gradient Boosting Machine) framework is adopted to train our prediction model and an algorithm-level approach of modifying the loss function is used to address the imbalance of classes to improve forecasting skill. Results show that the data-driven prediction model, namely DDFR (data driven forecast of freezing rain), out-performs the benchmark NWP, i.e., ECMWF IFS product. It's improvements in terms of TS score range from 120% to 258% depending on different forecast leading times, which range from 0 to 12 h. In addition, DDFR is applied in an operational NWP model of China. The problem of domain adaptation is tackled and transfer learning method is employed to adapt the original DDFR to this NWP model. The effectiveness of such adaptation has been demonstrated by its performance on both training and testing datasets.

1. Introduction

Freezing rain is a type of precipitation which freezes on contact with the ground, vegetation, cars, power lines and other surfaces and as a special type of winter precipitation, is one of the most damaging weathers in winter or early spring (Call 2010; Cortinas 2000; Yu et al., 2016; Ou et al., 2011). For example, the Great Ice Storm of 1998 caused massive damage to trees and electrical infrastructure throughout a relatively narrow swath of land from eastern Ontario to southern Quebec, New Brunswick and Nova Scotia in Canada, and bordering areas from northern New York to central Maine in the United States, leading to widespread long-term power outages. Not only millions were left in the dark and cold in the coldest part of the winter for up to several weeks, it also led to 34 fatalities, and a shutdown of activities in large

cities like Montreal and Ottawa in Canada. In the winter of 2020, a severe freezing rain event resulted in 9390 dead trees and more than 151, 188 broken of branches, with an estimated loss of street roads being over 222 million Yuan at Changchun, the capital of Jilin province, China (Liu et al., 2022a). Since the beginning of the 21st century, an increased trend in the frequency of freezing rain has been observed in many regions worldwide such as the South-central Canada, the Long Island in New York (Cortinas 2000). Thus, skillful freezing rain forecasting is vital for local transportation management, road maintenance, aviation ground deicing operation, electric power planning and infrastructure maintenance in winter as well as disaster loss control.

The formation mechanism of freezing rain has been extensively studied, yet it remains an ongoing research problem (Zerr 1997). Two primary microphysical mechanisms have been widely studied: the

---

* Corresponding author.
 E-mail address: dongquan@cma.gov.cn (Q. Dong).

melting mechanism or the ice-phase mechanism (IPM), and the super-cooled warm rain mechanism (WRM) (Huffman and Norman 1988; Rauber et al., 1994, 2001; Forbes et al., 2014). The IPM mechanism is characterized by existence of sufficient ice particles in the cloud, warm layer in which the ice particles melt to liquid as they fall through, and cold layer in which the liquid particles become supercooled. This mechanism has been observed in cases such as the U.S. following cold-air damming for the Appalachian Mountains, characterized by multiple 0 °C levels in the upper atmosphere (Bell and Bosart 1988) as well as cases in Jilin Province, China (Liu et al., 2022a). On the other hand, WRM mechanism describes the situation that the temperatures are not cold enough for ice nuclei to be active in a purely subfreezing profile. Instead, particles primarily exist as supercooled liquid that falls to the ground. This phenomenon is commonly observed in Southwest China (Dong et al., 2020a, 2020b; Reeves et al., 2014). Due to these two mechanisms, when freezing rain occurs at the surface under similar terrain conditions, the vertical thermal structure of the upper atmosphere may vary considerably. For example, in the Beijing area, there is similar chance for a freezing rain event to be associated with or without a melting layer in the upper atmosphere, associated with IPM and WRM respectively as reported by Yu et al. (2016) based on analysis using 53-year historical observations and NCEP reanalysis data.

Despite numerous studies focusing on microphysical precipitation processes and numerical modeling of hydrometeor microphysics to forecast various ground-level precipitation types using parameters like hydrometeor mixing ratios and temperature (Thériault and Stewart, 2010; Ikeda et al., 2013), there remains a notable absence of reliable quantitative modeling methods for precipitation types other than rain and snow. This gap is particularly evident in extreme precipitation types like freezing rain (Thériault et al., 2006). Freezing rain is sensitive to several factors including the content of ice particles in the cloud, the depth and intensity of warm layer and cold layer, the intensity of precipitation and their interactions with the thermodynamics of the atmosphere through latent heating and cooling (Thériault and Stewart 2010). In addition to the defect of microphysical parameters and limitation of spatial scale to simulate these processes, parameterization schemes are applied to predict the occurrence of freezing rain in NWP models recently. While IPM mechanism is more widely considered the consideration of WRM mechanism is rare (ECMWF 2016). As a result, the accuracy of NWP model forecast of freezing rain is not high (Gascón et al., 2018). In general, the main methodology in parameterization of freezing rain is an implicit method which uses environmental profiles of temperature and/or humidity to infer the precipitation type (e.g., Baldwin and Treadon 1994; Bourgouin 2000; Schuur et al., 2012; Elmore and Grams 2015; Chenard et al., 2015).

Common implicit methods include the Ramer algorithm (Ramer 1993), the Baldwin algorithm (Baldwin and Treadon 1994), the Bourgouin algorithm (Bourgouin 2000), the NSSL algorithm (Schuur et al., 2012). Numerous studies have revealed that snow and rain can be predicted with high skill, but freezing rain cannot be well predicted due to its complex physical nature and uncertainty effects involved in computation (Reeves et al., 2014; Reeves 2016; Dong et al., 2020a). The data-driven method based on machine learning (ML) to prediction which is less restricted by poorly understood physical processes and their interactions provides a new perspective to more skillful forecast of freezing rain. For example, freezing rain is affected by temperature, humidity, density of cloud drops et al. of the whole layer of atmosphere, from the top to the ground and can react to the atmosphere by vertically transporting heat as falling through the atmosphere. Recent operational numerical weather models cannot describe these physical processes quantitatively (ECMWF 2020). The data-driver method using the historical data can be trained to fit this physical process through estimating the parameters of ML.

In recent years, ML has been widely applied in earth system research, such as seasonal forecasting, long-range spatial telecommunication, and prediction of ENSO, etc. (Reichstein et al., 2019; Ham et al., 2019).

Efforts have been made using ML to forecast weather parameters where standard NWP products show more room for improvement. Recent studies have shown that forecasting precipitation amount can be improved using deep neural network models (Shi et al., 2017; Sonderby et al., 2020). Cloud and precipitation have been identified as high-potential areas where ML methodologies can make significant progresses (Bonavita et al., 2020; ECMWF 2020).

In this paper, a data-driven method named DDFR (Data-Driven Forecast of Freezing Rain) is developed for forecasting freezing rain in China. The study addresses the 4-class classification (rain, snow, rain-snow mix, freezing rain) problem of precipitation types. Focusing specifically on freezing rain. The paper is organized as follows: Section 2 introduces datasets used in this study; Section 3 describes the machine learning approach in DDFR, encompassing the mathematical model and the algorithmic framework; Section 4 evaluates the performance of DDFR; Section 5 integrates DDFR with an NWP system to enhance its parameterization for precipitation type forecast in China. It illustrates the performance through a recent freezing rain case study in 2021. Finally, Section 6 provides concluding remarks and discussions.

2. Data used in this study

2.1. Observation

Weather phenomenon data from 2 515 national weather stations in China, observed at 02:00, 05:00, 08:00, 11:00, 14:00, 17:00, 20:00 and 23:00 CST during the winter half-year (from October to March) of 2016–2018, were used in this study. The dataset has been quality controlled by weather observers and forecasters in Chinese weather stations based on consistency between precipitation type, temperature and precipitation amount, and widely used for their forecast and service operations Dong et al. (2020a; 2020b). The original observations are encoded based on SYNOP (surface synoptic observations) Code, referred to as FM-12 by the WMO. Observations containing precipitation were selected and categorized into four groups: rain (RA), snow (SN), rain-snow mix (RASN) and freezing rain (FZRA). The classification rule we applied is similar to that in Gascón et al. (2018), with the exception that we classify the precipitation type of ice pellets as FZRA. The impact of these changes is deemed negligible, given that we only encountered 8 cases of ice pellets out of the total 357560 samples.

In total, we have 1838 FZRA samples, and their spatial distribution is depicted in Fig. 1. It is evident from the figure that FRZA predominantly
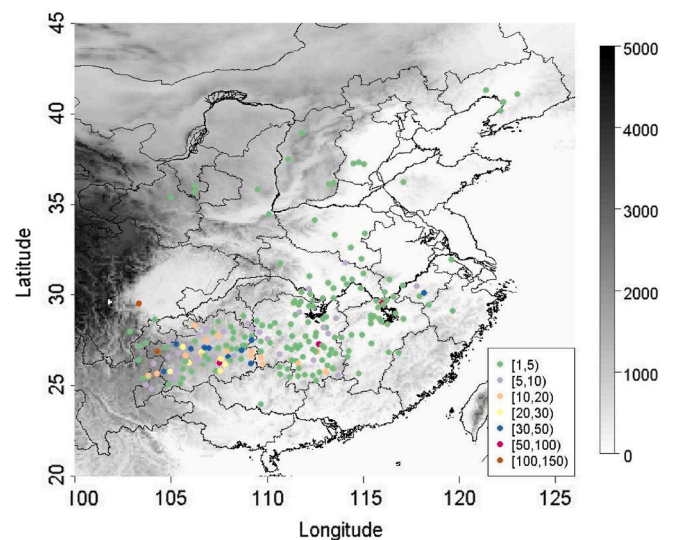

Fig. 1. Station distribution and number of freezing rain samples (color) with altitude shaded. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

occurs in the southern part of China, consistent with previous findings indicating that FZRA events in China are primarily concentrated south of the Yangtze River (Ou et al., 2011). Past studies have shown that 73% of FZRA events in China are attributed to the WRM mechanism, which is nearly three times more prevalent than those caused by IPM. FZRA occurrences associated with the IPM mechanism are primarily observed in northern China, while the WRM mechanism predominantly influences the incidence of FZRA in southern China (Ou et al., 2011).

2.2. ERA5 reanalysis data

ERA5 is the fifth generation of global reanalysis, covering period from 1979 to near real time. Produced by the European Centre for Medium-range Forecast (ECMWF), it uses the 4D-Var data assimilation and model forecasts within CY41R2 of the ECMWF Integrated Forecast System (IFS). ERA5 dataset consists of a high-resolution dataset (HRES) at a 31 km resolution, providing hourly data with 137 vertical pressure levels ranging from 1000hpa to 0.01hpa, which is the product used in this study. Recent assessment has indicated that, compared to its predecessor (ERA-Interim), ERA5 exhibits enhanced suitability in China, particularly regarding temperature and relative humidity at lower and medium levels, as well as the wind field and geopotential height at upper pressure levels (Meng et al., 2018). In this study, ERA5 serves as the pseudo-observations for the vertical profile of the atmosphere, forming the dataset of predictors for our prediction model. We obtained hourly atmospheric variables from the ERA5 pressure level dataset, encompassing 16 vertical levels between 500 hPa and 1000 hPa, cover all the national weather stations introduced in section 2.1, during the winter seasons from 2016 to 2019.

2.3. NWP data

The forecasting method proposed in this study can serve as an independent prediction model for FZRA when adequate observations of vertical atmosphere profiles are available. Additionally, it can be integrated into an NWP model, enhancing FZRA diagnosis capabilities. In this study, we initially compare the proposed data-driven prediction method with a state-of-the-art (SOTA) NWP model, namely the IFS high-resolution forecast generated by ECMWF. Subsequently, we integrate this method with a Chinese-developed NWP system (YHGSM) to enhance its FZRA forecasting performance. YHGSM is a global atmospheric spectral model exhibiting performance parameters akin to ECMWF-IFS. It features a horizontal resolution of 0.5° latitude/longitude and comprises 91 vertical model layers. Capable of generating numerical forecasts for up to 10 days, YHGSM provides updates twice daily.

To evaluate the performance of our proposed method, we selected a SOTA NWP, specifically the IFS (cycle 41r1) developed and operated by ECMWF, as the benchmark representing the current state of the art in FZRA forecasting. The ECMWF-IFS comprises various modules of operational global models, including an atmospheric general circulation model, an ocean general circulation model, an ocean wave model, a land surface model, and perturbation models used for data assimilation and generating forecast ensembles. This system generates forecasts covering time frames ranging from the intermediate range (up to 15 days) to seasonal intervals (up to 7 months). It employs a horizontal resolution of 0.125° (deterministic HRES) and a temporal resolution of 1 h. In version 41r1, ECMWF-IFS has improved its representation of physical processes related to clouds and precipitation, enabling enhanced prediction of FZRA by more realistically modeling the melting and refreezing of precipitation particles during the formation of IPM-type FZRA (Forbes et al., 2014). Precipitation type forecasts generated by the deterministic high-resolution run of ECMWF-IFS, denoted as PTYPE, were obtained for the same period as the observations, from October 1st, 2016, to March 31st, 2019, covering the entire China region. Grid forecasts were interpolated to stations using the nearest neighbor approach. The PTYPE

variable comprises six categories of weather phenomena associated with precipitation: rain, snow, wet snow, a mixture of rain and snow, freezing rain, and ice pellets. To facilitate comparison with our results, we recoded PTYPE into four types by combining snow and wet snow, as well as freezing rain and ice pellets into two separate types, respectively, following the approach used in the research conducted by Dong et al. (2020a).

3. Methodology of data-driven forecasting

3.1. The ML modeling framework

ML is adept at approximating functions in very high dimensions, making it particularly advantageous for problems affected by the 'curse of dimensionality,' which contributes to its widespread popularity. Let's consider one of the most fundamental and commonly used supervised learning problems in ML. Using notations similar to those used in Li et al. (2023), we denote the given dataset by: $S = (x_i, y_i)$, $i = 1, ..., N$. Here, S comprises inputs x_i and their corresponding labels y_i, and N is the size of the data. The underlying assumption is the existence of a deterministic model or function F^*, such that each y_i can be determined by x_i through F^*, denoted as $y_i = F^*(x_i)$. When the labels y_i take values in a continuum, such as in R (real numbers), it is considered as a regression problem. Conversely, if y_i assumes discrete values, it is considered a classification problem. The objective here is to construct a predictive model, F^*, and estimate its parameters.

Suppose there is no prior knowledge on the structure of F^*, a feasible approach is as follows: 1) consider a collection of functions that can be well represented on a computer; 2) from this collection, select an $\widetilde{F}$ that "best approximates" F^* in some sense, which then becomes our learned predictive model. This subjective collection of functions is termed the hypothesis space, denoted as **H**. But how do we select a specific function $\widetilde{F}$ to approximate F^*? This selection heavily depends on a precise definition of "best approximation", typically defined by the choice of loss functions. In practice, beginning with some initial point $\widetilde{F}_0$, optimizing the given loss function over the hypothesis space allows for the learning of the best estimate, $\widetilde{F}$. In this theoretical framework, both basis functions and loss functions play crucial roles in constructing a ML model. For instance, in a deep learning model with a hypothesis $\mathbf{H_m}$, the basis function $f(x, \theta)$ can be written as:

$$f(x, \theta) = \frac{1}{m} \sum_{j=1}^{m} a_j \sigma(< \omega_j, x_j >), \tag{1}$$

where $\theta = \{(a_j, \omega_j), j = 1, ..., m\}$ are parameters to be trained, and $\sigma(\cdot)$ is the activation function (Weinan 2020). A natural loss function would be the L^2 loss which measures how good the data fits the model:

$$L_n(\widehat{\theta}) = \frac{1}{n} \sum_{j=1}^{N} \left| f(x_j, \widehat{\theta}) - y_j \right|^2. \tag{2}$$

Another popular example is tree-based ML model, the basis function of which, namely CART (i.e. the classification and regression tree), has the following form:

$$f(x, \theta) = \sum_{j=1}^{m} c_j I(x_j \in R_k), \tag{3}$$

where $\theta = \{c_1, ..., c_m\}$ are parameters to be learned, and $\{R_k, k = 1, ..., K\}$ define K clusters of the entire feature space. Most ML models are additive models, i.e., model learned is a linear combination of basis models, since simplicity guarantees stronger representation ability in high-dimensional space.

Consider the problem of forecasting precipitation type at a given location. Firstly, we define the vertical and temporal environmental profile at this location to be ep, and let pt, a categorical variable taking

discrete values, to represent the type of all possible precipitation to be considered, according to the methodology of implicit modeling, define a forecast function F^* such that:

$$F^*(ep) = F^*(f(ep, \theta)) = pt, \tag{4}$$

where ep is the input features of the model following a probability distribution D on $\mathbf{R^m}$ and $f(\cdot)$ is the basis function of the hypothesis space $\mathbf{H_m}$. Obviously this is a classification problem since pt only takes discrete values. In definition, ep is a m-dimension tensor, which could include vertical dimension, temporal dimension, and variable dimension, each representing different characteristics of the atmospheric environmental profiles. The resolution chosen for each dimension varies according to the value space of pt. If the value space of pt only includes snow and rain, the resolution of each dimension could be coarse, for example, it's enough to only consider temperature and relative humidity in the variable space. But when it comes to freezing rain, ice pellets or hails, the finer the resolution presenting the environmental profile, the more precise forecast can be obtained. Thus, in order to tackle the problem of FZRA forecasting, dimensions of the forecast model have to be large.

In this work, tree model is used to approximate the forecast function F^*, i.e., adopting CART given in Equation (3) as the basis function (Chen and Guestrin 2016). Hyperplane and neural network are also checked in preliminary analysis, but numerical experiments in the Beijing-Tianjin-Hebei region of China show less promising predictive performance. CART is simple and easy to interpret, thus has strong ability in representing complex function. In addition, there are many popular and reliable ensemble learning framework to support its realization, for example, the bagging framework (i.e., random forest), and the boosting framework (i.e., XgBoost and LightGBM). We adopt the LightGBM (Ke et al., 2017) framework to train our forecast model in this work.

3.2. DDFR: a data-driven forecasting method for FZRA

The forecasting problem has been formulated as a tree-based classification problem under supervised learning framework in Section 3.1. To solve this problem, first, we need to collect and organize data to obtain the dataset $S = \{ep_i, pt_i\}$, $i = 1, \ldots, N$. We define:

$$pt = \begin{cases} 1, rain\ (RA) \\ 2, rain - snow - mix\ (RASN) \\ 3, snow\ (SN) \\ 4, freezing - rain\quad (FZRA) \end{cases} \tag{5}$$

Construction of ep is referred to as feature engineering in ML and each array in the tensor ep represents a feature. In this section, details in our feature engineering are introduced firstly, then the issues of optimization and its realization are tackled, respectively.

Instead of modeling the precipitation process in the atmosphere using physical laws governing explicit equations, physical mechanism-informed features on the precipitation process are chosen to construct such implicit ML method. In order to finely represent the atmospheric environment profile under different types of precipitation, we construct a 3- dimensional feature space including dimension of meteorological elements, vertical structures, and temporal variation. More specifically, in order to predict precipitation type at location S and time T, we have:

$$ep^{(S,T)} = ep^{(S,T)}(v, p, t), \tag{6}$$

where $v = 1, \ldots, 6$, $p = 1, \ldots, 16$, and $t = 1, \ldots, 24$. Here v represents number of selected meteorological elements as predictors; p represents number of pressure levels considered; and t represents time window used to forecast time T.

Selection of physical mechanism-informed or empirical meteorological variables is crucial for model performance and interpretability. Temperature and relative humidity are vital features to include, which have been widely used by both human forecasters, NWP when

diagnosing precipitation type, and previous objective forecast of precipitation type (Dong et al., 2013; Ramer 1993; Baldwin and Treadon 1994; Bourgouin 2000; Schuur et al., 2012). In addition, wind and vertical velocities also play important role in determining the type of particles and where they will fall, especially for freezing rain that formed when southwest wind at 700 hPa (Ou et al., 2011). Geopotential height of some pressure level such as 500hpa is also key variables considered by weather forecasters (Liu et al., 2022a), thus geopotential height is also included in our model (Fig. 3). The dimensions of the vertical structure for the feature space include 16 different pressure levels, ranging from 500 to 1000hpa; the temporal dimensions of the feature space involve an inverse time window with a length of 24 h staring from the forecasting time point T. The dimension of the input feature space is 2 304 in total. Value of each dimension for the predictive feature $ep(S, T)$ (v, p, t), are given in Equation (7) and further visualized by a set of 16*24 images with 6 channels as illustrated in Fig. 2.

The feature data used for modeling are extracted from the ERA5 dataset introduced in section 2.2.

$$\begin{cases} v^T = [t, r, w_u, w_v, v_v, z] \\ t(i) = T - 24 + i, i = 1, \ldots, 24 \\ p(i) = \begin{cases} 1000 - 25*(i-1), i = 1, \ldots 11 \\ 750 - 50*(i-11), i = 12, \ldots, 16 \end{cases} \end{cases}, \tag{7}$$

Data used for model training consists of observations introduced in section 2.1 as labels and the feature dataset derived from ERA5 reanalysis. We split the data into two parts: training and testing, with a ratio of 9:1, randomly but keeping the distribution of the proportions of each class constant in both datasets. That is for each of the four categories defined in equation (5), all samples are randomly arranged, using the random. shuffle(.) function in Python, then the first 90% of the samples are selected into the training dataset, while the rest remains in the testing dataset.

3.2.1. Model training by LightGBM

Given the dataset S, model training is the key step for solving the forecast function F^*, i.e., finding the minimum of the loss function. We adopt LightGBM as the basic framework for realization of model training. It's a gradient boosting framework for tree-based learning algorithm. The idea of boosting came from the "hypothesis boosting problem" proposed by Michael in 1988 (Kearns 1988), which "asks whether an efficient learning algorithm that outputs a hypothesis whose performance is only slightly better than random guessing implies the
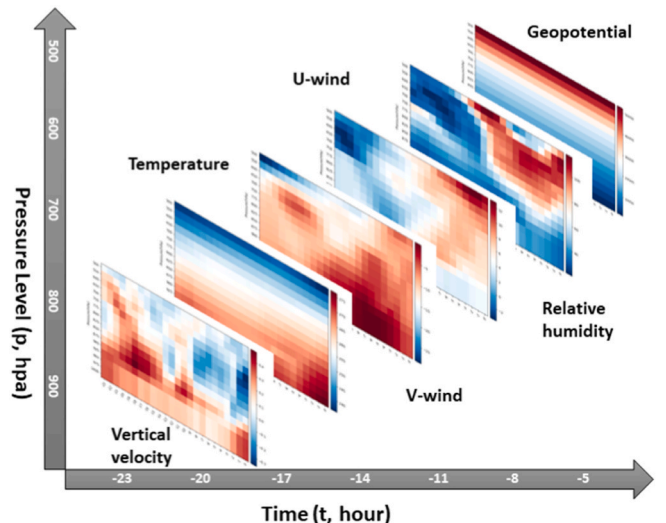

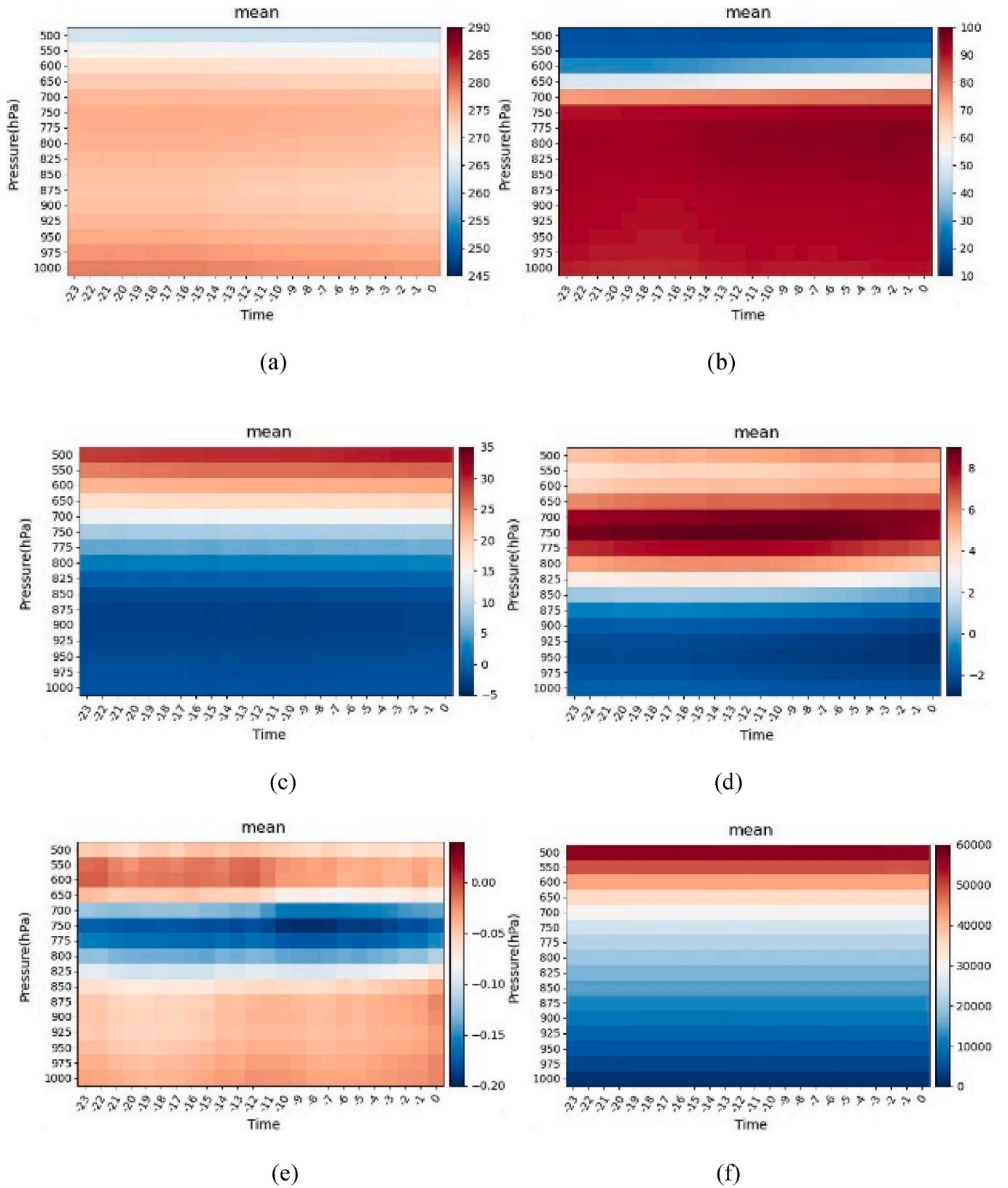**Fig. 2.** Illustration of the predictive feature $ep(v, p, t)$.

Fig. 3. Mean of features for freezing rain events: (a) temperature, (b) relative humidity, (c) *u*-wind, (d) *v*-wind, (e) vertical velocity, (f) geopotential heights.

existence of an efficient algorithm that outputs a hypothesis of arbitrary accuracy".

The idea was firstly realized as the Adaptive Boosting or AdaBoost algorithm in 1995 (Freund 1995), and later were recast under statistical framework by Brieman in 1997 (Breiman, 1997). In 1999, it was further developed under the framework of numerical optimization (Friedman, 1999). A gradient descent like procedure was introduced to minimize the loss of the model, and the well-known Gradient Boosting Decision

Tree (GBDT) was proposed as its computational realization. LightGBM is an upgrade framework of GBDT designed for massive data processing. It is designed to be distributed and efficient with the following advantages: faster training speed and higher efficiency, lower memory usage, better accuracy, supporting of parallel, distributed, and GPU learning, and capable of handling large-scale data (Ke et al., 2017). The target function to minimize in our problem can be written by:

$$L = \sum_{i=0}^{N} l(\widehat{pt} + \widetilde{F}(ep)) + \frac{1}{2}\lambda \sum_{j=1}^{M} w_j^2 + \alpha \sum_{j=1}^{M} |w_j| \qquad (8)$$

$$w_j = r \times \frac{\sum_{i \in I_j} \frac{\partial l}{\partial(\widehat{pt}=0)}}{\sum_{i \in I_j} \left(\frac{\partial^2 l}{\partial(\widehat{pt}=0)^2} + \lambda \right)} \qquad (9)$$

and the solution of our problem is given by:

$$F^* = \underset{F}{\mathrm{argmin}}\, L \qquad (10)$$

Here $l(\cdot)$ is a differentiable convex loss function; $\widetilde{F}(\cdot)$ is the weak learner, a function in the hypothesis space **H** with CART as the basis function; the latter two parts in Equation (8) are regularization terms which penalized for the complexities of the model and to smooth the learned weights to avoid model over-fitting. $\{M, \lambda, \alpha, r\}$ are hyper-parameters subjectively chosen and adjusted to optimize the performance of model.

The major innovations of LightGBM include two techniques during the optimization process: Gradient-based One Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) which fulfills the limitations of histogram-based algorithm that is primarily used in all GBDT frameworks. In GBDT, the information gain is usually measured by the variance after splitting. GOSS keeps those features with large gradients and only randomly drop those features with small gradients to retain the accuracy of information gain estimation. This procedure can lead to a more accurate gain estimation than uniformly random sampling, with the same target sampling rate, especially when the range of information gain is large. In addition, by adopting the EFB technique, the speed for model training is improved without sacrificing accuracy. This is achieved by assuming sparsity in the high-dimensional data. Specifically, in a sparse feature space, many features are mutually exclusive, i.e., they never take nonzero values simultaneously. The exclusive features can be safely bundled into a single feature, which is called an EFB. From the perspective of architecture, LightGBM splits the tree leaf-wise as opposed to other boosting algorithms that grow tree level-wise. It chooses the leaf with maximum delta loss to grow.

Since the leaf is fixed, the leaf-wise algorithm has lower loss compared to the level-wise algorithm. But it is noticed that leaf-wise tree growth might increase the complexity of the model and may lead to over-fitting in small datasets. More technical details on the realization of LightGBM can be found in Ke et al. (2017).

3.2.2. Challenge of class-imbalance

When dealing a classification problem using machine learning, a common underlying assumption is the distribution of class is close to uniform. However, this assumption can seldom be satisfied in real-world applications, and the problem of class-imbalance indicates a significant bias from this assumption. Ignoring such bias may result in misleading classification skill towards the majority class, and in extreme cases, may ignore the minority class altogether. As a matter of fact, class with less data is of greater interests under most situations (Megahed et al., 2021; Liu et al., 2022b). In our problem, out of the total 505300 samples, we only have 1838 FZRA samples, which indicate a severe class-imbalance problem.

Effective classification using class-imbalanced data is an active research topic in machine learning, which has many applications such as financial fraud detection and email spam filtering. To deal with class-

imbalance problem, techniques at data level or algorithm level or both are developed during the past decade (Johnson and Khoshgoftaar 2019). At data level, data resampling technique such as oversampling or undersampling is usually used to create a balanced dataset based on the original imbalanced one; at the algorithm level, modifications are usually made on the learner, i.e., the loss function or its output. Data-level techniques are more suitable for low degree of class imbalance, or for experimental purposes. Whether model trained using resampled dataset can still work in real application depends on how significant the effect of domain shift has on the robustness of model performance, i.e., the issue of model generalization from resampled data distribution to the original data distribution. Algorithm-level adaptation works better for scenarios with high class imbalance. The distribution of training data is not altered in algorithmic methods. Instead, the main idea is to adjust the learning process so that importance of minority class can be increased, and this is usually implemented by adding penalty or weight to each class in the loss function.

In our problem, the ratio of freezing rain occurrence to the total precipitation events is only 0.5%, indicating a critical and high class-imbalance issue. Data resampling technique would change the resampled data distribution significantly, resulting to inauthentic occurrence frequency of freezing rain in the prediction. Thus we adopt the algorithm-level approach of modifying the loss function by weighting each class using the inverse of its sample size to increase the weight of freezing rain fitting. Though focal loss is frequently used to address the problem of class-imbalance, it works better for two-class classification problem. More specifically, the loss function in our algorithm is defined as:

$$L = \sum_{i=0}^{N} \frac{1}{n_{\widehat{pt}}} l(\widehat{pt} + \widetilde{F}(ep)) + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 + \alpha \sum_{j=1}^{T} |w_j| \qquad (11)$$

$$w_j = r \times \frac{\sum_{i \in I_j} \frac{1}{n_{\widehat{pt}}} \frac{\partial l}{\partial(\widehat{pt}=0)}}{\sum_{i \in I_j} \left(\frac{1}{n_{\widehat{pt}}} \frac{\partial^2 l}{\partial(\widehat{pt}=0)^2} + \lambda \right)} \qquad (12)$$

Here $1/n_{\widehat{pt}}$ is the weight of loss function for sample i, and $n_{\widehat{pt}}$ is the number of samples with label $\widehat{pt}$ in the training set. Under such modification, the learning ability of the model for freezing rain is significantly increased.

The DDFR method workflow is shown in Fig. 4 and is programmed in Python 3.7.7 on a computing server with 4 Intel Skylake CPUs. The entire dataset is split into the training dataset and validation dataset with a ratio of 9:1. The total number of samples in our training dataset are 454770 and the total number of validation samples we used are 50530. Within each dataset, the ratio among four class remains the same, i.e., 208 : 5: 39 : 1, for RA, RASN, SN and FR. The training time is about 20 min in average. We applied the above-described method to obtained 5 forecasting models with different lead-time respectively, i.e., diagnostic model (0hr lead-time), 3hr lead-time, 6hr lead-time, 9hr lead-time and 12hr lead-time forecasting models.

4. Results

4.1. Forecast skill evaluation

Forecasting skill of precipitation type is often evaluated in terms of accuracy (ACC) and Heidke skill score (HSS) (Jolliffe and Stephenson, 2003; Dong et al., 2020a). The 4-category contingency table is given in Table 1.

The metric of accuracy in machine learning measures overall how often our model can achieve a correct classification, that is:

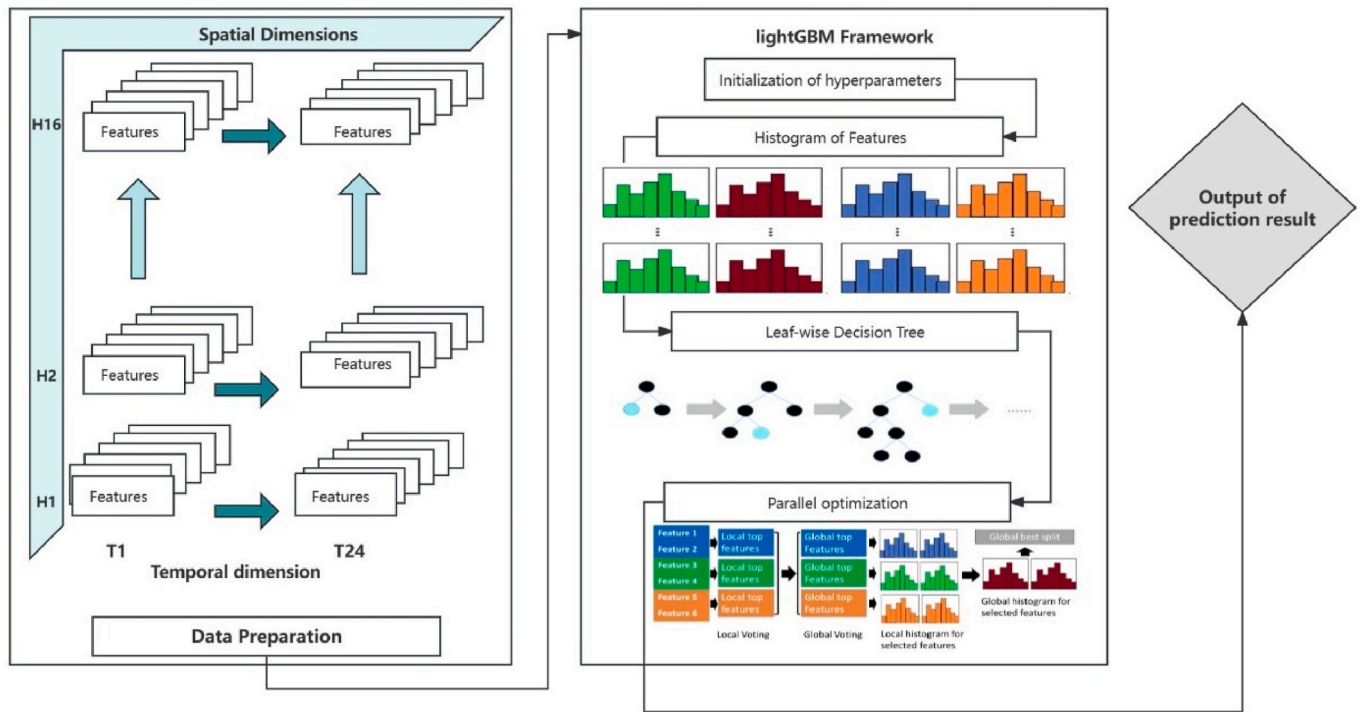$$ACC = \frac{a + b + c + d}{a + b + c + \ldots + h + q + u + v + z}$$

Fig. 4. Workflow of the proposed DDFR method.

Table 1
Contingency table for 4-category of winter Precipitation Types.

		Ground Truth			
		Rain	Sleet	Snow	Freezing Rain
Prediction	Rain	a	e	f	g
	Sleet	h	b	u	v
	Snow	z	l	c	m
	Freezing Rain	n	o	q	d

However, when class imbalance is not negligible in the dataset, as is the case in our study, ACC is a biased estimate of model performance which may provide results on class with fewer samples. Thus, we employ another metric, HSS, to take the problem of class imbalance into account. The HSS, also known as kappa score in Statistics, is "a skill score for categorical forecasts where the ACC measure is scaled with the reference value from correct forecasts due to chance" (Jolliffe and Stephenson 2003). Define the observed and predicted marginal distributions to be p_{0k} and p_{1k} respectively for each type $k \in \{1, 2, 3, 4\}$. By definition, HSS is calculated using the following formula:

$$HSS = \frac{ACC - (p_{01}p_{11} + p_{02}p_{12} + p_{03}p_{13} + p_{04}p_{14})}{1 - (p_{01}p_{11} + p_{02}p_{12} + p_{03}p_{13} + p_{04}p_{14})}$$

In addition, TS score is also employed to evaluate the forecasting skill on every single type of precipitation event (Jolliffe and Stephenson 2003). Forecasting skill of DDFR is compared to a SOTA NWP, i.e. ECMWF-HRD, the benchmark in this study. The benchmark results are obtained from Dong et al. (2020a), which are evaluated using the same observational dataset of our study. Overall performance of DDFR in terms of ACC and HSS are summarized in Tables 2 and 3 and further visualized in Figs. 4 and 5. Results for both training (T) and validation (V) dataset are presented with that from validation dataset better reflecting the generalization ability of DDFR in practice. Forecasting lead time ranges from 0 to 12 h, with a 3-h interval.

Results show DDFR outperforms the SOTA NWP across all metrics (Fig. 6). Specifically, there are improvements of 4.9% in ACC, 6.3% in HSS (Table 2), 3.6% in TS for rain forecasting, 8.8% in TS for snow forecasting, 223% in TS for rain-snow mix forecasting, and 101.4% in TS for freezing rain forecasting (Table 3). These findings highlight the effectiveness of ML-based data-driven models, particularly for forecasting high-impact weather phenomena where the physical mechanisms alone may not be sufficient or computable. Furthermore, upon reviewing Fig. 5 and Tables 3 and it becomes evident that DDFR's advantage in forecasting freezing rain remains consistent over time. Its improvements in TS range from 93.6% to 178.7% compared to the SOTA NWP, demonstrating its robustness in this aspect.

4.2. Analysis of feature importance

Within the LightGBM framework, the importance of each feature is computed using two methods: 'split,' which counts the number of times a feature is utilized in the model, and 'gain,' which measures the total gains of splits utilizing the feature. In our study, we investigated both methods and found minimal discrepancies in the results. The outcomes

Table 2
Overall performance of DDFR (T-traning, V-validation, B-benchmark).

	ACC-T	ACC-V	ACC-B	HSS-T	HSS-V	HSS-B
DDFR-0hr	0.996	0.965		0.986	0.887	
DDFR-3hr	0.990	0.960	0.915	0.9665	0.872	0.820
DDFR-6hr	0.995	0.964		0.9821	0.881	
DDFR-9hr	0.995	0.963		0.984	0.879	
DDFR-12hr	0.998	0.964		0.994	0.882	

Table 3
DDFR for freezing rain.

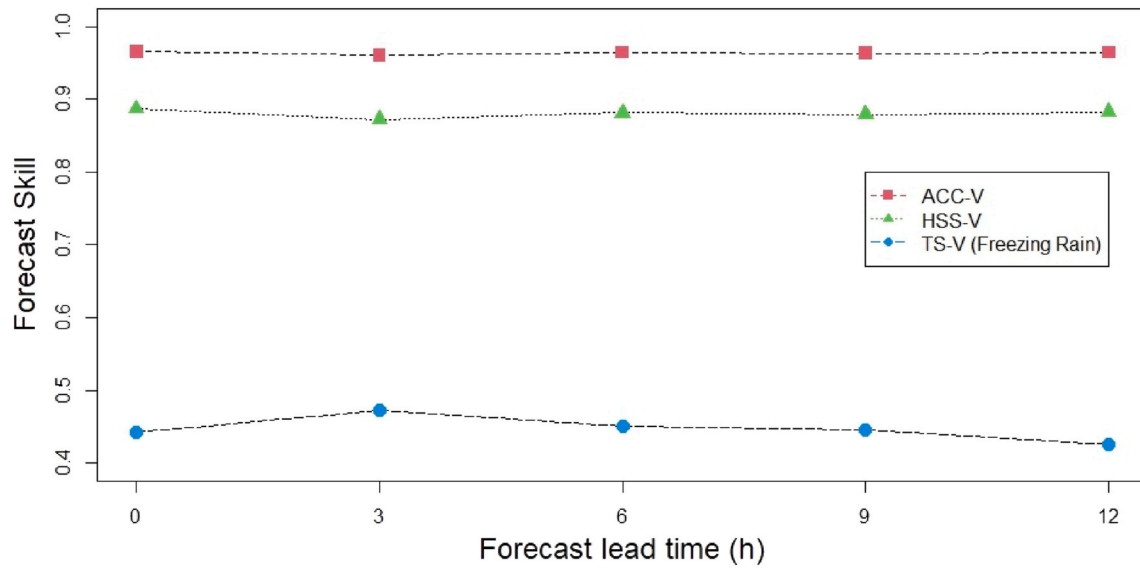|	TS-T	TS-V	TS-B
FZRA-0hr	0.931	0.443	
FZRA-3hr	0.962	0.473	0.20
FZRA-6hr	0.926	0.450	0.18
FZRA-9hr	0.913	0.446	0.16
FZRA-12hr	0.978	0.426	0.22

Fig. 5. Forecasting skill of DDFR w. r.t lead time (0–12hr).
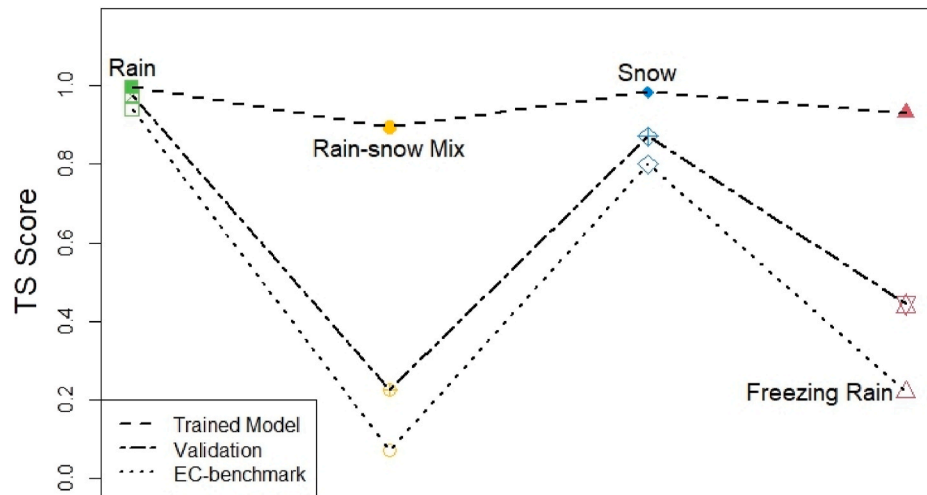

Fig. 6. Forecasting skill of DDFR in terms of TS score.

obtained using the 'split' method are illustrated in Fig. 7. An analysis of feature importance indicates that.

1. When considering the temporal-vertical profile, the average temperature ceases to be the most influential predictor. Among the six variables investigated in our study, temperature seems to rank ahead only of geopotential height, which consistently exhibits the lowest importance score. However, despite this, the significance of ground-level temperature and the near-time vertical structure of temperature remains notable. This finding aligns with prior research (Dong et al., 2013), underscoring their ongoing significance in the analysis.

2. Vertical velocity demonstrates the highest average importance score, with its significance particularly pronounced in the upper levels, specifically above 700 hPa. Additionally, it is observed that the importance of the V-component of wind surpasses that of the corresponding U-component above 700 hPa. These findings strongly indicate the crucial role played by upper-level variables in achieving more accurate forecasts of freezing rain, notably linked to the formation of warm layers (Ou et al., 2011).

3. The importance of variables typically increases over time, notably evident in the importance plot of relative humidity. Relative

humidity, overall, holds more significance compared to temperature and wind variables. Moreover, its vertical structure over the nearest 5-h interval proves to be particularly critical in the context of DDFR. This aligns with the IPM mechanism, where the presence of ice particles in higher levels of clouds is acknowledged as pivotal for FZRA formation (Zerr, 1997).

5. Application and case study

5.1. Domain adaptation

Initially, DDFR is trained using the ERA5 dataset. However, for practical applications, predictive inference is required using NWP data output. Assuming the probability distributions of these two datasets are identical becomes unreliable due to distinct data generation mechanisms. Moreover, the historical dataset of NWP lacks stability for ML model training, attributed to frequent model upgrades. Consequently, this introduces the typical domain adaptation problem in ML. Solutions to this issue generally fall into three categories: the discrepancy-based method, the adversarial-based method, and the reconstruction-based method (Kouw and Loog, 2018). In our study, we employ a
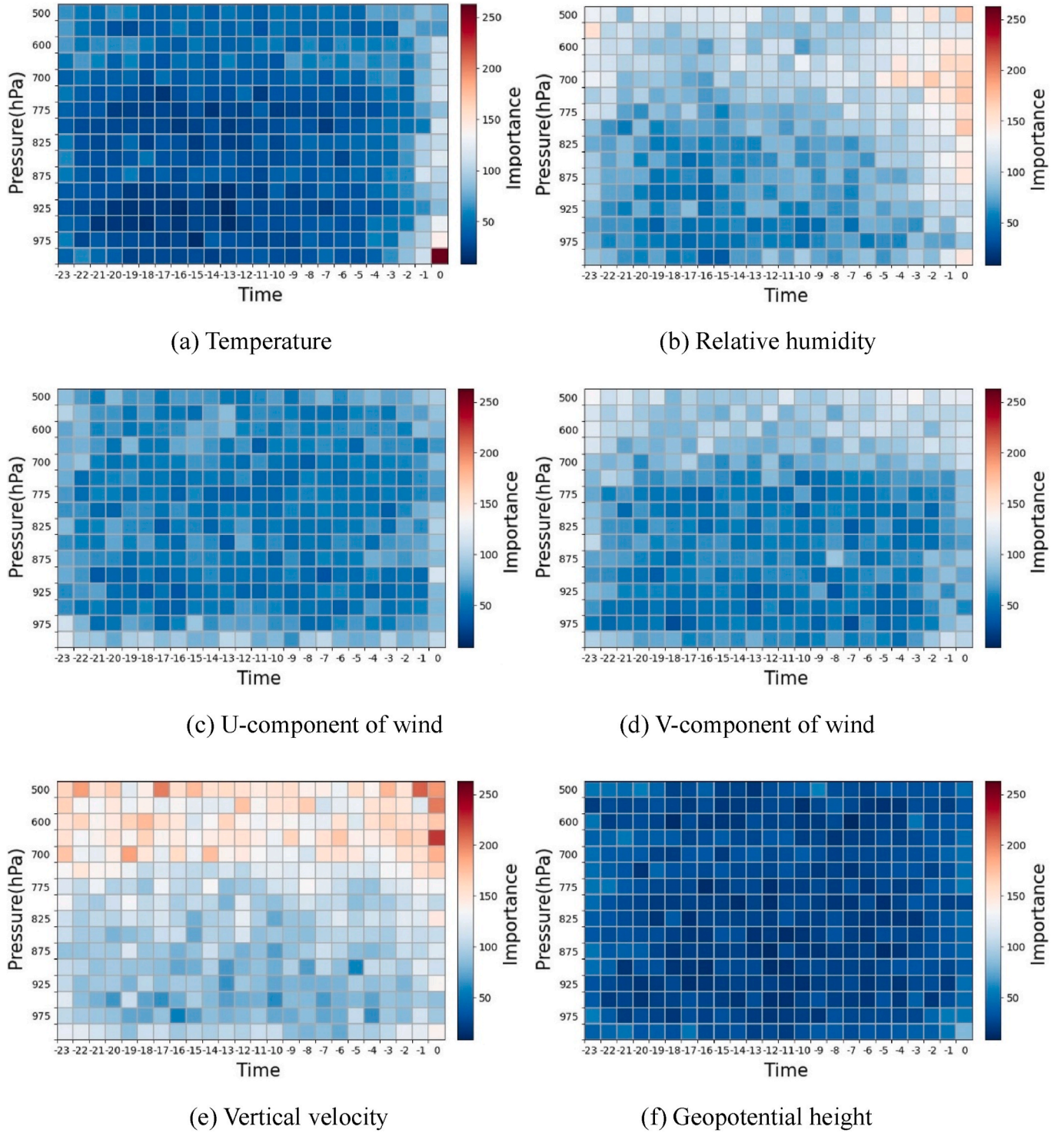
(a) Temperature

(b) Relative humidity

(c) U-component of wind

(d) V-component of wind

(e) Vertical velocity

(f) Geopotential height

Fig. 7. Feature importance visualization.

reconstruction-based method to adapt DDFR to the domain generated by the given NWP, referred to as ADDFR.

We collected data generated by our target NWP model throughout January to March 2019 across China, totaling 129,577 samples. Adopting the same feature space as used in DDFR, we reconstructed the training and validation datasets. The training dataset comprises 90% of the ERA5 data and 80% of the NWP data, while the remaining 10% of the ERA5 data forms the validation dataset. Finally, we tested the results using the remaining 20% of the NWP data to examine its performance in real applications.

The performance of ADDFR was evaluated using ACC, HSS, and TS, and the results are depicted in Figs. 8 and 9. Although the ACC and HSS of ADDFR were marginally lower than DDFR in the training dataset, its performance matched that of DDFR in the validation set. This suggests the strong adaptability of DDFR to the NWP data domain. Similar to DDFR, ADDFR also exhibited notable improvements in forecasting freezing rain compared to forecasting rain, snow, or rain-snow mix, as displayed in Fig. 9. These findings provide additional evidence supporting the potential of ML in enhancing predictions of extreme weather and climate events. ADDFR has been seamlessly integrated into the
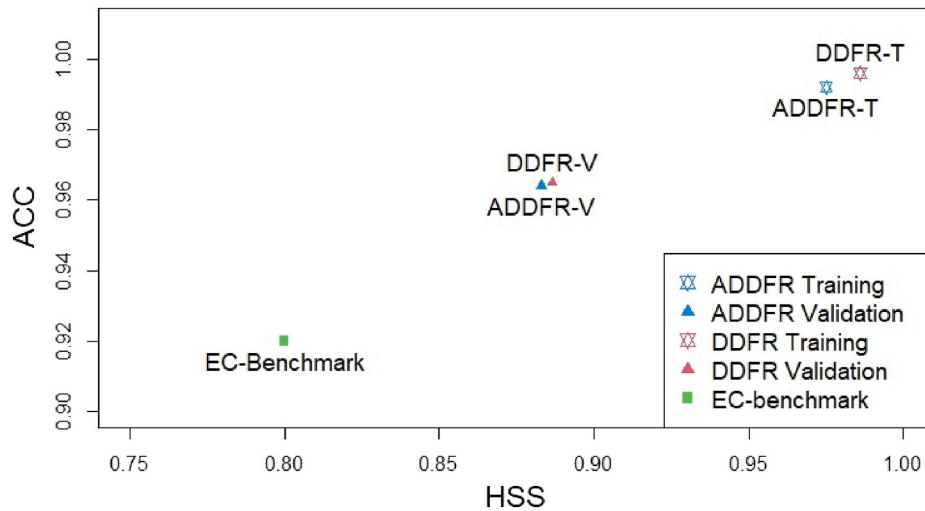
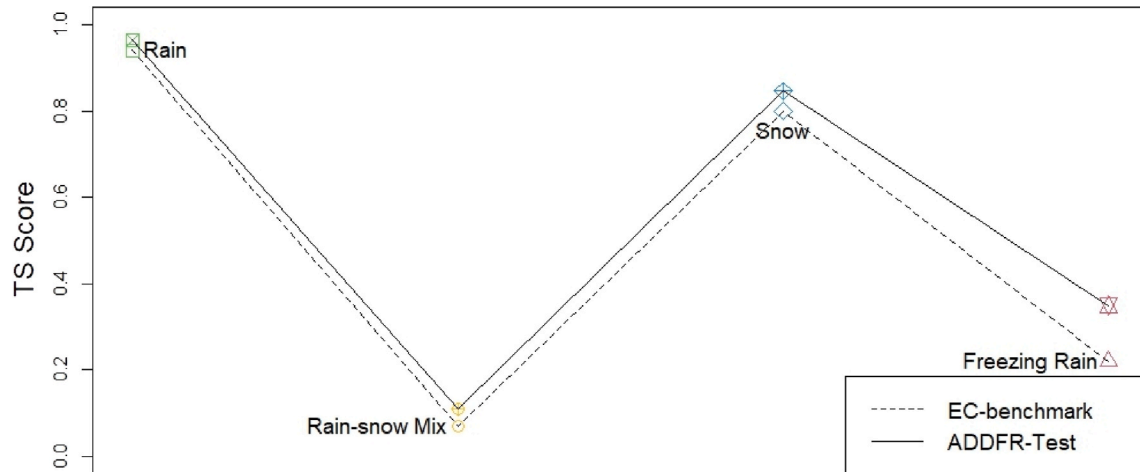**Fig. 8.** Overall performance of ADDFR (NWP Adapted DDFR).

Fig. 9. ADDFR Forecasting skill in terms of TS score.

target NWP model as an enhancement module, offering hourly forecast results on precipitation types in operational settings.

ADDFR has demonstrated substantial improvements in forecasting freezing rain in practical applications. Freezing rain stands as one of the most prevalent meteorological hazards during winters in Guizhou province, China. Therefore, more accurate forecasts hold the potential to alleviate its adverse impacts on agriculture, transportation, and the safety of lives and properties. For instance, during an early November 2021 cold wave event, ADDFR was employed to predict precipitation types. According to observations, it effectively predicted a freezing rain event occurring between 3 and 4 a.m. on November 9.

The selected predictive features for this freezing rain event are depicted in Fig. 10. It is evident that low-level temperature and relative humidity significantly contribute to freezing rain formation in Guizhou, associated with the presence of super-cooled raindrops (Dong et al., 2013). Additionally, the southerly wind near 700 hPa and the vertical wind near 500 hPa are also substantial, corresponding to two other factors influencing freezing rain formation in Guizhou: water vapor transportation from the south and the condensation of water vapor into cloud or raindrops (Gao et al., 2014).

6. Conclusions and discussions

Freezing rain stands as one of the most catastrophic weather

phenomena worldwide, causing disruptions in traffic, agriculture, power grids, and various other sectors. In the past February of 2024, the southern China suffered from several extreme precipitation events including freezing rains, which caused great losses to transportation, residents' lives, and government management. More accurate forecast of freezing rain events and other extreme types of precipitation are in urgent demand. Nevertheless, the formation of freezing rain is influenced by numerous dynamic, thermodynamic, and physical factors, where even slight changes in any of these factors can lead to precipitation as rain or snow. Hence, achieving precise forecasts of freezing rain poses a challenge globally.

Machine learning provides more powerful modeling tools to face these challenges. Though forecast of precipitation types using machine learning methods such as random forest or supporting vector machine have been proposed recently based on various sources of data such as reanalysis data and dual-polarization radar measurements, type of freezing rain is seldomly included in these works (Półrolniczak et al., 2021; Shin et al., 2022; Zhang et al., 2023). In this study, a ML method called DDFR is employed to forecast freezing rain in China. This method utilizes data from ERA5, NWP models, and surface station observations which were well quality controlled.

The results highlight that the data-driven prediction model, DDFR, surpasses the benchmark NWP (ECMWF IFS) product. The improvement in TS ranges from 120% to 258% across different forecast leading times,
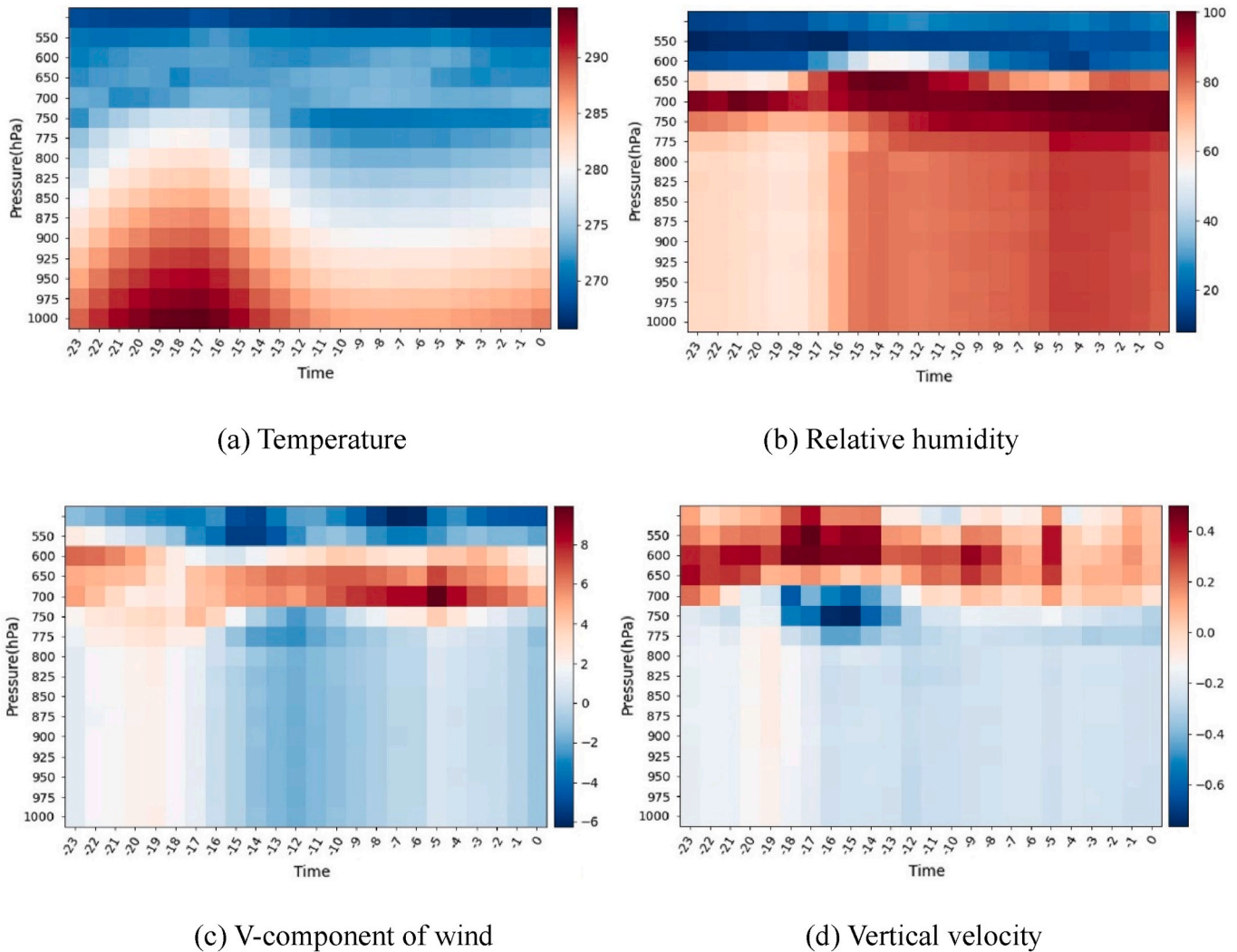
(a) Temperature

(b) Relative humidity

(c) V-component of wind

(d) Vertical velocity

Fig. 10. Predictive feature visualization.

spanning from 0 to 12 h. Following the resolution of the domain adaptation challenge and employing transfer learning methods, the original DDFR is successfully adapted to a Chinese NWP model, showcasing effectiveness across both training and testing datasets.

However, certain limitations exist for the DDFR model. Firstly, the algorithm's development and evaluation solely rely on datasets from China, without verifying its generalization performance in other global regions. Secondly, the model doesn't account for precipitation rate, a significant factor in determining precipitation type (Reeves et al., 2014). Given the inherent uncertainties in the atmosphere, an ensemble or probabilistic forecast for precipitation type might prove beneficial, especially for rare events such as freezing rain in North and Northeast China, or snow in South China. Therefore, exploring data-driven probabilistic forecasts for precipitation type stands as a potential avenue for future research.

CRediT authorship contribution statement

Qiuzi Han Wen: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Dingyu Wan:** Writing – original draft, Visualization, Software, Investigation, Formal analysis, Data curation. **Quan Dong:** Writing – review & editing, Supervision, Funding acquisition, Data curation. **Yan Yan:** Writing –

original draft, Resources, Formal analysis, Data curation. **Pingwen Zhang:** Formal analysis, Data curation.

Declaration of competing 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.

Data availability

The authors do not have permission to share data.

References

Baldwin, M.R., Treadon, C.S., 1994. Precipitation type prediction using a decision tree approach with NMC's mesoscale Eta Model. In: Preprints, 10th Conf. On Numerical Weather Prediction. Amer. Meteor. Soc., Portland, OR, pp. 30–31.

Bell, G.D., Bosart, L.F., 1988. Appalachian cold-air damming. Mon. Weather Rev. 116, 137–161. https://doi.org/10.1175/1520-0493(1988)116<0137:ACAD>2.0.CO;2.

Bonavita, M., Arcucci, R., Carrassi, A., et al., 2020. Machine learning for earth system observation and prediction. Bull. Am. Meteor. Soc. 102, 1–13. https://doi.org/10.1175/BAMS-D-20-0307.1.

Bourgouin, P., 2000. A method to determine precipitation type. Weather Forecast. 15, 583–592.

Breiman, L., 1997. Arcing the edge. In: Statistics Department, vol. 486. University of California at Berkeley. Technical Report.

Call, D.A., 2010. Changes in ice storm impacts over time: 1886-2000. Wea. Climate Soc 2, 23–35.

Chen, T.Q., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chenard, M., Schumacher, P.N., Reeves, H.D., 2015. Determining precipitation type from maximum temperature in the lower atmosphere. In: Preprints, 23rd Conf. On Numerical Weather Prediction/27th Conf. on Weather Analysis and Forecasting, 6B2. Amer. Meteor. Soc., Chicago, IL.

Cortinas, J., 2000. A climatology of freezing rain in the great lakes region of North America. Mon. Weather Rev. 128, 3574–3588. https://doi.org/10.1175/1520-0493(2001)129<3574:ACOFRI>2.0.CO;2.

Dong, Q., Hu, N., Zong, Z.P., 2020a. Application and verification of the ECMWF precipitation type forecast product (PTYPE). Meteor. Mon. 46 (9), 1210–1221 (in Chinese).

Dong, Q., Huang, X.Y., Zong, Z.P., 2013. Comparison of artificial neural network and linear regression methods in forecasting precipitation types. Meteor. Mon. 39 (3), 324–332 (in Chinese).

Dong, Q., Zhang, F., Zong, Z.P., 2020b. Objective precipitation type forecast based on ECMWF ensemble prediction product. Journal of Applied Meteorological Science. 31 (5), 527–542. https://doi.org/10.11898/1001-7313.20200502 (in Chinese).

ECMWF, 2016. ECMWF IFS Documentation Cy43R1, Part IV: Physical Processes, pp. p118–p120.

ECMWF, 2020. ECMWF STRATEGY 2021-2030. https://doi.org/10.21957/s21ec694kd.

Elmore, K.L., Grams, H.M., 2015. Using mPING Data to drive a forecast precipitation type algorithm. In: 13th Conf. On Artificial Intelligence. Amer. Meteor. Soc., Phoenix, AZ.

Forbes, R., Tsonevsky, I., Hewson, T., et al., 2014. Towards Predicting High-Impact Freezing Rain Events. ECMWF Newsletter, pp. 15–21. https://www.ecmwf.int/node/17334.

Freund, Y., 1995. Boosting a weak learning algorithm by majority. Inf. Comput. 121 (2), 256–285.

Friedman, J.H., 1999. Greedy function approximation: a gradient boosting machine. Technical Discussion: Foundations of TreeNet(tm).

Gao, S.T., Zhang, X., Wang, J., et al., 2014. The environmental field and ensemble forecast method for the formation of freezing rain over Guizhou Province. Chin. J. Atmos. Sci. 38 (4), 645–655 (in Chinese).

Gascón, E., Hewson, T., Haiden, T., 2018. Improving predictions of precipitation type at the surface: description and verification of two new products from the ECMWF ensemble. Weather Forecast. 33, 89–108. https://doi.org/10.1175/WAF-D-17-0114.1.

Ham, Y.G., Kim, J.H., Luo, J.J., 2019. Deep learning for multi-year ENSO forecasts. Nature 573, 568–572. https://doi.org/10.1038/s41586-019-1559-7.

Huffman, G.J., Norman, Jr.G.A., 1988. The supercooled warm rain process and the specification of freezing precipitation. Mon. Weather Rev. 116, 2172–2182.

Ikeda, K.M., Steiner, J.P., Alexander, C., 2013. Evaluation of cold-season precipitation forecasts generated by the hourly updating high-resolution rapid refresh model. Weather Forecast. 28 (4), 921–939.

Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. J Big Data 6, 27. https://doi.org/10.1186/s40537-019-0192-5.

Jolliffe, I.T., Stephenson, D.B., 2003. Forecast Verification a Practitioner's Guide in Atmospheric Science. The Atrium. John Wiley & Sons Ltd, Southern Gate, Chichester, West Sussex PO19 8SQ, England.

Ke, G.L., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. NIPS.

Kearns, M., 1988. Thoughts on hypothesis boosting. Machine Learning Class Project.

Kouw, W.M., Loog, M., 2018. An Introduction to Domain Adaptation and Transfer Learning. https://doi.org/10.48550/arXiv.1812.11806 arXiv.

Li, Q., Lin, T., Shen, Z., 2023. Deep learning via dynamical systems: an approximation perspective. J. Eur. Math. Soc. 25 (5), 1671–1709.

Liu, J.B., Liu, D.W., Gao, C.L., et al., 2022a. The causes of rare snow and freezing rain weather in Jilin Province on November 18, 2020, and its impacts on transmission lines. In: Proc. SPIE 12348, 2nd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2022), 1234827. https://doi.org/10.1117/12.2641406 (10 November 2022).

Liu, Z.N., Wei, P.F., Wei, Z.P., et al., 2022b. Handling Inter-class and Intra-class Imbalance in Class-Imbalanced Learning arXiv. 2111.12791.

Megahed, F.M., Chen, Y.J., Megahed, A., et al., 2021. The class imbalance problem. Nat. Methods 18, 1270–1272. https://doi.org/10.1038/s41592-021-01302-4.

Meng, X.G., Guo, J.J., Han, Y.Q., 2018. Preliminarily assessment of ERA5 reanalysis data. Journal of Marine Meteorology 38 (1), 91–99. https://doi.org/10.19513/j.cnki.issn2096-3599.2018.01.011 (in Chinese).

Ou, J.J., Zhou, Y.Q., Yang, Q., et al., 2011. Analyses on spatial-temporal distributions and temperature-moisture structure of freezing rain in China. Plateau Meteor. 30 (3), 692–699 (in Chinese).

Półrolniczak, M., Kolendowicz, L., Czernecki, B., et al., 2021. Determination of surface precipitation type based on the data fusion approach. Adv. Atmos. Sci. 38, 387–399. https://doi.org/10.1007/s00376-020-0165-9.

Ramer, J., 1993. An empirical technique for diagnosing precipitation type from model output. Preprints. Fifth Int. Conf. On Aviation Weather Systems. Amer. Meteor. Soc., Vienna, VA, pp. 227–230.

Rauber, R.M., Ramamurthy, M.K., Tokay, A., 1994. Synoptic and mesoscale structure of a severe freezing rain event: the St. Valentine's Day ice storm. Weather Forecast. 9, 183–208.

Rauber, R.M., Ramamurthy, M.K., Tokay, A., 2001. Further investigations of a physically based, nondimensional parameter for discriminating between locations of freezing rain and ice pellets. Weather Forecast. 16, 185–191.

Reeves, H.D., 2016. The uncertainty of precipitation-type observations and its effect on the validation of forecast precipitation type. Weather Forecast. 31, 1961–1971. https://doi.org/10.1175/WAF-D-16-0068.1.

Reeves, H.D., Elmore, K.L., Ryzhkov, A., Schuur, T., Krause, J., 2014. Sources of uncertainty in precipitation-type forecasting. Weather Forecast. 29, 936–953. https://doi.org/10.1175/WAF-D-14-00007.1.

Reichstein, M., Camps-Valls, G., Stevens, B., et al., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Schuur, T.J., Park, H.S., Ryzhkov, A.V., et al., 2012. Classification of precipitation types during transitional winter weather using the RUC model and polarimetric radar retrievals. J. Appl. Meteorol. Climatol. 51, 763–779.

Shi, X.J., Gao, Z.H., Lausen, L., et al., 2017. Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. https://doi.org/10.48550/arXiv.1706.03458 arXiv e-prints.

Shin, K., Kim, K., Song, J.J., Lee, G., 2022. Classification of precipitation types based on machine learning using dual-polarization radar measurements and thermodynamic fields. Rem. Sens. 14, 3820. https://doi.org/10.3390/rs14153820.

Sonderby, K.S., Espeholt, L., Heek, J., et al., 2020. MetNet: A Neural Weather Model for Precipitation Forecasting. ArXiv, abs/2003.12140.

Thériault, J.M., Stewart, R.E., 2010. A parameterization of the microphysical processes forming many types of winter precipitation. J. Atmos. Sci. 67 (5), 1492–1508.

Thériault, J.M., Stewart, R.E., Milbrandt, J.A., Yau, M.K., 2006. On the simulation of winter precipitation types. J. Geophys. Res. 111, D18202 https://doi.org/10.1029/2005JD006665.

Weinan, E., 2020. Machine learning and computational mathematics. Commun. Comput. Phys. 28 (5), 1639–1670. https://doi.org/10.4208/cicp.OA-2020-0185.

Yu, B., Du, J., Zhang, L.N., 2016. Characteristics of freezing rain in Beijing from 1960 to 2013. Journal of Meteorology and Environment 32 (4), 113–118 (in Chinese).

Zerr, R.J., 1997. Freezing rain: an observational and theoretical study. J. Appl. Meteorol. Climatol. 36, 1647–1661. https://doi.org/10.1175/1520-0450(1997)036<1647:FRAOAT>2.0.CO;2.

Zhang, L., Wen, H.Q., Yu, B., Li, S., Wang, Y., 2023. Forecast of winter precipitation type based on machine learning method. Entropy 25 (1), 138. https://doi.org/10.3390/e25010138.