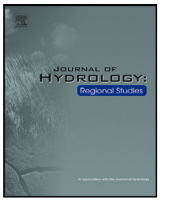
Journal of Hydrology: Regional Studies 64 (2026) 103243

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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



Applying machine learning for precipitation forecasting in uneven rainfall regions of Taiwan☆

Fan-Ming Chiu [a], Lawrence Jing-Yueh Liu [b], Uswatun Hasanah [c], Chuan-Ming Liu [d,*]

[a] Architecture Design Department, Phison Electronics Corporation, Zhunan Township, Miaoli County 35059, Taiwan
[b] Department of Climate, Meteorology and Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana IL 61801, United States
[c] College of Electrical Engineering and Computer Science, National Taipei University of Technology, Taipei 10618, Taiwan
[d] Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei 10618, Taiwan

ARTICLE INFO

Keywords:
Machine learning
Precipitation forecasting
Regression model
Uneven data

ABSTRACT

Study region: Taiwan, with its steep topography, high spatiotemporal rainfall variability, and frequent extreme weather events, which make it difficult to predict the weather accurately. Predictability of rainfall in the short term is essential in flood prevention, hydrological modeling, and water resource management in this area.

Study focus: This paper introduces a superior machine learning model in precipitation prediction with surface observation data. Even though the conventional method of Numerical Weather Prediction (NWP) models is still in use, it is computationally demanding and time-consuming. The machine learning models offer a more data-driven option, but heavy rainfall is hard to predict because precipitation data is non-Gaussian and highly skewed. In order to overcome this shortcoming, the Zero-Order Hold (ZOH) method is used to preprocess temporal data, and a regression-based ZD-XGBoost model is suggested. The objective function of the model is meant to fit the uneven distribution of rainfalls and enhance extreme precipitation prediction.

New hydrological insights for the region: The suggested ZD-XGBoost model maintains the accuracy of the light rainfall and minimizes the errors of heavy rainfall prediction by about 20% and 40% respectively, on hourly and daily data. These findings indicate that the distribution-conscious machine learning with surface observations can be effectively used to reduce the problem of data imbalance and improve the reliability of rainfall prediction in topographically complex and rainfall-heterogeneous areas like Taiwan.

1. Introduction

The weather and climate impact everyday life, ranging from daily routine activities to the fields of agriculture and water resources. Therefore, weather and climate forecasts are often the guidance for hazardous weather prevention, irrigation scheduling, reservoir management, and urban planning for flood control. Conventionally, the Numerical Weather Prediction (NWP) is the common approach to weather forecasting, based on numerical modeling with physical laws to predict future weather systems (Lynch, 2008). The NWP in the past decades has been undoubtedly successful, making the guidance useful for society to make decisions. This lies in the detailed description of the weather and in the greater accuracy with greater computer resources.

☆ This article is part of a Special issue entitled: 'AI in Regional Hydrology' published in Journal of Hydrology: Regional Studies.

* Corresponding author.
E-mail address: cmliu@ntut.edu.tw (C.-M. Liu).

https://doi.org/10.1016/j.ejrh.2026.103243
Received 23 October 2025; Received in revised form 2 February 2026; Accepted 9 February 2026
Available online 21 February 2026

However, NWP faces some challenges. First, the governing equations are complex and thus require large computational resources to predict the weather. The top-notch weather predictions from the European Centre for Medium-Range Weather Forecasts (ECMWF) take around an hour to produce a global 10-day weather forecast (Rasp et al., 2020). Second, NWP is highly sensitive to errors in the initial condition due to the chaotic nature of the atmosphere, commonly referred to as the butterfly effect (Lorenz, 1963). To obtain an optimal estimate of the initial state, data assimilation is employed to combine unevenly distributed observations with imperfect numerical models (Kalnay, 2003). Nevertheless, biases and uncertainties remain due to intrinsic atmospheric variability and observational limitations. Third and finally, the precipitation from either global weather models (e.g. HRES (Rasp et al., 2020)) or high-resolution limited area models (e.g. HRRR; Dowell et al. (2022)) is based on parameterizations due to the subgrid issue (unresolvable process due to the limit of resolutions). Microphysical parameterizations, for example, are highly uncertain due to the assumption of unresolvable physical phenomena (Morrison et al., 2020). These caveats, ranging from limited computational resources to the inaccuracy in the current NWP, make machine learning-based weather predictions an alternative to improve weather forecasts in both accuracy and timeliness.

ECMWF suggests different approaches for using machine learning in predicting weather (ECMWF, 2024). These approaches are: (1) Predictions from analysis; (2) Predict observations targets from analysis input; (3) Learn the analysis; (4) Prediction from observations in model space; (5) Predict future observations from observations. Which approach is most suitable for weather prediction remains an open and actively debated question within the community. Most of the machine learning weather predictions aim to simulate the full-field atmospheric conditions, so the "Predictions from analysis" is chosen. For example, Pangu-weather (Bi et al., 2023), FourCastNet (Pathak et al., 2022), and GraphCast (Lam et al., 2023b) use the same ECMWF ERA5 data to train in different ML configurations. They achieve tremendous success in showing that the ML model can perform similarly or even outperform the traditional NWP models. However, the learning dataset is the reanalysis, meaning the model inherits the bias in the model. Although the reanalysis contains the observation information in its data assimilation process, the full-field approach remains limited in its ability to faithfully represent real-world atmospheric variability. Meanwhile, the precipitation prediction in ERA5 is known to have significant bias (Lavers et al., 2022). The precipitation is unevenly distributed and a very non-Gaussian variable. With the aforementioned subgrid issues, this "Predictions from analysis" approach is not the optimal choice in the context of precipitation predictions. Therefore, the "Predict future observations from observations" is a more reasonable approach to reduce the bias from data assimilation and physical model.

There are several ways to do the "Predict future observations from observations" approach. In precipitation forecasts, the radar nowcasting (Shi et al., 2017; Sønderby et al., 2020; Ravuri et al., 2021; Espeholt et al., 2022) using a radar emulator inherits the concept of radar data extrapolation (Germann and Zawadzki, 2002; Seed, 2003; Woo and Wong, 2017). Although this approach cannot resolve the full-field atmospheric condition, observation-based ML models are advantageous in terms of computational cost and timeliness by directly leveraging observational data to generate precipitation forecasts at short lead times. Nevertheless, these models do not incorporate information from ground-based observations. Since the radar reflectivity does not have a one-to-one relationship with the rainfall rate (Ignaccolo and Michele, 2020), the actual rainfall value cannot be directly inferred from the radar reflectivity pattern. Therefore, the ground-based observations are an alternative to acquiring the precipitation value directly from the observation. Additionally, the ground-based observations can provide variables other than precipitation, such as temperature, wind, and humidity. This makes ML forecasts from the ground-based observations a low-cost and effective choice.

Ground-based observational data have been used in machine learning over the past few years, especially in predicting meteorological variables. The need for guidance for the community from such a model is mainly precipitation and temperature. For temperature predictions, significant advances have been made in hourly (Anjali et al., 2019; Hou et al., 2022), daily (Sari et al., 2022) and monthly (Hou et al., 2023) predictions. However, the prediction of precipitation is challenging due to spatial and regional inhomogeneity and unpredictable seasonal variations, i.e., non-Gaussian nature.

The predictive capabilities of various deep-learning architectures have been recently tested in various settings. Yen et al. have used a Deep Echo State Network (DeepESN) to predict hourly precipitation in southern Taiwan, an area characterized by high variability between long dry periods and torrential rainfall events (Yen et al., 2019). Despite the positive value of root mean square error (RMSE) reported by the authors, the model could not forecast extreme precipitation events. Similarly, this constraint was also noted by Barrera-Animas et al. when a Bidirectional Long Short-Term Memory (Bi-LSTM) network was utilized to predict precipitation in five regions in the United Kingdom (Barrera-Animas et al., 2022). The nature of low total annual precipitation and rare high-intensity rainfall in the UK area results in a low RMSE. Still, the model could not predict extreme rainfall. Regarding the daily precipitation, Markuna et al. used a Random Forest model to predict precipitation in India, and Wang et al. improved a CNN-BiLSTM network with ICEEMDAN-VMD and attention mechanisms in China (Markuna et al., 2023; Wang et al., 2023). Ehteram et al. introduced a hybrid model that integrates CNN, Support Vector Regression (SVR), and Gaussian Process Regression (GPR) in Malaysia (Ehteram et al., 2023). All these studies showed positive results in predicting daily precipitation.

However, torrential precipitation still cannot be predicted, and one main reason is the different climatic characteristics between different regions. Such models face a challenge to be generalized. For example, those trained in regions with moderate mean precipitation cannot apply to regions with more extreme and variable rainfall patterns. Interestingly, the relatively stable annual precipitation and the low number of zero-precipitation days in Malaysia allowed more accurate modeling of precipitation dynamics.

A number of recent studies have discussed the trend in precipitation prediction at monthly scales. The empirical study carried out in China shows that such analysis provides positive results. Empirical studies conducted in China have shown that such analyses can yield positive results, as evidenced by the periodic nature of monthly rainfall and the associated model learning (Li et al., 2022; Zhang et al., 2023). Conceptually, the design and testing of precipitation forecasting systems using ground-based observational data

should be carried out with consideration of the unique climatic characteristics and precipitation patterns of the region of interest. Significantly, the choice of models for different regions should be carefully inspected, especially the extreme rainfall areas.

To address the ways to tackle the highly imbalanced data, especially precipitation, the techniques for imbalanced regression problems are introduced here. In the framework of imbalanced regression applications, researchers emphasize the necessity of thorough data preprocessing and model calibration. As opposed to classification scenarios where the imbalance in the classes can be addressed by oversampling or undersampling, the regression results are continuous. This makes the identification and correction of rare cases more complex. Therefore, Yang et al. proposed Label Distribution Smoothing (LDS) and Feature Distribution Smoothing (FDS). These techniques minimize the errors in predictions by regulating the target value distribution (Yang et al., 2021). In parallel, Kavianpour et al. showed that the addition of the concept of zero-order hold (ZOH) to a CNN-BiLSTM model with an attention mechanism significantly increased the accuracy of forecasting monthly earthquake occurrences by reducing the impact of zero values in the data (Kavianpour et al., 2023).

On the architecture of the model, the optimization of the loss function has become a viable approach to improving the prediction of extreme values. Steininger et al. have introduced the Dense Loss function that uses kernel density estimation to build a weighting scheme based on the target-value distributions and outperformed other methods like SMOGN (Steininger et al., 2021; Branco et al., 2017). Ribeiro et al. proposed SERA, a parameter-free method that uses an association function to assign weights to improve predictive performance (Ribeiro and Moniz, 2020; Silva et al., 2022). Kou et al. argued that SERA ignores intrinsic data characteristics in similar target ranges. They suggested the weighting scheme in SERA can be improved either by refinements of the association function of SERA or by inverse normalization in combination with kernel-density-based weighting (Kou and Fu, 2023). These improvements enhance flexibility to various data characteristics, thus increasing learning efficiency and accuracy of predictions. Despite these developments, further exploration is needed to develop the forecasting of rare yet significant events, such as hazard weather prevention, that lies in weather forecasting.

Machine Learning can be applied in different research areas that need analysis from a large set of data. For example, Basirian et al. (2026) integrated the satellite-based remote sensing with AI models to monitor the water quality. Amiranipour et al. (2026) utilize the drought indices with machine learning to provide early drought warning. However, precipitation is a non-Gaussian variable that requires special treatment in its prediction. In areas where rainfall distributions are very irregular, such as in Taiwan, there is an ongoing challenge because zero rainfall values can produce a significant effect on the overall model performance. This study aims to alleviate the inadequacy of surface observation data in predicting heavy rainfall by developing a model that increases the accuracy of the prediction, especially in regions with uneven rainfall. The study develops an optimized regression model by systematically analyzing hourly and daily precipitation data in the cities of Taipei, Keelung, and Tainan in Taiwan. The results show that the proposed model maintains high accuracy for light rainfall and substantially improves the prediction of extreme precipitation events.

2. Datasets and preprocessing

2.1. Sources and description

The data used in the examination of this research was downloaded through the Climate Observation Data Inquire Service of the Taiwan Central Weather Administration (CODiS Climate Observation Data Inquire Service, 2024). Taiwan is situated in a subtropical monsoon climate region and has an average rainfall three times the average in the whole world, and it is exposed to frequent extreme rainfall incidences. These attributes make the island an ideal location for studying heavy rainfall forecasting.

The meteorological stations with different geographic and climatic characteristics were selected to examine climate performance under different precipitation regimes as shown in Fig. 1. The capital Taipei is characterized by strong urban heat island effects due to the high degree of urbanization. Keelung is a coastal city that has plenty of rainfall all year round and it is one of the rainiest areas in Taiwan. Tainan, not during the summer, is comparatively arid with large seasonal variations in rainfall. The existence of such a variety of climates allows determining the proposed method in various precipitation conditions.

In order to further demonstrate the geographic and climatic diversity of the chosen stations, an elevation map of Taiwan is presented in Fig. 2. Taiwan has a high level of topographic variability with low-lying coastal plains and a central mountain range of over 3000 m. Such differences in elevation exert a powerful orographic effect on precipitation especially during monsoon and typhoon conditions. The spatial distribution of the sampling points in various topographical and climatic conditions is clearly illustrated by the overlaying of the meteorological stations on the elevation map, which contributes to assessing the model performance during various precipitation regimes.

2.2. Data preprocessing

An extensive data set was created by taking meteorological observations on an hourly and daily basis of every station. The hourly data comprises five years of data (2019 through 2023), which is greater than 43,824 observations based on 17 quantitative variables, and the daily data has 20 years of data (2004 through 2023) with 7305 observations and 35 features.

After preprocessing, non-numerical variables were removed. The hourly data therefore had 16 numerical variables left: station pressure (hPa), sea level pressure (hPa), temperature (°C), dew point temperature (°C), relative humidity (percent), wind speed (m/s), wind direction (°), maximum gust speed (m/s), maximum gust direction (°), precipitation (mm), precipitation duration (hours), sunshine duration (hours), global solar radiation (J/m2), visibility (km), ultraviolet index, and total cloud cover (percent).

Fig. 1. Locations of the meteorological stations (CODiS Climate Observation Data Inquire Service, 2024).

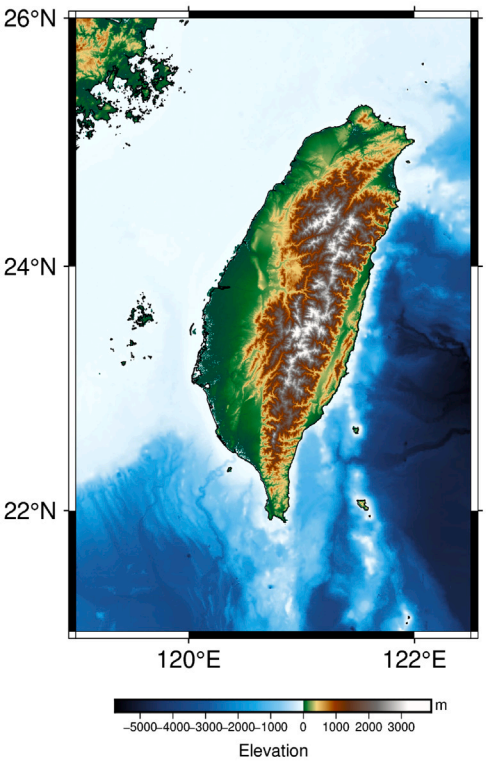


Fig. 2. Elevation map of Taiwan.

In the daily data, the 16 hourly variables were complemented by the following nine extra measures: maximum station pressure (hPa), minimum station pressure (hPa), maximum temperature (°C), minimum temperature (°C), minimum relative humidity (%), maximum 10-minute precipitation (mm), maximum 60-minute precipitation (mm), sunshine rate (%), A-type evaporation (mm). The ultraviolet index was restructured as the daily maximum ultraviolet index, which added 25 features of daily characteristics. Since A-type evaporation is not recorded in Tainan, there are 24 attributes in the daily set of the station.

The bidirectional interpolation method was used to fill in the missing data, which maintained continuity of data. About 70%–90% of hourly records and 50%–80% of daily records were recorded as zero precipitation (Fig. 3). The percentages of such records differed among stations with the lowest percentage reported in Keelung, then followed by Taipei and Tainan. Although with less pronounced rainfall recorded, Tainan recorded a high precipitation of 500 mm, which is as in the other stations.

In order to avoid distortion of the model due to large values, we use min–max normalization to scale each of the features to [0, 1] range, which facilitates both fairness and efficiency during training. The dataset is then split into training, validation and test subsets in proportion of 2:1:1. Since the validation set is used to optimize hyperparameters, its size has to be large enough to

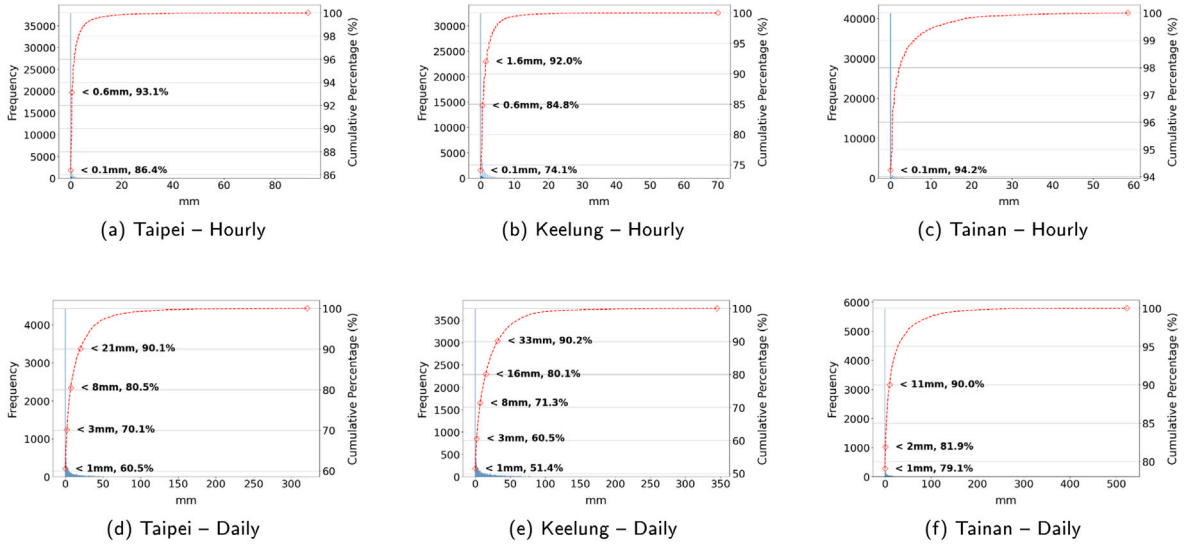


Fig. 3. Cumulative distribution functions of hourly and daily precipitation for Taipei, Keelung, and Tainan.

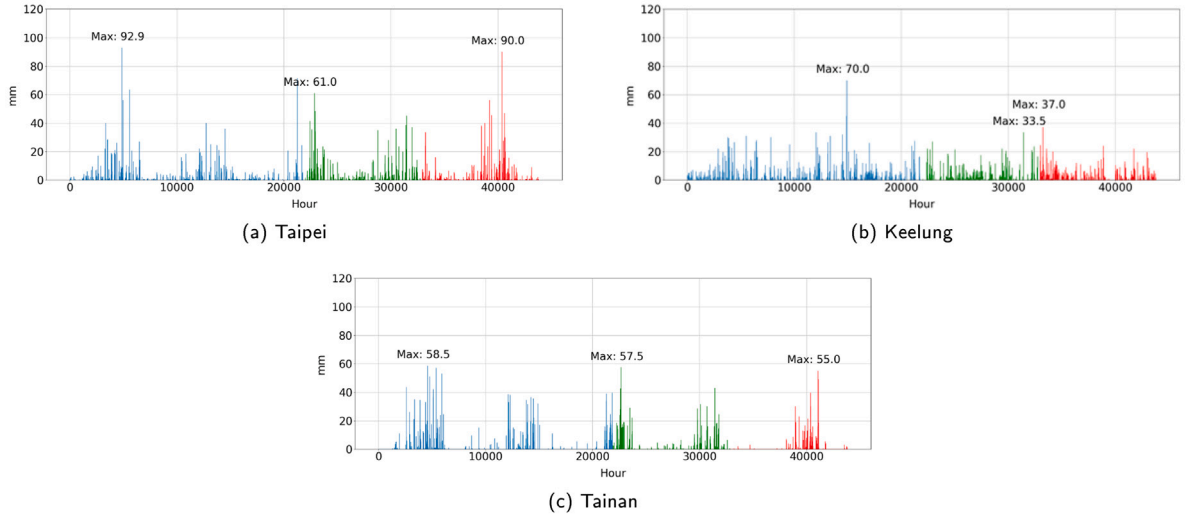


Fig. 4. Hourly dataset splits for Taipei, Keelung, and Tainan: training (blue), validation (green), and test (red).

properly capture the diversity of the dataset, particularly the occurrence of extreme rainfall events. Figs. 4 and 5 show the hourly and daily splits of Taipei, Keelung and Tainan respectively.

3. Machine learning models

In order to evaluate the effectiveness of the suggested method, the empirical studies were conducted using a wide range of representative models. Those models included a Support Vector Regression (SVR) model, two tree-based models, Random Forest Regression and Extreme Gradient Boosting (XGBoost), and two deep-learning models, Stacked Long Short-Term Memory (LSTM) and Bidirectional LSTM. All these models provided a comprehensive assessment and verification of the proposed method.

The SVR and Random Forest (RF) were selected as the representative baseline models of machine learning because they are commonly used in hydrological and meteorological prediction problems. SVR is a regression model which builds a margin-based decision function that balances the model between prediction accuracy and regularization, and supports the use of nonlinear relationships based on feature mapping by the use of a kernel. RF, on the other hand, is a tree-based ensemble technique that averages the predictions of a set of regression trees constructed through random sampling of data and input variables, which makes it more robust and generalized and less adversed to variance. Collectively, these two models can be viewed as complementary in

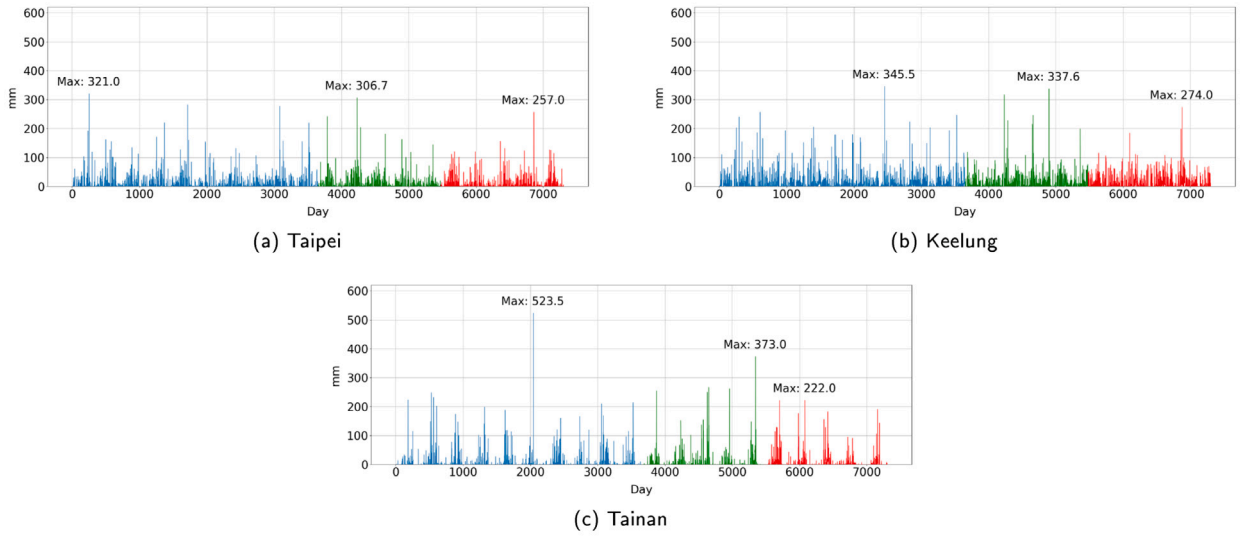


Fig. 5. Daily dataset splits for Taipei, Keelung, and Tainan: training (blue), validation (green), and test (red).

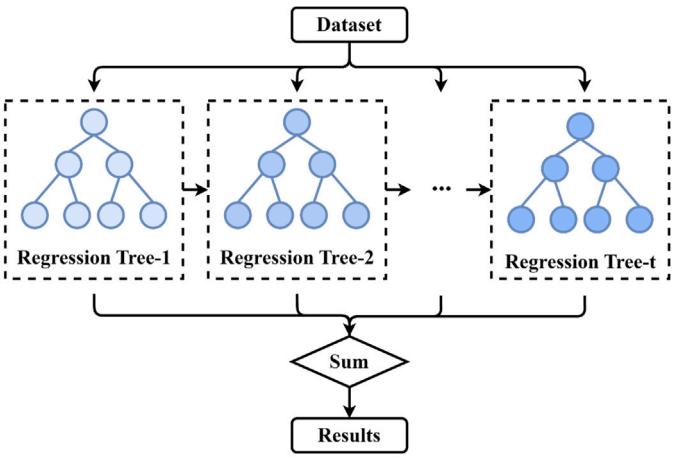


Fig. 6. Extreme Gradient Boosting.

that SVR focuses on the control of margins and regularity within the framework of shallow learning, and RF makes use of ensemble averaging to represent nonlinear relationships in high-dimensional data. This inclusion can provide a justifiable comparison with more complicated models, including XGBoost and deep learning structures, in cases of uneven rainfall. The detailed architectures and learning mechanisms of XGBoost and LSTM-based models are described in the following subsections.

3.1. Extreme gradient boosting

XGBoost (Chen and Guestrin, 2016) is a Gradient Boosting Decision Trees (GBDT) based ensemble learning algorithm, which is applicable in classification and regression tasks. This is in contrast to Random Forest where each tree is built separately, whereas in XGBoost the trees are built sequentially, minimizing the residuals that are left after a previous tree has been fitted, and hence reducing the error in the prediction, and increasing the accuracy (see Fig. 6). XGBoost also uses quantile sketch to efficiently split nodes, and considers sparse data, and also uses a caching mechanism to improve computational efficiency, making it especially well-suited to large-scale data.

In the context of XGBoost, the loss function $L^{(t)}$ is used to represent the total loss after building the tth tree. This is a function of an error term and regularization term as shown in (1).

$$L^{(t)} = \sum_{i=1}^{N} \left[l(y_i, \hat{y}_i^{(t)}) + \Omega(f_t) \right] \tag{1}$$

The error refers to the difference between the true value y_i and the predicted value $\hat{y}_i^{(t)}$ after the tth iteration. The new tree f_t is supposed to improve the residuals of the previous model, so that the expression $\hat{y}_i^{(t)}$ can be rewritten as $\hat{y}_i^{(t-1)} + f_t(x_i)$. A Taylor expansion is assumed to simplify the calculations in which only the first-order g_i and second-order h_i derivatives are used and constant terms are ignored (2).

$$\begin{aligned}\sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t)}) &= \sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) \\ &\approx \sum_{i=1}^{N} \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] \\ &\approx \sum_{i=1}^{N} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right]\end{aligned} \tag{2}$$

The regularization term $\Omega(f_t)$ in XGBoost consists of the number of leaf nodes T and the output values υ, which are all regulated by the regularization parameters γ and λ, to prevent overfitting (3).

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \|\upsilon\|^2 \tag{3}$$

The data points are mapped using a mapping function q to a leaf of the tree f_t. The collection of data points of the jth leaf node is called I_j (4). The simplified error and regularization terms can be added to get the optimal loss function (5). Optimal weight and score of each leaf node is then calculated according to the quadratic optimization formula (6).

$$I_j = \{ i \mid q(x_i) = j \}, \quad j = 1, 2, \ldots, T \tag{4}$$

$$\begin{aligned}L^{(t)} &= \sum_{i=1}^{N} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \left[\gamma T + \frac{1}{2}\lambda \|\upsilon\|^2 \right] \\ &= \sum_{i=1}^{N} \left[g_i \upsilon_q(x_i) + \frac{1}{2} h_i \upsilon_q^2(x_i) \right] + \left[\gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \upsilon_j^2 \right] \\ &= \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) \upsilon_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \upsilon_j^2 \right] + \gamma T\end{aligned} \tag{5}$$

$$\begin{aligned}\upsilon_j^* &= -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \\ L^{(t)}(q) &= -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T\end{aligned} \tag{6}$$

In the context of XGBoost, the node splitting algorithm is executed by implementing a greedy or an approximation algorithm that identifies the best split point that can maximize the gain. The gain is usually defined as the arithmetic difference between the score of the left and right branches after the split and the score of the whole node before the split, minus the complexity cost γ of adding a new branch (7).

$$L_{\text{split}} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \tag{7}$$

XGBoost is constructed by a sequence of decision trees and the key hyperparameters are the number of trees, the maximum depth of the tree, the minimum numbers of samples per split and the number of features taken into account during each split. The trees are sequentially trained to correct errors of the previous trees and that is why it is computationally intensive as compared to independent tree models. More boosting rounds and deeper trees enable the model to acquire complex patterns, however, at the cost of overfitting. This complexity can be controlled using regularization parameters and efficient node splitting and sparse data processing minimizes computation time. In general, XGBoost is an effective and versatile method of both regression and classification, which can operate with large and high-dimensional datasets.

3.2. Long short-term memory

LSTM (Hochreiter and Schmidhuber, 1997) (see Fig. 7) is a deep-learning model based on Recurrent Neural Networks (RNN) and designed to resolve vanishing and exploding gradient problems faced by RNNs in the case of long sequences. Due to this ability, LSTM is regularly used in time-series analysis, language modeling, and speech recognition, where its ability to maintain reliable predictions over long periods of time is of primary concern.

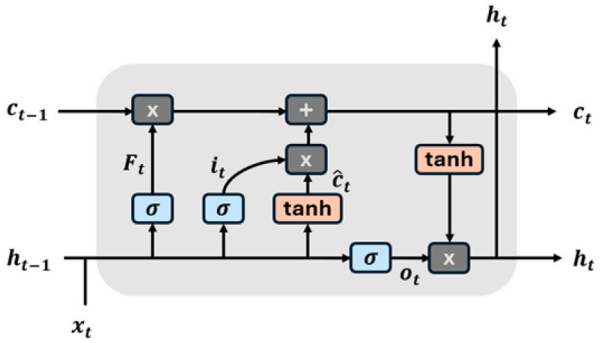


Fig. 7. Long Short-Term Memory Network.

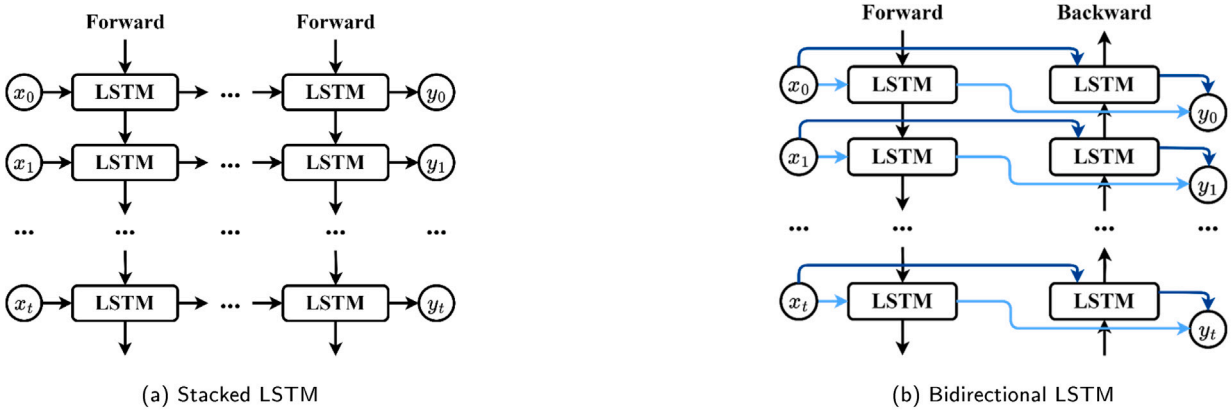


Fig. 8. Types of Long Short-Term Memory Network.

In LSTM, the cell state c_t stores and hands the information to the next time step, and the hidden state h_t stores the domain-specific prediction at time step t. The main operation is based upon three gating units, namely, forget gate F_t, input gate i_t, and output gate o_t, which control the flow of information within the cell and enhance the performance of the model when working with complex sequential data (8) and (9).

$$c_t = F_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t, \quad h_t = o_t \cdot \tanh(c_t) \tag{8}$$

$$\begin{aligned} F_t &= \sigma(w_f \cdot [h_{t-1}, x_t] + b_F) \\ i_t &= \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{c}_t &= \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \\ o_t &= \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \end{aligned} \tag{9}$$

To improve performance, the LSTM can be augmented to two versions, Stacked LSTM (S-LSTM) and Bidirectional LSTM (Bi-LSTM) (Hou et al., 2023). as shown in Fig. 8. S-LSTM enhances the feature extraction process gradually through stacking layers, but Bi-LSTM uses information flowing in both directions to increase the accuracy of the model.

LSTM networks are made up of memory cells which keep a state of the cell and a hidden state, and three gates-forget, input and output- which determine what is stored, added or produced at each step. Some of the key parameters that influence the complexity of the model are the number of layers, the number of hidden units in a layer, and stacked or bidirectional network. Stacked LSTMs tend to enhance the complexity of feature extraction whereas bisirectional LSTMs tend to process sequences on both directions effectively adding more parameters and computational cost. The dimension of the hidden state, as well as the length of the input sequences, also affect the trainable parameters and the memory usage. Generally, LSTM and its variants are effective in dealing with sequential data, as they offer the same capability to learn long-term dependencies and the complexity is easy to handle.

4. Model optimization

To enhance predictive accuracy, we used a set of objective functions to address imbalanced regression tasks. To have an overall understanding of the model performance in various aspects the model was trained using Dense Loss, Squared Error Relevance Area

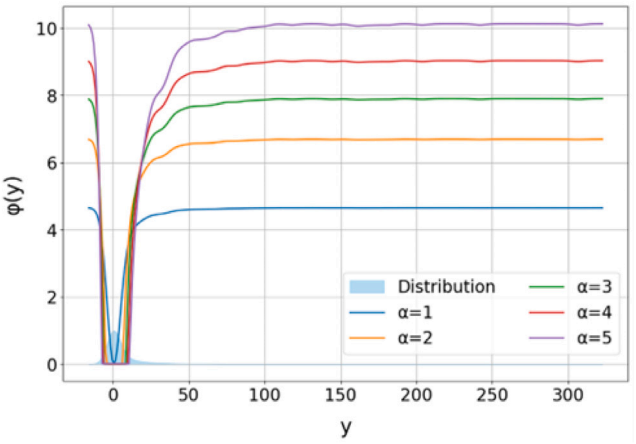


Fig. 9. Effect of different α values on the weight function in Dense Loss.

(SERA) and two variants of Adapted Squared Error Relevance (ASER), and the result was compared with the standard Mean Squared Error (MSE). In addition, Zero-Order Hold (ZOH) preprocessing was also applied to avoid prediction errors in case of heavy rainfalls.

4.1. Dense loss

Dense Loss (DENSE) (Steininger et al., 2021) is a learning algorithm specifically aimed at improving the results on imbalanced data, by giving increased weight to rare cases. The process can be outlined in the following way:

4.1.1. Kernel density estimation

Use Kernel Density Estimation (KDE) to describe the empirical distribution of target values with a Gaussian kernel K, and bandwidth h parameter selected in line with the Silverman rule of thumb (10).

$$p(y) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{y - y_i}{h}\right) \tag{10}$$

4.1.2. Normalization of the density function

Normalize the resulting density to get an interpretable measure (11).

$$p'(y) = \frac{p(y) - \min(p(Y))}{\max(p(Y)) - \min(p(Y))} \tag{11}$$

4.1.3. Weight function definition

The weight function, $f_w(\alpha, y)$, raising the weight of rare data points as α grows, is defined and shown in Fig. 9. A small constant ϵ is added to avoid the zero weight assignments on common instances (12).

$$f_w(\alpha, y) = \frac{\max(1 - \alpha p'(y), \epsilon)}{\frac{1}{N}\sum_{i=1}^{N} \max(1 - \alpha p'(y), \epsilon)} \tag{12}$$

4.1.4. DENSE calculation

Dense Loss depends on the weight function f_w and the chosen evaluative measure M which is usually MSE to include the total disparity of the weighted data and the real values (13).

$$\text{DenseLoss}(\alpha) = \frac{1}{N}\sum_{i=1}^{N} f_w(\alpha, y) \cdot M(y_i, \hat{y}_i) \tag{13}$$

4.2. Squared error relevance area

Squared Error Relevance Area (SERA) (Ribeiro and Moniz, 2020; Silva et al., 2022) is a cost-sensitive learning algorithm based on the concept of relevance and provides a non-parametric measure of imbalanced data. The calculation is performed as follows:

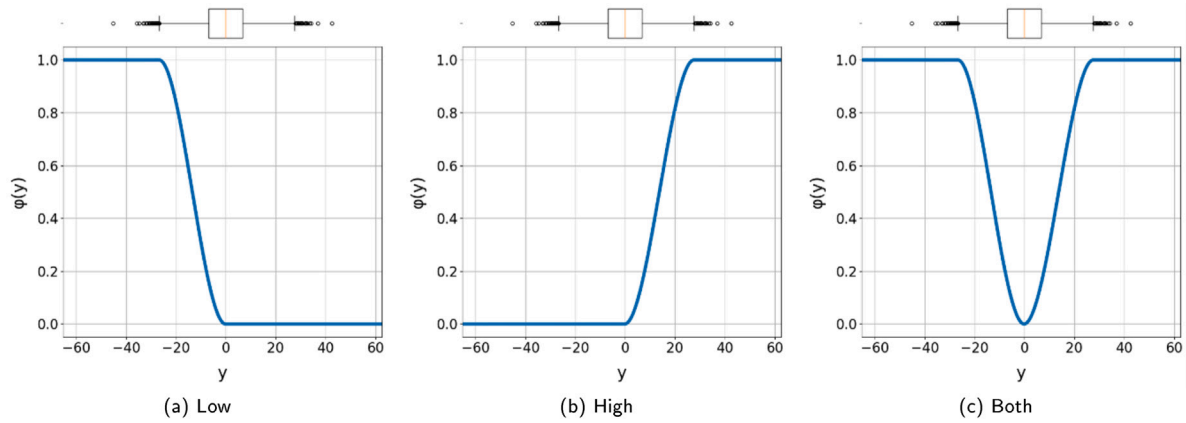


Fig. 10. Types of Squared Error Relevance Area.

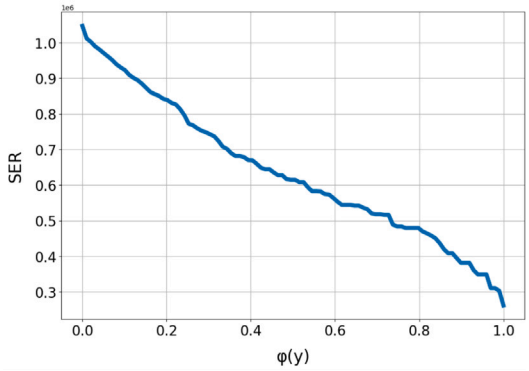


Fig. 11. Squared Error Relevance Area.

4.2.1. Control point identification

The control points are established through a box plot, which is supplemented by the Medcouple technique (Hubert and Vandervieren, 2008) to take into consideration both variability and skewness to support outliers (14). The Medcouple value, MC, is applied to adjust the lower and upper values against the interquartile range (IQR) (15). The interquartile range is the difference between the third quartile (Q_3) and the first quartile (Q_1) that defines the range of the data containing the middle 50% of data.

$$MC = \frac{(y_j - Q_2) - (Q_2 - y_i)}{y_j - y_i}, \quad \text{for } y_i \le Q_2 \le y_j \tag{14}$$

$$\begin{aligned} &\text{if } MC \ge 0, \quad [Q_1 - 1.5e^{-4MC} IQR,\ Q_3 + 1.5e^{3MC} IQR] \\ &\text{if } MC < 0, \quad [Q_1 - 1.5e^{-3MC} IQR,\ Q_3 + 1.5e^{4MC} IQR] \end{aligned} \tag{15}$$

4.2.2. Relevance function definition

The above obtained control points are then interpolated with a Cubic Hermite Interpolating Polynomial to remap the target values into the range of [0, 1]. The resulting relevance function can focus on the low, high, or both values depending upon the properties of the data (see Fig. 10). In this research, both extremes are considered to improve the prediction of heavy and light rainfall.

4.2.3. SERA calculation

SERA is calculated as the squared error over the relevance function ϕ and is then combined with Mean Squared Error (MSE) based on the formula in (16), which makes the metric especially useful with imbalanced data sets and reduces prediction errors in extreme values without the need of further parameter optimization (see Fig. 11).

$$\begin{aligned} &\text{SERA} = \int_0^1 \sum_{i \in D^t} (y_i - \hat{y}_i)^2 \, dt, \\ &D^t = \{\langle x_i, y_i \rangle \in D \mid \phi(y_i) \ge t\} \end{aligned} \tag{16}$$

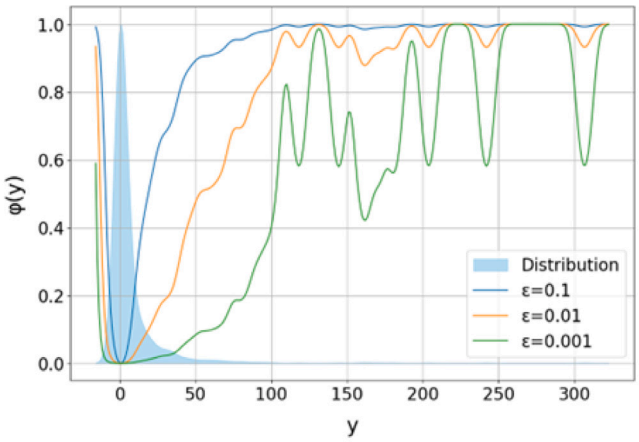


Fig. 12. Effect of different ϵ values on the weight function in Adapted Squared Error Relevance.

4.3. Adapted squared error relevance

Kou and Fu (2023) proposed Adapted Squared Error Relevance (ASER), which aims to improve the SERA method by considering asymmetries in distributions of data even in cases when identical target variables are used. Two different mechanisms are suggested to assign data-point weights in a more accurate manner.

4.3.1. Method 1: Exponential weighting

The initial approach provides a boost to the relevance feature of SERA. A relevance value greater than a threshold t_w is multiplied by a weight factor k and exponentially scaled, and its impact on the calculation is increased (17):

$$w'(y_i) = \frac{\exp[k(\phi(y_i) - t_w)]}{\frac{1}{N}\sum_{i=1}^{N}\exp[k(\phi(y_i) - t_w)]} \tag{17}$$

4.3.2. Method 2: KDE-based weighting

The second one uses Kernel Density Estimation (KDE) to estimate a probability density function $p(y)$ of the target variable. The result is then normalized and adjusted by a small constant ϵ to avoid numerical instability in the case that the denominator is zero (18). This normalized density is re-normalized to give the new relevance function $\phi(y)$ (19). The strength of weighting can be varied by the choice of ϵ, with noticeable deviations of the weighting to SERA (see Fig. 12), with the weighting function being given in (20).

$$p'(y) = \frac{p(y) - \min(p(Y))}{\max(p(Y)) - \min(p(Y))} + \epsilon \tag{18}$$

$$\phi(y) = \frac{\frac{1}{p'(y)} - \min\left(\frac{1}{p'(Y)}\right)}{\max\left(\frac{1}{p'(Y)}\right) - \min\left(\frac{1}{p'(Y)}\right)} \tag{19}$$

$$w'(y_i) = \frac{1 + k/p(y_i)}{\frac{1}{N}\sum_{i=1}^{N}\left(1 + k/p(y_i)\right)} \tag{20}$$

4.3.3. ASER calculation

The calculation of ASER using these two weighting schemes is possible as indicated in (21) using their respective relevance functions. These two techniques will be referred to as ASER-1 and ASER-2 respectively in this paper to make it convenient. Using the same data as SERA, results plots (see Fig. 13) show that ASER provides more detailed information about the heterogeneity of data than SERA does, although the resulting curves have less smooth behavior than those obtained under SERA on the same data.

$$\begin{aligned} \text{ASER} &= \int_0^1 \sum_{i \in D^t} w_i'(y_i - \hat{y}_i)^2 \, dt, \\ D^t &= \{\langle x_i, y_i\rangle \in D \mid \phi(y_i) \geq t\} \end{aligned} \tag{21}$$

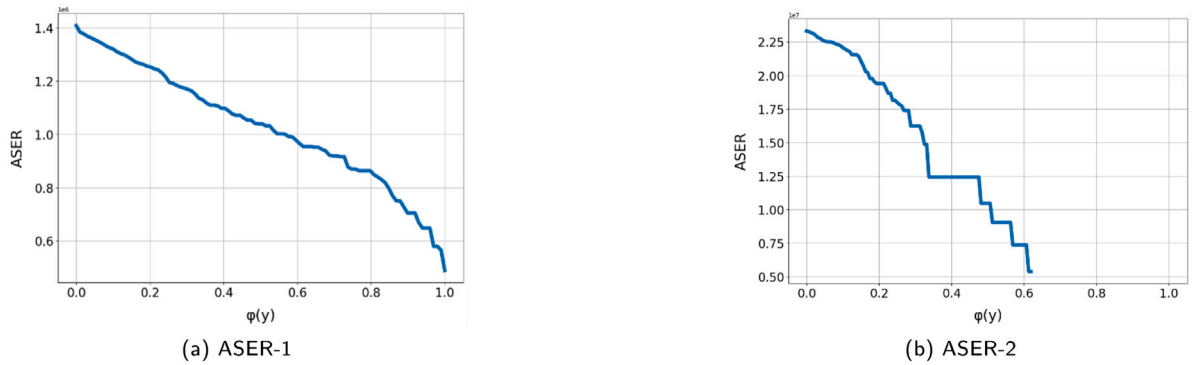


Fig. 13. ASER-1 and ASER-2 Results.

4.4. Zero-order hold

Zero-Order Hold (ZOH) (Moir, 2022) is a typical signal processing technique that transforms discrete signals to continuous ones and maintains stability within the sampling duration (see Fig. 14). In this study, we generalized the idea of ZOH to data preprocessing using the algorithm used by Kavianpour et al. (2023) to predict monthly earthquake events in China.

To be more precise, in the training set we replaced the observation of zero rainfall with the previous non-zero observation to ensure that we did not have too many zero observations in precipitation data. Such a calibration can help the model identify extreme rain events better. This technique is especially useful in continuous variables that have a high proportion of zeroes, like earthquake occurrences and rainfall, and is beneficial when compared to conventional preprocessing methods, especially in improving the accuracy of predictions of extreme events.

It should be noted that this zero-rainfall replacement strategy is applied only to the training data, while the validation and test datasets remain unchanged to ensure unbiased model evaluation. The ZOH-based preprocessing does not generate artificial rainfall events but instead preserves temporal continuity in the learning process when zero values dominate the series. Alternative approaches, such as replacing zero values with a small constant (e.g., 0.001), were not adopted because they introduce artificial precipitation signals and may bias cumulative rainfall statistics and extreme-event learning. Therefore, the use of ZOH in this study is intended to enhance training stability and address data imbalance, rather than to alter the physical interpretation of precipitation observations.

Moreover, regarding the concern that replacing zero rainfall with a previous extreme value may contradict real-time precipitation variability, we observe that heavy precipitation events in Taiwan are usually inherently persistent and event-driven (e.g. when typhoons pass by), as opposed to isolated instant events. As such, it is physically possible to continue rainfall intensity in neighboring time steps. In additional efforts to guarantee that the ZOH preprocessing does not unnaturally enhance extreme rainfall, a series of robustness tests were done and it was determined that the approach mainly enhances the stability of learning and underrepresentation of severe rainfall without artificially inflating the extreme values.

5. Experimental methods

5.1. Methodology

The workflow of the study is presented in Fig. 15 and corresponds to preprocessing six datasets described in Chapter V. These data include hourly and daily weather records taken by the Taipei, Keelung and Tainan weather stations. The Optuna framework is then used to optimize five learning models, SVR, RF, XGBoost, S-LSTM, and Bi-LSTM as discussed in Chapter III. The optimization procedure targets the loss functions mentioned in Chapter IV: DENSE, SERA, ASER-1, ASER-2, and the baseline MSE, and all models use MSE as the main loss function.

In the case of DENSE objective function, 0.001 is used as the value of the parameter ϵ, and the weighting factor α is varied between 1 and 5. The model that performs best on the validation set in terms of predicting heavy rainfall is assigned as the final outcome of DENSE. Similarly, in the case of ASER-1 and ASER-2, the weighting factor k is optimized between 1 and 5 and the model with the best results is selected.

The five models and the five loss functions give a total of 25 results per dataset. MSE is utilized as the objective measure to determine the effectiveness of these results. The study also observes whether the non-MSE results of 20 are better than MSE and estimates the station-specific performance of the five loss functions of the model based on a model-based perspective.

Another series of experiments are performed to generate more results. In this case, Zero-Order Hold (ZOH) method is used in preprocessing, when the value of precipitation in the training set equal to zero is substituted with the preceding non-zero value. The models are retrained using the identified best hyperparameters in the previous experiments and the performance of the models with ZOH is compared with that of the models without ZOH.

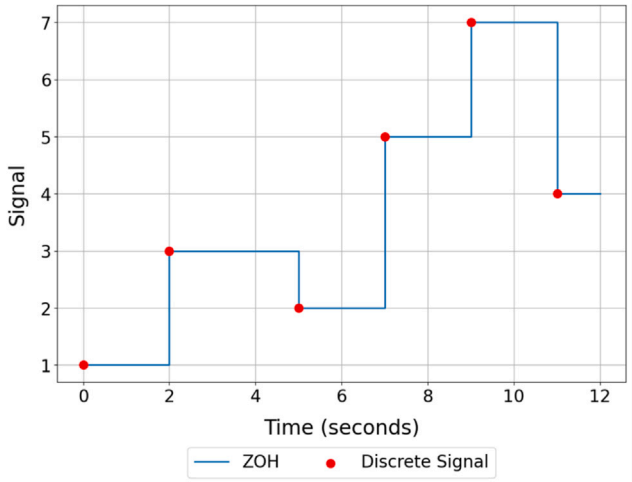


Fig. 14. Traditional Zero-Order Hold.

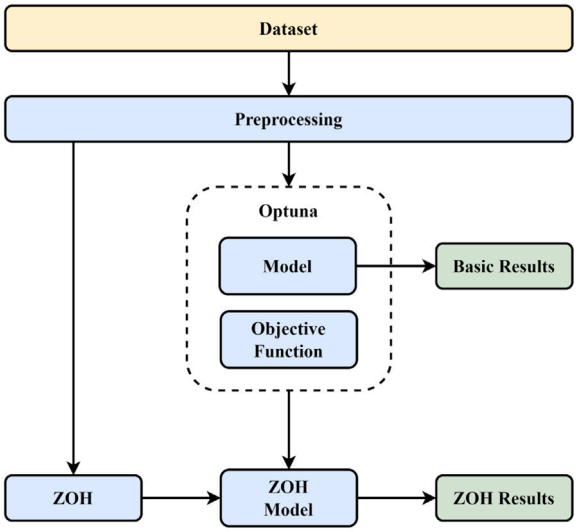


Fig. 15. Experimental workflow.

This workflow aims to explore how different loss functions and machine-learning models affect the accuracy of predictions, and at the same time assess the usefulness of ZOH in solving the problem of data imbalance. In the process, the process provides more credible methods of predicting highly unbalanced phenomena- in this case, precipitation.

5.2. Experimental setup

The experiments were performed using an Ubuntu 22.04.3 LTS operating system. The hardware configuration included 13th Gen Intel(R) Core(TM) i7-13700 and dual-channel DDR4 3200 32 GB memory, where the Support Vector Regression (SVR), Random Forest Regression (RF), and eXtreme Gradient Boosting (XGBoost) models were run. To speed up computations, an NVIDIA GeForce RTX 4080 GPU with 16 GB memory was used in the Stacked Long Short-Term Memory (S-LSTM) and Bidirectional LSTM (Bi-LSTM) models.

Model development was done using Python 3.10.12. The SVR and RF models were applied to the `scikit-learn` 1.3.0 package, and the XGBoost model was created using the `xgboost` 1.7.6 package. S-LSTM and Bi-LSTM architectures were trained on deep learning models with the CUDA 12.2 environment, which runs on `tensorflow` 2.11.1 and `keras` 2.11.1.

Table 1
SVR hyperparameter search space.

Hyper-parameter	Type	Range	Notes
kernel	Categorical	rbf, poly	Kernel function
gamma	Categorical	scale, auto	Kernel width
	Float	10^{-5}–10^{0} (log step)	
C	Float	10^{-3}–10^{3} (log step)	Penalty factor
epsilon	Float	10^{-3}–10^{0} (log step)	Tolerance margin
degree	Integer	1–5	Degree of polynomial kernel
coef0	Float	−5–5	Constant term in polynomial kernel
max_iter	Fixed Value	10,000	Maximum iterations

Table 2
RF hyperparameter search space.

Hyper-parameter	Type	Range	Notes
n_estimators	Integer	1–1000	Number of trees
max_features	Integer	1–10	Max features for splitting
max_depth	Integer	1–30	Max depth of each tree
min_samples_ split	Integer	2–100	Min samples required to split
min_samples_ leaf	Integer	1–20	Min samples per leaf node
max_samples	Float	0.1–1	Max sample proportion per tree

Table 3
XGBoost hyperparameter search space.

Hyper-parameter	Type	Range	Notes
n_estimators	Integer	100–1000 (step=50)	Number of trees
max_depth	Integer	3–10	Max depth of each tree
min_child_ weight	Integer	1–10	Min sum of weights per leaf
gamma	Float	0–5 (step=0.1)	Min loss reduction for split
subsample	Float	0.5–1.0 (step=0.1)	Subsample ratio of training instances
colsample_ bytree	Float	0.5–1.0 (step=0.1)	Feature subsampling ratio
learning_rate	Categorical	10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 0.2, 0.3	Learning rate
reg_alpha	Float	0–5 (step=0.1)	L1 regularization term
reg_lambda	Float	0–5 (step=0.1)	L2 regularization term

Table 4
S-LSTM and Bi-LSTM hyperparameter search space.

Hyper-parameter	Type	Range	Notes
optimizer	Categorical	SGD, Adam	Optimizer
learning_rate	Float	10^{-5}–10^{-2} (log step)	Learning rate
batch_size	Integer	64–2048	Batch size
epochs	Integer	30–100	Number of epochs
n_layers	Integer	1–4	Number of layers
units	Integer	32–512	Units per layer, adjustable per layer

5.3. Hyperparameter optimization

We used Optuna, open-source hyperparameter optimization tool specific to machine-learning models, in order to maximize the performance of the model (Akiba et al., 2019). Optuna makes the process much easier since all the user needs to do is specify the search space and objective function; the system will automatically find the optimal hyperparameters.

We used a fixed random seed of 1226 in every Optuna run and performed 100 search trials to stabilize the values in order to maintain consistency and fairness across experiments. Specifications of hyperparameter search space and the specifications of each model are provided in Tables 1, 2, 3, and 4. These hyperparameters are categorized into three categories, which include Categorical, Integer, and Float. In the case of Integer types, the step value is not given defaultly, the search continues with the step of 1. In Float types the search is done on a continuous range, unless a step is passed. When a step value is logarithmic, the search is performed in logarithmic space, thus minimizing the search space and also efficiency. Fixed values represent predetermined settings that are used in the experiments.

5.4. Evaluation metrics

In order to evaluate the performance in different rainfall conditions, we used a rainfall classification scheme that consists of four categories as shown in Table 5. The Central Weather Administration of Taiwan only gives rainfall categories above heavy rain;

Table 5
Rainfall classification definitions.

Category	Hourly	Daily
Light	0 mm ≤ Rainfall < 2.5 mm	0 mm ≤ Rainfall < 10 mm
Moderate	2.5 mm ≤ Rainfall < 10 mm	10 mm ≤ Rainfall < 40 mm
Heavy	Rainfall ≥ 10 mm	Rainfall ≥ 40 mm

Table 6
Selected weight scaling factors for objective functions.

		DENSE		ASER-1		ASER-2	
		Hourly	Daily	Hourly	Daily	Hourly	Daily
Taipei	SVR	3	5	1	1	1	1
	RF	4	1	2	5	1	1
	XGBoost	3	3	2	1	1	1
	S-LSTM	4	1	1	1	1	1
	Bi-LSTM	2	5	4	3	1	1
Keelung	SVR	2	5	5	1	1	1
	RF	4	5	2	5	1	1
	XGBoost	4	5	1	1	1	1
	S-LSTM	3	4	2	5	1	1
	Bi-LSTM	2	1	1	4	1	1
Tainan	SVR	2	2	1	1	1	1
	RF	2	4	1	4	1	1
	XGBoost	3	4	2	3	1	1
	S-LSTM	1	2	3	3	1	1
	Bi-LSTM	1	1	1	1	1	1

thus, we have relied on the classification criteria of the People Republic of China to classify light rain. In the case of heavy rain, to overcome the problem of data scarcity, we used the cumulative distribution curves of rainfall at the three stations (see Fig. 16) to determine the inflection points as the thresholds. Fig. 17 shows the percentages of light, moderate and heavy rain in the training, validation and test sets of the three stations.

The performance will be measured through the Mean Squared Error (MSE) of each category of rainfall as given in (22). smaller MSE indicates better predictive accuracy.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{22}$$

6. Results and discussion

6.1. Comparison of objective functions

6.1.1. Weight analysis

DENSE, ASER-1 and ASER-2 involve modification of the weight scaling factor. Each function was evaluated with scaling factors of 1 to 5 and the one that yielded the best validation-set MSE under heavy rain conditions was chosen. When several scaling factors produced the same accuracies, the smallest factor was used. Table 6 gives the corresponding scale factors of each station and each function.

ASER-2 had consistently optimal hyperparameter results at any scaling value. Since the density function in question is very concentrated, i.e., see Eq. (20), its weight function is not very sensitive to changes in the scaling parameter. Thus, the factor of 1 was selected as the smallest one in ASER-2.

6.1.2. Results analysis

The performance of DENSE, SERA, ASER-1, and ASER-2 was tested in light, moderate, and heavy rainfall in the validation and the test sets. The percentage of cases when these covariate-adjustment techniques performed better than MSE is shown in Tables 7 and 8.

The findings indicate that no significant benefit over mean-squared error (MSE) criterion was found in the case of light rainfall. In moderate rainfall, there were modest improvements but they were not statistically significant. Conversely, in the case of heavy rainfall, nearly all combinations showed an improvement over MSE, in fact, many cases showed a 100% improvement.

Further examination of the regression models indicates that DENSE exhibited the highest overall stability, achieving consistent performance improvements in more than 85% of the experimental trials. SERA demonstrated strong predictive capability at the hourly scale, however, its performance was less consistent for daily sequences. ASER-1 marginally outperformed SERA in daily precipitation prediction, suggesting its suitability for larger temporal datasets. In contrast, ASER-2 provided the least improvement across all datasets and time scales, indicating limited effectiveness in capturing precipitation variability.

Fig. 16. Cumulative distribution functions (CDFs) of hourly and daily rainfall.

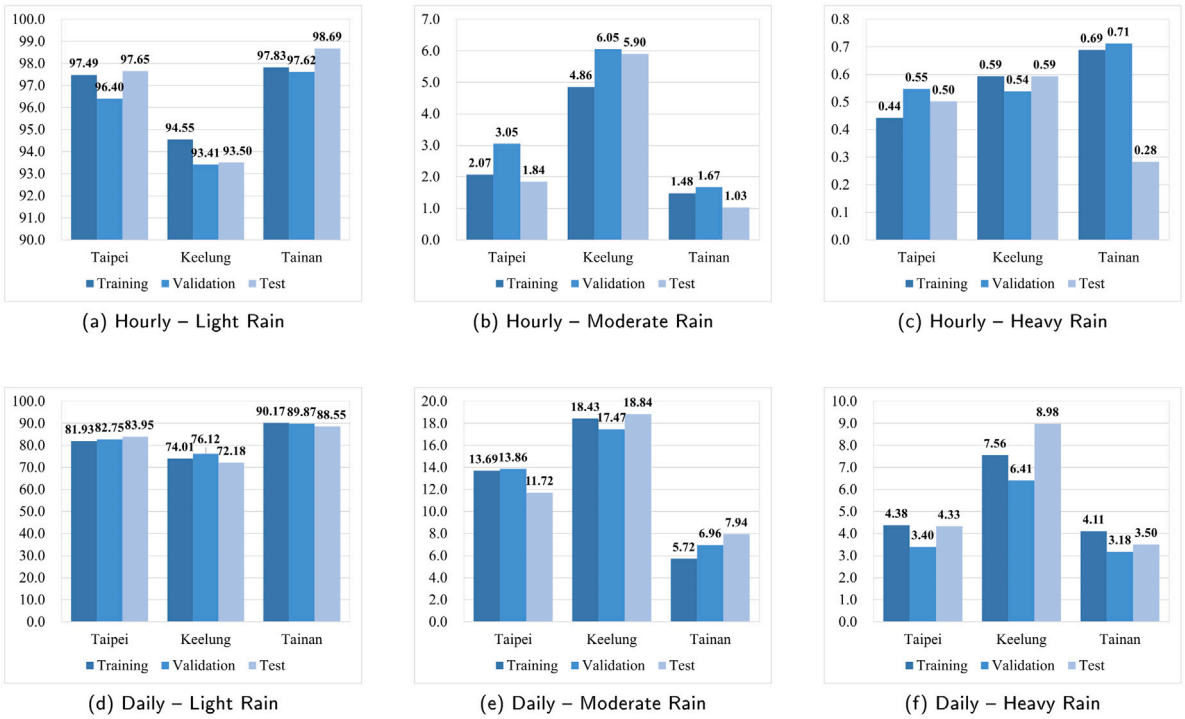


Fig. 17. Rainfall proportions in the dataset by category for hourly and daily time scales.

To sum up, DENSE is the most stable and effective objective function, followed by SERA and ASER-1. ASER-2 offered the minimum improvement. In order to give a complete analysis of the performance of each objective function compared to MSE, the mean changes of MSE are shown in Tables 9 and 10. Reductions in MSE are written in bold. The changes in MSE are also given as ratios in Tables 11 and 12 to further analyze the effect of the different stations on the objective functions, with reductions again being highlighted. It should be noted that MSE is used as the primary evaluation metric in this study to ensure consistent and fair comparison across all regression models and objective functions. Additional statistical indicators such as R, MAPE, Mean Bias Error (MBE), and Scatter Index (SI) could not be consistently recomputed for all models due to reproducibility limitations associated with stochastic learning algorithms under different computational environments. Nevertheless, the relative performance trends and comparative conclusions among the objective functions remain valid, as all models were evaluated under the same experimental protocol.

A critical analysis of hourly rainfall measurements indicates that DENSE, SERA and ASER-1 together provided an average of 30 improvement in mean-squared error (MSE) under heavy rainfall, but the improvement in MSE under light rainfall was relatively small, about 0.5. This analysis focuses on the relative improvement patterns across rainfall intensities rather than absolute error magnitudes, allowing a robust comparison of objective functions under highly imbalanced precipitation conditions. In a number of instances, ASER-2 produced the largest improvement in heavy rainfall, but significantly degraded performance in light rainfall, particularly in Taipei.

Table 7
Percentage of improvement compared to MSE (hourly data).

		Validation set				Test set			
		DENSE	SERA	ASER-1	ASER-2	DENSE	SERA	ASER-1	ASER-2
Taipei	Light	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Moderate	0.00	0.00	0.00	0.00	40.00	60.00	0.00	20.00
	Heavy	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Keelung	Light	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Moderate	60.00	60.00	40.00	0.00	60.00	60.00	60.00	20.00
	Heavy	**100.00**	**100.00**	**100.00**	**100.00**	80.00	**100.00**	**100.00**	80.00
Tainan	Light	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Moderate	20.00	20.00	20.00	0.00	20.00	20.00	40.00	0.00
	Heavy	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	80.00	40.00
Overall	Light	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Moderate	26.67	26.67	20.00	0.00	40.00	46.67	33.33	13.33
	Heavy	**100.00**	**100.00**	**100.00**	**100.00**	93.33	**100.00**	93.33	73.33

Table 8
Percentage of improvement compared to MSE (daily data).

		Validation set				Test set			
		DENSE	SERA	ASER-1	ASER-2	DENSE	SERA	ASER-1	ASER-2
Taipei	Light	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Moderate	40.00	20.00	20.00	20.00	40.00	0.00	0.00	0.00
	Heavy	**100.00**	**100.00**	**100.00**	80.00	80.00	**100.00**	**100.00**	60.00
Keelung	Light	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Moderate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Heavy	**100.00**	80.00	**100.00**	80.00	80.00	20.00	40.00	40.00
Tainan	Light	0.00	40.00	0.00	0.00	0.00	60.00	0.00	0.00
	Moderate	0.00	40.00	60.00	0.00	0.00	20.00	0.00	0.00
	Heavy	**100.00**	60.00	80.00	80.00	**100.00**	80.00	80.00	**100.00**
Overall	Light	0.00	13.33	0.00	0.00	0.00	20.00	0.00	0.00
	Moderate	13.33	20.00	26.67	6.67	13.33	6.67	0.00	0.00
	Heavy	**100.00**	80.00	93.33	80.00	86.67	66.67	73.33	66.67

Table 9
MSE change for hourly data.

		Validation set				Test set			
		DENSE	SERA	ASER-1	ASER-2	DENSE	SERA	ASER-1	ASER-2
Taipei	Light	0.64	0.68	0.56	7.21	0.50	0.45	0.30	7.12
	Moderate	2.18	2.39	3.62	14.73	0.54	0.44	1.48	6.01
	Heavy	**−52.29**	**−52.88**	**−45.37**	**−65.82**	**−43.64**	**−57.14**	**−60.01**	**−79.33**
Keelung	Light	0.35	0.42	0.55	0.90	0.29	0.35	0.47	0.75
	Moderate	**−0.23**	**−0.69**	0.10	2.63	**−0.52**	**−0.84**	**−0.45**	2.24
	Heavy	**−18.57**	**−14.86**	**−18.59**	**−17.32**	**−14.86**	**−10.47**	**−15.27**	0.00
Tainan	Light	0.44	0.56	0.61	0.73	0.33	0.43	0.48	0.59
	Moderate	3.16	2.54	1.63	10.41	2.19	2.99	1.48	11.07
	Heavy	**−33.17**	**−35.40**	**−28.44**	**−28.48**	**−5.69**	**−24.93**	**−14.94**	8.31
Overall	Light	0.48	0.55	0.57	2.94	0.37	0.41	0.42	2.82
	Moderate	1.70	1.41	1.78	9.26	0.73	0.86	0.84	6.44
	Heavy	**−34.67**	**−34.38**	**−30.80**	**−37.21**	**−21.40**	**−30.85**	**−30.07**	**−23.68**

The outcome of the daily rainfall records also indicates that DENSE exploits reliably better performance in both validation and test sets. SERA and ASER-1 are also satisfactory, and ASER-2 is less effective and sometimes increases errors, especially when heavy rainfall occurs in Tainan.

The comparison of the stations shows that Taipei is performing better than both Keelung and Tainan in both temporal resolutions. This fact can be easily explained by the different rainfall distributions presented in Fig. 17: in daily observations, Taipei has 16% to 18% moderate and heavy rainfall events, Keelung has 24% to 28%, and Tainan has only 10% to 12%. Objective functions therefore work best when the proportion of extreme events is in this moderate range, but performance deteriorates when the proportion is high, as in Keelung.

Table 10
MSE change for daily data.

		Validation set				Test set			
		DENSE	SERA	ASER-1	ASER-2	DENSE	SERA	ASER-1	ASER-2
Taipei	Light	273.22	156.14	176.51	3186.25	330.50	157.36	179.10	4651.31
	Moderate	26.22	6.79	34.55	3404.34	168.63	76.41	108.18	2990.98
	Heavy	**−1255.27**	**−1149.93**	**−1316.00**	2326.38	**−995.74**	**−867.21**	**−938.80**	2148.73
Keelung	Light	339.77	347.15	499.38	1407.11	333.78	370.89	495.38	2894.55
	Moderate	132.61	344.19	435.75	1042.55	228.63	468.23	471.83	1415.42
	Heavy	**−1075.40**	**−500.36**	**−856.18**	224.73	**−475.10**	187.69	**−205.08**	736.55
Tainan	Light	52.59	6.70	18.20	296.48	71.29	9.40	38.34	394.68
	Moderate	64.88	26.76	**−5.39**	713.97	118.68	37.95	89.84	858.28
	Heavy	**−827.29**	**−214.15**	**−399.80**	**−702.31**	**−797.97**	**−264.27**	**−495.70**	**−557.93**
Overall	Light	221.86	170.00	231.36	1629.95	245.19	179.22	237.61	2646.85
	Moderate	74.57	125.91	154.97	1720.29	171.98	194.20	223.28	1754.89
	Heavy	**−1052.65**	**−621.48**	**−857.33**	616.26	**−756.27**	**−314.60**	**−546.53**	775.78

Table 11
MSE change ratios for hourly data.

		Validation set				Test set			
		DENSE	SERA	ASER-1	ASER-2	DENSE	SERA	ASER-1	ASER-2
Taipei	Light	2.09	2.18	1.85	29.10	2.50	2.26	1.55	39.64
	Moderate	0.23	0.25	0.35	1.43	0.05	0.03	0.12	0.48
	Heavy	**−0.14**	**−0.15**	**−0.12**	**−0.18**	**−0.09**	**−0.12**	**−0.13**	**−0.17**
Keelung	Light	1.49	1.71	2.73	5.19	1.24	1.45	2.41	4.61
	Moderate	**−0.03**	**−0.11**	0.02	0.45	**−0.08**	**−0.13**	**−0.07**	0.35
	Heavy	**−0.13**	**−0.10**	**−0.13**	**−0.12**	**−0.12**	**−0.09**	**−0.13**	0.00
Tainan	Light	1.67	2.14	2.48	3.31	1.76	2.33	2.92	4.03
	Moderate	0.20	0.16	0.12	0.63	0.16	0.19	0.12	0.69
	Heavy	**−0.14**	**−0.15**	**−0.11**	**−0.12**	**−0.01**	**−0.07**	**−0.04**	0.02
Overall	Light	1.75	2.01	2.35	12.53	1.83	2.01	2.29	16.09
	Moderate	0.13	0.10	0.16	0.84	0.04	0.03	0.06	0.51
	Heavy	**−0.14**	**−0.13**	**−0.12**	**−0.14**	**−0.08**	**−0.09**	**−0.10**	**−0.05**

Table 12
MSE change ratios for daily data.

		Validation set				Test set			
		DENSE	SERA	ASER-1	ASER-2	DENSE	SERA	ASER-1	ASER-2
Taipei	Light	8.16	4.04	4.32	161.61	12.44	4.61	4.97	258.12
	Moderate	0.07	0.04	0.18	11.46	0.67	0.36	0.55	11.32
	Heavy	**−0.19**	**−0.18**	**−0.21**	0.31	**−0.22**	**−0.20**	**−0.21**	0.43
Keelung	Light	6.23	8.04	9.32	37.39	6.30	7.74	8.81	69.24
	Moderate	0.52	1.30	1.81	3.53	0.94	2.00	2.12	5.34
	Heavy	**−0.20**	**−0.09**	**−0.16**	0.03	**−0.17**	0.06	**−0.09**	0.22
Tainan	Light	3.06	0.36	0.62	35.01	1.78	0.33	0.72	19.43
	Moderate	0.15	0.07	**−0.01**	1.75	0.41	0.14	0.32	2.99
	Heavy	**−0.09**	**−0.02**	**−0.04**	**−0.07**	**−0.15**	**−0.05**	**−0.10**	**−0.10**
Overall	Light	5.82	4.14	4.75	78.00	6.84	4.22	4.84	115.60
	Moderate	0.25	0.47	0.66	5.58	0.67	0.83	1.00	6.55
	Heavy	**−0.16**	**−0.10**	**−0.14**	0.09	**−0.18**	**−0.06**	**−0.13**	0.18

6.2. Regression model comparison

The comparative analysis of the five models, including SVR, RF, XGBoost, S-LSTM, and Bi-LSTM, at three stations (Taipei, Keelung, Tainan) in four objective functions (DENSE, SERA, ASER-1, ASER-2) is shown in Figs. 18 and 19. The performance of each model is ranked based on its results on light rainfall, moderate rainfall and heavy rainfall in the validation and test sets; the rankings are shown in each dataset. The models are color-coded as follows: blue SVR, orange RF, red XGBoost, cyan S-LSTM, and green Bi-LSTM. The models have numbers that show the frequency of each rank. The general scores are tabulated in Table 13 with a lower score representing a better performance.

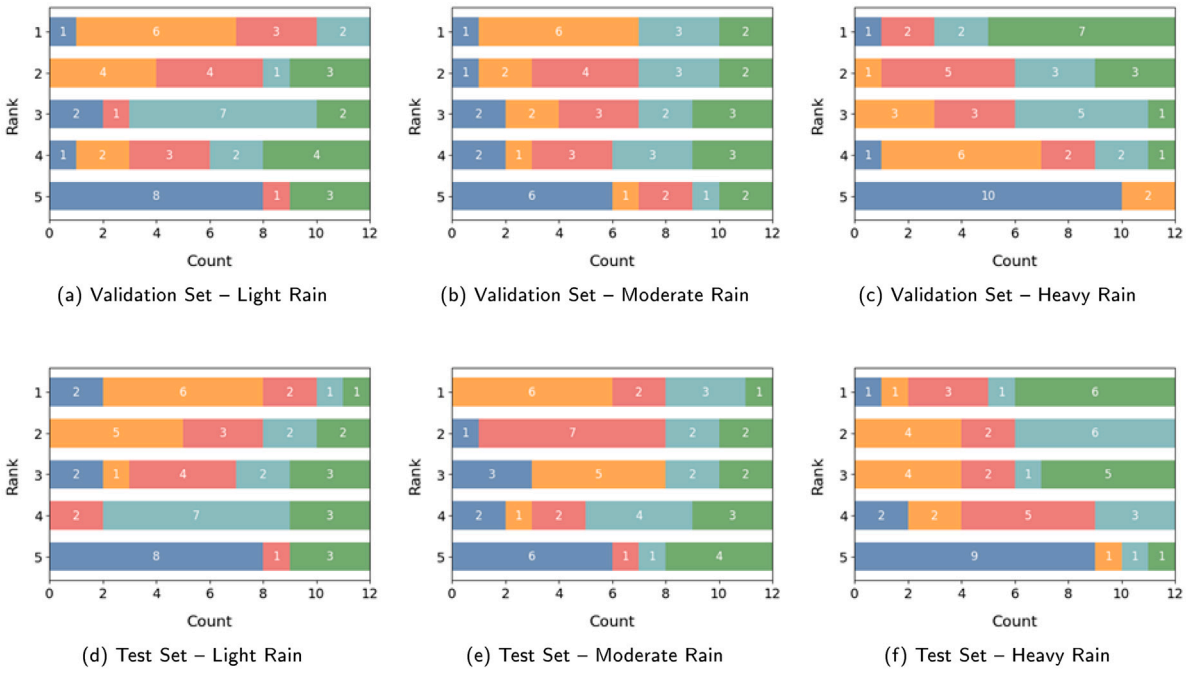


Fig. 18. Hourly Data Model Performance Rankings for Different Rainfall Categories. The models are color-coded as follows: blue—SVR, orange—RF, red—XGBoost, cyan—S-LSTM, and green—Bi-LSTM. The numbers within each bar indicate the frequency of each rank.

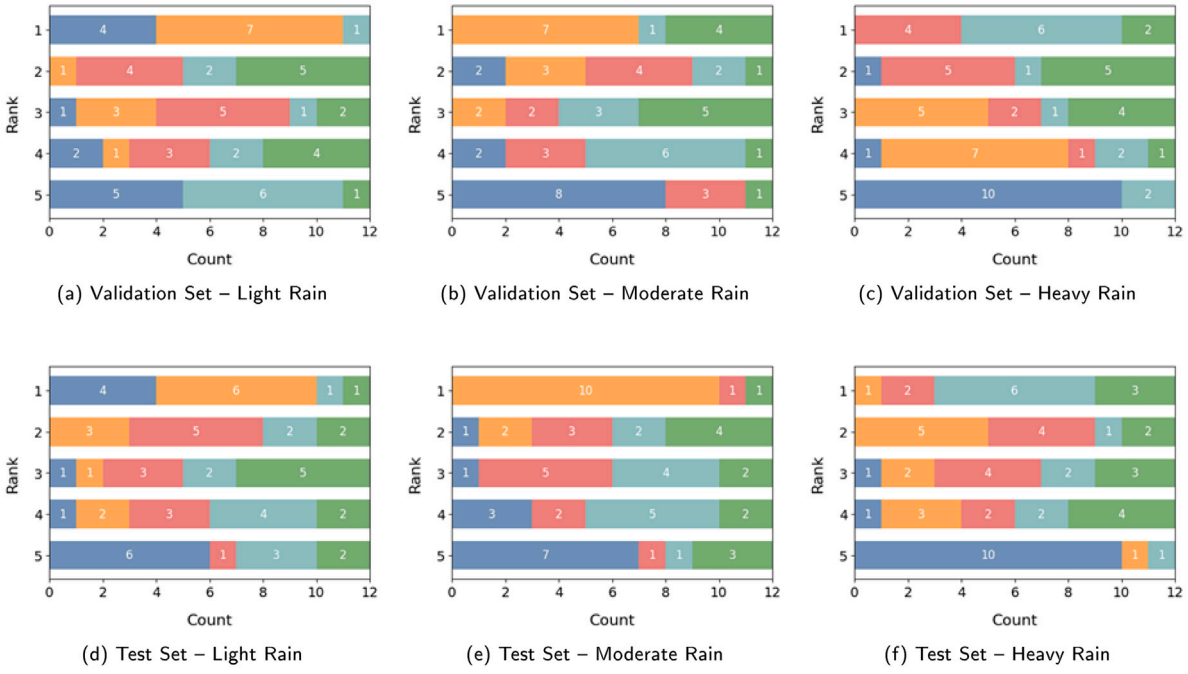


Fig. 19. Daily Data Model Performance Rankings for Different Rainfall Categories. The models are color-coded as follows: blue—SVR, orange—RF, red—XGBoost, cyan—S-LSTM, and green—Bi-LSTM. The numbers within each bar indicate the frequency of each rank.

Based on the analysis of the figures and tables, SVR performed significantly below the optimum in all the rainfall categories, especially during heavy rainfall. This can be explained by the fact that it is oriented to the minimization of the global mean-squared error, which makes it less suitable to work with unbalanced datasets.

Table 13
Model performance scores.

Model	Validation set			Test set		
	Light	Moderate	Heavy	Light	Moderate	Heavy
Hourly Data						
SVR	51 (5)	47 (5)	55 (5)	48 (5)	49 (5)	54 (5)
RF	**22 (1)**	**25 (1)**	45 (4)	**19 (1)**	**25 (1)**	34 (4)
XGBoost	31 (2)	39 (4)	29 (2)	33 (2)	29 (2)	33 (2)
S-LSTM	33 (3)	32 (2)	31 (3)	39 (3)	34 (3)	33 (2)
Bi-LSTM	43 (4)	37 (3)	**20 (1)**	41 (4)	43 (4)	**26 (1)**
Daily Data						
SVR	40 (4)	52 (5)	56 (5)	41 (4)	52 (5)	57 (5)
RF	**22 (1)**	**19 (1)**	43 (4)	**23 (1)**	**14 (1)**	34 (4)
XGBoost	35 (2)	41 (4)	**24 (1)**	36 (2)	35 (2)	30 (2)
S-LSTM	46 (5)	38 (3)	29 (3)	42 (5)	41 (4)	**27 (1)**
Bi-LSTM	37 (3)	30 (2)	28 (2)	38 (3)	38 (3)	32 (3)

Table 14
Average change in performance by objective function after ZOH.

		Validation set			Test set		
		Light	Moderate	Heavy	Light	Moderate	Heavy
Hourly	MSE	0.17	**−0.48**	**−8.19**	0.12	**−0.40**	**−14.19**
	DENSE	0.00	**−0.63**	14.30	**−0.02**	**−0.50**	3.23
	SERA	**−0.11**	**−0.02**	15.18	**−0.10**	**−0.18**	6.97
	ASER-1	0.12	0.34	5.28	0.10	**−0.40**	**−2.60**
	ASER-2	**−0.08**	**−0.66**	15.47	0.63	1.21	**−9.16**
Daily	MSE	45.48	37.41	**−369.70**	45.02	28.10	**−366.29**
	DENSE	91.46	108.26	**−35.19**	87.57	66.64	**−83.48**
	SERA	173.41	201.86	**−40.49**	182.33	129.90	**−125.33**
	ASER-1	104.90	148.50	311.18	116.30	91.37	48.00
	ASER-2	**−307.98**	**−305.91**	**−14.92**	**−756.22**	**−450.78**	**−187.29**

RF demonstrated the best levels of accuracy during light and moderate rainfall periods, but it failed to perform well during heavy rainfall situations. This difference seems to be caused by the averaging process which levels out the extremes, thus reducing accuracy of highly skewed samples.

XGBoost showed strong results during both light and heavy rainfall periods, but the validation scores were significantly variable in the moderate-rainfall range. Such fluctuations can be attributed to the regularization methods used by XGBoost to prevent overfitting, but also add more uncertainty to the model performance.

S-LSTM consistently ranked in the middle layer of most rainfall categories but scored the highest in daily heavy rainfall- a fact that demonstrates its suitability in time-series analysis.

Bi-LSTM similar to S-LSTM, and it was more effective on the restricted-range datasets and equally effective when predicting heavy rainfall.

In conclusion, XGBoost can be identified as the most consistent in both light and heavy rainfall conditions. SVR is not suitable to imbalanced data, and RF had lower accuracy in extreme rainfall events. S-LSTM and Bi-LSTM performed worse in low and moderate precipitation but excel in situations of heavy rainfall, which makes them especially suitable to extreme weather conditions.

6.3. Impact of ZOH on model performance

The effects of Zone of Hyperactivity (ZOH) were tested in all the five models and the five objective functions in each of the three stations. To this end, hourly data and daily data with light, moderate, and heavy rainfalls were analyzed. Two evaluation models were used. First, the mean improvement in performance across the five models and three stations of each objective function was calculated giving 15 evaluations per objective function (Table 14). Second, another 15 evaluations per model were taken into account by measuring the performance of each of the models on the three stations and five objective functions (Table 15). Negative values in both tables are printed in bold. In general, the findings indicate that because the hyperparameters were trained on non-ZOH data, some overfitting was present. Consequently, the performances of the validation set were mostly lower than those of the test set when ZOH was used.

In the analytic context of objective functions, the empirical combination of MSE and ZOH produced empirically better results, significantly reducing errors in the case of heavy rainfall and also reducing the increase in error in the case of light rainfall. DENSE and SERA showed small improvements in the hourly series, where the range of precipitation is smaller, but they worked well in the daily subset. ASER-1 showed no visible enhancement following ZOH and often increased the error magnitudes at various rainfall

Table 15
Average change in performance by model after ZOH.

		Validation set			Test set		
		Light	Moderate	Heavy	Light	Moderate	Heavy
Hourly	SVR	0.10	2.34	**−8.13**	0.75	4.46	**−18.95**
	RF	0.10	1.50	**−6.84**	0.05	**−0.22**	**−10.78**
	XGBoost	0.16	0.50	6.56	0.06	**−0.11**	**−11.79**
	S-LSTM	**−0.08**	**−2.99**	29.65	**−0.03**	**−0.02**	24.97
	Bi-LSTM	**−0.17**	**−2.80**	20.79	**−0.10**	**−4.38**	0.81
Daily	SVR	**−107.00**	**−55.48**	43.73	**−514.86**	**−223.09**	**−157.34**
	RF	46.84	71.88	**−292.76**	41.29	64.97	**−247.57**
	XGBoost	36.05	98.47	132.72	19.99	59.55	**−156.75**
	S-LSTM	126.90	56.70	**−81.27**	122.75	3.10	**−92.18**
	Bi-LSTM	4.50	18.54	48.46	5.84	**−39.30**	**−60.53**

intensities. ASER-2 was second behind MSE, and it caused a significant decrease in heavy rainfall errors and it was the only objective function to enhance even light rainfall accuracy.

According to the modeling perspective, ZOH combined with SVR and RF significantly minimized the errors of heavy rainfall, and XGBoost took the third level. S-LSTM and Bi-LSTM showed a bit oscillating trend in the hourly sequence of data but still reduced the error of predicting heavy rainfall in the daily frequency.

To sum up, ZOH reduces the excessive weight of the excessive zero-rainfall events, thus, allowing to learn the precipitation distribution more correctly and significantly increases the accuracy of heavy rainfall predictions. When ZOH was applied, the objective functions ranking was MSE ≥ ASER-2 > SERA ≥ DENSE > ASER-1. In the hierarchy of modeling architectures, the hierarchy was RF ≥ SVR > XGBoost > S-LSTM ≥ Bi-LSTM.

6.4. Comparison of methods and analysis

6.4.1. Comparison with baseline models

The most efficient algorithm to improve heavy-rainfall forecasting and maintain light-rainfall accuracy is XGBoost, as stated in Section VII-B. As a result, XGBoost is selected to be compared with other approaches. The baseline model (Base) uses MSE as the objective function without ZOH; the optimized one, ZD-XGBoost, uses ZOH with DENSE as the objective function. Table 14 shows empirical evidence that XGBoost has the advantage of using ZOH to enhance predictions of heavy-rainfall. Section VII-A justifies the use of DENSE by showing that it reduces the errors of prediction of heavy rainfall consistently.

The Figs. 20 and 21 show the overlap of the predicted and observed values of rainfall in Base and ZD-XGBoost models. The blue line indicates the actual rainfall and the orange line indicates the predicted rainfall. The parentheses show percentage decreases in prediction errors of heavy rainfall using ZD-XGBoost. The findings show that ZD-XGBoost can decrease heavy-rainfall prediction errors by 20% (hourly) and 40% (daily) and at the same time preserve accuracy on light rainfall.

6.4.2. Comparison with previous studies

We also compared our method to two recent papers: the Bi-LSTM network of Barrera-Animas et al. on the hourly data (Barrera-Animas et al., 2022) and RF model of Markuna et al. on the daily data (Markuna et al., 2023) to assess the differences in performance on the highly imbalanced datasets. Figs. 22 and 23 show the mean-squared-error (MSE) results of each rainfall category obtained by the three models. In the case of hourly data, ZD-XGBoost minimizes the errors in predicting heavy rainfall significantly, except in the Taipei dataset, where it performs worse than the Bi-LSTM model proposed by Barrera-Animas et al. However, Fig. 24 indicates that in some instances, the replacement of XGBoost with Bi-LSTM leads to better results, which is why it is crucial to choose the right objective function when working with the imbalanced regression problem. Furthermore, ZD-XGBoost shows the highest performance on the prediction of heavy rainfall and maintains the errors of light and moderate rainfall within an acceptable range in the daily data. These results show that the suggested technique is useful in improving the prediction of strong precipitation without sacrificing accuracy in weaker precipitation.

To sum up, ZD-XGBoost is more effective in most situations, significantly reducing the error of predicting heavy rainfall without a significant decrease in the accuracy of light and moderate rainfall. These results highlight the importance of choosing the right objective function in facing imbalanced data as the choice can significantly increase the precision of extreme-value predictions.

7. Conclusion and future work

This report contributes to the existing study of predicting heavy rainfalls with the use of ground-based observational data, especially in the very unbalanced surroundings of Taipei, Keelung and Tainan. The ensemble of ZD-XGBoost, built in the present work, minimized errors in predicting hourly heavy rainfall by about 20% and daily errors by about 40%. The role of objective functions and prediction models was also investigated, as well as the role of the zero-order term (ZOH). These findings demonstrate the practical applicability of the proposed framework for improving short-term precipitation forecasting in regions with highly uneven rainfall distributions, which is critical for flood risk management and early warning systems.

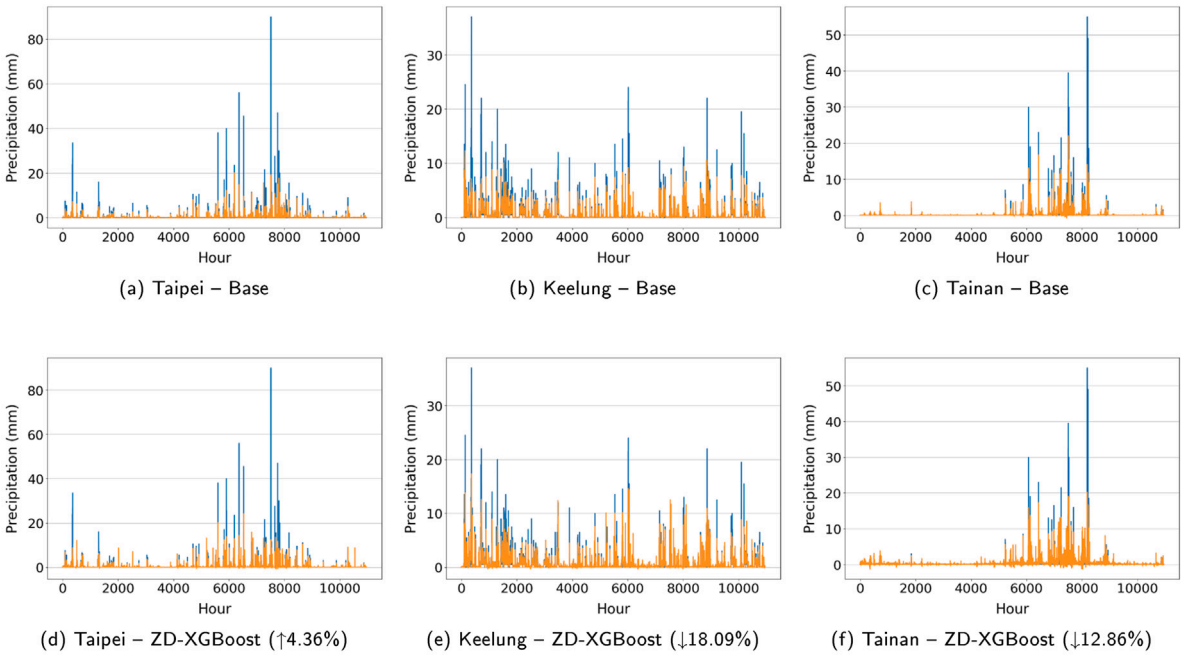


Fig. 20. Comparison of Base and ZD-XGBoost models on hourly test data across three stations.

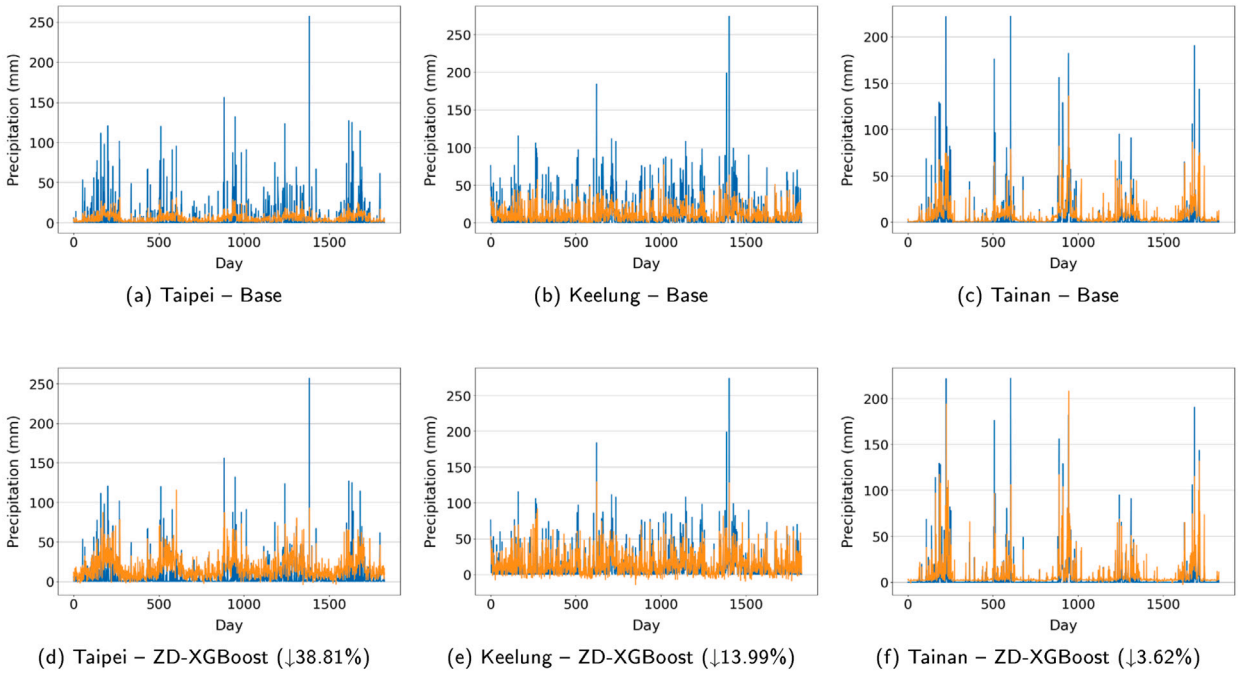


Fig. 21. Comparison of Base and ZD-XGBoost models on daily test data across three stations.

Regarding objective functions, DENSE showed the most evident enhancement in the prediction of heavy rain, SERA and ASER-1 showed the second most improvement, whereas ASER-2 tended to increase forecast errors in light rain and generally increased error rates. This observation highlights the importance of making trade-offs in the choice of objective functions used in various precipitation intensities. This result indicates that the effectiveness of an objective function is strongly dependent on rainfall intensity, highlighting the need to select loss functions that are tailored to extreme-event prediction rather than overall average performance.

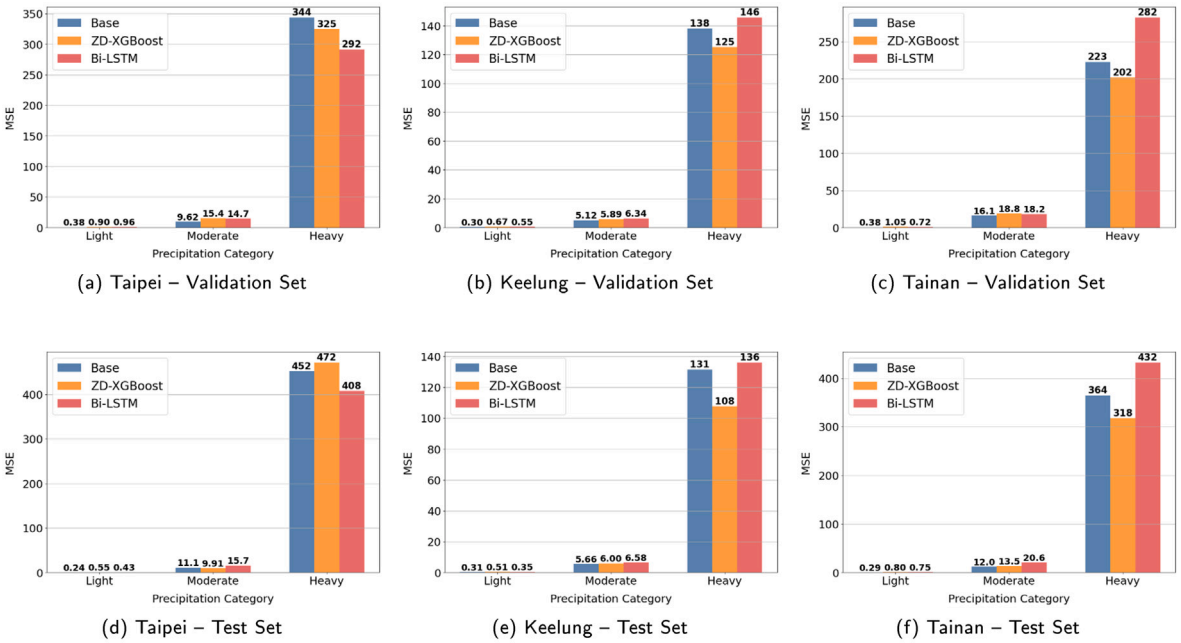


Fig. 22. Performance comparison on hourly data for Taipei, Keelung, and Tainan across validation and test sets.

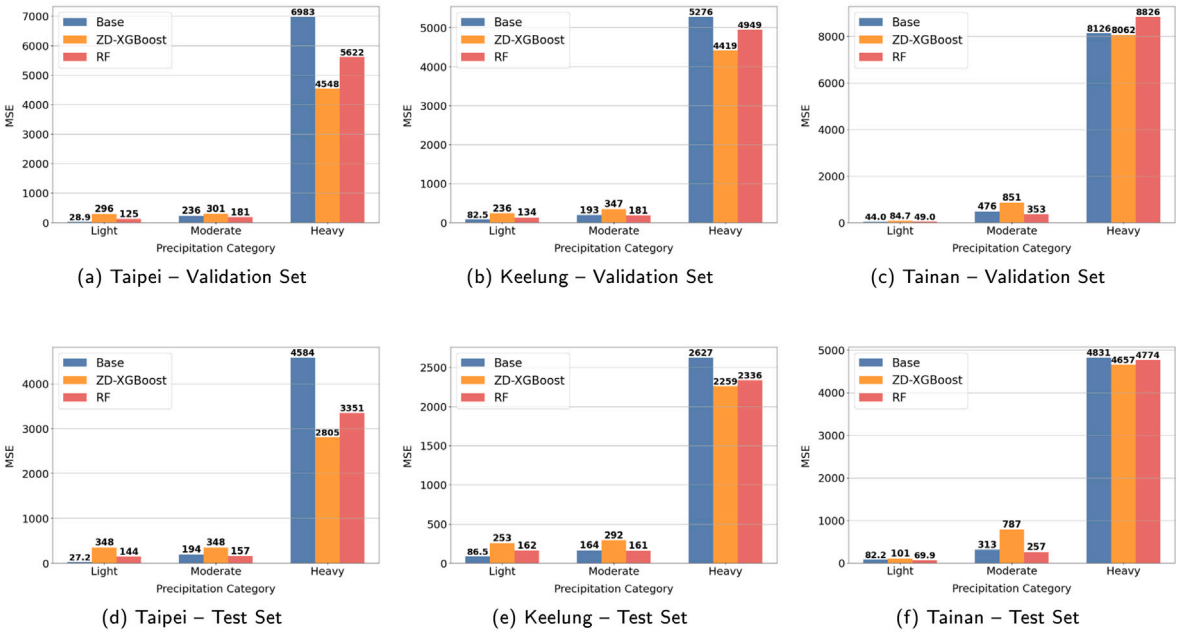


Fig. 23. Performance comparison on daily data for Taipei, Keelung, and Tainan across validation and test sets.

XGBoost was the most consistent of the predictive models in improving predictions of heavy rainfall and maintaining accuracy in light rainfall. The RF outperformed light and moderate rainfall, but both S-LSTM and Bi-LSTM enhanced the forecasts of heavy rainfall at the cost of high errors in the light rainfall. SVR performed poorly in all measures of evaluation. Overall, these comparisons reveal a clear trade-off between improving heavy-rainfall prediction and maintaining stability during light rainfall, explaining why no single predictive model dominates across all rainfall regimes. These observations further emphasize the importance of combining suitable prediction models with appropriate training strategies to address rainfall imbalance. The effectiveness of the ZOH method was confirmed, especially in combination with MSE and ASER-2, which significantly reduced the error in forecasting when it rained heavily. In comparison, the use of ASER-1 produced insignificant or adverse results. It should be noted that the ZOH

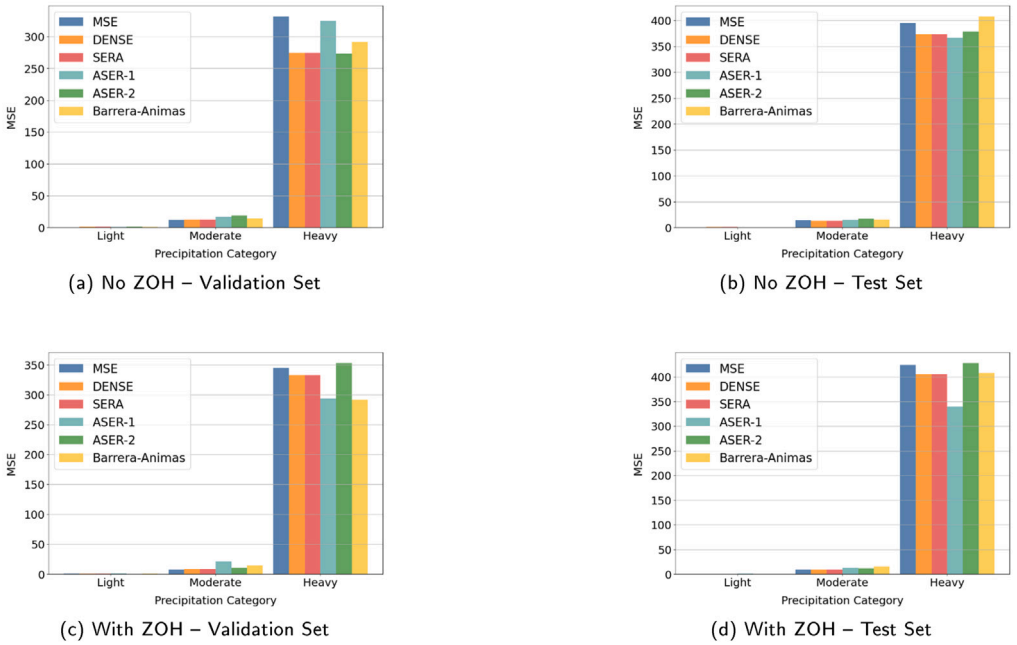


Fig. 24. Bi-LSTM performance comparison on Taipei hourly data before and after applying ZOH preprocessing.

technique primarily improves training stability under imbalanced rainfall conditions and does not alter the physical interpretation of precipitation measurements.

To sum up, the ZD-XGBoost model, which is an ensemble that combines XGBoost with ZOH and DENSE as the objective function, was the most efficient, retaining the stability of prediction in light-rainfall and significantly reducing the error of heavy-rainfall estimations. The results show a potential trend of predicting precipitation based on ground observations and can offer a useful lesson on how to tackle the imbalanced regression issue. The conclusions drawn in this study are based mainly on MSE-based evaluation and data from three stations in Taiwan, and therefore the generalization of the findings to other climatic regions should be investigated further. One area of future study should be to include uncertainty estimation, e.g. predicting the range of precipitation instead of its point value, to make meteorological forecasting more flexible and applicable to the real world.

CRediT authorship contribution statement

Fan-Ming Chiu: Writing – original draft, Methodology, Conceptualization. **Lawrence Jing-Yueh Liu:** Writing – original draft, Methodology, Investigation. **Uswatun Hasanah:** Writing – review & editing, Writing – original draft. **Chuan-Ming Liu:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that there are no competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Acknowledgments

This work is supported by National Science and Technology Council, R.O.C. under the grants NSTC 113-2221-E-027-051 and NSTC 114-2221-E-027-067.

Data availability

No new datasets were generated or analyzed during this study. The research utilized publicly available meteorological observation data from Central Weather Bureau of Taiwan open data platform.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining. KDD'19, Anchorage, AK, USA, pp. 2623–2631.

Amiranipour, M., Najafzadeh, M., Mohamadi, S., 2026. Meteorological and agricultural drought assessments using satellite imagery and machine learning models. Adv. Space Res. 77 (1), 549–572. http://dx.doi.org/10.1016/j.asr.2025.10.014.

Anjali, T., Chandini, K., Anoop, K., Lajish, L.V., 2019. Temperature prediction using machine learning approaches. In: Proc. 2019 2nd Int. Conf. Intell. Comput., Instrum. Control Technol.. ICICICT, pp. 1264–1268.

Barrera-Animas, A.Y., Oyedele, L.O., Bilal, M., Akinosho, T.D., Delgado, J.M.D., Akanbi, L.A., 2022. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Mach. Learn. Appl. 7, 100204.

Basirian, S., Najafzadeh, M., Demir, I., 2026. Water quality monitoring for coastal hypoxia: Integration of satellite imagery and machine learning models. Marine Poll. Bull. 222, 118735. http://dx.doi.org/10.1016/j.marpolbul.2025.118735, pt. 2.

Bi, K., Xie, L., Zhang, H., et al., 2023. Accurate medium-range global weather forecasting with 3D neural networks. Nature 619, 533–538. http://dx.doi.org/10.1038/s41586-023-06185-3.

Branco, P., Torgo, L., Ribeiro, R.P., 2017. SMOGN: A pre-processing approach for imbalanced regression. In: Proc. 1st Int. Workshop Learn. Imbal. Domains: Theory Appl.. Porto, Portugal.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining. KDD'16, pp. 785–794.

CODiS Climate Observation Data Inquire Service, 2024. Central weather bureau of taiwan. (Accessed August 27 2024) [Online]. Available: https://codis.cwa.gov.tw.

Dowell, D.C., et al., 2022. The high-resolution rapid refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. Weather. Forecast. 37, 1371–1395. http://dx.doi.org/10.1175/WAF-D-21-0151.1.

ECMWF, 2024. Red sky at night... producing weather forecasts directly from observations. In: ECMWF Newsletter. no. 178, Winter 2024, [Online]. Available: https://www.ecmwf.int/en/newsletter/178/earth-system-science/red-sky-night-producing-weather-forecasts-directly.

Ehteram, M., Ahmed, A.N., Khozani, Z.S., El-Shafie, A., 2023. Convolutional neural network-support vector machine model-Gaussian process regression: A new machine model for predicting monthly and daily rainfall. Water Resour. Manag. 37 (9), 3631–3655.

Espeholt, L., et al., 2022. Deep learning for twelve hour precipitation forecasts. Nat. Commun. 13, 5145.

Germann, U., Zawadzki, I., 2002. Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. Mon. Weather Rev. 130 (12), 2859–2873. http://dx.doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hou, J., Wang, Y., Hou, B., Zhou, J., Tian, Q., 2022. Prediction of hourly air temperature based on CNN–LSTM. Geomatics, Nat. Hazards Risk 13 (1), 1962–1986.

Hou, J., Wang, Y., Hou, B., Zhou, J., Tian, Q., 2023. Spatial simulation and prediction of air temperature based on CNN-LSTM. Appl. Artif. Intell. 37 (1), 2166235.

Hubert, M., Vandervieren, E., 2008. An adjusted boxplot for skewed distributions. Comput. Statist. Data Anal. 52 (12), 5186–5201.

Ignaccolo, M., Michele, C. De, 2020. One, no one, and one hundred thousand: The paradigm of the Z–R relationship. J. Hydrometeorol. 21, 1161–1169. http://dx.doi.org/10.1175/JHM-D-19-0177.1.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge Univ. Press, Cambridge, U. K.

Kavianpour, P., Kavianpour, M., Jahani, E., Ramezani, A., 2023. A CNN-bilstm model with attention mechanism for earthquake prediction. J. Supercomput. 79 (17), 19194–19226.

Kou, Y., Fu, G.H., 2023. ASER: Adapted squared error relevance for rare cases prediction in imbalanced regression. J. Chemom. 37 (11), 3515.

Lam, R., et al., 2023b. Learning skillful medium-range global weather forecasting. Science 382, 1416–1421. http://dx.doi.org/10.1126/science.adi2336.

Lavers, D.A., Simmons, A., Vamborg, F., Rodwell, M.J., 2022. An assessment of ERA5 reanalysis precipitation over the United Kingdom. Q. J. R. Meteorol. Soc. 148, 3152–3165.

Li, Z., Wang, Y., Xie, J., 2022. Application of the ARIMA-LSTM model in monthly precipitation prediction. In: 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference. IMCEC, pp. 1434–1441.

Lorenz, E.N., 1963. Deterministic nonperiodic flow. J. Atmos. Sci. 20, 130–141. http://dx.doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lynch, P., 2008. The origins of computer weather prediction and climate modeling. J. Comput. Phys. 227 (7), 3431–3444.

Markuna, S., Kumar, P., Ali, R., et al., 2023. Application of innovative machine learning techniques for long-term rainfall prediction. Pure Appl. Geophys. 180 (1), 335–363.

Moir, T.J., 2022. Rudiments of Signal Processing and Systems. Springer, Switzerland.

Morrison, H., van Lier-Walqui, M., Fridlind, A.M., Grabowski, W.W., Harrington, J.Y., Hoose, C., et al., 2020. Confronting the challenge of modeling cloud and precipitation microphysics. J. Adv. Model. Earth Syst. 12, e2019MS001689. http://dx.doi.org/10.1029/2019MS001689.

Pathak, J., et al., 2022. FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv:2202.11214.

Rasp, S., Dueben, P.D., Scher, S., Weyn, J.A., Mouatadid, S., Thuerey, N., 2020. WeatherBench: A benchmark data set for data-driven weather forecasting. J. Adv. Model. Earth Syst. 12, e2020MS002203. http://dx.doi.org/10.1029/2020MS002203.

Ravuri, S., et al., 2021. Skilful precipitation nowcasting using deep generative models of radar. Nature 597, 672–677.

Ribeiro, R.P., Moniz, N., 2020. Imbalanced regression and extreme value prediction. Mach. Learn. 109, 1803–1835.

Sari, Y., Arifin, Y.F., Novitasari, N., Faisal, M.R., 2022. Deep learning approach using the GRU-LSTM hybrid model for air temperature prediction on a daily basis. Int. J. Intell. Syst. Appl. Eng. 10 (3), 430–436.

Seed, A.W., 2003. A dynamic and spatial scaling approach to advection forecasting. J. Appl. Meteorol. 42 (3), 381–388. http://dx.doi.org/10.1175/1520-0450(2003)042<0381:ADASSA>2.0.CO;2.

Shi, X., et al., 2017. Deep learning for precipitation nowcasting: A benchmark and a new model. In: Advances in Neural Information Processing Systems (NeurIPS), Vol. 30, pp. 5617–5627.

Silva, A., Ribeiro, R.P., Moniz, N., 2022. Model optimization in imbalanced regression. In: Proc. Int. Conf. Discov. Sci.. Montpellier, France, pp. 3–21.

Sønderby, C.K., et al., 2020. MetNet: A neural weather model for precipitation forecasting. arXiv:2003.12140.

Steininger, M., Kobs, K., Davidson, P., Krause, A., Hotho, A., 2021. Density-based weighting for imbalanced regression. Mach. Learn. 110, 2187–2211.

Wang, S., Meng, P., Kong, L., Qin, S., Gui, X., 2023. Precipitation prediction based on ICEEMDAN-VMD and CNN-BiLSTM-AT models: A case study in Lanzhou, Gansu. In: 2023 IEEE International Conference on Sensors, Electronics and Computer Engineering. ICSECE, pp. 710–717.

Woo, W.-C., Wong, W.-K., 2017. Operational application of optical flow techniques to radar-based rainfall nowcasting. Atmosphere 8 (3), 48. http://dx.doi.org/10.3390/atmos8030048.

Yang, Y., Zha, K., Chen, Y., Wang, H., Katabi, D., 2021. Delving into deep imbalanced regression. In: Proc. 38th Int. Conf. Mach. Learn.. Virtual Event, pp. 11842–11851.

Yen, M.H., Liu, D.W., Hsin, Y.C., Lin, C.E., Chen, C.C., 2019. Application of deep learning for the prediction of rainfall in southern Taiwan. Sci. Rep. 9 (1), 12774.

Zhang, X., Zheng, Z., Li, H., Liu, F., Yin, Q., 2023. Deep learning precipitation prediction models combined with feature analysis. Env. Sci. Pollut. Res. 30 (58), 121948–121959.