ISSN 2959-6157

Comparative Analysis of Machine Learning Models for Weather Forecasting: A Heathrow Case Study

Muyan Xu

Department of Mathematics, Queen Mary University of London, London, UK
Corresponding author: muyan.xu@se22.qmul.ac.uk

Abstract:

Accurate weather forecasting, especially temperature prediction, is fundamental for various segments within the UK, including agriculture, energy, and policy planning, as the nation adapts to the effects of climate change. This study addresses the limitations of conventional linear models in capturing the complex, non-linear relationships within meteorological data by comparing the effectiveness of different Machine Learning (ML) strategies. This study evaluates the performance of baseline ML models, such as Linear Regression (LR), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN). It also examines advanced ensemble and boosting models such as Decision Tree (DT), Random Forest (RF), XGBoost (XGB), LightGBM (LGBM), and CatBoost, using a comprehensive dataset from Heathrow Airport. Detailed preprocessing, model training, and optimization through cross-validation were conducted, with performance assessed using Mean Squared Error (MSE) and Coefficient of Determination (R^2) metrics. The results demonstrate that ensemble methods, particularly XGB and LGBM, offer superior predictive accuracy for weather forecasting tasks, highlighting their potential to enhance predictive models in meteorological applications.

Keywords: Weather forecasting, Temperature prediction, Machine Learning, Random Forest, XGBoost.

1. Introduction

While many studies have explored the application of ML techniques to weather forecasting, a critical gap remains in understanding their effectiveness in specific localized contexts, such as Heathrow Airport, where accurate predictions are essential for operational safety, efficiency, and planning. This study is motivated by the aim to bridge the existing knowledge gap by examining the performance of different ML models for temperature prediction at Heathrow in depth.

A thorough understanding of weather forecasting is vital for decision-making in sectors like agriculture, energy, and policy planning, especially as the UK faces increasing impacts from climate change. Accurate temperature predictions are essential for supporting climate adaptation efforts, ensuring resilience across critical sectors, and maintaining safety and efficiency in operations at strategic locations like Heathrow [1-5]. Although ML techniques have shown promise as alternatives to conventional models, which often fail to capture complex, non-linear relationships within meteorological data, comprehensive evaluations of these models for temperature prediction in specific localized settings remain limited.

This study seeks to fill this gap by assessing the performance of various ML models for temperature prediction at Heathrow, including baseline models such as LR, SVR, and KNN, as well as advanced ensemble and boosting models like DT, RF, XGB, LGBM, and CatBoost. Recent studies have highlighted the potential of ML algorithms in enhancing forecasting accuracy where traditional models underperform [6-8]. However, identifying the most effective models for localized weather forecasting and providing insights into their applicability in meteorology remains a priority.

The methodology adopted involves preprocessing, normalization, transformation, and handling of missing values. The models are trained and optimized using cross-validation techniques, with performance evaluated using MSE and R^2 metrics. By employing a multi-model approach and integrating various ML methodologies, this research contributes to the field of weather forecasting by addressing the gaps in the current literature and optimizing predictive models for specific climatic conditions.

2. Literature References

Extensive research has shown that ML algorithms greatly improve weather forecasting by accurately identifying intricate and non-linear correlations in meteorological data that are often overlooked by conventional models. Tahsin

et al. [6] demonstrated the efficacy of various ML techniques, such as DT, RF, SVM, and ANN, in forecasting temperature in Chittagong, Bangladesh. However, in contrast to the current research that only examines Heathrow, their analysis was conducted within a broader regional framework.

El Hafyani et al. [7] utilized a multi-view stacking technique, merging XGB and LSTM networks to enhance precipitation forecasting in Morocco. This study showcases the benefits of ensemble approaches in wider scenarios, as opposed to the focused strategy of employing DT-based models for localised weather forecasting specifically at Heathrow. Furthermore, Kaya et al. [5] presented the adaptability of RF and other models in different fields, emphasizing its significance in predicting air temperature.

Ferchichi et al. [8] discovered that advanced models such as RF and ANN provide superior performance compared to standard models when it comes to predicting sea surface temperatures in Canada. This finding supports the current study's decision to utilize sophisticated models. Contrary to Ferchichi et al.'s focus on coastal sea surface temperature (SST), this study utilizes gradient boosting techniques (XGB and LGBM) to analyze unique meteorological conditions at Heathrow.

Ahmed et al. [3] emphasized the importance of using MMEs to decrease uncertainties in temperature predictions in Pakistan. Although the current study shares similarities with other studies in its use of a multi-model approach, its main objective is to optimize models specifically for the setting of Heathrow. Anjali et al. [2] and Wolff et al. [1] also investigated the efficacy of decision tree-based models, such as RF and XGB, for more extensive temperature forecasting tasks, demonstrating their proficiency in handling non-linear data efficiently.

This research aims to enhance the accuracy of weather forecasting models by optimizing ML models specifically for predicting localized temperatures at Heathrow. It builds upon previous findings and addresses the gaps highlighted in broader studies.

3. Methodology

3.1 Data Preprocessing

The dataset consists of daily weather observations from Heathrow, which include meteorological variables such as cloud cover, humidity, pressure, radiation, precipitation, sunshine, and temperature measures (mean, minimum, and maximum). To use this data effectively, preprocessing steps standardized and prepared the data. All categorical variables were converted to numerical values utilizing techniques such as one-hot encoding to enhance the processing efficiency of ML algorithms. The numerical features were normalized using min-max normalization, which rescaled each feature to a range of [0, 1]. This prevented features with bigger magnitudes from having a disproportionate influence on the model's learning process. The min-max normalization formula is defined as:

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Where x_i is the original value, and x_{min} and x_{max} are the minimum and maximum values of the feature, respectively.

3.2 Model Implementation

Nine models were implemented: LR, SVR, KNN, DT, RF, XGB, LGBM, and CatBoost. Each model's implementation details are as follows:

3.2.1 Linear Regression

LR aims to find the best-fit line that minimizes the residual sum of squares between observed targets and predicted values. The model is represented as:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{2}$$

Where $\hat{y}$ is the predicted value, x_i are the input features, and $\beta_i are the coefficients.$

3.2.2 Support Vector Regression

SVR is a supervised learning model that finds a hyperplane in a higher-dimensional space to fit data points within a threshold. The objective function of SVR is:

$$\text{minimize } \frac{1}{2}\left| \ |w| \ \right|^2 + C\sum_{i=1}^{n} max\left(0, |y_i - \hat{y}_i| - \epsilon\right) \tag{3}$$

Where w is the weight vector, C is the penalty parameter, and ϵ is the margin of tolerance.

3.2.3 K-Nearest Neighbors

KNN classifies a data point based on its K nearest neighbors. The prediction for regression is:

$$\hat{y} = \frac{1}{K}\sum_{i=1}^{K} y_i \tag{4}$$

Where $\hat{y}$ is the predicted value, and y_i are the values of the K nearest neighbors.

3.2.4 Decision Tree

DT is a non-parametric algorithm using a tree-like model of decisions. The splitting criterion is:

$$SplitCriterion = argmin\left(Impurity\left(L\right) + Impurity\left(R\right)\right) \tag{5}$$

Where L and R represent the subsets resulting from the split.

3.2.5 Random Forest

RF is an ensemble method constructing multiple DTs and

averaging their predictions. The prediction $\hat{y}$ is:

$$\hat{y} = \frac{1}{M}\sum_{m=1}^{M}\hat{y}_m \tag{6}$$

Where $\hat{y}_m$ is the prediction from the m-th tree, and M is the total number of trees.

3.2.6 XGBoost

XGB is an optimized gradient boosting algorithm. The objective function is:

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K}\Omega(f_k) \tag{7}$$

Where l represents the loss function, and $\Omega(f_k)$ is the regularization term.

3.2.7 LightGBM

LGBM uses a leaf-wise growth strategy and an optimized histogram-based DT algorithm. Its objective function is:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \lambda\sum_{j=1}^{J}(w_j^2 + b_j^2) \tag{8}$$

Where λ is the regularization parameter, and w_j and b_j are weights and biases.

3.2.8 CatBoost

CatBoost handles categorical variables with a gradient boosting algorithm. The transformation for categorical features is:

$$x'_{cat} = \frac{\sum_{i\in C_{train}} I(y_i = C) + \alpha}{|C_{train}| + \beta} \tag{9}$$

Where α and β are smoothing parameters.

3.3 Performance Evaluation Metrics

Models were evaluated using MSE and R^2. Residuals were analyzed for systematic patterns in prediction errors. The MSE is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

Where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

The R^2 is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{11}$$

Where $\bar{y}$ is the mean of the actual values.

4. Results

4.1 Dataset Splitting

The dataset from the European Climate Assessment Dataset (ECAD) [9, 10] comprises daily meteorological observations from Heathrow, including variables like temperature, cloud cover, wind speed, humidity, and more, collected over 3654 days. The data underwent rigorous cleaning and transformation to retain the original units. An 80-20 split was implemented, assigning 80% to training and 20% to testing. The training set, (X_{train}, y_{train}), was utilized for model development, while the testing set, (X_{test}, y_{test}), was employed to evaluate performance on unseen data, ensuring unbiased estimates. This strategic preprocessing and partitioning enabled a thorough evaluation of model performance tailored to Heathrow's distinct meteorological characteristics. This method guarantees robust validation, enhancing the applicability of the findings to real-world scenarios.

4.2 Learning Curve Analysis

Learning curves (Fig. 1) for all models, except for SVR, reveal convergence as training data increases, indicated by declining error rates. Models like RF, XGB, LGBM, and CatBoost demonstrated significant improvement with additional data, reflecting their proficiency in generalizing and managing intricate meteorological patterns. In contrast, SVR exhibited substantial discrepancies between training and validation errors (Fig. 2), pointing to underfitting and limited generalizability. These findings are consistent with prior research that questions SVR's capacity to handle complex, non-linear relationships frequently encountered in climate prediction tasks [1, 2]. The notable performance of ensemble models like RF and XGB illustrates their capability to accurately capture the complexity of meteorological data.

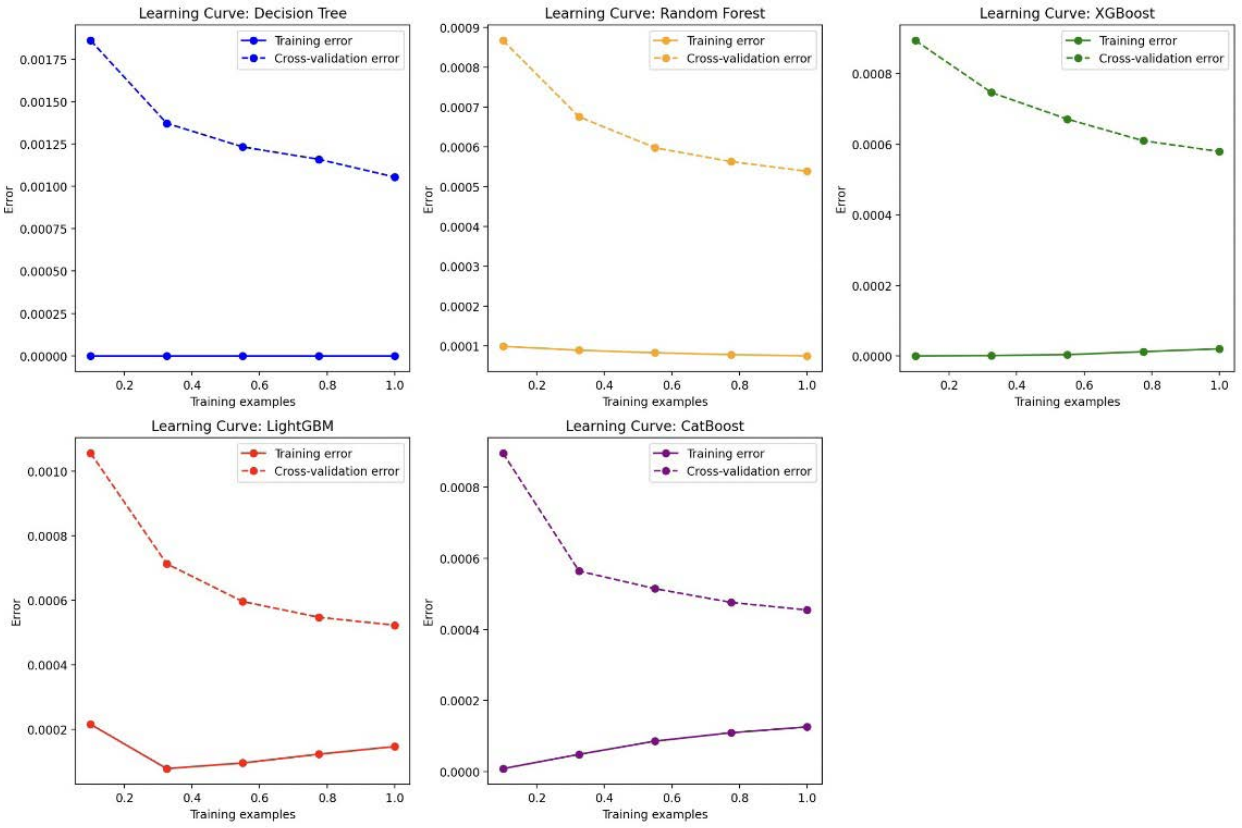


Fig. 1 Learning curves for the different models

4.3 Model Performance Comparison

The assessment of multiple ML models was conducted using MSE and R2 metrics. The ensemble approaches, notably RF and XGB, delivered notable outcomes, achieving the minimal MSE of 0.0006 and the maximal R^2 value of 0.9822. These results confirm their proficiency in detecting intricate data patterns, as demonstrated in Figure 3. The significant performance of RF and XGB aligns with earlier studies highlighting the effectiveness of tree-based ensemble techniques in meteorological forecasting [3, 5]. In contrast, SVR and KNN underperformed, evidenced by greater MSE and lower R^2 values (Table 1). The R^2 of 0.0000 for SVR suggests underfitting, while the R^2 of 0.8651 for KNN reveals its limitations in capturing non-linear dynamics. This evaluation suggests that although simpler models may offer computational simplicity, they often lack the necessary precision for applications where high forecasting accuracy is critical.

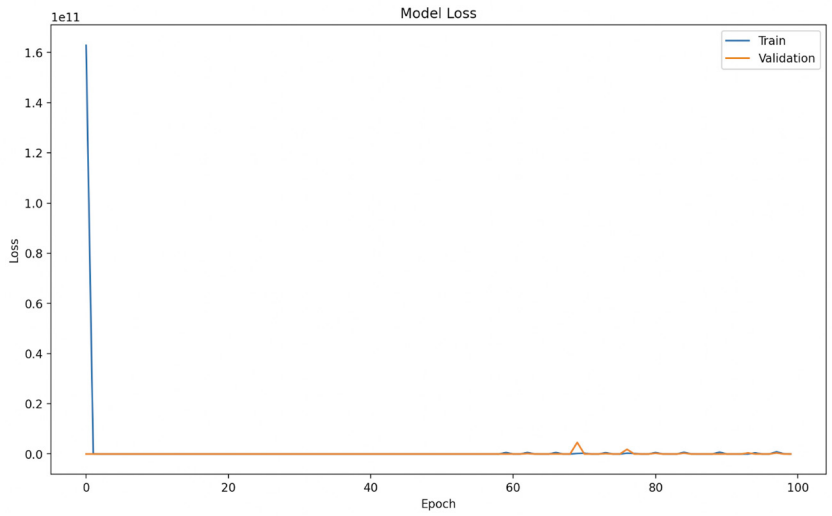


Fig. 2 Training and Validation Loss for SVR

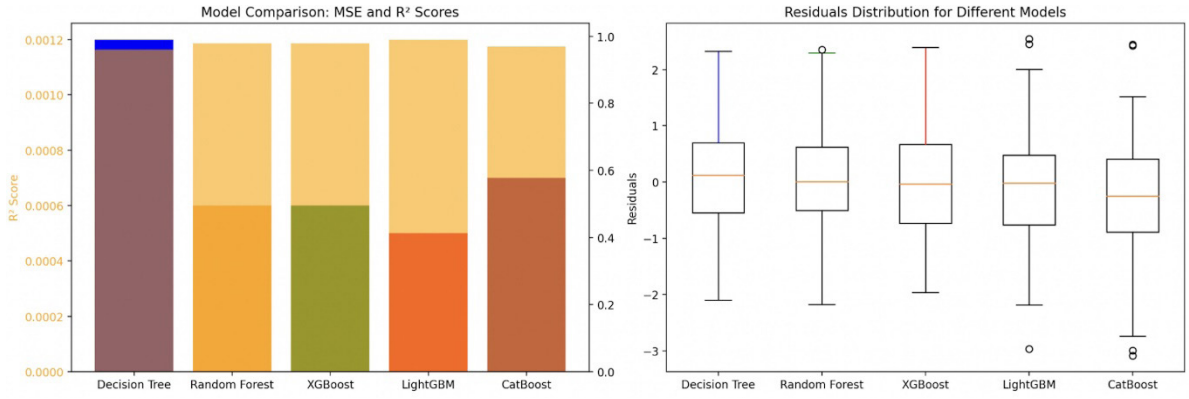


Fig. 3 Model Comparison of MSE and R^2 Scores

Table 1. Performance Metrics of Different Models

Model	MSE	R^2
LR	0.0006	0.9817
SVR	0.0320	0.0000
KNN	0.0043	0.8651
DT	0.0006	0.9810
RF	0.0006	0.9822
XGB	0.0006	0.9822
LGBM	0.0008	0.9805
CatBoost	0.0007	0.9800

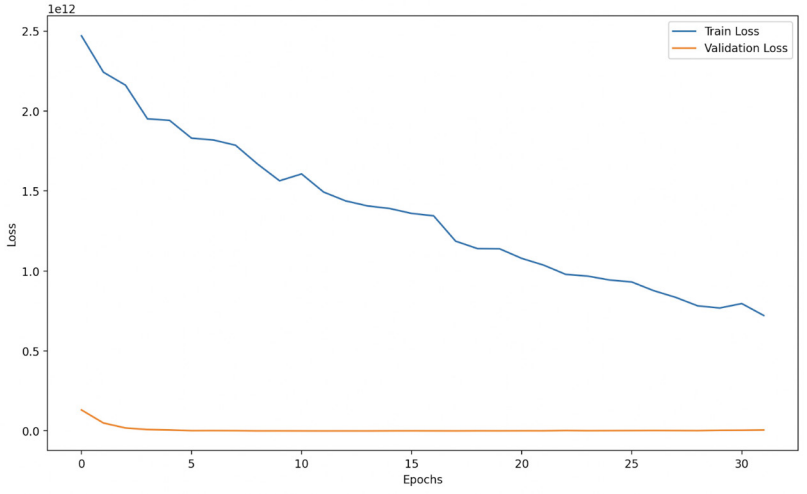
4.4 Residuals Analysis

The residual distribution (Fig. 4) indicates that RF and XGB residuals are closely aligned around zero, reflecting high predictive accuracy and minimal error margins. Conversely, SVR and KNN exhibit broader residual distributions, signifying elevated error rates and reduced reliability. These findings underscore the benefits of ensemble models in reducing prediction errors and enhancing forecast reliability, crucial for ensuring operational safety and strategic planning in high-risk con- texts like Heathrow.

4.5 Scatter Plot Analysis

Fig. 5 presents a comparative analysis of different models against actual temperature values. RF and XGB exhibit

the closest alignment with actual values, suggesting their advanced predictive capabilities, while SVR and KNN demonstrate less accuracy. This analysis further substantiates the conclusion that ensemble methods, particularly RF and XGB, are more competent at managing complex data structures, especially in the context of weather forecasting.



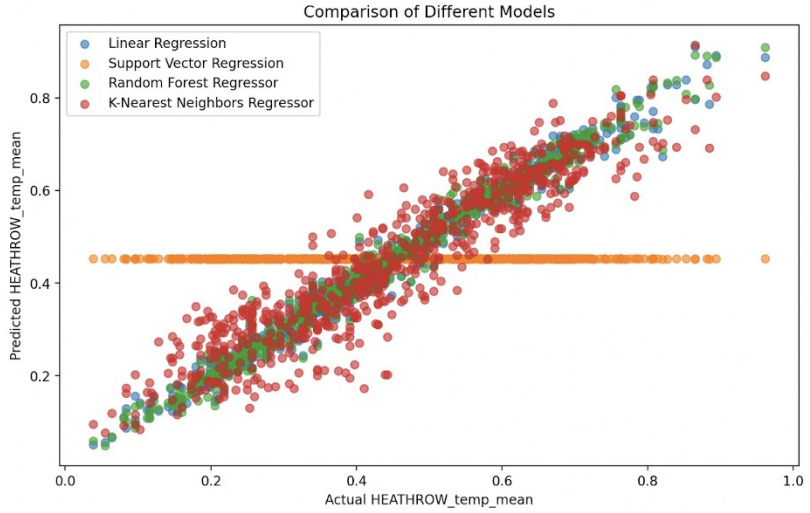


Fig. 4 Residuals Distribution for Different Model

Fig. 5 Comparison of Different Models on Predicted vs. Actual Temperatures

4.6 Model Comparison Discussion

The results of this research are consistent with earlier findings that identified RF and XGB as the most adept models for weather prediction tasks, attributed to their capacity to handle intricate and non-linear data relationships [6, 7]. This study extends the application of these models by demonstrating their effectiveness in specific environments such as Heathrow Airport, characterized by its unique weather patterns and operational demands.

Contrarily, the lower performance of SVR and KNN aligns with previous research that highlighted their limitations in dealing with highly variable and complex datasets [1, 2]. These findings suggest that attaining high accuracy necessitates prioritizing ensemble methods. Moreover, the findings hold practical significance for sectors reliant on precise weather forecasting, including aviation and agriculture. Future research should continue to explore advanced ML techniques across varied climatic regions to improve the adaptability and generalization of weather prediction models.

5. Conclusion

This study evaluated a diverse range of ML models, encompassing basic methods such as LR, SVR, and KNN, alongside advanced ensemble approaches like DT, RF, XGB, LGBM, and CatBoost, aimed at forecasting tem-

peratures at Heathrow. The findings indicate that ensemble methods, specifically RF, XGB, and LGBM, demonstrate enhanced predictive abilities, affirming their potential in weather forecasting. This research offers a theoretical foundation for deploying advanced ensemble methods in meteorological modeling and supports industries that depend on precise weather predictions, including agriculture, energy, and policy formulation. Future studies should focus on applying diverse datasets and expanding evaluation criteria to enhance model application and validity.

References

[1] Wolff S, O'Donncha F, Chen B. Statistical and machine learning ensemble modelling to forecast sea surface temperature[J]. Journal of Marine Systems, 2020, 208: 103347.

[2] Anjali T, Chandini K, Anoop K, et al. Temperature prediction using machine learning approaches[C]//2019 2nd International conference on intelligent computing, instrumentation and control technologies (ICICICT). IEEE, 2019, 1: 1264-1268.

[3] Ahmed K, Sachindra D A, Shahid S, et al. Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms[J]. Atmospheric Research, 2020, 236: 104806.

[4] Wang F, Liu R, Yan H, et al. Ground visibility prediction using tree-based and random-forest machine learning algorithm: Comparative study based on atmospheric pollution and atmospheric boundary layer data[J]. Atmospheric Pollution Research, 2024, 15(11): 102270.

[5] Kaya H, Guler E, Kırmacı V. Prediction of temperature separation of a nitrogen-driven vortex tube with linear, kNN, SVM, and RF regression models[J]. Neural Computing and Applications, 2023, 35(8): 6281-6291.

[6] Tahsin M S, Abdullah S, Al Karim M, et al. A comparative study on data mining models for weather forecasting: A case study on chittagong, Bangladesh[J]. Natural Hazards Research, 2024, 4(2): 295-303.

[7] El Hafyani M, El Himdi K, El Adlouni S E. Improving monthly precipitation prediction accuracy using machine learning models: a multi-view stacking learning technique[J]. Frontiers in Water, 2024, 6: 1378598.

[8] Ferchichi H, St-Hilaire A, Ouarda T B M J, et al. Prediction of coastal water temperature using statistical models[J]. Estuaries and Coasts, 2022, 45(7): 1909-1927.

[9] Klein Tank A M G, Wijngaard J B, Können G P, et al. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment[J]. International Journal of Climatology: A Journal of the Royal Meteorological Society, 2002, 22(12): 1441-1453.

[10] Huber F, van Kuppevelt E D, Steinbach P, et al. Will the sun shine?-An accessible dataset for teaching machine learning and deep learning[C]//The Third Teaching Machine Learning and Artificial Intelligence Workshop. PMLR, 2023: 27-31.