American Journal of Environmental Science and Engineering
2024, Vol. 8, No. 3, pp. 79-85
https://doi.org/10.11648/j.ajese.20240803.13

Research Article

Evaluating the Performance of a Stacking-Based Ensemble Model for Daily Temperature Prediction

Qiwei Xu[1], Anqi Guo[2], Wangzhi Yu[3, *], Chenfei He[4]

[1]Nottingham University Business School China, University of Nottingham Ningbo China, Ningbo, China
[2]Kelley School of Business, Indiana University Bloomington, Bloomington, United States
[3]School of Information Technology & Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou, China
[4]School of Economics & Finance, Zhejiang University of Finance and Economics, Hangzhou, China

Abstract

Temperature, as a critical element of weather forecasting, has consistently attracted extensive public attention. Accurate daily temperature prediction is essential for mitigating economic losses, preventing casualties, and maintaining public safety. However, traditional temperature prediction methods often fail to forecast the temperature promptly and effectively. To achieve more accurate daily temperatures prediction, researchers have turned to the recent advancement of artificial intelligence. This study aims to address the prediction of daily temperature in Algiers, by developing a stacking-based ensemble model. Firstly, the data normalization method is employed to preprocess the raw temperature data of Algiers in the experiment. Secondly, Decision Tree, K-Nearest Neighbors, Linear Regression, Random Forest, Recurrent Neural Network, and Support Vector Regression are selected as base models to predict the daily temperature. Finally, a stacking-based ensemble model with Recurrent Neural Network as the meta regressor (S-RNN) is applied for further accurate prediction. The experiment involves evaluating multiple metrics on the dataset to assess the performance of the model in predicting daily temperatures in Algiers. The experimental results indicate that the ensemble model outperforms other base models in addressing the challenges of daily temperature prediction. Meanwhile, this study confirms the significant potential in the application of stacking-based ensemble learning in the field of daily temperature prediction.

Keywords

Ensemble Model, Stacking, Daily Temperature, Prediction

1. Introduction

In recent years, the rapid development of modern industry has significantly contributed to climate change and global warming, which are among the most critical global changes observed and projected for the 21st century. These phenomena have become particularly pronounced over the past 65 years [1]. As a fundamental component of climate forecasting, daily temperature prediction plays a significant role in ensuring public safety [2]. Therefore, to protect human life and

*Corresponding author: yuwangzhi1@zufe.edu.cn (Wangzhi Yu)

Received: 21 August 2024; **Accepted:** 23 September 2024; **Published:** 29 September 2024

property, providing accurate prediction of daily temperature is a matter of urgency.

In the past, traditional statistical methods were commonly used to predict daily temperature. However, these methods often faced challenges due to the nonlinearity and seasonality of temperature changes [3, 4]. In addition, their high computational costs also limit their predictive ability [2]. Recently, advancements in artificial intelligence technology have facilitated the development of advanced prediction models capable of accurately monitoring and predicting daily temperature fluctuations, effectively capturing complex nonlinear relationships. In this study, a multi-model fusion method through stacking-based ensemble is proposed for predicting daily temperature accurately. Firstly, after normalizing the data, different basic prediction factors are obtained through six basic prediction models: Decision Tree (DT), K-Nearest Neighbors (KNN), Linear Regression (LR), Random Forest (RF), Recurrent Neural Network (RNN), and Support Vector Regression (SVR). Subsequently, the multi-model fusion method is employed to integrate these foundational predictors into a stacking-based ensemble model with Recurrent Neural Network as the meta regressor (S-RNN). In the experiment, the performance of the ensemble model and six base models is compared on a dataset of daily temperatures from Algiers. The experimental results indicate that the proposed model performs better than other base models in predicting daily temperature.

The remainder of this paper is organized as follows. Section 2 reviews the research achievements of previous researchers on daily temperature prediction and ensemble learning. Section 3 explores the process of data preprocessing and the method of integrating the proposed model. In section 4, the experimental setup is detailed, and the experimental results are analyzed. Section 5 summarizes the conclusions drawn from the experiment and provides potential research directions for the future.

2. Related Work

2.1. Daily Temperature Prediction

The significance of daily temperature prediction has driven researchers to consistently explore and develop various research methods. Traditionally, daily temperature prediction depends on Numerical Weather Prediction (NWP) [3] and Model Output Statistics (MOS) [4]. However, with the enhancement of forecast resolution, traditional temperature prediction methods have incurred higher costs and greater time demands. Simultaneously, the difficulty of these prediction methods necessarily results in deviations from traditional statistical methods that rely on historical observation data [5]. These methods typically predict future daily temperatures by describing the linear relationship between historical daily temperature observation data and future daily temperatures, often overlooking the consideration of the complex nonlinear relationship between them.

With the rapid development of artificial intelligence technology, machine learning methods have been widely applied in temperature prediction due to their ability to model complex nonlinear relationships and significantly improve prediction accuracy. For instance, Abdel-Aal and Elhadidy [6] proposed the abductory induction mechanism (AIM) to predict the daily maximum temperature in Dharan, Saudi Arabia, achieving an accuracy 77% higher than the local official prediction. Paniagua-Tineo et al. [7] applied SVR to the daily maximum temperature prediction problem within a 24-hour range and obtained accurate results. Krenn et al. [8] discovered that pure machine learning methods have enormous potential to predict the future of artificial intelligence.

Previous literatures have demonstrated that machine learning methods can effectively predict daily temperatures. In this study, six traditional machine learning models with good predictive performance are selected to predict daily temperatures. To further enhance prediction accuracy, the stacking-based multi-model fusion method is adopted to integrate the six base predictors to construct an ensemble model. Furthermore, a RNN is employed as the meta-regressor due to its high flexibility and adaptability.

2.2. Ensemble Method

In machine learning, the ensemble method has gained wide attention through, combining several base models to form a more powerful model than its constituents [9]. The most widely used ensemble techniques include averaging, bagging, random forest, boosting, and stacking [10, 11]. Among them, stacking appears to be particularly efficient, as it can combine heterogeneous base models and optimize predictions through training meta-models [12]. Compared to other ensemble methods, stacking generally demonstrates excellent performance. Consequently, this technique is extensively applied in addressing a variety of prediction challenges, including mineral exploration prediction, temporal network link prediction, and GDP growth forecasting [13-15].

In recent years, numerous studies have focused on utilizing ensemble methods for predicting temperatures. For example, Jose et al. [16] achieved excellent performance in predicting daily precipitation and temperature by using six machine learning models for multi-model ensemble. Li et al. [17] proposed a machine learning ensemble model to provide technical support for temperature prediction at twelve meteorological stations in Ontario. Bihlo [18] advanced an integrated weather forecasting system based on deep learning models and successfully predicted the seasonal temperature changes in the town of Innsbruck.

Accurate prediction of daily temperature can provide valuable information for social and economic activities, ensuring public health, and enhancing current production efficiency. However, few studies have employed the stacking-based ensemble method to daily temperature prediction studies.

Considering the superior performance of the stacking ensemble method, this study adopts the stacking ensemble method to integrate the selected six base prediction factors to effectively forecast the daily temperature of Algiers.

3. Methodology

3.1. Dataset Description

The dataset used in this study consists of daily temperature records detected in Algiers, including 9262 records detected between 1 January 1995 and 9 May 2020. This dataset is available on the data science and machine learning platform Kaggle[1].

3.2. Dataset Preprocessing

Due to missing values in the dataset, mean imputation was used to maintain data integrity by replacing the missing values of daily temperature with the mean. Additionally, data normalization, as a preprocessing method, was applied to scale the values of a dataset to a common range in order to reduce feature bias [19]. This study uses Z-score normalization to convert data into a standard normal distribution σ with zero mean μ and unit standard deviation using the following Eq. (1).

$$X(normalized) = \frac{x-\mu}{\sigma} \tag{1}$$

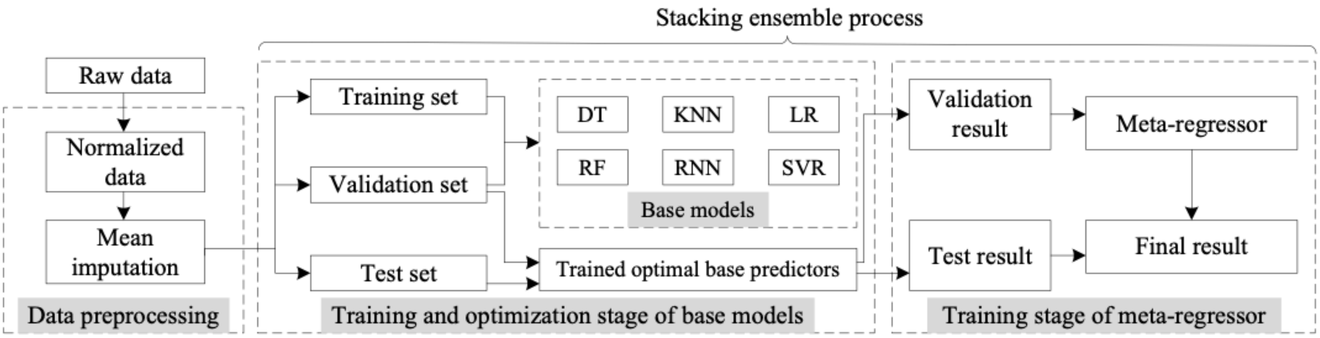


Figure 1*. Framework of the proposed model.*

3.3. Ensemble Model

The workflow of the proposed ensemble model in this study is shown in Figure 1.

(1) Data partitioning is a fundamental operation used to manage and process data on a computing cluster [20]. After preprocessing the dataset, to rigorously evaluate model performance, the dataset was partitioned into three subsets: 64% for training, 16% for validation, and 20% for testing. The training set was used to construct the base model and validation set to optimize parameters, while the test set is used to assess the performance of the proposed ensemble model.

(2) After considering the characteristics of daily temperature data in Algiers, this study chooses six machine learning models including DT, KNN, LR, RF, RNN, and SVR, for training and optimization.

(3) In the initial phase of the stacking-based ensemble model, predictions generated and optimized by six base models were concatenated. Subsequently, a meta-regressor was employed in the second phase to integrate previous trained optimal base predictors. Then the previous predicted data was used as input to test the training of the meta-regressor and evaluate the output results. The RNN with excellent flexibility and high continuity was selected as the meta-regressor.

4. Experiments

This section describes five evaluation metrics for assessing the predictive performance of base models and ensemble learning model. In addition, it analyzes and discusses the experimental results of daily temperature prediction across all models. All models and methods were implemented using the Python programming language.

4.1. Evaluation Metrics

Three statistical metrics have been proposed to evaluate the base models and the stacking-based ensemble model. These metrics can accurately reflect the predictive ability of the models, comprehensively capture various aspects of performance, and are calculated using the following Eqs. (2-6). In the equations, $\mathbf{Y}_i$ $(1 \le i \le n)$ and $\hat{\mathbf{Y}}_i$ $(1 \le i \le n)$ respectively represent the observed and predicted daily temperature data of Algiers within the same time range, where n denotes the size of test samples. In addition, the lower value obtained by the equations indicates better prediction performance.

(1) Mean Absolute Error (MAE) represents the average absolute difference between the predicted and actual

values, as illustrated in Eq. (2).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{\mathbf{Y}}_i - \mathbf{Y}_i\right| \tag{2}$$

(2) Mean Absolute Percentage Error (MAPE) represents the average absolute percentage difference between the predicted and actual values, as shown in Eq. (3).

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{\mathbf{Y}}_i - \mathbf{Y}_i}{\mathbf{Y}_i}\right| \times 100\% \tag{3}$$

(3) Mean Square Error (MSE) represents the average of the squared differences between the predicted and actual values, as illustrated in Eq. (4).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathbf{Y}}_i - \mathbf{Y}_i)^2 \tag{4}$$

(4) Root Mean Square Error (RMSE) represents the square root of the average of the squared differences between the predicted and actual values, as shown in Eq. (5).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathbf{Y}}_i - \mathbf{Y}_i)^2} \tag{5}$$

(5) R-squared (R2) represents the accuracy of the model's prediction on a scale of 0 to 1, as illustrated in Eq. (6).

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(\mathbf{Y}_i - \hat{\mathbf{Y}}_i)^2}{\sum_{i=1}^{n}(\mathbf{Y}_i - \bar{\mathbf{Y}})^2} \tag{6}$$

4.2. Experimental Results and Discussion

Using the above evaluation metrics, Table 1 provides a detailed comparison of the test sets of seven machine learning models including DT, KNN, LR, RF, RNN, and SVR and the proposed S-RNN model. The bold font indicates the optimal value. To present the evaluation results more clearly, the prediction results of all models are depicted in the form of horizontal bar charts, as displayed in Figure 3.

Table 1. *Evaluation results of seven models on daily temperature dataset.*

Metrics	MAE	MAPE	MSE	RMSE	R^2
KNN	4.6125	0.0356	243.1730	15.5940	0.3268
DT	4.3914	0.0305	273.1942	16.5286	0.2436
RF	4.2121	0.0312	232.4634	15.2467	0.3564
SVR	4.2982	0.0349	318.5404	17.8477	0.1181
RNN	6.6293	0.0677	272.2388	16.4997	0.2463
LR	4.5740	0.0416	251.9320	15.8724	0.3025
S-RNN	4.1078	0.0304	230.4045	15.1791	0.3621

In addition, the radar chart illustrates the performance differences between various models. To more clearly represent the performance differences, the R2 values in Table 1 were normalized within the range of [0.2, 1], and the data for the other four indicators were first calculated by taking the reciprocal and then normalized to the range of [0.2, 1]. The processing results are shown in Figure 2.

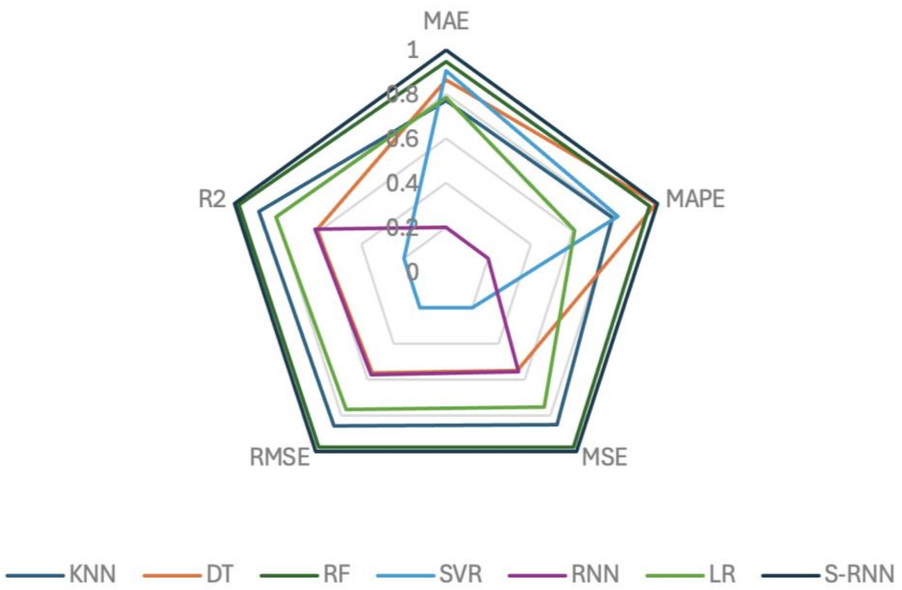
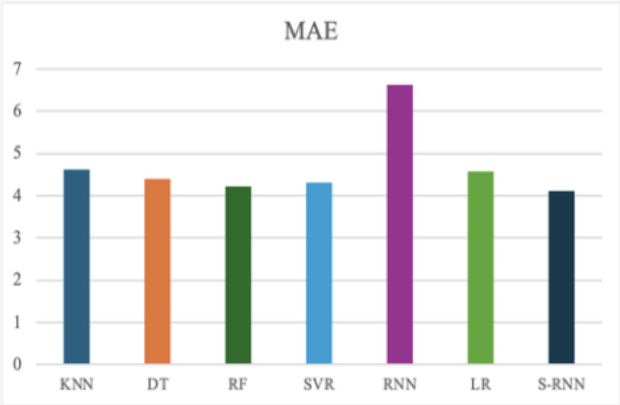
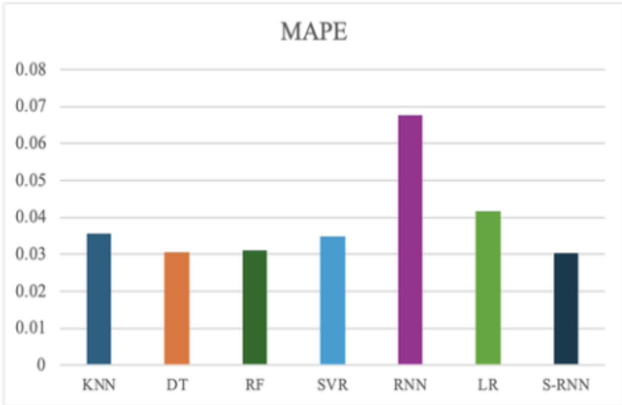


Figure 2. *Radar map.*

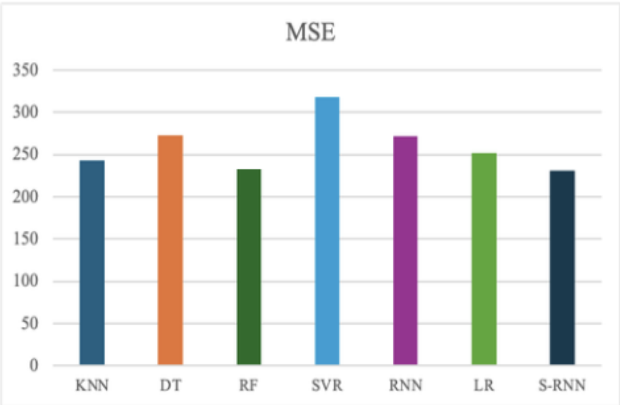
Through Table 1, Figure 2, and Figure 3, the experimental results can be visualized. Under the five evaluation metrics of MAE, MAPE, MSE, RMSE and R^2, the stacking-based ensemble model outperformed the other six base models in predicting daily temperature in Algiers. Therefore, it can be concluded that the multi-model fusion method through stacking-based ensemble demonstrates superior accuracy and stability upon evaluation against the base models.



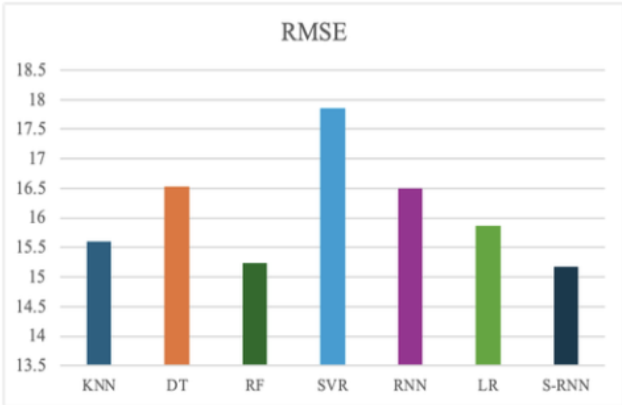
(a) Evaluation results using MAE



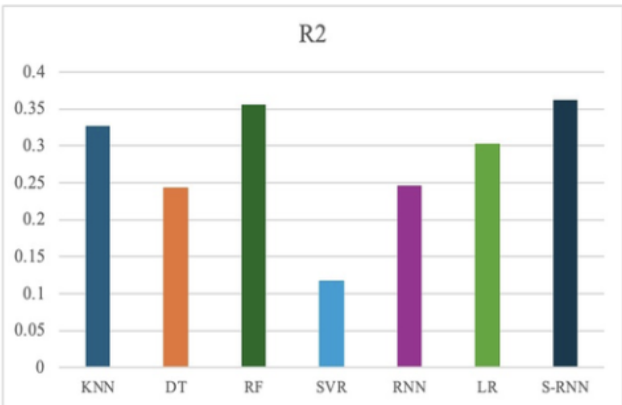
(b) Evaluation results using MAPE



(c) Evaluation results using MSE



(d) Evaluation results using RMSE



(e) Evaluation results using R^2

Figure 3. *Evaluation results.*

5. Conclusion

Accurate daily temperature prediction can provide strong assurances for production, daily life, and public safety. This study proposes a multi-model fusion method through stacking-based ensemble. The experimental verification of daily temperature in Algiers according to multiple evaluation indicators demonstrates that the proposed S-RNN model outperforms the basic models in predicting daily temperature.

However, the current research also highlights the necessity for further investigation and refinement in certain areas. Firstly, in the selection of base models, several deep learning models such as Multilayer Perceptron, Convolutional Neural Network, and Long Short-Term Memory can be appropriately added to improve the scalability of predictions. Secondly, during the training and optimization stages of the base predictors, Bayesian optimization methods can be applied to optimize the hyperparameters of the base models. Thirdly, more evaluation metrics can also be considered to assess the performance of the predictors, so as to achieve more comprehensive evaluation results. Moreover, the proposed muti-model fusion method through stacking-based ensemble can also be employed to forecast other climate phenomena such as rainfall, wind speed, and pressure or to address other prediction tasks in other fields including air pollution prediction, traffic flow prediction, and stock price prediction. Meanwhile, considering the scarcity and high cost of labeled data, exploring the application of unsupervised learning (e.g., self-encoder, cluster analysis) and semi-supervised learning techniques in predictive models can effectively utilize unlabeled data to improve the generalization ability and learning efficiency of models.

Abbreviations

DT	Decision Tree
KNN	K-Nearest Neighbors
LR	Linear Regression
RF	Random Forest
RNN	Recurrent Neural Network
SVR	Support Vector Regression
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
R^2	R-squared

Author Contributions

Qiwei Xu: Writing - original draft, Methodology & Visualization

Anqi Guo: Conceptualization, Writing - review & editing

Wangzhi Yu: Writing - review & editing, Software, Project administration

Chenfei He: Validation, Writing - review & editing

Data Availability Statement

The data which support the findings of this study can be found at:
https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities/data.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Abbass, K., Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., and Younis, I. A review of the global climate change impacts, adaptation, and sustainable mitigation measures. Environmental Science and Pollution Research. 2022, 29, 42539-42559. https://doi.org/10.1007/s11356-022-19718-6

[2] An, H. Y., Li, Q. L., Lv, X. Y., Li, G. X., Qian, Q. F., Zhou, G. B., Nie, G. Z., Zhang, L. J., and Zhu, L. W. Forecasting daily extreme temperatures in Chinese representative cities using artificial intelligence models. Weather and Climate Extremes. 2023, 42, 100621. https://doi.org/10.1016/j.wace.2023.100621

[3] Bauer, P., Thorpe, A., and Brunet, G. The quiet revolution of numerical weather prediction. Nature. 2015, 525(7567), 47-55. https://doi.org/10.1038/nature14956

[4] Glahn, H. R., and Lowry, D. A. The use of Model Output Statistics (MOS) in objective weather forecasting. Journal of Applied Meteorology. 1972, 11(8), 1203-1211. https://doi.org/10.1175/1520-0450(1972)011

[5] Glahn, B. Determining an optimal decay factor for Bias-Correcting MOS temperature and Dewpoint forecasts. Weather and Forecasting. 2014, 29(4), 1076-1090. https://doi.org/10.1175/waf-d-13-00123.1

[6] Abdel-Aal, R. E., and Elhadidy, M. A. Modeling and forecasting the daily maximum temperature using abductive machine learning. Weather and Forecasting. 1995, 10(2), 310-325. https://doi.org/10.1175/1520-0434(1995)010

[7] Paniagua-Tineo, A., Salcedo-Sanz, S., Casanova-Mateo, C., Ortiz-García, E., Cony, M., and Hernández-Martín, E. Prediction of daily maximum temperature using a support vector regression algorithm. Renewable Energy. 2011, 36(11), 3054-3060. https://doi.org/10.1016/j.renene.2011.03.030

[8] Krenn, M., Buffoni, L., Coutinho, B., Eppel, S., Foster, J. G., Gritsevskiy, A., Lee, H., Lu, Y., Moutinho, J. P., Sanjabi, N., Sonthalia, R., Tran, N. M., Valente, F., Xie, Y., Yu, R., and Kopp, M. Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network. Nature Machine Intelligence. 2023, 5(11), 1326-1335. https://doi.org/10.1038/s42256-023-00735-0

[9] Kumar, V., Aydav, P. S. S., and Minz, S. Multi-view ensemble learning using multi-objective particle swarm optimization for high dimensional data classification. Journal of King Saud University. Computer and Information Sciences. 2022, 34(10), 8523-8537. https://doi.org/10.1016/j.jksuci.2021.08.029

[10] Mienye, I. D., and Sun, Y. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Access. 2022, 10, 99129-99149. https://doi.org/10.1109/access.2022.3207287

[11] Mohammed, A., and Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University. Computer and Information Sciences. 2023, 35(2), 757-774. https://doi.org/10.1016/j.jksuci.2023.01.014

[12] Chatzimparmpas, A., Martins, R. M., Kucher, K., and Kerren, A. StackGenVIS: alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. IEEE Transactions on Visualization and Computer Graphics. 2021, 27(2), 1547-1557. https://doi.org/10.1109/tvcg.2020.3030352

[13] Hajihosseinlou, M., Maghsoudi, A., and Ghezelbash, R. Stacking: A novel data-driven ensemble machine learning strategy for prediction and mapping of Pb-Zn prospectivity in Varcheh district, west Iran. Expert Systems with Applications. 2024, 237, 121668. https://doi.org/10.1016/j.eswa.2023.121668

[14] He, X., Ghasemian, A., Lee, E., Clauset, A., and Mucha, P. J. Sequential stacking link prediction algorithms for temporal networks. Nature Communications. 2024, 15, 1364. https://doi.org/10.1038/s41467-024-45598-0

[15] Cui, S., Yin, Y. Q., Wang, D. J., Li, Z. W., and Wang, Y. Z. A stacking-based ensemble learning method for earthquake casualty prediction. Applied Soft Computing. 2021, 101, 107038. https://doi.org/10.1016/j.asoc.2020.107038

[16] Jose, D. M., Vincent, A. M., and Dwarakish, G. S. Improving multiple model ensemble predictions of daily precipitation and temperature through machine learning techniques. Scientific Reports. 2022, 12, 4678. https://doi.org/10.1038/s41598-022-08786-w

[17] Li, X. Y., Li, Z., Huang, W., and Zhou, P. X. Performance of statistical and machine learning ensembles for daily temperature downscaling. Theoretical and Applied Climatology. 2020, 140(1-2), 571-588. https://doi.org/10.1007/s00704-020-03098-3

[18] Bihlo, A. A generative adversarial network approach to (ensemble) weather prediction. Neural Networks.2021, 139, 1-16. https://doi.org/10.1016/j.neunet.2021.02.003

[19] Singh, D., and Singh, B. Investigating the impact of data normalization on classification performance. Applied Soft Computing. 2020, 97, 105524. https://doi.org/10.1016/j.asoc.2019.105524

[20] Mahmud, M. S., Huang, J. Z., Salloum, S., Emara, T. Z., and Sadatdiynov, K. A survey of data partitioning and sampling methods to support big data analysis. Big Data Mining and Analytics. 2020, 3(2), 85-101. https://doi.org/10.26599/bdma.2019.9020015

1 https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities/data