Adaptive Ensemble of XGBoost and LSTM for Temperature Forecasting

Mingcheng Ye

Beijing Institute of Technology, School of Computer Science and Technology, Beijing, 102401, China
ye.mc@bit.edu.cn

Abstract. Accurate weather prediction is crucial for numerous applications, ranging from daily decision-making to emergency response and disaster mitigation. This study introduces an adaptive ensemble method for temperature forecasting that integrates two distinct machine learning algorithms. The ensemble framework dynamically adjusts the weights of each individual model based on their performance characteristics, ensuring that the most reliable predictions are prioritized. The method was tested on historical weather data from five major European cities, consistently demonstrating superior performance compared to standalone models. The results show that the adaptive ensemble achieved R^2 values exceeding 0.99 across all locations, indicating a high degree of predictive accuracy. Notably, the geographic location of each city significantly influenced the weight allocation within the ensemble, suggesting that spatially-dependent feature interactions play a more dominant role than temporal patterns in determining temperature variations in these regions. These findings highlight the potential of adaptive ensemble strategies in enhancing the robustness and precision of weather forecasting models across diverse climatic and geographical contexts.

Keywords: Weather prediction, ensemble methods, temperature forecasting, adaptive weighting

1 Introduction

In the atmospheric field, weather prediction is a challenging obstacle with important implications in different situations. Traditional weather prediction models face difficulties in dealing with low computation resources and accuracy for temporal forecasts. The improvement of machine learning (ML) and deep learning (DL) algorithms has provided new opportunities by obtaining complex non-linear relationships inherent in meteorological conditions.

At the first stage, ML applications in weather predictions basically use conventional statistical models and basic artificial neural networks (ANNs). Chattopadhyay et al. demonstrated the effectiveness of DL approaches for climate modeling, establishing that neural networks can successfully address complex weather systems and laying the foundation for DL-based meteorological forecasting [1].

The evolution of ensemble learning has significantly advanced meteorological forecasting capabilities. Chen & Guestrin introduced eXtreme Gradient Boosting (XGBoost) in 2016, which rapidly became a preferred tool for tabular data applications due to its computational efficiency and accuracy [2]. XGBoost has since been successfully applied to various weather prediction tasks with remarkable results. Ensemble approaches have also shown consistent promise in weather forecasting applications. Deep Karan Singh et al. in 2023 implemented multi-model systems and XGBoost consistently outperformed other individual models in weather predictions, which got 0.83 R-squared scores and 0.66 root mean square errors (RMSE) [3].

To address temporal dependencies in weather data, specialized architectures such as the Convolutional Long Short-Term Memory network (ConvLSTM), proposed by Shi et al., achieved notable success in precipitation nowcasting by effectively capturing spatial-temporal features [4]. Further work by Grover et al. integrated sequence-to-sequence LSTM architectures for temperature forecasting, demonstrating the model's capability to capture long-term dependencies and improve forecasting accuracy [3].

Hybrid modeling paradigms have emerged in recent years, combining conventional statistical and ensemble techniques with neural networks. Rasp & Lerch enhanced continuous ranked probability score (CRPS) metrics by combining ensemble weather prediction with deep learning post-processing [5]. More sophisticated hybrid models, like the CNN-LSTM by Liu et al. and the combination of LSTM and Random Forest by Kim et al., further enhanced performance in precipitation prediction and extreme disaster forecasting, reporting high accuracy across meteorological events [6, 7].

Adaptive ensemble procedures are becoming more important in the atmospheric science field, according to systematic reviews like the review of McGovern et al. Robust model evaluation and adaptive ensemble procedures are becoming increasingly important in weather predictions, according to systematic reviews like the review of McGovern et al. [8].

By creating an adaptive ensemble system that blends XGBoost and LSTM neural networks for temperature prediction, this paper fills these gaps. This development presents a viable avenue for enhancing forecast precision in meteorological applications.

2 Methodology

2.1 Related Work

Long Short-Term Memory Networks. A specific type of recurrent neural network, LSTM networks, were first presented by Hochreiter and Schmidhuber in 1997. They were created to solve the vanishing gradient issue that conventional RNNs had [9]. Three gating mechanisms—the forget gate, input gate, and output gate—are incorporated into the LSTM design to control the flow of information throughout the network.

Extreme Gradient Boosting. Chen and Guestrin made XGBoost in 2016. It is an improved version of gradient boosting decision trees that has done quite well in several machine learning applications [2]. The algorithm builds a set of weak learners (decision trees) one at a time, with each tree trying to fix the mistakes produced by the trees that came before it.

XGBoost's effectiveness stems from its ability to handle missing values, built-in regularization mechanisms, and efficient parallel processing capabilities. The algorithm's tree-based structure excels at capturing complex feature interactions and non-linear relationships within tabular data.

2.2 Architecture and Framework

Overall System Design. The adaptive framework combines three key components to leverage the strengths of different modeling paradigms.

LSTM Architecture. The LSTM component employs a dual-pathway design optimized for weather sequence processing. Pathway processes historical meteorological sequences through multiple LSTM layers equipped with Dropout regularization to prevent overfitting. This pathway captures long-term temporal dependencies and seasonal patterns inherent in weather data.

Feature Engineering. The XGBoost modulation extracts comprehensive statistical features from historical data, including moving averages, extrema values, trend indicators, and recent pattern summaries. These engineered features are combined with current meteorological conditions to create a rich feature space that captures complex inter-variable relationships and non-linear dependencies.

Adaptive Weight Learning Module. A unique feature of this study is the weight learning module, which uses a neural network architecture to dynamically identify the best model combinations. This module creates real-time weight allocations between the LSTM and XGBoost predictions by analyzing past statistical summaries and current meteorological conditions.

$$w_{lstm} = \frac{|\varepsilon_{xgb}|}{|\varepsilon_{lstm}| + |\varepsilon_{xgb}|}, w_{xgb} = \frac{|\varepsilon_{lstm}|}{|\varepsilon_{lstm}| + |\varepsilon_{xgb}|} \tag{1}$$

w'_{lstm}: LSTM model's weight, $w'_{xgboost}$: XGBoost model's weight, ε_{lstm}: LSTM model's training error, ε_{xgb} : XGBoost model's training error. By adjusting to changing meteorological circumstances and model performance characteristics, the weight learning method overcomes a basic drawback of fixed ensemble approaches. The system emphasizes XGBoost contributions when feature interactions become more important, while increasing LSTM influence during times when temporal patterns predominate.

Training Methodology. Three steps make up the training procedure, which is intended to maximize model performance on an individual basis while guaranteeing successful ensemble integration. The LSTM and XGBoost models are first trained separately on historical data, which enables each part to gain specialized knowledge in its own field.

Retrospective model performance analysis is then used to determine the best weight distributions for every historical case. This procedure generates a thorough training dataset for the weight learning module by identifying scenarios in which each model has exceptional predictive power.

The final stage involves training the weight allocation network to replicate optimal weighting decisions, enabling the system to generalize weight selection strategies to new atmospheric conditions. This approach ensures that the ensemble framework can adapt to previously unseen weather patterns while maintaining predictive accuracy.

$$\hat{y}_{final} = w_{lstm} \times \hat{y}_{lstm} + w_{xgb} \times \hat{y}_{xgb} \tag{2}$$

$\hat{y}_{final}$: final prediction result, w_{lstm}: weight of LSTM model, $\hat{y}_{lstm}$: LSTM models prediction result, w_{xgb}: weight of XGBoost model, $\hat{y}_{xgb}$: XGBoost models prediction result.

3 Experiment

3.1 Test Data

The experimental validation utilizes a comprehensive meteorological dataset sourced from Kaggle, encompassing daily weather observations from five strategically selected European cities: Düsseldorf, Basel, De Bilt, Dresden, and Budapest. This dataset spans 3,650 days (approximately 10 years) of continuous measurements, providing substantial temporal coverage for robust model training and evaluation [10]. The dataset incorporates seven key meteorological variables that collectively characterize atmospheric conditions in Table 1.

Table 1. Atmosphere conditions

Variable	unit
Cloud Cover	%
Humidity	0-1
Pressure	1000 hPa
Global Radiation	100 W/m²
Precipitation	10mm
Sunshine Hours	Hour
Mean/Min/Max Temperature	°C

The selected cities represent diverse climatic zones across Central and Western Europe, ensuring broad applicability of the developed framework. Basel and Düsseldorf exhibit continental climate characteristics, while De Bilt demonstrates oceanic influences. Dresden represents continental European conditions, and Budapest showcases transitional climate patterns between oceanic and continental regimes. This geographic distribution provides essential variability for testing model robustness across different atmospheric dynamics.

The dataset's reliability stems from several key factors. The decade-long observation period captures multiple seasonal cycles, including various weather extremes and anomalous conditions essential for comprehensive model training. The multi-variable approach ensures that complex atmospheric interactions are adequately represented, avoiding oversimplified single-parameter models.

The evaluation is based on results like RMSE, R^2 (how much of the temperature change it explains), and MSE (mean square error), and the five results in Table 2.

Table 2. Models results

City name \Metrics	RMSE	R^2	MSE
Basel	0.5191	0.9950	0.4141
De Bilt	0.5361	0.9925	0.3968
Düsseldorf	0.5475	0.9923	0.4274
Dresden	0.6304	0.9930	0.4392
Budapest	1.0187	0.9849	0.7834

3.2 Result

In Figs. 1-5, the first column is a prediction hitting map, the second column represents weather prediction curves, and the third column is prediction error distributions. Fig. 1-5 have the same structures and meaning.

Basel delivered the strongest predictive performance among the five European cities studied (Fig. 1), has achieved exceptionally low errors, indicating it captured nearly all temperature variation. Significantly, its optimal model weighting heavily favored XGBoost over the LSTM component, suggesting Basel's temperature patterns are more strongly driven by complex interactions among contemporaneous atmospheric features than by temporal sequences alone. Close behind Basel, De Bilt (Fig. 2) and Düsseldorf (Fig. 3) demonstrated similarly high accuracy with comparable weightings favoring XGBoost, reinforcing the approach's effectiveness across distinct locations. Dresden (Fig. 4) showed a slight but notable increase in reliance on the LSTM compared to the top performers, possibly reflecting a stronger time-dependent element in its local temperature dynamics. Budapest (Fig. 5) had higher errors and the most balanced weighting.

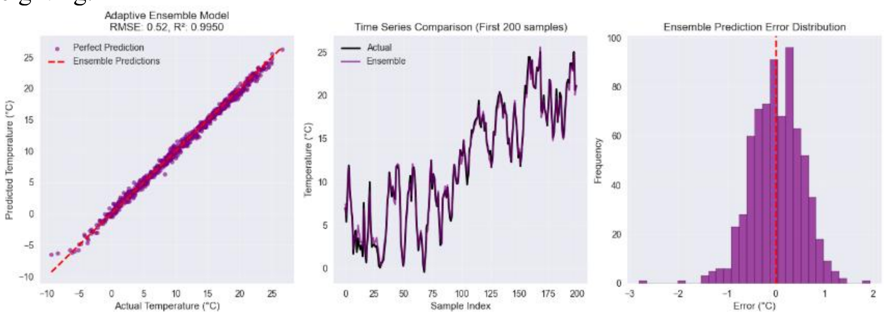

Fig. 1. Basel result (Photo/Picture credit: Original).

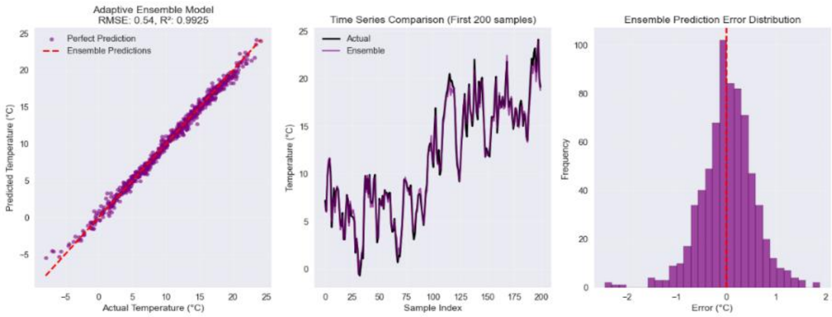

Fig. 2. De Bilt (Photo/Picture credit: Original).

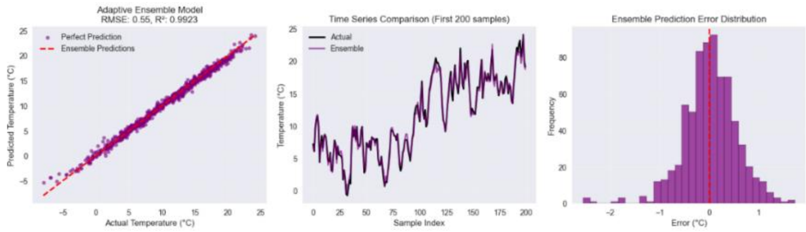

Fig. 3. Düsseldorf (Photo/Picture credit: Original).

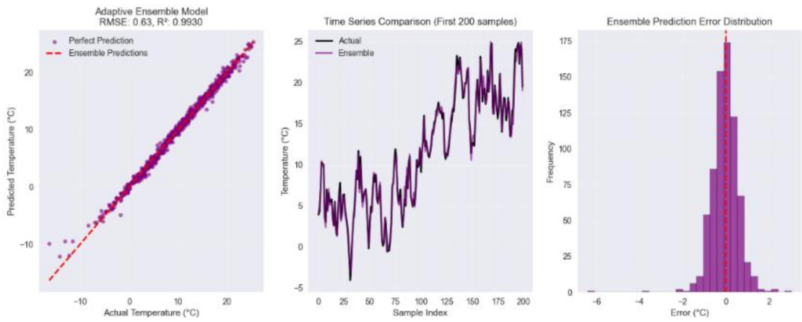

Fig. 4. Dresden (Photo/Picture credit: Original).

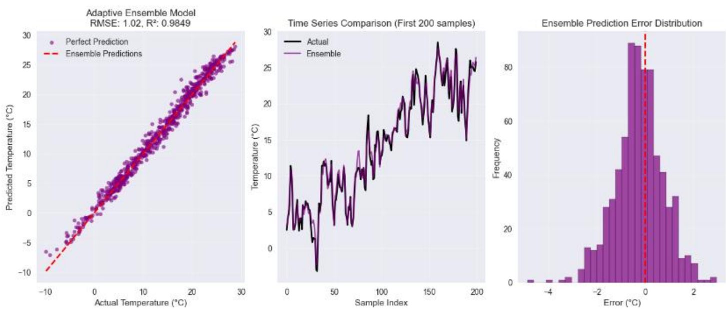

Fig. 5. Budapest (Photo/Picture credit: Original).

The remarkable consistency in performance across Basel, De Bilt, Düsseldorf, and Dresden highlights the model's particular suitability for mid-latitude European continental climates. This success likely stems from the inherent stability of these regional weather patterns. Compared to more coastal or extreme climates, central European temperatures experience more predictable seasonal progressions and relatively moderate, less volatile short-term fluctuations. The dominant influence of complex feature interactions (captured effectively by XGBoost) over intricate long-term temporal dependencies (the LSTM's focus) in these settings allows the adaptive weighting scheme to find a stable, highly effective configuration favoring XGBoost.

Table 3. Basel's testing result

	LSTM	XGBoost	Adaptive (The author)
RMSE	0.9825	0.4523	0.5191
R^2	0.9645	0.9923	0.9950

Further analysis of the results supports this robustness (Table 3). Inspection of the prediction error distributions across cities reveals a tight clustering near zero with minimal bias (no consistent over- or under-shooting). The temporal prediction curves (Fig. 2, etc.) show the ensemble model adeptly tracking both gradual seasonal shifts and more rapid weather changes with minimal lag. This consistent accuracy across diverse conditions within the European context signifies a model capable of reliable, operational forecasting year-round. The stable weighting preferences observed – consistently between 59.6% and 64.4% for XGBoost – across the top-performing cities (excluding Budapest) further validate the approach, indicating it identifies physically meaningful configurations rather than relying on arbitrary optimization.

4 Discussion

4.1 Model Analysis

Model Advantages. The proposed framework demonstrates several key advantages over conventional approaches. The dynamic weighting mechanism eliminates the need for manual parameter tuning, automatically adapting to local weather patterns and seasonal variations. This adaptability addresses a fundamental limitation identified in previous ensemble methods that rely on fixed combination strategies. The complementary nature of XGBoost and LSTM components proves particularly effective. Chen et al. demonstrated that XGBoost excels at capturing complex feature interactions in tabular data, achieving superior performance in scenarios with rich feature spaces [2].

Model Limitations. Despite its effectiveness, the current framework exhibits certain limitations. The reliance on historical data for weight optimization may reduce performance in regions with limited meteorological records. Additionally, the ensemble approach increases computational complexity compared to individual models, though this overhead remains manageable for operational deployment.

The framework's performance is constrained by the quality of input features and the representativeness of training data. Extreme weather events, which are inherently rare in historical datasets, may not be adequately captured by the current training methodology.

4.2 **Optimization Suggestions**

Several optimization strategies could enhance the framework's performance and applicability. Transfer learning techniques could address the limitation of insufficient historical data by leveraging knowledge from well-monitored stations to improve predictions in data-sparse regions. Future research could consider incorporating uncertainty quantification mechanisms to provide confidence intervals alongside point predictions.

4.3 Experimental Design Limitations

The current study focuses exclusively on temperature prediction across European cities with temperate climates. This geographical constraint limits the generalizability of findings to tropical, arctic, or desert climates where different atmospheric dynamics may prevail. The relatively short evaluation period may not capture long-term climate variations or detect potential model degradation over extended operational periods.

The evaluation methodology relies primarily on statistical metrics (RMSE, MSE, R^2) without considering practical implications such as forecast value for specific applications or user satisfaction. Additionally, the study does not address model interpretability, which is increasingly important for operational weather services.

4.4 Future Research Directions

Future investigations could consider multi-variable prediction capabilities to simultaneously forecast temperature, precipitation, humidity, and wind patterns. Quangao Liu et al. demonstrated that multi-output ensemble models can capture inter-variable dependencies more effectively than separate single-output models, achieving 12-18% improvement in overall prediction accuracy across multiple meteorological variables [11].

The integration of additional modeling paradigms represents another promising direction. Research could explore incorporating transformer architectures or graph neural networks to better capture spatial relationships between meteorological stations.

5 Conclusion

This research presents a novel adaptive ensemble framework that combines LSTM and XGBoost predictions through an intelligent weighting mechanism, achieving significant improvements in temperature forecasting accuracy.

The experimental results provide compelling evidence that intelligent model combination strategies substantially outperform individual model approaches. The consistent high performance across different urban environments establishes the framework's reliability and positions it among the most accurate deterministic machine learning forecasting methods currently available. These achievements represent a significant advancement in the field of weather prediction and establish a new performance benchmark for ensemble-based approaches.

Several promising directions emerge for future research. The framework could be extended to simultaneously predict multiple meteorological variables, potentially capturing cross-variable dependencies that enhance overall forecasting accuracy in different area and different weather conditions. The adaptive ensemble framework provides a robust and versatile foundation for next-generation weather prediction systems. The demonstrated accuracy improvements represent a substantial advancement toward more reliable and precise meteorological forecasting capabilities.

References

1. Chattopadhyay, A., Hassanzadeh, P., Subramanian, D.: Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods. Nonlinear Processes in Geophysics **27**(3), 373–389 (2020)
2. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
3. Grover, A., Kapoor, A., Horvitz, E.: A deep hybrid model for weather forecasting. In: Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 379–386 (2015)
4. Rasp, S., Lerch, S.: Neural networks for postprocessing ensemble weather forecasts. Monthly Weather Review **146**(11), 3885–3900 (2018)
5. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems, vol. 28, pp. 802–810 (2015)
6. Kim, S., Hong, S., Joh, M., Song, S.K.: DeepRain: ConvLSTM network for precipitation prediction using multichannel radar data. Geoscientific Model Development **13**(2), 631–647 (2020)
7. Liu, Y., Racah, E., Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., et al.: Application of deep convolutional neural networks for detecting extreme weather in climate datasets. arXiv:1605.01156, https://arxiv.org/abs/1605.01156, last accessed 2024/08/04
8. McGovern, A., Elmore, K.L., Gagne, D.J., Haupt, S.E., Karstens, C.D., Lagerquist, R., et al.: Using artificial intelligence to improve real-time decision-making for high-impact weather. Bulletin of the American Meteorological Society **100**(10), 2073–2090 (2019)
9. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997)
10. Liu, Q., Li, R., Jiang, M., Yang, W., Liang, C., Pang, L., Zou, Z.: MSTVI: Multi-scale time-variable interaction for multivariate time series forecasting. Knowledge-Based Systems 319, 113551 (2025)
11. Florian-Huber: Weather prediction dataset. https://github.com/florian-huber/weather_prediction_dataset, last accessed 2024/08/04