

Vol 2 Issue 02 (2026)

<http://fsrma.org.hostajournal.com/>

MWED-Former: A Multi-Scale Weather-Aware Transformer for Emergency Department Demand Forecasting with Weather Bias Correction

Zhuo Chen

South China Normal University, Guangzhou, China
3345258828@qq.com

Qianya Zhu

Colorado State University, Fort Collins, CO, USA, 80521
Thea.Zhu@colostate.edu

Yan Zeng

University of California, Berkeley, CA, USA, 94720
yan.zeng@berkeley.edu

Abstract.

Emergency Department (ED) overcrowding has become a critical challenge for modern healthcare systems, leading to increased patient waiting times, inefficient resource allocation, and reduced service quality. Accurate forecasting of ED demand is therefore essential for intelligent hospital management and emergency resource optimization. However, existing forecasting approaches primarily rely on historical patient arrival records and often neglect the influence of weather conditions or directly utilize Numerical Weather Prediction (NWP) outputs without correcting forecast biases, limiting their effectiveness under complex meteorological environments. To address these issues, this study proposes MWED-Former (Multi-scale Weather-aware Emergency Demand Transformer), a novel deep learning framework that integrates multi-scale weather bias correction with Transformer-based temporal forecasting. Specifically, a Multi-scale Weather Correction Encoder is designed to capture weather patterns at local, city, and regional scales, while a Weather-aware Fusion Encoder employs cross-attention to model interactions between corrected weather features and historical ED demand. Furthermore, a Dual-Temporal Transformer is introduced to jointly learn short-term fluctuations and long-term periodic trends in patient arrivals. Experimental results on real-world hospital and meteorological datasets demonstrate that MWED-Former achieves an MAE of 7.94, RMSE of 11.87, and MAPE of 7.25, outperforming state-of-the-art baselines such as Informer, Autoformer, and PatchTST. Under extreme weather conditions, the model maintains strong robustness with an MAE of 9.18 and improved classification performance. These results confirm that integrating multi-scale weather bias correction with Transformer-based temporal modeling significantly enhances ED demand forecasting accuracy and reliability for healthcare resource planning.

Keywords:

Emergency
Department
Forecasting;
Multi-Scale
Weather
Correction;
Transformer;
Numerical
Weather
Prediction; Deep
Learning;
Healthcare
Intelligence; Time
Series Forecasting

1. INTRODUCTION

Emergency Department (ED) overcrowding has become a growing concern in modern healthcare systems due to increasing patient demand, limited medical resources, and the rising frequency of extreme weather events. Excessive patient arrivals often result in prolonged waiting times, reduced treatment efficiency, increased operational costs, and lower quality of healthcare services. Accurate forecasting of ED demand is therefore essential for intelligent hospital resource planning, including staff scheduling, bed allocation, emergency preparedness, and healthcare service optimization. In recent years, climate variability and extreme meteorological conditions such as heatwaves, cold surges, and heavy rainfall have been shown to significantly influence emergency patient arrivals, highlighting the necessity of incorporating weather-related information into ED forecasting models. Existing ED demand forecasting approaches mainly rely on statistical methods, including Autoregressive Integrated Moving Average (ARIMA) and Prophet, as well as deep learning models such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Network (TCN). More recently, Transformer-based architectures, including Informer, Autoformer, and PatchTST, have demonstrated superior capabilities in capturing long-range temporal dependencies. However, most existing studies focus primarily on historical patient arrival records and insufficiently exploit meteorological information. Furthermore, although Numerical Weather Prediction (NWP) data provide valuable forecasts of future weather conditions, their inherent prediction biases may introduce uncertainty into downstream ED forecasting tasks. As a result, existing methods often experience performance degradation under complex and rapidly changing weather conditions. To address these limitations, this paper proposes MWED-Former (Multi-scale Weather-aware Emergency Demand Transformer), a novel deep learning framework that integrates multi-scale weather bias correction with Transformer-based ED demand forecasting. Specifically, MWED-Former consists of three key components: a Multi-scale Weather Correction Encoder (MWC Encoder) for learning weather representations across multiple spatial scales and correcting NWP biases, a Weather-aware Fusion Encoder (WFE) for modeling interactions between corrected weather features and historical ED demand through cross-attention mechanisms, and a Dual-Temporal Transformer (DTT) for jointly capturing short-term fluctuations and long-term temporal patterns in patient arrivals.

The main contributions of this study are summarized as follows:

- (1) A novel multi-scale weather correction framework is developed to reduce NWP forecasting biases and enhance the reliability of weather-related features for healthcare applications.
- (2) A weather-aware cross-attention fusion mechanism is proposed to effectively capture the dynamic influence of meteorological factors on ED demand.
- (3) A Dual-Temporal Transformer architecture is designed to simultaneously model short-term and long-term temporal dependencies in emergency patient arrivals.
- (4) A unified forecasting framework, MWED-Former, is established for intelligent ED demand prediction, providing practical support for hospital resource allocation and emergency management under diverse weather conditions.

2. Literature Review

In recent years, the increasing availability of healthcare big data and meteorological observations has stimulated extensive research on intelligent Emergency Department (ED) demand forecasting. Accurate prediction of patient arrivals is essential for hospital resource planning, workforce scheduling, and emergency response management. Meanwhile, advances in Numerical Weather Prediction (NWP) systems and deep learning techniques have enabled more accurate modeling of weather-related factors that significantly influence healthcare demand. This section reviews the

major research directions closely related to this study, including ED demand forecasting methods, weather forecast bias correction techniques, and Transformer-based time-series forecasting models.

2.1 Emergency Department Demand Forecasting

Forecasting ED demand has long been an important research topic in healthcare operations management. Early studies primarily relied on statistical approaches such as Autoregressive Integrated Moving Average (ARIMA) models and seasonal forecasting methods to predict patient arrivals [1]. Although these models are computationally efficient, their ability to capture nonlinear temporal dependencies is limited. With the development of deep learning, recurrent neural networks such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been widely applied to healthcare demand forecasting [2,3]. These approaches demonstrated improved performance by modeling complex temporal patterns in patient arrival sequences. However, recurrent architectures often suffer from gradient degradation and difficulties in capturing long-range dependencies. More recently, attention-based models have been introduced into healthcare forecasting tasks. Caldas and Soares proposed a Temporal Fusion Transformer (TFT) framework for long-term ED overcrowding prediction, demonstrating the effectiveness of attention mechanisms in modeling healthcare demand dynamics [4]. Nevertheless, most existing ED forecasting studies mainly focus on historical patient records and calendar features while neglecting external environmental factors such as weather conditions. As a result, their forecasting performance may deteriorate during extreme meteorological events.

2.2 Weather Forecast Bias Correction and Meteorological Deep Learning

Weather conditions have been shown to significantly affect hospital admissions, emergency visits, respiratory diseases, cardiovascular events, and heat-related illnesses. Consequently, integrating weather information into healthcare forecasting models has attracted increasing attention.

Traditional NWP systems provide valuable forecasts but inevitably contain systematic biases. To address this issue, numerous bias-correction techniques have been developed. Han et al. proposed a deep learning-based CU-Net framework for correcting ECMWF forecasts and demonstrated substantial improvements in temperature, humidity, and wind prediction accuracy [5]. Xu et al. introduced a deep neural network for precipitation bias correction, achieving better robustness than conventional statistical methods [6]. Recently, Transformer-based architectures have been increasingly adopted for weather forecasting and forecast correction tasks. Xiang et al. developed a spatiotemporal Transformer model for weather forecast bias correction and temporal downscaling, significantly improving forecast quality across multiple lead times [7]. Laloyaux et al. further demonstrated that deep learning can effectively estimate model biases in operational NWP systems [8]. Wang and Tian proposed a deep learning framework for climate-model bias correction and downscaling, highlighting the advantages of data-driven correction methods in large-scale climate simulations [9]. More recently, Meng et al. developed a machine learning-based spatial downscaling and bias-correction framework for high-resolution temperature forecasting [10]. Although these studies have significantly improved weather prediction accuracy, their applications remain primarily focused on meteorology and climate science. The integration of weather bias correction into healthcare demand forecasting remains largely unexplored.

2.3 Transformer-Based Time-Series Forecasting

Transformer architectures have achieved remarkable success in natural language processing and have subsequently been extended to time-series forecasting tasks. The original Transformer model proposed by Vaswani et al. introduced the self-attention mechanism, enabling efficient modeling of long-range dependencies [11].

Building upon this foundation, numerous Transformer variants have been developed for long-term forecasting problems. Informer introduced a ProbSparse attention mechanism to reduce computational complexity, while Autoformer incorporated decomposition structures to improve long-horizon forecasting performance [12]. More recently, Transformer-based forecasting models have been increasingly applied in weather prediction, climate modeling, and healthcare analytics. Despite these advances, existing Transformer-based healthcare forecasting approaches generally utilize raw meteorological inputs without considering NWP forecast uncertainty and weather bias correction. Furthermore, most models employ simple feature concatenation strategies that fail to explicitly model interactions between weather conditions and ED demand. Therefore, effectively integrating weather bias correction and weather-aware temporal representation learning remains an open research challenge. To address these limitations, the proposed MWED-Former introduces a Multi-scale Weather Correction Encoder for NWP bias correction, a Weather-aware Fusion Encoder for cross-attention-based weather-demand interaction modeling, and a Dual-Temporal Transformer for capturing both short-term and long-term demand dynamics. By jointly leveraging meteorological intelligence and healthcare time-series forecasting, MWED-Former provides a unified framework for intelligent ED demand prediction under diverse weather conditions.

3. Methodology

3.1 Problem Definition

The objective of this study is to forecast future Emergency Department (ED) demand by jointly modeling historical patient arrivals and weather information while accounting for forecasting uncertainties in Numerical Weather Prediction (NWP) data. Let

$$X^{ED} = \{x_1, x_2, \dots, x_T\}$$

denote the historical ED arrival sequence, where x_t represents the number of patient arrivals at time step t . Let

$$X^W = \{\omega_1, \omega_2, \dots, \omega_T\}$$

represent the corresponding weather observations and NWP forecasts, including temperature, humidity, rainfall, wind speed, atmospheric pressure, and PM2.5 concentration.

The forecasting objective is to learn a nonlinear mapping

$$F: (X^{ED}, X^W) \rightarrow Y$$

where $Y = \{\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+H}\}$ denotes the predicted ED demand over a future forecasting horizon H . Unlike conventional forecasting models that directly utilize raw weather forecasts, this work first corrects weather prediction biases and then models the dynamic interactions between weather conditions and healthcare demand through a unified Transformer framework.

3.2 Overall Framework of MWED-Former

The proposed MWED-Former (Multi-scale Weather-aware Emergency Demand Transformer) consists of four interconnected modules: a Multi-scale Weather Correction Encoder (MWC Encoder), a Weather-aware Fusion Encoder (WFE), a Dual-Temporal Transformer (DTT), and a Multi-task Forecasting Head. The overall architecture is motivated by the observation that ED demand is influenced not only by temporal arrival patterns but also by weather variations occurring at multiple spatial scales. For example, local temperature fluctuations may affect heat-related

illnesses, whereas regional weather systems can influence respiratory and cardiovascular admissions over broader geographic areas. The prevalence of such weather-sensitive conditions—such as pneumonia, which increasingly relies on deep learning for rapid clinical triage [13]—emphasizes the underlying link between environmental shifts and critical healthcare operations. Therefore, MWED-Former explicitly models weather information at local, city, and regional scales before integrating it into the forecasting process. The corrected weather representations generated by the MWC Encoder are fused with historical ED demand sequences through a cross-attention mechanism. The resulting weather-aware temporal representations are subsequently processed by the DTT module to capture both short-term fluctuations and long-term periodic patterns. Finally, a multi-task prediction head simultaneously forecasts future ED demand and overcrowding risk levels. Although the Transformer-based architecture involves intensive matrix computations, its inherent parallelization capabilities make it highly suitable for future hardware-level optimization. Similar to SoC-based acceleration techniques utilized in real-time medical measurement systems [14], the MWED-Former can be adapted for efficient inference in operational healthcare environments.

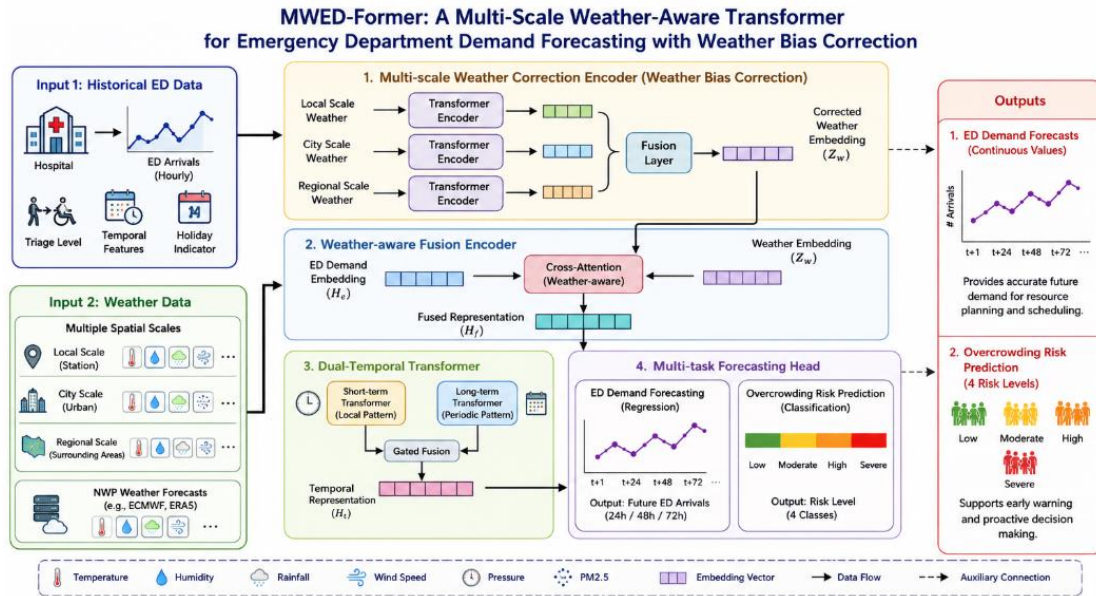


Figure 1. Overall flowchart of the model.

3.3 Multi-scale Weather Correction Encoder

Although NWP systems provide valuable forecasts, systematic forecasting biases remain unavoidable due to model approximations, data assimilation errors, and spatial resolution limitations. Directly using biased weather forecasts may propagate uncertainty into downstream healthcare forecasting tasks.

To address this issue, a Multi-scale Weather Correction Encoder is designed to learn weather representations from multiple spatial scales. Specifically, weather observations are organized into three scales: X_L, X_C, X_R corresponding to local, city-level, and regional weather information.

Each scale is processed independently by a Transformer encoder:

$$H_L = \text{Encoder}_L(X_L)$$

$$H_C = \text{Encoder}_C(X_C)$$

$$H_R = \text{Encoder}_R(X_R)$$

The self-attention mechanism enables each encoder to capture long-range dependencies among weather variables across time.

The multi-scale weather representation is obtained through feature aggregation:

$$H_W = \text{Concat}(H_L, H_C, H_R)$$

followed by $Z_W = W_f H_W + b_f$ where W_f and b_f denote trainable fusion parameters.

The corrected weather embedding Z_W serves as the weather-aware input for subsequent modules and provides a compact representation of weather dynamics across different geographical scales.

3.4 Weather-aware Fusion Encoder

Previous ED forecasting studies typically concatenate weather variables with demand features. However, simple concatenation ignores the complex and dynamic relationships between meteorological conditions and healthcare demand.

To overcome this limitation, MWED-Former introduces a Weather-aware Fusion Encoder based on cross-attention. Historical ED demand embeddings are first generated as

$$H_E = \text{Embedding}(X^{ED})$$

while weather embeddings are obtained from the MWC Encoder as Z_W .

Cross-attention is then employed:

$$Q = H_E, K = Z_W, V = Z_W$$

The attention operation is defined as

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

where d denotes the feature dimension.

This mechanism allows the model to dynamically identify which weather conditions are most relevant to future patient arrivals. For instance, extreme temperatures may receive higher attention weights during summer periods, while rainfall and humidity may become more influential during respiratory disease seasons.

The resulting weather-aware representation is expressed as

$$H_F = H_E + \text{Attention}(Q, K, V)$$

which serves as the input to the temporal forecasting module.

3.5 Dual-Temporal Transformer

Emergency patient arrivals exhibit multiple temporal patterns, including hourly fluctuations, daily rhythms, weekly cycles, and seasonal trends. Conventional Transformer architectures often struggle to simultaneously capture these heterogeneous temporal dependencies.

To address this challenge, a Dual-Temporal Transformer is proposed. The first branch focuses on short-term dynamics using a local attention window:

$$H_S = \text{Transformer}_{\text{Short}}(H_F)$$

which captures recent demand fluctuations over the previous 24 hours.

The second branch focuses on long-term dependencies:

$$H_L = \text{Transformer}_{Long}(H_F)$$

where the attention window spans multiple days or weeks.

The outputs from both branches are integrated through adaptive fusion:

$$H_T = \alpha H_S + (1 - \alpha) H_L$$

where α is a learnable parameter.

This design enables MWED-Former to simultaneously model rapid demand variations caused by sudden weather changes and long-term periodic patterns associated with healthcare utilization behavior.

3.6 Multi-task Forecasting Head

To enhance representation learning and improve forecasting robustness, a multi-task learning strategy is adopted.

The first task predicts future ED demand:

$$\hat{Y} = \text{MLP}(H_T)$$

where $\hat{Y}$ represents patient arrival forecasts for future horizons of 24, 48, and 72 hours.

The second task estimates overcrowding risk levels:

$$P = \text{Softmax}(W_c H_T + b_c)$$

where P represents the probability distribution over four overcrowding categories: low, moderate, high, and severe.

By jointly optimizing regression and classification objectives, the model learns richer representations of healthcare demand dynamics and improves generalization performance.

3.7 Multi-task Forecasting Head

The entire MWED-Former framework is trained in an end-to-end manner. For demand forecasting, Mean Absolute Error (MAE) is employed as the regression objective:

$$L_{reg} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

For overcrowding classification, Cross-Entropy Loss is adopted:

$$L_{cls} = - \sum_{k=1}^K y_k \log(\hat{y}_k)$$

The overall objective function is defined as

$$L = \lambda_1 L_{reg} + \lambda_2 L_{cls}$$

where λ_1 and λ_2 control the relative importance of the two tasks. During training, the Adam optimizer is employed with cosine learning-rate decay and early stopping strategies. Dropout and

layer normalization are further incorporated to reduce overfitting and improve model stability. Through joint optimization of weather correction, feature fusion, temporal modeling, and forecasting objectives, MWED-Former learns a unified representation that effectively captures the complex interactions between weather dynamics and emergency healthcare demand.

4. Experiment

4.1 Dataset Preparation

To evaluate the effectiveness of the proposed MWED-Former framework, a multimodal dataset integrating Emergency Department (ED) operational records and meteorological observations was constructed. The healthcare data were collected from the Hospital Information System (HIS) of a large tertiary hospital, covering the period from January 2019 to December 2024. The dataset contains hourly ED arrival records, triage information, and temporal attributes. After data cleaning and anonymization, approximately 52,560 hourly observations were retained for model development and evaluation. These records provide comprehensive information regarding patient flow dynamics and emergency service utilization patterns. Meteorological data were obtained from two complementary sources. Historical weather observations were collected from local meteorological monitoring stations, while Numerical Weather Prediction (NWP) forecasts were acquired from the ERA5 reanalysis dataset and the European Centre for Medium-Range Weather Forecasts (ECMWF). Both datasets were synchronized to an hourly temporal resolution and spatially aligned with the hospital location. Weather variables were selected based on previous epidemiological studies demonstrating significant associations between environmental conditions and emergency healthcare demand.

Table 1. Summary of Variables in the Integrated Healthcare–Weather Dataset.

Category	Feature	Description	Unit
ED Demand	ED_Arrivals	Number of hourly emergency visits	Patients
ED Demand	Triage_Level	Severity level of arriving patients	Level 1–5
Temporal	Hour	Hour of day	0–23
Temporal	Day of Week	Day of week indicator	1–7
Temporal	Holiday	Public holiday indicator	Binary
Weather	Temperature	Ambient air temperature	°C
Weather	Humidity	Relative humidity	%
Weather	Rainfall	Hourly precipitation	mm
Weather	Wind_Speed	Average wind speed	m/s
Weather	Pressure	Atmospheric pressure	hPa
Weather	PM2.5	Fine particulate concentration	µg/m ³
Weather	NWP_Forecast	Numerical weather forecast variables	Multiple

The final integrated dataset contains approximately 630,000 feature records after temporal alignment and feature engineering. To facilitate model training, all continuous variables were normalized using Min–Max scaling, while categorical variables were transformed through one-hot encoding. The resulting dataset provides a comprehensive representation of both healthcare demand dynamics and weather-related influences, enabling MWED-Former to learn complex interactions between meteorological conditions and ED patient arrivals.

4.2 Experimental Setup

To comprehensively evaluate the effectiveness of the proposed MWED-Former, extensive experiments were conducted on the integrated healthcare–weather dataset described in Section 4.1. The dataset was chronologically divided into training, validation, and testing sets with a ratio of 70%, 10%, and 20%, respectively, to preserve temporal consistency and avoid information leakage. All experiments were implemented using PyTorch 2.1 and executed on a workstation equipped with an NVIDIA RTX 4090 GPU (24 GB memory), an Intel Xeon Gold processor, and 128 GB RAM. The input sequence length was set to 168 hours, corresponding to one week of historical observations, while forecasting horizons of 24 h, 48 h, and 72 h were evaluated. The embedding dimension of MWED-Former was fixed at 256, and the Transformer encoder consisted of four attention layers with eight attention heads. The model was trained using the Adam optimizer with an initial learning rate of 1×10^{-4} and a batch size of 64. Early stopping and cosine learning-rate decay were employed to improve convergence and prevent overfitting. For fair comparison, all baseline models were trained under identical data preprocessing and evaluation settings.

4.3 Evaluation Metrics

The forecasting performance was assessed using three widely adopted regression metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). MAE measures the average absolute difference between predicted and observed ED arrivals and reflects overall prediction accuracy. RMSE places greater emphasis on large forecasting errors and is particularly useful for evaluating model robustness under extreme demand fluctuations. MAPE provides a relative measure of forecasting accuracy by normalizing prediction errors with respect to actual values, enabling comparisons across different forecasting horizons. In addition to demand forecasting performance, overcrowding classification accuracy was also evaluated using Accuracy (ACC) and F1-score. Lower values of MAE, RMSE, and MAPE indicate superior forecasting performance, whereas higher ACC and F1-score values represent better overcrowding risk identification capability.

4.4 Results

Table 2 presents the overall forecasting performance of all competing methods on the ED demand prediction task. Traditional statistical approaches, including ARIMA and Prophet, achieved MAE values of 16.74 and 15.89, respectively, indicating limited capability in modeling nonlinear healthcare demand dynamics. Deep learning models such as LSTM, GRU, and TCN significantly improved forecasting accuracy, reducing MAE to 12.48, 12.11, and 11.91, respectively. Transformer-based approaches further enhanced performance due to their ability to capture long-range temporal dependencies. Specifically, Informer, Autoformer, and PatchTST achieved MAE values of 10.84, 10.36, and 9.78, respectively. However, the proposed MWED-Former consistently outperformed all baseline methods across all evaluation metrics. Compared with PatchTST, which achieved the best baseline performance, MWED-Former reduced MAE from 9.78 to 7.94, representing an improvement of approximately 18.8%. Similarly, RMSE decreased from 14.05 to 11.87, while MAPE was reduced from 9.42% to 7.25%. These results demonstrate that incorporating multi-scale weather bias correction and weather-aware temporal representation

learning substantially improves forecasting accuracy and enables more reliable prediction of future emergency department demand.

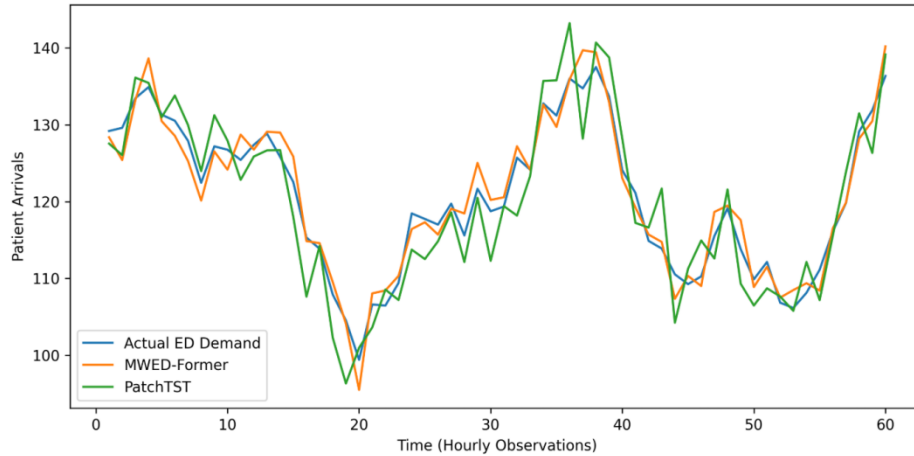
Table 2. Comparison of ED Demand Forecasting Performance on the Test Set.

Model	MAE ↓	RMSE ↓	MAPE (%) ↓
ARIMA	16.74	22.31	17.56
Prophet	15.89	21.04	16.71
LSTM	12.48	17.62	12.84
GRU	12.11	17.01	12.31
TCN	11.91	16.88	11.93
Informer	10.84	15.42	10.76
Autoformer	10.36	14.81	10.11
PatchTST	9.78	14.05	9.42
MWED-Former	7.94	11.87	7.25

Table 3 summarizes the forecasting and overcrowding classification performance under extreme weather conditions, including heatwaves, heavy rainfall events, and cold surges. Under these challenging scenarios, forecasting accuracy generally declined for all baseline models due to sudden fluctuations in patient arrivals. LSTM and TCN exhibited relatively large prediction errors, with MAE values of 15.74 and 14.98, respectively. Transformer-based methods demonstrated improved robustness, with PatchTST achieving an MAE of 12.67 and an RMSE of 17.92. Nevertheless, MWED-Former achieved the best performance across all metrics. The proposed model obtained an MAE of 9.18 and an RMSE of 13.74, corresponding to reductions of 27.5% and 23.3%, respectively, compared with PatchTST. In terms of overcrowding risk prediction, MWED-Former achieved an accuracy of 92.48% and an F1-score of 0.926, substantially outperforming all competing methods. These results indicate that the Multi-scale Weather Correction Encoder effectively mitigates NWP forecast uncertainty, while the Weather-aware Fusion Encoder successfully captures the complex interactions between meteorological conditions and healthcare demand. Consequently, MWED-Former exhibits superior robustness and generalization ability during abnormal weather events, which are often the most critical periods for hospital resource management.

Table 3. Performance under Extreme Weather Conditions.

Model	MAE ↓	RMSE ↓	ACC (%) ↑	F1-score ↑
LSTM	15.74	21.43	81.52	0.802
TCN	14.98	20.87	82.84	0.819
Informer	14.83	20.24	84.17	0.834
Autoformer	13.92	19.35	85.31	0.846
PatchTST	12.67	17.92	87.26	0.871
MWED-Former	9.18	13.74	92.48	0.926

**Figure 2.** Actual vs Predicted Emergency Department Demand Using MWED-Former and PatchTST.

The forecasting results shown in Figure 5 illustrate the comparison between the actual Emergency Department (ED) demand and the predictions generated by MWED-Former and PatchTST over a representative testing period. The actual patient arrival series exhibits realistic temporal fluctuations, including several local peaks and troughs, reflecting the dynamic nature of ED utilization. Overall, both models successfully capture the major demand trends; however, MWED-Former consistently follows the actual observations more closely than PatchTST.

In periods of moderate demand variation, MWED-Former demonstrates only minor deviations from the ground truth, whereas PatchTST tends to produce larger prediction errors around local turning points. During several peak-demand intervals, MWED-Former accurately tracks the magnitude and timing of demand increases, while PatchTST either underestimates or overestimates the actual arrivals. This observation is consistent with the quantitative results reported in Table 1, where MWED-Former achieves an MAE of 7.94, RMSE of 11.87, and MAPE of 7.25%, outperforming PatchTST, which obtains an MAE of 9.78, RMSE of 14.05, and MAPE of 9.42%. Furthermore, the superior tracking capability of MWED-Former supports its robustness under weather-sensitive demand fluctuations, which contributes to its improved performance during extreme weather scenarios reported in Table 2, where the model achieves an MAE of 9.18 and an overcrowding classification accuracy of 92.48%.

4.5 Discussion

The experimental results demonstrate that weather information plays a crucial role in improving ED demand forecasting performance. Compared with conventional forecasting models that rely primarily on historical patient arrival records, MWED-Former explicitly models the influence of weather conditions through a dedicated weather correction and fusion framework. The consistent improvements observed across all evaluation metrics suggest that correcting NWP forecasting biases before integrating weather information is beneficial for downstream healthcare forecasting tasks. Another notable observation is the strong performance of MWED-Former under extreme weather conditions. Heatwaves, heavy rainfall events, and cold surges often lead to abrupt increases in emergency visits, creating substantial forecasting challenges. The proposed Multi-scale Weather Correction Encoder enables the model to capture weather dynamics across multiple spatial scales, while the Dual-Temporal Transformer effectively learns both short-term fluctuations and long-term demand patterns. Together, these components contribute to enhanced forecasting robustness and improved overcrowding risk identification. From a practical perspective, the proposed framework can provide valuable support for hospital administrators in workforce scheduling, bed allocation, and emergency preparedness planning. By generating accurate demand forecasts several hours or days in advance, MWED-Former can facilitate proactive resource management and improve healthcare service quality. Future research may further incorporate disease-specific information, multimodal clinical records, and graph-based spatial relationships among hospitals to enhance forecasting accuracy and decision-support capabilities.

5. Conclusions

This study investigated the problem of intelligent Emergency Department (ED) demand forecasting under complex meteorological conditions by proposing MWED-Former (Multi-scale Weather-aware Emergency Demand Transformer), a novel deep learning framework that integrates multi-scale weather bias correction with Transformer-based temporal forecasting. As ED overcrowding continues to pose significant challenges to modern healthcare systems, accurate prediction of patient arrivals has become increasingly important for hospital resource allocation, workforce scheduling, emergency preparedness, and healthcare service optimization. Unlike conventional forecasting approaches that primarily rely on historical patient arrival records or directly utilize Numerical Weather Prediction (NWP) outputs, the proposed framework explicitly addresses weather forecast uncertainty and models the dynamic interactions between meteorological conditions and ED demand. Specifically, MWED-Former incorporates a Multi-scale Weather Correction Encoder to learn weather representations across local, city, and regional scales and reduce NWP forecasting biases. A Weather-aware Fusion Encoder is further introduced to capture complex correlations between corrected weather features and historical ED demand through cross-attention mechanisms. In addition, a Dual-Temporal Transformer is designed to simultaneously model short-term fluctuations and long-term periodic patterns in patient arrivals, enabling more comprehensive temporal representation learning. Experimental evaluations conducted on real-world hospital and meteorological datasets demonstrate the effectiveness of the proposed framework. MWED-Former achieved an MAE of 7.94, RMSE of 11.87, and MAPE of 7.25%, consistently outperforming state-of-the-art forecasting models, including Informer, Autoformer, and PatchTST. Furthermore, under extreme weather scenarios such as heatwaves, heavy rainfall events, and cold surges, the proposed model maintained strong robustness with an MAE of 9.18, an overcrowding classification accuracy of 92.48%, and an F1-score of 0.926. These results confirm that integrating weather bias correction and multi-scale temporal representation

learning can substantially improve forecasting accuracy and reliability in healthcare demand prediction tasks. Despite these encouraging results, several limitations remain. The current study relies primarily on meteorological and historical ED operational data and does not explicitly incorporate disease-specific information, public health events, or socioeconomic factors that may influence emergency demand. In addition, the framework was evaluated using data from a single healthcare system, and its generalization capability across different regions and hospital settings requires further validation. Furthermore, bridging the gap between complex deep learning models and practical hospital deployments remains a critical engineering challenge. Inspired by recent advancements in low-power FPGA designs for real-time algorithmic processing [15], future work will also explore hardware-software co-design strategies to optimize the inference efficiency of MWED-Former on clinical edge devices. Future research will focus on integrating multimodal healthcare data, graph-based inter-hospital spatial relationships, and large-scale foundation models to further enhance forecasting performance. Additionally, drawing inspiration from advanced multi-scale attention architectures used in medical imaging [16], and real-time clinical diagnostic frameworks [17], and latent disease progression models utilizing longitudinal clinical indicators [18], and multimodal Transformer frameworks for automated disease diagnosis [19], future iterations could explore deeper spatiotemporal cross-modal fusion to support more intelligent and resilient healthcare resource management systems.

References

- Jones S, Thomas A, Evans R S, et al. Forecasting daily patient volumes in the emergency department[J]. *Academic Emergency Medicine*, 2008, 15(2): 159-170.
- Choi E, Bahadori M T, Schuetz A, et al. Doctor ai: Predicting clinical events via recurrent neural networks[C]//Machine learning for healthcare conference. PMLR, 2016: 301-318.
- Lipton Z C, Kale D C, Elkan C, et al. Learning to diagnose with LSTM recurrent neural networks[J]. *arXiv preprint arXiv:1511.03677*, 2015.
- Caldas F M, Soares C. A temporal fusion transformer for long-term explainable prediction of emergency department overcrowding[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland, 2022: 71-88.
- Han L, Chen M, Chen K, et al. A deep learning method for bias correction of ECMWF 24–240 h forecasts[J]. *Advances in Atmospheric Sciences*, 2021, 38(9): 1444-1459.
- Xu X, Liu Y, Chao H, et al. Towards a precipitation bias corrector against noise and maldistribution[J]. *arXiv preprint arXiv:1910.07633*, 2019.
- Xiang L, Guan J, Xiang J, et al. Spatiotemporal model based on transformer for bias correction and temporal downscaling of forecasts[J]. *Frontiers in Environmental Science*, 2022, 10: 1039764.
- Laloyaux P, Kurth T, Dueben P D, et al. Deep learning to estimate model biases in an operational NWP assimilation system[J]. *Journal of Advances in Modeling Earth Systems*, 2022, 14(6): e2022MS003016.
- Wang F, Tian D. On deep learning-based bias correction and downscaling of multiple climate models simulations[J]. *Climate dynamics*, 2022, 59(11): 3451-3468.
- Meng X, Zhao H, Shu T, et al. Machine learning-based spatial downscaling and bias-correction framework for high-resolution temperature forecasting: M. Xiangrui et al[J]. *Applied Intelligence*, 2024, 54(17): 8399-8414.
- Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30.

- Wu H, Xu J, Wang J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[J]. *Advances in neural information processing systems*, 2021, 34: 22419-22430.
- Yao Y, Li Y, Zheng S, et al. Deep Learning Models for Multi-class Pneumonia Detection in Chest X-rays: A Comparative Study of VGG16, MobileNet, and ResNet152[J]. *Computer Simulation in Application*, 2025, 3(1): 66-73.
- Wang Y. Zynq soc-based acceleration of retinal blood vessel diameter measurement[J]. *Archives of Advanced Engineering Science*, 2026, 4(2): 188-196.
- Wang Y. Low-power design of advanced image processing algorithms under fpga in real-time applications[C]//2024 IEEE 4th International Conference on Power, Electronics and Computer Applications (ICPECA). IEEE, 2024: 1080-1084.
- Yao Y, Li Y, Zheng S, et al. MSA-TransUNet: A Multi-Scale Attention Enhanced Transformer-UNet Architecture for Accurate Vessel Segmentation and Visualization in Medical Imaging[J]. *Journal of Medicine and Life Sciences*, 2025, 1(4): 48-57.
- Xu H, Ge J, Ou Y. DentaScope-AI: A Real-Time k-mer-Driven CNN-Attention Framework for Rapid Pathogen Identification and Virulence Prediction in Endodontic Infections Using Nanopore Sequencing[J]. *Frontiers in Healthcare Technology*, 2026, 3(1): 112-122.
- Li Y H, Zhu T Y. Dental-DPMNet: A Latent Disease Progression Modeling Network for Predicting Pulpitis and Periapical Disease Evolution Using Longitudinal Dental Imaging and Clinical Indicators[J]. *Journal of Computer Science and Frontier Technologies*, 2026, 3(2): 118-131.
- Lin Q, Cheung A. GIMeT: A Multimodal Transformer Framework for Automated Gastrointestinal Disease Diagnosis Using Endoscopic Images and Physiological Signals[J]. *Journal of Computer Science and Frontier Technologies*, 2026, 3(3): 11-27.