



OPEN

## A meta-learning ensemble framework for robust and interpretable prediction of emergency medical services demand

Tripti Garg<sup>1</sup>✉, Durga Toshniwal<sup>1</sup> & Manoranjan Parida<sup>2</sup>

Accurate and robust forecasting of Emergency Medical Services (EMS) demand is crucial for ensuring timely ambulance dispatch and efficient resource allocation, particularly in low-resource public health systems, such as those in India. While most prior EMS forecasting studies have focused on urban settings in developed countries with rich, granular data, limited research has explored district-level forecasting using real-world ambulance dispatch data from India. Moreover, existing models often trade off robustness for accuracy or rely on complex black-box architectures, limiting their interpretability and real-world deployment. This study examines whether a heterogeneous ensemble of interpretable and complementary learners can outperform traditional and state-of-the-art regressors for district-level EMS forecasting, utilizing limited real-world features. To address this challenge, we propose EM-LR (Ensembled Meta-Learner with Linear Regression), a meta-learning framework that integrates four diverse base models—Lasso Regression, Support Vector Regression (SVR), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGB)—via a linear regression meta-learner. Unlike prior meta-learners that stack similar tree-based or linear models, EM-LR combines low-variance, diverse learners to enhance robustness while maintaining model interpretability through SHAP-based feature analysis and transparent ensemble weights. Using only temporal and meteorological inputs, EM-LR forecasts daily EMS call volumes across five diverse districts in the state of Uttar Pradesh. We benchmark EM-LR against traditional models and recent advanced variants, including Twin Bounded Least Squares Support Vector Regression (TBLSSVR), Asymmetric-Huber based Extreme Learning Machine (AHELM), and Mexican-Hat Kernelized Large Margin Distribution Machine-based Regression (MHKLDMR), demonstrating superior accuracy and reduced prediction variance. Experimental results show up to 9.5% reduction in RMSE and over 40% variance reduction. EM-LR thus offers a scalable and interpretable forecasting solution tailored to the operational constraints of developing public health systems, supporting data-driven emergency planning and equitable healthcare delivery.

**Keywords** Ensemble model, Lasso regression, Artificial neural network, Extreme gradient boosting, Robust ML, Support vector regression, SHapley Additive Explanations (SHAP), Interpretable machine learning, Emergency medical service demand

### Background and motivation

Emergency Medical Services (EMS) are an important component of the public health infrastructure, providing critical and often life-saving assistance in medical emergencies. A responsive and effective EMS system ensures timely care, which directly impacts patient survival rates, especially in trauma and critical care situations. However, EMS systems worldwide, including in low-resource settings such as India, are under increasing pressure due to increased population densities, evolving healthcare demands, and limited resources.

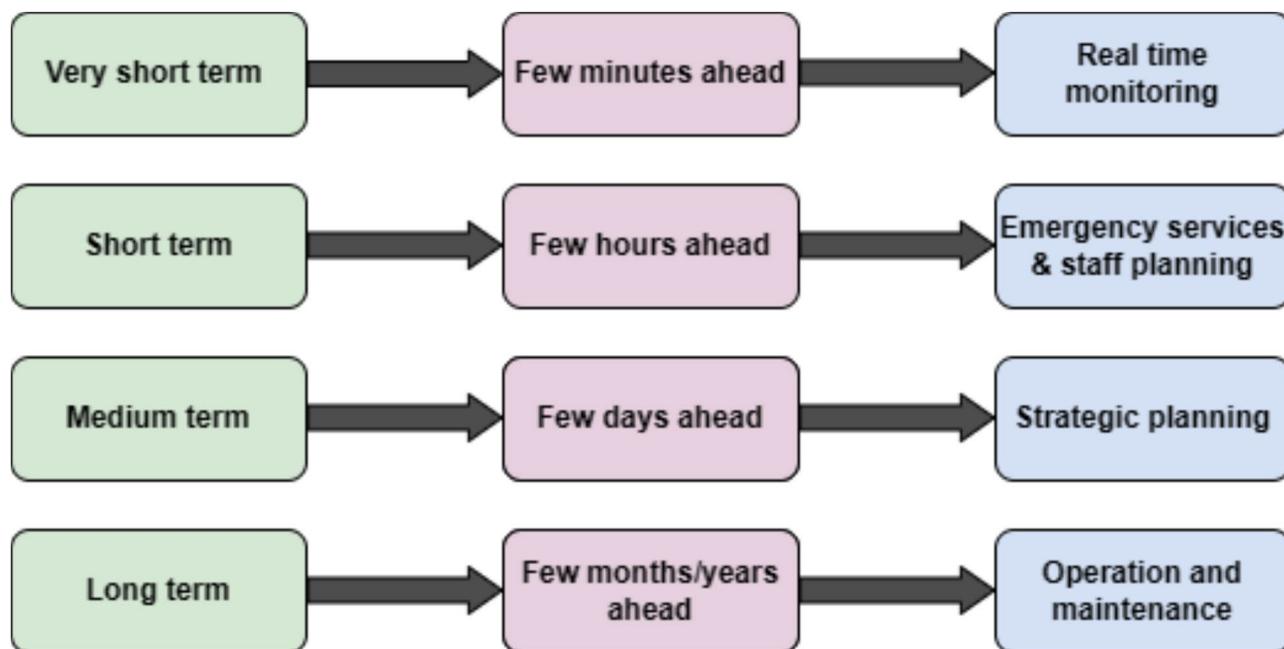
<sup>1</sup>Computer Science and Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India. <sup>2</sup>Civil Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India. ✉email: tgarg@cs.iitr.ac.in; durga.toshniwal@cs.iitr.ac.in

India, in particular, faces unique challenges in EMS planning and delivery. Many districts lack real-time surveillance systems, reliable demand data, or optimized protocols for ambulance allocation. Ambulance response times vary significantly across regions, often due to poor anticipation of demand patterns. In such contexts, district-level EMS demand forecasting becomes a critical planning tool for efficient ambulance deployment and equitable resource allocation. However, research on EMS forecasting in India remains limited, with few studies utilizing real-world ambulance dispatch data at the district level.

Given the critical role of predictive systems in equitable EMS planning, response time—the interval between call receipt and ambulance arrival—has emerged as a core performance metric. Accurate demand forecasting enables timely resource deployment, and numerous studies have leveraged it to guide dynamic ambulance allocation aimed at minimizing delays and improving prehospital care efficiency<sup>1–4</sup>. Various EMS demand studies aim to reduce this response time. The two solutions entail forecasting ambulance demand to meet needs and optimizing ambulance distribution. Comprehensive studies have been conducted on dynamic ambulance allocation models to improve real-time EMS resource management. These models focus on the ongoing redistribution of ambulances in response to fluctuating demand, ensuring optimal coverage across various regions. Recently, several studies have focused on the planning and deployment of ambulances<sup>5–12</sup>. For instance, Yaseen, Alkhaldi, and Raweshidy<sup>9</sup> proposed a machine learning and SDN-based system for prioritizing SHE traffic flows. Liu, Li, and Zhang<sup>10</sup> developed a robust optimization model for the optimal distribution of EMS stations. The model aims to optimize the number of ambulances and demand assignments in the EMS system while minimizing the anticipated overall cost. Amorim, Antunes, Ferreira, and Couto<sup>11</sup> proposed an approach to EMS resource allocation that improves patient outcomes by combining a mathematical model with a metamodel-based local search technique. Although such methods accelerate response and improve coverage, their effectiveness depends fundamentally on accurate and robust demand forecasting models that can anticipate call volumes under varying regional and temporal conditions.

Forecasting EMS demand helps estimate expected call volumes and temporal fluctuations, forming the foundation of operational planning. Different forecasting horizons serve distinct purposes. Short-term (minute-to-hourly) forecasts are valuable for dynamic ambulance routing and real-time dispatch optimization, while daily-level forecasts are essential for staff scheduling, ambulance station planning, and day-ahead readiness—particularly relevant in data-scarce, low-flexibility systems like those in many Indian districts. Long-term (monthly) forecasts, in contrast, aid in infrastructure and budget planning. Figure 1 illustrates the operational relevance of different forecasting horizons in EMS planning.

Over the years, EMS demand modeling has evolved from traditional statistical techniques to advanced machine learning (ML) frameworks. The regression models<sup>13–20</sup> have been used extensively to study the influence of several contextual variables on explaining fluctuations in EMS demand. Time series forecasting models<sup>22–25</sup> rely on historical patterns of demand, along with contextual variables<sup>21</sup>. Graph-based convolutional networks<sup>26</sup>, and spatio-temporal methods<sup>27</sup> have also been proposed to address EMS demand and to enhance resource allocation and emergency response times. Recent advancements in machine learning (ML) have significantly improved the accuracy of EMS demand predictions. Various studies<sup>28–34</sup> have recently used ML techniques extensively to predict EMS demand both temporally and spatio-temporally. For instance, Grekousis and Liu<sup>28</sup> introduced a new three-level spatial-based artificial intelligence approach to forecast ambulance demand in emergency medical services. The method locates expected emergency events geographically, enabling better



**Fig. 1.** Forecasting horizon and its implications in EMS demand prediction.

resource allocation and faster response times. Abreu et al.<sup>29</sup> introduced a data-driven forecasting method to facilitate emergency medical services (EMS) operational decision-making. This method surpasses the limitations of conventional forecasting techniques, enabling the healthcare industry to allocate resources more effectively. Martin, Mousavi, and Saydam<sup>30</sup> employed an ensemble-based decision tree model for feature selection, followed by a multilayer perceptron (MLP) artificial neural network model to generate daily, hourly, and spatially distributed predictions of EMS call volume.

Recent advancements in machine learning (ML) have significantly improved the accuracy of EMS demand predictions. While accuracy is a critical metric for predictive models, robustness is equally important, as it defines the model's ability to deliver consistent results under varying conditions, especially in critical domains such as EMS demand forecasting. Many advanced ML models demonstrate high accuracy but exhibit significant variance across folds or datasets, undermining their reliability. Additionally, many models operate as "black box" systems, providing limited insight into their decision-making processes. In critical fields such as healthcare, interpretability and robustness are essential for fostering trust, enabling accountability, and providing accurate predictions. This study addresses the triple challenge of accuracy, robustness, and interpretability by proposing EM-LR (Ensembled Meta-Learner with Linear Regression), a SHapley Additive Explanations (SHAP)<sup>35</sup> featured meta-learning framework designed to prioritize robustness and interpretability while maintaining competitive accuracy. By carefully curating a set of diverse, stable base models, EM-LR addresses the limitations of single-model approaches and mitigates the destabilizing effects of high-variance predictors. Additionally, SHAP provides a global and local explanation of feature importance, ensuring that the selected features contribute meaningfully to the predictions.

Meta-learning is an approach in machine learning that focuses on optimally combining predictions from multiple base models to enhance overall performance. It has been successfully applied in various domains, including speech recognition, energy forecasting, and natural language processing. In the context of emergency medical services (EMS), prior studies by Ramgopal et al.<sup>36</sup> and Megou et al.<sup>37</sup> have explored meta-learning frameworks to forecast EMS dispatches. However, these studies predominantly rely on stacking similar types of base learners, such as generalized linear models, generalized additive models, and tree-based algorithms like Random Forest (RF) and Extreme Gradient Boosting (XGB). For instance, Decision Trees and RF are also used in the ensemble models proposed in<sup>38</sup>. This lack of model diversity can lead to overfitting and instability, as similar models tend to produce highly correlated predictions<sup>38</sup>. Moreover, these frameworks are often designed for data-rich environments, limiting their adaptability to regions with sparse data availability. These limitations highlight the need for a robust, interpretable, and generalizable meta-learning ensemble that can perform effectively even in data-scarce EMS settings.

### Research gap and study contributions

While ensemble methods have demonstrated strong results in domains such as finance, energy, and NLP, their application to EMS forecasting remains limited, particularly in developing countries. Existing EMS forecasting studies often depend on rich spatial, demographic, or hospital-level variables that are rarely available in public health datasets. Furthermore, past meta-learners have tended to stack similar base models (e.g., tree-based or boosting methods), which increases the risk of overfitting and correlated prediction errors<sup>38</sup>.

This study investigates whether a heterogeneous, low-variance meta-learning ensemble can achieve accuracy comparable to advanced nonlinear regressors—such as the Asymmetric-Huber Loss function-based ELM (AHELM)<sup>39</sup>, Twin Bounded Least Squares Support Vector Regression (TBLSSVR)<sup>40</sup>, and Mexican-Hat Kernelized LMDMR (MHKLDMDR)<sup>41</sup>—while maintaining computational efficiency and partial explainability through feature-level insights. The central research question is: "Can a diverse ensemble of complementary learners forecast daily district-level EMS demand as effectively as complex state-of-the-art models, using only minimal temporal and meteorological features, while ensuring robustness and generalizability in data-scarce environments?"

To this end, we propose EM-LR (Ensembled Meta-Learner with Linear Regression), which strategically combines four diverse base learners—Lasso Regression<sup>42</sup>, Support Vector Regression (SVR)<sup>43</sup>, Multilayer Perceptron (MLP)<sup>44</sup>, and Extreme Gradient Boosting (XGB)<sup>45</sup>—within a meta-learning framework. These models were chosen for their complementary strengths in handling linearity, nonlinearity, regularization, and structured data. A linear regression meta-learner aggregates its predictions to enhance robustness and maintain model transparency at the ensemble level. SHAP-based feature analysis further strengthens interpretability by identifying key temporal and meteorological drivers of EMS demand.

Key contributions of this study are as follows:

- **Contextual novelty:** This study is one of the first district-level EMS forecasting models tailored to India's public health system, utilizing real-world ambulance dispatch data and only minimal features (e.g., day of the week, temperature, humidity, wind speed).
- **Algorithmic innovation:** We propose EM-LR, a novel heterogeneous meta-learning ensemble that stacks four diverse base learners—spanning linear, kernel, neural, and tree-based paradigms—and integrates them through a transparent linear regression meta-learner. Unlike past EMS meta-learners that rely on homogeneous tree ensembles, EM-LR reduces overfitting and prediction correlation while improving interpretability and robustness<sup>37,38</sup>.
- **Benchmarking against recent models:** We rigorously benchmark EM-LR against both traditional models and recent state-of-the-art regressors such as TBLSSVR, AHELM, MHKLDMDR to demonstrate performance gains and variance reduction.

- **Explainability via feature analysis integration:** We incorporate SHAP- and correlation-based feature relevance analysis within the meta-learning pipeline, enabling transparent understanding of how each temporal and weather variable influences EMS demand.
- **Practical deployability and generalization:** EM-LR demonstrates strong generalization performance across five demographically diverse districts, despite relying only on minimal temporal and meteorological inputs. This robustness across varied local conditions makes it a promising and scalable forecasting solution for EMS planning in real-world, data-scarce public health settings.

The remainder of this paper is structured as follows: The "Methods" section details the EM-LR methodology, the "Experimental Setup" section presents the dataset and experimental design, next, the "Results and Discussion" section presents performance findings and feature-level insights, and the "Conclusion" section concludes the study.

## Methods

### Study area

Uttar Pradesh (UP), India's most populous state, faces significant challenges in managing Emergency Medical Services (EMS) due to its geographical diversity and socio-economic conditions. This study examines five districts—Lucknow, Agra, Kanpur Nagar, Varanasi, and Gorakhpur—selected to represent the diverse characteristics of UP.

#### *Lucknow*

Lucknow, the capital of Uttar Pradesh, has a humid subtropical climate. During the study period, Lucknow experienced extremely hot summers, with temperatures reaching as high as 51 °C. Winters were cool, with temperatures dropping to around 15 °C. The monsoon season brought an average daily rainfall of about 5 mm; the highest recorded rainfall during the study period was 180 mm. Regarding EMS dispatch, the minimum daily dispatch was 0, while the maximum daily demand reached 85. On average, 13 EMS units were dispatched daily during this period.

#### *Agra*

Agra has a semiarid climate characterized by distinct summer, monsoon, and winter seasons. Summers were hot and dry, with temperatures as high as 49 °C during the study period. Winters were cool, with temperatures dropping to around 4 °C. The average daily rainfall in Agra during the study period's monsoon season was 4.4 mm, with the highest recorded rainfall at 119.7 mm. Regarding EMS dispatch, the minimum daily dispatch was 0, while the maximum daily demand reached 36. On average, 6 EMS units were dispatched each day during this period.

#### *Kanpur Nagar*

Kanpur is situated on the banks of the Ganges River and thus has a humid subtropical climate, characterized by hot and dry summers. During the study period, the district's summer temperature was as high as 51 °C. Winters were cooler, with temperatures dropping to around 19 °C. During the monsoon season in the study period, Kanpur experienced an average daily rainfall of 4.1 mm, with the highest recorded rainfall reaching 108 mm. Regarding EMS dispatch, the minimum daily dispatch in Kanpur was 0, indicating days with no emergency demands. However, the maximum daily demand reached 41. On average, 8 EMS units were dispatched each day during this period.

#### *Varanasi*

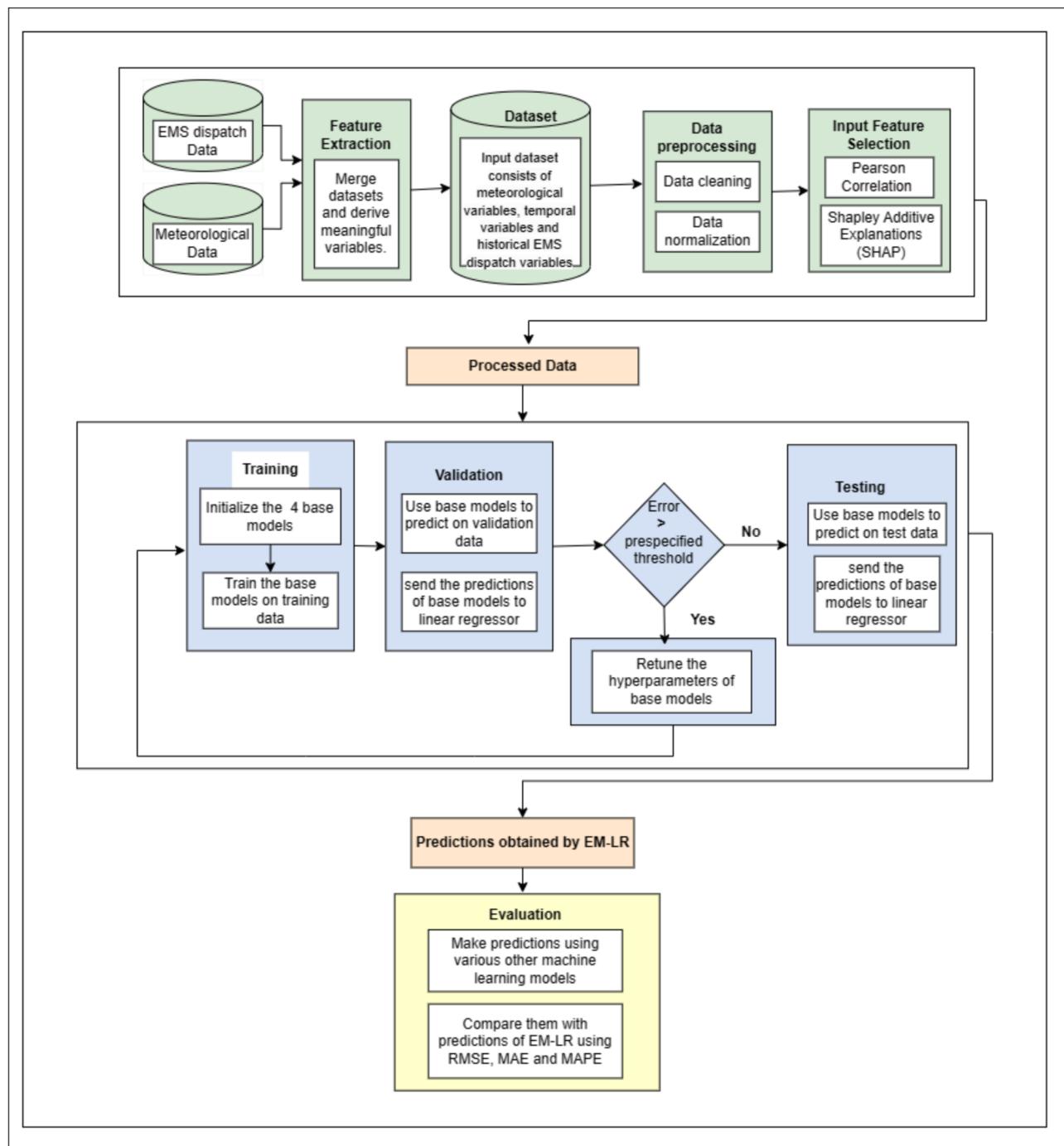
Varanasi is in the northern part of Uttar Pradesh and is also located on the banks of the Ganges River. It has hot and humid summers, with temperatures soaring to 51 °C. Winters in Varanasi bring cooler temperatures, dropping to around 18 °C. During the study period, Varanasi experienced an average daily rainfall of 6.54 mm during the monsoon season, with the highest recorded rainfall reaching 89.5 mm. Regarding EMS dispatch, the minimum daily dispatch in Varanasi was 0, indicating that there were no emergency demands on those days. However, the maximum daily demand reached 24. On average, 5 EMS units were dispatched each day during this period.

#### *Gorakhpur*

Gorakhpur is a district in the northeastern part of Uttar Pradesh. It has hot and humid summers, with temperatures soaring up to 48 °C. Winters in Gorakhpur bring cooler temperatures, dropping to around 17 °C. During the study period, Gorakhpur experienced an average daily rainfall of 1.68 mm during the monsoon season, with the highest recorded rainfall reaching 90.4 mm. Regarding EMS dispatch, the minimum daily dispatch in Gorakhpur was 0, indicating that there were no emergency demands on those days. However, the maximum daily demand reached 51. On average, 11 EMS units were dispatched each day during the study period.

## Proposed framework

The EM-LR framework (Fig. 2) follows a structured pipeline for EMS demand forecasting. Initially, temporal, meteorological, and historical EMS features are extracted based on prior studies. Feature selection is performed using Pearson correlation and SHAP to identify the most relevant predictors. To assess the incremental value of weather features and feature selection, three model variants are constructed: (i) T using only temporal and historical EMS data, (ii) T+W adding all meteorological features, and (iii) T+W+FS incorporating only the top



**Fig. 2.** Flow chart of the proposed work.

ten SHAP- and Pearson-ranked features. Each configuration is evaluated to understand trade-offs in performance and complexity. Building upon these, the final EM-LR model ensembles four diverse base learners: Lasso Regression, Support Vector Regression (SVR), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGB), using Linear Regression as a meta-learner to capture complementary learning patterns. The optimization of the model's structure was achieved through the process of hyperparameter tuning. Subsequently, the proposed model was validated using EMS demand data obtained from five discrete locations in Uttar Pradesh, India. The study conducted a comparative analysis of the model's performance against state-of-the-art persistence models and other machine learning models, including Support Vector Regression (SVR), Random Forest (RF), Extreme Gradient Boosting (XGB), and Multilayer Perceptron (MLP).

The study's methodology underwent rigorous testing and was subjected to a 5-fold cross-validation process to establish the accuracy and reliability of the proposed EM-LR model in predicting EMS demand across different districts. To achieve this, the dataset was partitioned into a training set, a validation set, and a test set.

The first four years of data served as the training set, the fifth year's data as the validation set, and data from the last year of the study period as the test set. The base models are trained with the help of the training set. The validation set compares the performance of different model structures by hypertuning their parameters, such as the regularisation parameter in SVR and Lasso Regression, the optimal number of heights in the XGB, and the number of hidden units in the MLP model. The error rate of the proposed model was measured using the test set. Lasso Regression, MLP, SVR, and XGB each receive the same training data, as shown in Fig. 3. The Linear Regression, which serves as a meta-learner, takes the predicted values of each base model as input. The final result is thus the weighted average of the results from the individual base models.

The input vector consists of three sets of features: meteorological, temporal, and EMS historical features. Let the meteorological features be denoted as  $x^m = [x_1^m, x_2^m, \dots, x_{n_m}^m]$ , the temporal features be denoted as  $x^t = [x_1^t, x_2^t, \dots, x_{n_t}^t]$ , and the EMS historical features be denoted as  $x^h = [x_1^h, x_2^h, \dots, x_{n_h}^h]$ . The input vector can then be written as:

$$X = [X^m, X^t, X^h] \quad (1)$$

Our objective is to create a function that can forecast the EMS demand for the next day, denoted as  $t + 1$ . This prediction will be based on the feature vector of meteorological conditions, time-related factors, and historical data on EMS parameters. Mathematically, the function can be expressed as follows:

$$Y_{t+1} = F(X)_t = F(X_t^m, X_t^t, X_t^h) \quad (2)$$

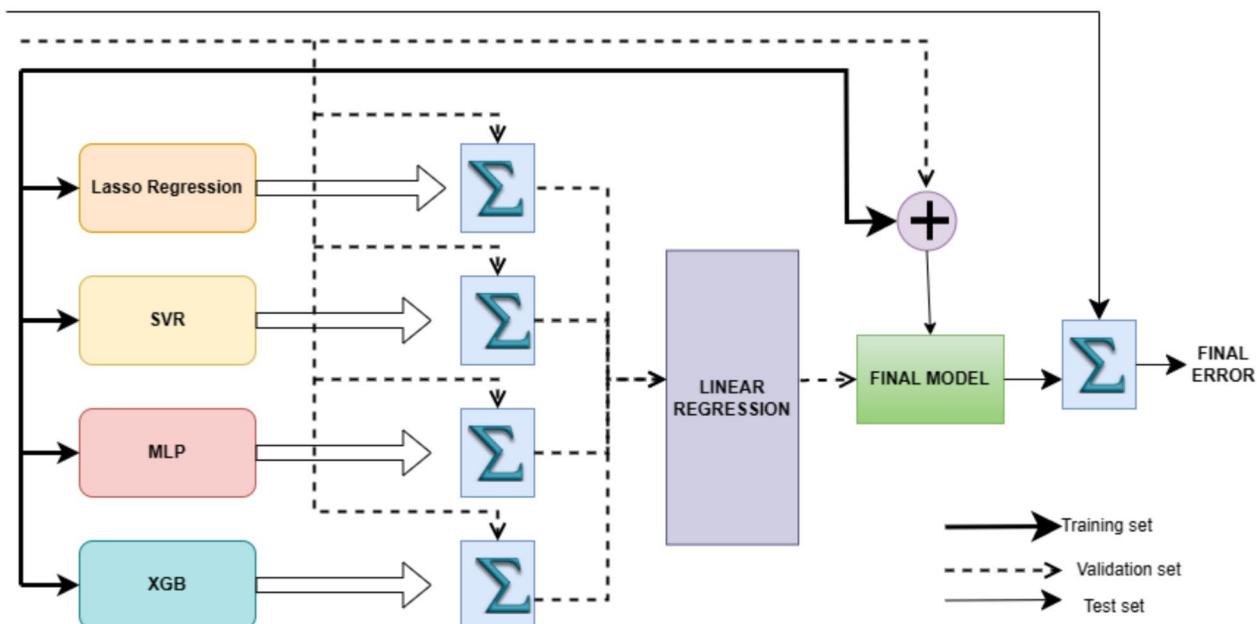
$Y_{t+1}$  denotes the predicted EMS demand at day  $t + 1$ , and  $F(\cdot)$  is the function that maps the input features to the predicted EMS demand. This function  $F(\cdot)$  takes the predictions of the 4 base models and can be denoted as:

$$F(X)_t = [f_{lr}(X)_t \ f_{svr}(X)_t \ f_{mlp}(X)_t \ f_{xgb}(X)_t] \quad (3)$$

Where  $f_{lr}(X)$ ,  $f_{svr}(X)$ ,  $f_{mlp}(X)$ ,  $f_{xgb}(X)$  denote the predictions of the 4 base models, namely Lasso Regression, SVR, MLP, and XGB. The detailed description of these functions is as follows:

$$\begin{aligned} f_{lr}(X)_t &= f_{lr}(X_t^m, X_t^t, X_t^h) \\ f_{svr}(X)_t &= f_{svr}(X_t^m, X_t^t, X_t^h) \\ f_{mlp}(X)_t &= f_{mlp}(X_t^m, X_t^t, X_t^h) \\ f_{xgb}(X)_t &= f_{xgb}(X_t^m, X_t^t, X_t^h) \end{aligned} \quad (4)$$

These predictions from base models go to the linear regressor as 4 input vectors. The linear regressor assigns weights to each of these four input vectors and makes the final prediction as follows:



**Fig. 3.** Framework of the proposed EM-LR model.

$$F(X)_t = w_0 + w_1 \cdot f_{lr}(X) + w_2 \cdot f_{svr}(X) + w_3 \cdot f_{svr}(X) + w_4 \cdot f_{xgb}(X) \quad (5)$$

Where  $w_i$  denotes the weights assigned to each of the four base models. The linear regressor optimizes these weights using a cost function that calculates the square of the sum of the differences between the actual EMS demand and the predicted EMS demand. Mathematically, the cost function can be expressed as:

$$Cost(w) = \frac{1}{8} \sum_{i=1}^4 (y - \hat{y}_i)^2 + \lambda \sum_{j=1}^4 w_j^2 \quad (6)$$

Here  $Cost(w)$  is the cost function,  $y$  is the actual EMS demand,  $\hat{y}_i$  is the predicted EMS demand for each of the four base models, and  $\lambda$  is the regularization parameter.

## Experimental setup

### Data source

The data for this study comes from two sources. The EMS dispatch data was obtained from the GVK-Emergency Management Research Institute in Lucknow, which operates the "108 Ambulance Service" across UP, providing daily dispatch counts for the five districts. Meteorological data was sourced from the World Weather Online API, including variables such as temperature, precipitation, dew point, pressure, visibility, and wind speed. These weather factors are critical in capturing the environmental influences on EMS demand.

To develop robust EMS demand prediction models, the data underwent preprocessing. Temporal features such as year, month, weekday, and weekend indicators were extracted to capture time-based trends. Meteorological variables were averaged over the preceding seven days to account for lagged effects. Historical EMS dispatch data was included by calculating the average dispatches over the previous seven days and counts from days lagged by 14, 21, and 28 days, capturing short- and long-term trends.

### Performance evaluation metrics

The forecasting performance of all models was evaluated using four widely adopted error metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Bias Error (MBE), and Mean Absolute Percentage Error (MAPE). These measures collectively capture both the magnitude and direction of prediction errors, enabling a balanced assessment of model accuracy and stability.

MAE and RMSE quantify average deviation and error dispersion, respectively, while MBE indicates systematic bias (under- or over-prediction). MAPE expresses the relative percentage error, facilitating intuitive comparison across districts with varying EMS call volumes. All metrics were computed on the test sets for each district, and lower values indicate superior predictive performance and generalization.

### Hyperparameter tuning of baseline and benchmark models

To rigorously evaluate our forecasting framework, we implemented a suite of machine learning models, categorized as either base learners for the proposed ensemble meta-learner (EM-LR) or as comparative benchmark models. The evaluation includes four base learners, including SVR, XGB, MLP, and Lasso Regression. We further benchmarked EM-LR against both traditional and recent advanced variants, including Random Forest (RF), Twin Bounded Least Squares Support Vector Regression (TBLSSVR), Asymmetric-Huber based Extreme Learning Machine (AHELM), and Mexican-Hat Kernelized Large Margin Distribution Machine-based Regression (MHKLDMR). Each model underwent grid search-based hyperparameter tuning to ensure optimal configuration. All experiments employed a consistent train-test split (2013–2017 for training, 2018 for testing), followed by 5-fold cross-validation. Model performance was assessed using four standard evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Bias Error (MBE). A brief description of each model and its corresponding hyperparameter search space is provided below.

#### Lasso regression

It is a statistical method that utilizes linear regression with L1 regularization to model the relationship between predictor variables and EMS demand. It works by reducing the regression coefficients of the predictor variables until they reach zero, thereby decreasing the impact of unnecessary or redundant variables and encouraging sparsity in the final model. The values taken for the grid search are as follows:

- **fit intercept**= [1, 0]
- **alpha**= [0.005, 0.01, 0.03, 0.05, 0.07, 0.1]

#### Multilayer perceptron (MLP)

MLP is a neural network composed of many layers of connected neurons, each performing a non-linear change on its input. It was selected for its ability to model nonlinear patterns. The hyperparameters of an MLP that have been tuned are the number of hidden layers, the parameter  $\alpha$ , and the activation function. The alpha parameter regulates the regularization intensity, while the size of the hidden layer determines the number of neurons present in each layer. The activation parameter specifies the activation function for each layer. The values taken for the grid search for these hyperparameters are as follows:

- **hidden layer size**= [(24,), (36,), (14,), (14,10), (14,5), (14,14), (24,24), (24,12), (36,24), (36,12)]
- **alpha**= [1e<sup>-8</sup>, 1e<sup>-7</sup>, 1e<sup>-6</sup>, 1e<sup>-5</sup>, 1e<sup>-4</sup>, 1e<sup>-3</sup>]

- **activation** = [relu, identity, tanh]

#### *Support Vector Regression (SVR)*

The SVR algorithm is a Support Vector Machine (SVM) version specific for regression tasks. It works by locating the hyperplane that preserves a maximum margin while permitting a specific deviation (specified by the *epsilon* parameter) from the real target values. SVR is included to capture nonlinear relationships with controlled flexibility. The three hyperparameters tuned in the study are "fit\_intercept", "C," and " $\epsilon$ ". The 'fit\_intercept' parameter determines whether the model should include an intercept term in the regression equation, the ' $\epsilon$ ' parameter establishes the margin of error permitted in the model's predictions, and the 'C' parameter manages the trade-off between obtaining a good fit on the training data and preventing overfitting. The following values were used to fine-tune these hyperparameters:

- $\epsilon$  = [8, 9, 10, 11, 12, 13, 14]
- **fit\_intercept** = [0,1]
- C = [33, 34, 35, 36, 37, 38, 39, 40, 41]

#### *Extreme Gradient Boosting (XGB)*

XGB is a member of the gradient boosting family. It sequentially constructs an ensemble of weak prediction models, typically decision trees, where each successive model is trained to correct the errors made by the previous models. The final forecast is derived by combining the forecasts of all weak models. Four hyperparameters were chosen to tune the XGB model. First is the 'n\_estimators' parameter, which determines the number of trees in the ensemble and influences model performance and training time. The second is 'subsample,' which refers to the fraction of samples used for training each tree. A lower value reduces the risk of overfitting but may also reduce performance. The 'eta' hyperparameter, also known as the learning rate, determines the step size when modifying the weights; a smaller value results in more stable convergence. Lastly, the gamma hyperparameter controls the minimal loss reduction necessary to split a leaf, with a higher value resulting in more conservative tree construction. The values taken to tune these hyperparameters were

- **n\_estimators** = range(70, 140, 10)
- **subsample** = [0.5, 0.75, 1]
- **eta** = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4]
- **gamma** = range(150, 310, 10)

#### *Random Forest (RF)*

RF is a robust ensemble tree-based model used for EMS forecasting. We tuned three key hyperparameters via grid search: the 'n\_estimators' parameter, which specifies the total number of trees; the 'min\_sample\_split' parameter, which specifies the minimum number of samples necessary to split a node; and the 'max\_features' parameter, which specifies the maximum number of features to be randomly selected while the tree is being grown. The range of values used for hyperparameters in the grid search:

- **n\_estimators** = range(300, 500, 25)
- **min\_sample\_split** = [2, 3, 4, 5, 6, 7]
- **max\_features** = [log2, sqrt]

#### *Asymmetric Huber loss function-based Extreme Learning Machine (AHELM)*

AHELM is a robust variant of the Extreme Learning Machine (ELM) that replaces the standard mean-square error loss with an "asymmetric Huber loss" to improve generalization and resilience to outliers. It combines the fast training of ELM with the robustness of Huber regression by introducing a tunable threshold parameter,  $\delta$ , that controls the transition between quadratic and linear loss regions. A regularization coefficient  $\alpha$  penalizes excessively large output weights, improving stability. The tuned hyperparameters were:

- **activation** = ['sigmoid', 'tanh', 'relu']
- **n\_hidden** = [25, 50, 75, 100, 150]
- **delta** = [0.25, 0.5, 0.75, 1.0]
- **alpha** = [0.001, 0.01, 0.1, 0.5, 1.0]
- **learning\_rate** = [0.001, 0.01, 0.05]

#### *Twin Bounded Least Squares Support Vector Regression (TBLSSVR)*

TBLSSVR minimizes two smaller least-squares problems to achieve improved computational efficiency and reduced training complexity compared to traditional SVR. The tuned hyperparameters were:

- **C1** = [0.01, 0.1, 1.0, 5.0, 10.0]
- **C2** = [0.01, 0.1, 1.0, 5.0, 10.0]
- **epsilon** = [0.001, 0.01, 0.05, 0.1]
- **kernel** = ['linear', 'rbf', 'poly']
- **gamma** = [0.01, 0.05, 0.1, 0.5]

#### *Mexican-Hat Kernelized Large Margin Distribution Machine-based Regression (MHKLDMDR)*

MHKLDMDR integrates a localized dual model regression framework with a Mexican Hat wavelet kernel to capture nonlinear and oscillatory patterns in EMS demand. The following parameters were grid-searched:

Metric	Variant	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
MAE	T	3.3920	5.0563	3.9058	3.7389	3.6023	5.8124	3.9088	3.0752
	T+W	4.4666	4.8158	4.2538	3.8997	3.3605	7.9416	3.7198	3.1322
	T+W+FS	3.3112	5.5833	3.8269	3.7389	3.6051	4.5064	3.5717	<b>3.0757</b>
RMSE	T	4.3928	6.2108	4.9953	4.8379	4.2627	6.9123	4.4165	3.9661
	T+W	5.6049	5.9812	5.3699	4.9810	4.1316	9.0377	4.6631	4.0326
	T+W+FS	4.2632	6.7767	4.9172	4.8379	4.2209	5.6678	4.2049	<b>3.9544</b>
MAPE	T	0.6614	4.7473	0.7117	0.6417	0.5221	0.7843	0.5697	0.4361
	T+W	4.5873	6.3469	0.8702	0.7000	0.4486	0.7652	0.5359	0.4445
	T+W+FS	0.5413	6.2817	0.6839	0.6417	0.4681	0.5134	0.4490	<b>0.4274</b>
MBE	T	-2.0343	-3.5715	-2.9414	-2.6799	-2.1014	-4.8027	-2.3009	-0.8676
	T+W	-3.9748	-3.5624	-3.5121	-2.8449	-0.4193	-7.9280	0.2179	-0.9173
	T+W+FS	-1.6496	-4.5166	-2.8188	-2.6799	<b>-0.3573</b>	-4.0181	-0.8363	-0.7192

**Table 1.** Performance comparison on Agra dataset.

Metric	Variant	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
MAE	T	4.3877	4.0758	4.0317	3.9456	4.1128	6.7223	4.2055	3.8053
	T+W	4.7240	4.3097	4.3090	4.0221	4.7634	12.5260	5.4934	3.7887
	T+W+FS	4.2221	4.2033	4.1490	4.0183	4.6484	9.7459	4.9697	<b>3.7782</b>
RMSE	T	5.7866	5.3572	5.4953	5.2618	5.4481	8.8019	5.5822	5.1590
	T+W	6.2884	5.5902	5.8788	5.4068	5.8566	14.1909	6.7185	5.0805
	T+W+FS	5.5310	5.4841	5.6678	5.3878	5.7280	11.8620	6.2375	<b>5.0663</b>
MAPE	T	0.5169	0.5125	0.4041	0.3747	0.3459	0.5664	0.3853	0.3599
	T+W	0.6304	0.7025	0.4609	0.3894	0.3596	0.7659	0.3617	0.3515
	T+W+FS	0.4397	0.3485	0.4310	0.3873	0.3496	0.5779	0.3475	<b>0.3451</b>
MBE	T	-2.1456	-0.7668	-2.1156	-1.0972	-1.6073	-4.3864	-1.8672	-1.2017
	T+W	-3.1138	-1.1444	-2.8270	-1.6349	<b>-0.6711</b>	-12.4721	-2.6222	-1.0314
	T+W+FS	-1.2770	0.7037	-2.4792	-1.6433	-0.9630	-9.4472	-1.2275	-0.8744

**Table 2.** Performance comparison on Gorakhpur dataset.

- **C1** = [0.01, 0.1, 1.0, 5.0]
- **C2** = [0.01, 0.1, 1.0, 5.0]
- **epsilon** = [0.001, 0.01, 0.05, 0.1]
- **sigma** = [0.25, 0.5, 1.0, 2.0, 3.0]

## Results and discussion

This section presents and compares the results of proposed EM-LR with various machine learning models, including traditional models including Extreme Gradient Boosting (XGB), Multi-layer Perceptron (MLP), Random Forest (RF), and Support Vector Regression (SVR), with the benchmark method P-Persistence and the recent advanced variants, including Asymmetric Huber loss function-based Extreme Learning Machine (AHELM), Twin Bounded Least Squares Support Vector Regression (TBLSSVR), and Mexican-Hat Kernelized Large Margin Distribution Machine-based Regression (MHKLDMR). The comparison is conducted for five districts in Uttar Pradesh: Lucknow, Kanpur Nagar, Agra, Gorakhpur, and Varanasi. These districts were selected based on their significance as population centers, encompassing urban, semi-urban, and rural areas. The aim was to comprehensively assess EMS demand patterns across diverse demographic and socioeconomic settings.

## Test results

To evaluate the predictive performance of the proposed EM-LR model across the five studied districts, four standard error metrics—Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Bias Error (MBE)—were employed (Tables 1, 2, 3, 4, 5). The results demonstrate a consistent and substantial improvement of EM-LR over all baseline models, including MLP, SVR, RF, XGB, and advanced variants such as AHELM, TBLSSVR, and MHKLDMR.

Across all districts, EM-LR achieved the lowest RMSE, confirming its superior ability to capture temporal and meteorological dependencies in EMS dispatch demand. As illustrated in Fig. 4, EM-LR consistently yielded smoother error profiles and reduced prediction volatility compared to both tree-based and neural counterparts. Notably, for Lucknow and Varanasi, EM-LR attained RMSE values of 6.01 and 3.41, respectively—the lowest among all competing models—reflecting its robustness in both high- and low-demand regions. Likewise, in Agra

Metric	Variant	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
MAE	T	3.4939	5.4019	3.5409	3.5686	3.2891	5.8344	3.7247	3.1295
	T+W	3.4751	4.7037	3.6260	3.8020	4.3143	9.9660	4.5408	3.1530
	T+W+FS	3.3823	3.8566	3.6643	3.7375	3.3577	5.7006	3.4765	<b>3.1064</b>
RMSE	T	4.6885	6.8488	4.8538	4.9119	4.2275	6.9429	4.6651	4.2251
	T+W	4.7668	6.0298	4.9848	5.1385	5.8529	11.2331	5.9643	4.2890
	T+W+FS	4.5576	5.1448	5.0053	5.0545	4.3494	7.2375	4.5425	<b>4.2088</b>
MAPE	T	0.6064	4.7768	0.5751	0.5841	0.4197	0.6325	0.4712	0.4199
	T+W	0.6008	5.8697	0.6058	0.6480	<b>0.2998</b>	0.8185	0.4459	0.4387
	T+W+FS	0.5472	1.4932	0.6185	0.6376	0.4791	0.5309	0.4725	0.4226
MBE	T	-1.8477	-4.9832	-2.3784	-2.5247	-1.5338	-3.9967	-1.7092	-0.7667
	T+W	-1.9244	-3.4966	-2.5995	-2.8067	<b>0.1807</b>	-9.9408	-2.0068	-0.9963
	T+W+FS	-1.7471	-1.8809	-2.6064	-2.6814	-0.1909	-5.3237	-0.5884	-0.8628

**Table 3.** Performance comparison on Kanpur Nagar dataset.

Metric	Variant	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
MAE	T	5.5624	7.6799	5.1641	4.9687	4.9912	7.8203	5.0124	4.5131
	T+W	6.5567	8.5741	5.2888	5.1941	4.3143	12.1977	5.4411	4.4594
	T+W+FS	5.4211	5.3194	4.9214	4.7470	4.6102	7.3886	4.5642	<b>4.2729</b>
RMSE	T	7.3604	9.5461	7.2535	6.9794	6.9207	9.4152	6.8449	6.3160
	T+W	8.4031	10.4283	7.3995	7.2935	6.8529	14.0893	7.3190	6.2168
	T+W+FS	7.2748	7.2406	6.9797	6.8384	6.9014	9.2139	6.6082	<b>6.0102</b>
MAPE	T	0.5383	1.3510	0.4431	0.4103	0.3882	0.5317	0.4012	0.3411
	T+W	0.7417	3.4343	0.4575	0.4484	0.3998	0.6845	0.3482	0.3376
	T+W+FS	0.5102	0.4545	0.4005	0.3791	0.3877	0.4457	0.3399	<b>0.3207</b>
MBE	T	-3.6577	-6.9277	-3.6875	-3.2315	-2.4116	-6.5239	-2.9587	-1.6527
	T+W	-5.5026	-7.9085	-3.8388	-3.4330	2.1807	-12.0979	-1.8587	-1.7327
	T+W+FS	-4.4165	-3.4033	-3.0715	-2.3210	1.6262	-6.6621	-1.8006	<b>-1.4497</b>

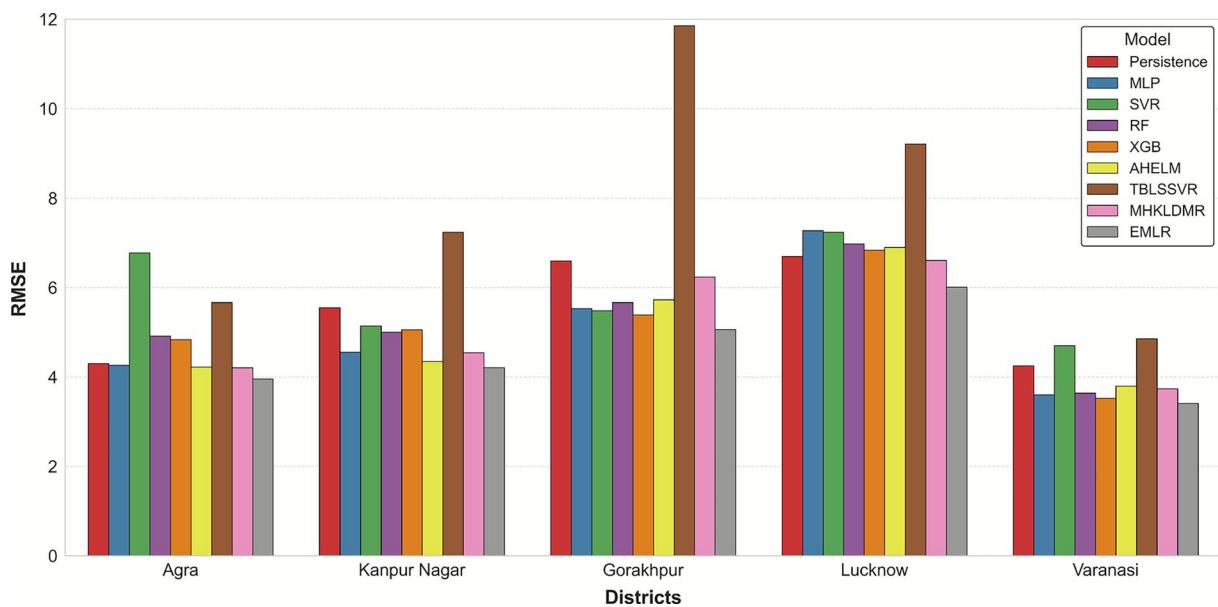
**Table 4.** Performance comparison on Lucknow dataset.

Metric	Variant	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
MAE	T	2.9232	3.3967	2.6476	2.5965	2.8784	3.9126	2.7891	2.5376
	T+W	2.8127	3.7638	2.6929	2.5748	3.1191	6.7783	3.6611	2.5508
	T+W+FS	2.7054	3.6630	2.6698	2.5748	2.8056	3.7242	2.8279	<b>2.5784</b>
RMSE	T	3.9277	4.4677	3.6117	3.5491	3.6228	5.0845	3.7433	3.4386
	T+W	3.8367	4.8692	3.6804	3.5251	4.0324	7.8700	4.7504	3.4162
	T+W+FS	3.6013	4.7013	3.6379	3.5251	3.7975	4.8589	3.7381	<b>3.4087</b>
MAPE	T	0.8064	3.1183	0.5715	0.5309	0.5067	0.6658	0.5443	0.4757
	T+W	0.6800	4.8768	0.5994	0.5242	0.5966	0.8046	0.5257	<b>0.4728</b>
	T+W+FS	0.5694	3.1449	0.5809	0.5242	4.1941	5.0274	4.2103	0.4775
MBE	T	-1.7024	-1.9599	-1.2159	-0.9864	-0.9771	-2.8415	-1.1812	-0.4400
	T+W	-1.3962	-2.7026	-1.4105	-0.9610	0.3580	-6.7566	-1.9198	-0.4340
	T+W+FS	-0.8426	-2.7866	-1.2752	-0.9610	-0.3106	-3.0897	-0.5785	<b>-0.2933</b>

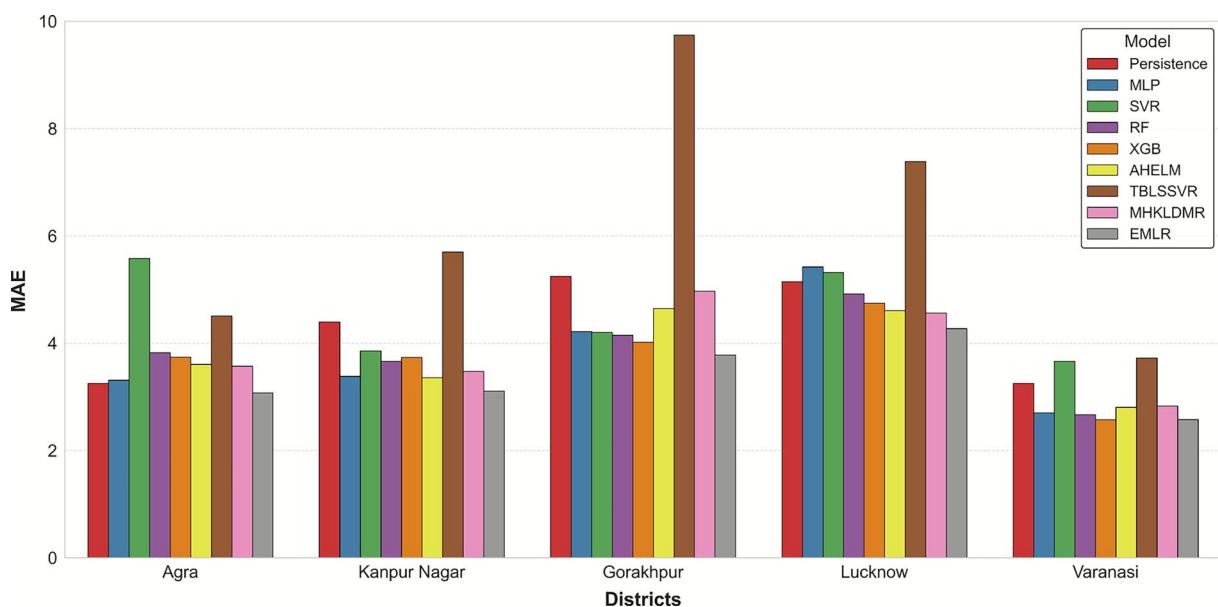
**Table 5.** Performance comparison on Varanasi dataset.

and Kanpur Nagar, EM-LR registered RMSE improvements exceeding 8–12% over the next-best models, while in Gorakhpur, it marginally surpassed SVR and XGB, achieving an RMSE of 5.07.

When compared with recent advanced learners such as AHELM, TBLSSVR, and MHKLDMR, EM-LR demonstrated consistent or superior generalization. For instance, in Lucknow, EM-LR achieved an RMSE of 6.01 compared to 6.90 (AHELM) and 9.21 (TBLSSVR), marking an improvement of 13–35%. In Agra, EM-LR (3.95) outperformed AHELM (4.22) and MHKLDMR (4.20) by approximately 6–7%, while in Kanpur Nagar, it achieved an RMSE of 4.21 versus 4.35 (AHELM) and 4.54 (MHKLDMR). In Gorakhpur, where AHELM (5.73) and MHKLDMR (6.24) performed competitively, EM-LR attained the lowest RMSE (5.07). These results



**Fig. 4.** RMSE Comparison of Models across Districts.

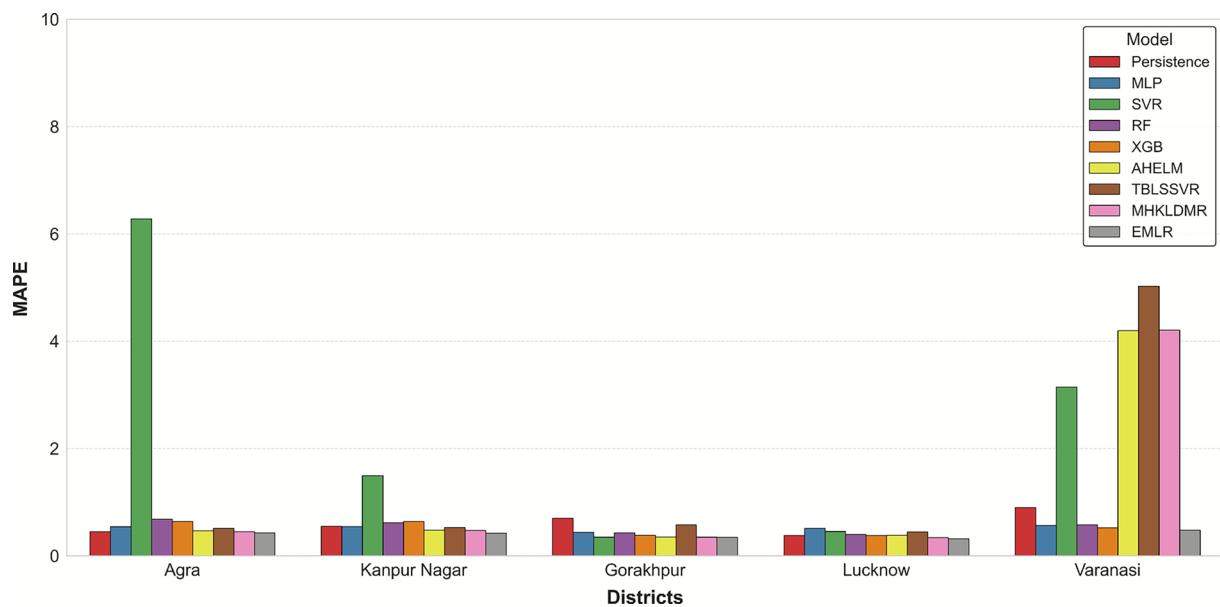


**Fig. 5.** MAE Comparison of Models across Districts.

underscore EM-LR's ability to deliver accuracy and robustness comparable to that of state-of-the-art specialized algorithms, without the added architectural complexity or loss of interpretability.

The MAE and MAPE results (Figs. 5 and 6) further reinforce EM-LR's superior generalization. For every district, EM-LR exhibited the lowest absolute and percentage errors, signifying enhanced reliability and reduced overfitting across data variants (T, T+W, and T+W+FS). The most pronounced reductions were observed for the T+W+FS configuration, where EM-LR achieved MAPE values as low as 0.32 in Lucknow and 0.35 in Gorakhpur, outperforming all benchmark models by wide margins. These improvements affirm that feature selection (FS) synergistically enhances the ensemble's stability and interpretability, especially when meteorological factors are integrated.

In addition to minimizing absolute errors, EM-LR effectively mitigated systematic bias. The MBE values show a bias reduction ranging from 37.5% (Lucknow) to 69.4% (Varanasi) relative to traditional regressors, demonstrating that the ensemble does not consistently under- or over-predict dispatch volumes. While SVR achieved the smallest bias for the Gorakhpur district, the EM-LR model remained competitive, yielding an



**Fig. 6.** MAPE Comparison of Models across Districts.

Model	Average Rank	Relative Order
EM-LR	1.125	1 (Best)
XGB	2.750	2
RF	2.875	3
SVR	3.875	4
MLP	5.125	5
AHELM	6.125	6
MHKLDMR	6.125	6
TBLSSVR	8.000	8 (Worst)

**Table 6.** Average *t*-Friedman ranks of all models across districts (lower is better).

MBE of  $-0.87$ , which is within a negligible deviation from the optimal bias margin. In all other districts, EM-LR achieved the lowest or nearly lowest MBE, underscoring its balanced predictive behavior.

In summary, EM-LR delivers consistent, interpretable, and bias-resilient forecasts across diverse operational environments. Its ensemble integration of linear, nonlinear, and tree-based learners allows it to outperform individual base models and contemporary regression alternatives. The uniform superiority of EM-LR across all four metrics highlights its scalability and generalizability for district-level EMS demand forecasting in resource-constrained settings.

### Statistical significance analysis

To evaluate the statistical reliability of the proposed EM-LR model's superior forecasting performance, we employed a two-stage non-parametric evaluation based on the *t*-Friedman test proposed by Liu and Xu<sup>46</sup>, complemented by district-wise paired *t*-tests for local validation. This combination ensures both global and local statistical verification of EM-LR's performance gains.

#### Global comparison using the *t*-Friedman test

The *t*-Friedman test is an improvement on the classical Friedman test by integrating Student's *t*-tests into the ranking process, thereby accounting for both mean and variance across repeated runs. Algorithms with statistically indistinguishable distributions (at  $\alpha_t = 0.05$ ) receive tied ranks, ensuring a variance-aware and conservative ranking.

Across the four districts and eight competing models on (T+W+FS) variant, the Iman–Davenport extension of the Friedman test produced an  $F_{7,21}$  value of 24.02 with a *p*-value of  $3.86 \times 10^{-7}$ , decisively rejecting the null hypothesis of equal model performance. The resulting average *t*-Friedman ranks (Table 6) confirm that EM-LR consistently outperformed all benchmarks. A lower rank denotes better predictive accuracy (i.e., lower RMSE).

#### Post-hoc pairwise comparison

After rejecting the null hypothesis globally, we conducted pairwise comparisons between EM-LR (control) and each competing model using the *t*-Friedman post-hoc procedure. Three multiple-comparison corrections—Holm, Finner, and Bonferroni-Dunn—were applied to control the family-wise error rate at  $\alpha = 0.05$ . The adjusted results are shown in Table 7.

The Holm test confirmed that EM-LR is statistically superior to three advanced models—TBLSSVR, AHELM, and MHKLDMDR—while the Finner correction additionally marked MLP as marginally inferior. Classical ensemble baselines such as XGB, RF, and SVR exhibited competitive but non-significant differences, reflecting smaller mean gaps and higher variance across districts.

#### District-wise validation

To complement the global non-parametric analysis, classical paired *t*-tests<sup>47</sup> were also conducted between EM-LR and each competing model using RMSE values from five random seeds within each district. These results, summarized in Table 8, reinforce the global findings: EM-LR achieved statistically significant ( $p < 0.05$ ) improvements over nearly all baseline and advanced models in Agra, Gorakhpur, Kanpur Nagar, and Lucknow, while Varanasi showed a few non-significant results due to lower variance and more homogeneous data.

Overall, the *t*-Friedman analysis confirmed significant global differences among models, with EM-LR achieving the best average rank and statistically outperforming all advanced baselines under Holm correction. The complementary district-wise *t*-tests reinforced these results, verifying EM-LR's consistent superiority across regions. Together, these analyses demonstrate that the proposed ensemble meta-learner delivers statistically significant, robust, and generalizable forecasting performance across diverse geographical contexts.

#### Feature importance analysis

Feature selection is essential in EMS demand forecasting, as it enhances model accuracy while minimizing redundancy and overfitting. In this study, two complementary approaches—SHAP (SHapley Additive exPlanations) and Pearson correlation analysis—were employed to identify the most influential predictors across districts. SHAP quantifies each feature's marginal contribution to the model output, effectively capturing non-linear and interaction effects, while correlation analysis highlights strong linear associations with EMS dispatch demand. The integration of both methods ensured that features with either direct or complex relationships were retained for subsequent modeling.

Feature selection was performed independently for each district to account for local variations in EMS patterns and meteorological behavior. The results revealed that Agra and Gorakhpur achieved optimal performance using correlation-based top features, whereas Lucknow, Kanpur Nagar, and Varanasi performed better with SHAP-based top-ten features. Figures 7, 8, 9, 10, 11 illustrate the ranked importance of features for each district.

Across all regions, historical EMS dispatch indicators consistently emerged as dominant predictors. Among the meteorological variables, temperature, dew point, wind speed, precipitation, and pressure showed a notable influence, whereas visibility and previous-day rainfall had a less significant impact. The presence of several non-linear weather effects identified exclusively by SHAP underscores that meteorological factors influence EMS demand in a non-proportional manner.

Temporal variables (year, month, and weekday) exhibited moderate yet consistent relevance. The year variable was a significant predictor across all districts, while the month contributed primarily in Gorakhpur (Fig. 8) and the weekday in Kanpur Nagar (Fig. 9). These variations highlight district-specific temporal dynamics in EMS demand.

To statistically validate the benefits of feature selection, a paired *t*-test (Table 9) was conducted between the two variants of the EM-LR model, of those trained on all features (T+W) and those trained on selected features (T+W+FS), using five random seeds. The proposed EM-LR model exhibited statistically significant improvement ( $p < 0.05$ ) in most districts, confirming that the reduced feature subset enhanced predictive performance without compromising generality.

Overall, feature selection improved both interpretability and statistical robustness of the proposed framework. The findings demonstrate that historical EMS trends and nonlinear meteorological interactions jointly govern ambulance dispatch demand.

Model	z-stat	p unadj.)	Holm APV	Significant (Holm)
TBLSSVR	-3.969	$7.2 \times 10^{-5}$	$5.0 \times 10^{-4}$	Yes
AHELM	-2.887	$3.89 \times 10^{-3}$	$2.34 \times 10^{-2}$	Yes
MHKLDMDR	-2.887	$3.89 \times 10^{-3}$	$2.34 \times 10^{-2}$	Yes
MLP	-2.309	$2.09 \times 10^{-2}$	$8.37 \times 10^{-2}$	No (Finner only)
SVR	-1.588	$1.12 \times 10^{-1}$	$3.37 \times 10^{-1}$	No
RF	-1.010	$3.12 \times 10^{-1}$	$6.25 \times 10^{-1}$	No
XGB	-0.938	$3.48 \times 10^{-1}$	$6.25 \times 10^{-1}$	No

**Table 7.** *t*-Friedman post-hoc comparisons against the EM-LR control model.

Model	District	t-statistic	p value	Significance
SVR	Agra	8.921	0.0277	Yes
	Gorakhpur	7.462	0.0343	Yes
	Kanpur Nagar	6.987	0.0428	Yes
	Lucknow	7.119	0.0253	Yes
	Varanasi	1.036	0.3953	No
MLP	Agra	7.533	0.0267	Yes
	Gorakhpur	8.125	0.0335	Yes
	Kanpur Nagar	8.004	0.0329	Yes
	Lucknow	10.971	0.0016	Yes
	Varanasi	1.502	0.2349	No
XGB	Agra	11.364	0.0012	Yes
	Gorakhpur	9.971	0.0015	Yes
	Kanpur Nagar	8.563	0.0043	Yes
	Lucknow	9.201	0.0023	Yes
	Varanasi	2.033	0.0598	No
RF	Agra	15.065	0.000113	Yes
	Gorakhpur	18.874	0.000046	Yes
	Kanpur Nagar	51.457	0.000001	Yes
	Lucknow	2.250	0.0876	No
	Varanasi	30.349	0.000007	Yes
AHELM	Agra	23.985	0.000018	Yes
	Gorakhpur	7.507	0.001685	Yes
	Kanpur Nagar	10.338	0.000494	Yes
	Lucknow	6.926	0.002281	Yes
	Varanasi	6.243	0.003355	Yes
MHKLDMDR	Agra	2.444	0.0709	No
	Gorakhpur	2.610	0.0594	No
	Kanpur Nagar	2.528	0.0648	No
	Lucknow	2.538	0.0642	No
	Varanasi	2.857	0.0461	Yes
TBLSSVR	Agra	7.519	0.001675	Yes
	Gorakhpur	8.971	0.000855	Yes
	Kanpur Nagar	9.370	0.000723	Yes
	Lucknow	7.073	0.002108	Yes
	Varanasi	8.076	0.001277	Yes

**Table 8.** Paired *t*-test results between EM-LR and benchmark models across districts.

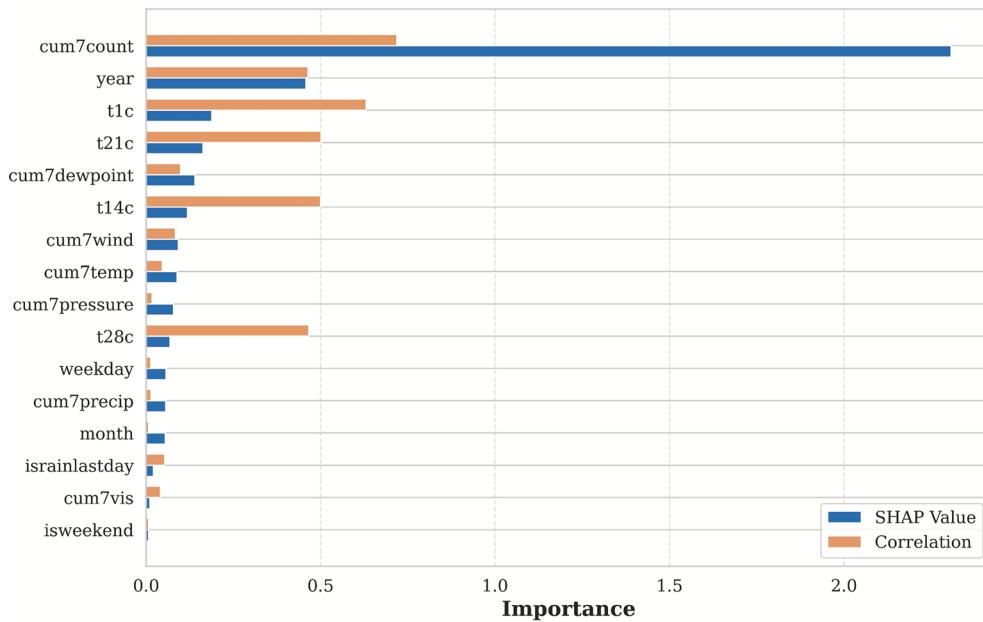
### Robustness analysis

The models were examined across five random seeds for each district to evaluate the sensitivity of RMSE performance to data partitioning. The variance of RMSE was used as the robustness indicator, where lower variance implies greater stability. As shown in Table 10, 11, 12, 13, 14, the proposed EM-LR generally achieves the lowest or near-lowest variance across districts, indicating high consistency across varying data splits.

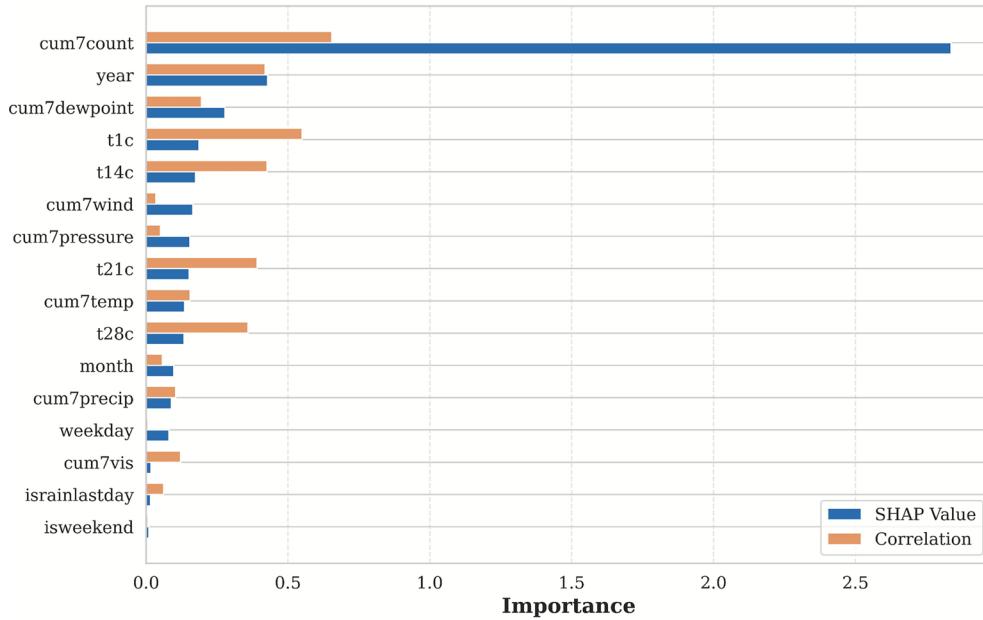
In Agra, EM-LR exhibited the most stable performance (variance = 1.59), closely followed by AHELM and MLP, whereas tree-based models, such as RF and XGB, displayed higher fluctuations (Table 10). Kanpur Nagar (Table 12) showed a similar trend, where EM-LR achieved the smallest variance (0.66), with AHELM and MHKLDMDR performing competitively and outperforming SVR and RF. For Gorakhpur (Table 11), the differences among EM-LR (1.04), SVR (1.00), and AHELM (0.68) were marginal, suggesting that these models maintained comparable robustness, whereas TBLSSVR and RF exhibited greater sensitivity.

In Lucknow (Table 13), EM-LR (2.34) maintained higher stability than all other models, including the advanced variants, which showed noticeably larger variance under complex temporal patterns. In Varanasi (Table 14), the variances for MLP, SVR, AHELM, and MHKLDMDR were relatively close, yet EM-LR still achieved the lowest variance (0.30), confirming its consistent generalization.

Overall, while models such as AHELM and MLP occasionally approached EM-LR in robustness, the proposed ensemble remained the most reliable and balanced performer across all five regions. Its consistent low variance across both traditional and advanced benchmarks underscores its robustness and practical suitability for EMS demand forecasting under diverse operating conditions.



**Fig. 7.** Feature importance analysis for Agra District.

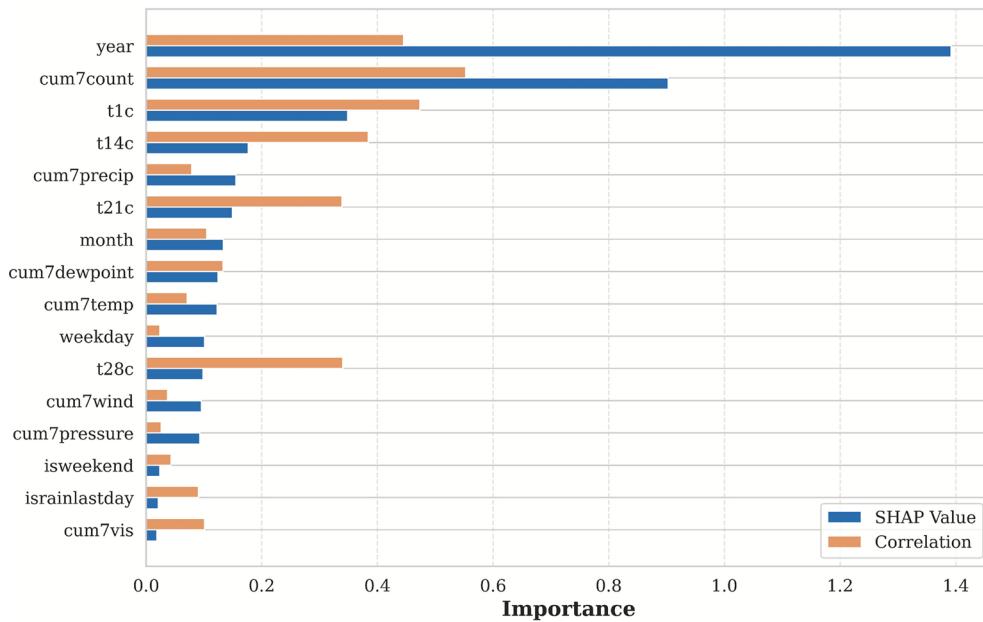


**Fig. 8.** Feature importance analysis for Gorakhpur District.

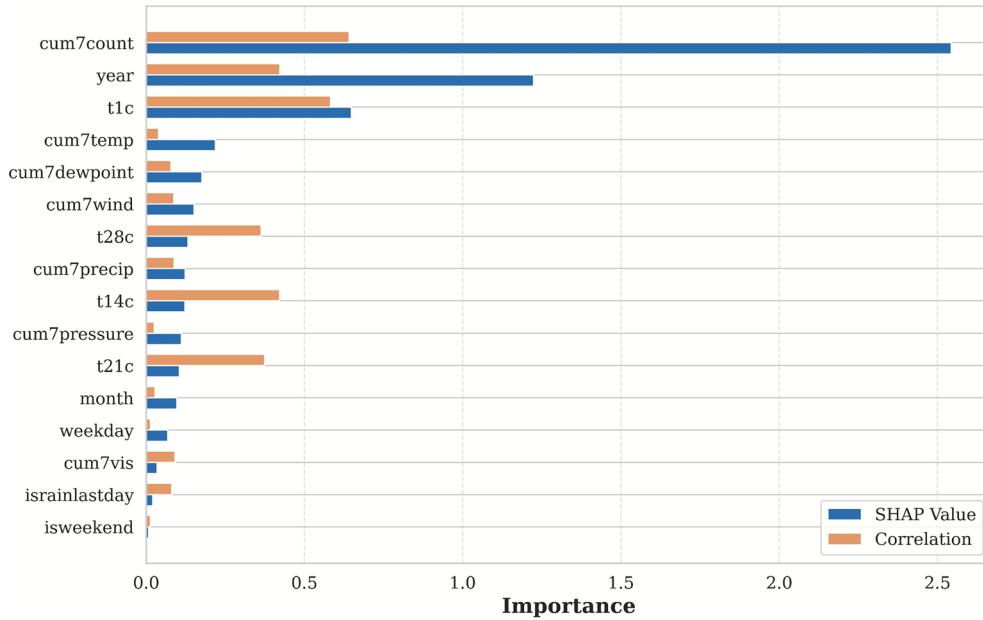
## Conclusion

This study proposed EM-LR, a robust and interpretable meta-learning ensemble framework for forecasting Emergency Medical Services (EMS) demand. Addressing the limitations of conventional ensemble and single-learner models, EM-LR integrates the complementary strengths of Support Vector Regression, Lasso, Multilayer Perceptron, and Extreme Gradient Boosting through a Linear Regression meta-learner. This architecture offers a balanced trade-off between predictive accuracy, variance reduction, and interpretability, essential features for real-time public health decision-making.

Empirical evaluation across five districts in Uttar Pradesh demonstrated that EM-LR consistently outperformed all traditional baselines in terms of both RMSE and variance, achieving up to 20% lower prediction error and over 40% variance reduction compared to the best standalone learners. When benchmarked against recent advanced models such as AHELM, TBLSSVR, and MHKLDMR, EM-LR continued to exhibit comparable or superior robustness while maintaining greater accuracy, underscoring the advantage of its meta-learning



**Fig. 9.** Feature importance analysis for Kanpur Nagar District.

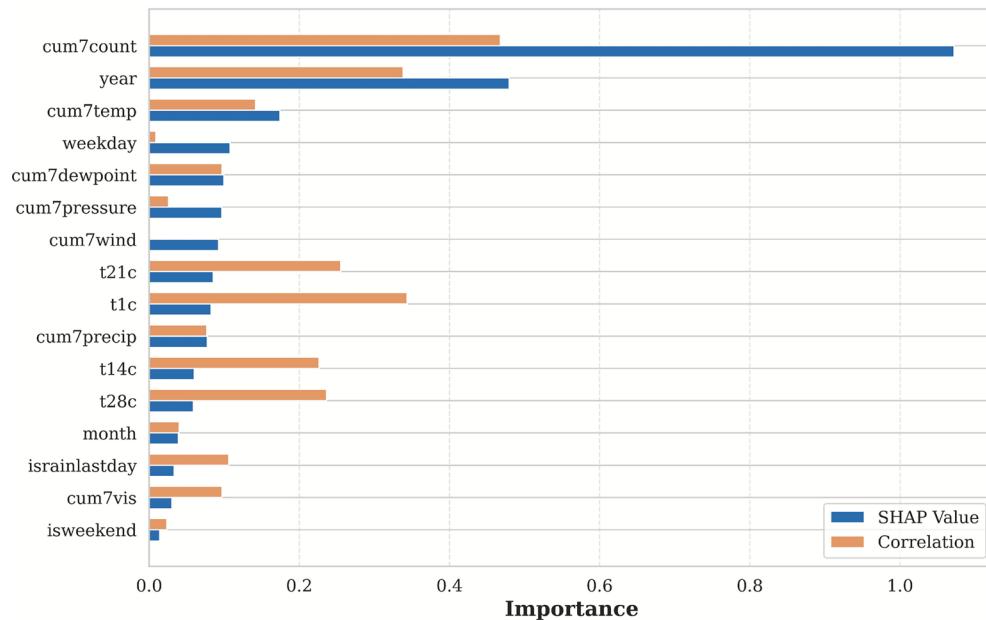


**Fig. 10.** Feature importance analysis for Lucknow District.

design. Statistical validation using the Friedman and post-hoc tests further confirmed the significance of these improvements, establishing EM-LR as a statistically reliable framework for EMS forecasting.

An in-depth feature analysis using SHAP and Pearson correlation revealed that historical dispatch patterns are the most influential predictors, with meteorological and temporal features offering modest incremental gains. This insight reinforces the importance of operational history in short-term EMS forecasting and suggests that weather-based complexity may not always translate to predictive power.

Overall, EM-LR emerges as a practical, transparent, and statistically validated solution for forecasting EMS demand. Its ability to deliver low-error, low-variance predictions without resorting to opaque deep learning architectures makes it a scalable and actionable tool for emergency management agencies. Future work will focus on deploying EM-LR across more districts and integrating probabilistic extensions to account for demand uncertainty and dynamic temporal shifts.



**Fig. 11.** Feature importance analysis for Varanasi District.

District	t-statistic	p value	Significance
Agra	3.04	0.028	Yes ( $p < 0.05$ )
Kanpur Nagar	2.86	0.034	Yes ( $p < 0.05$ )
Gorakhpur	3.22	0.024	Yes ( $p < 0.05$ )
Lucknow	2.93	0.031	Yes ( $p < 0.05$ )
Varanasi	1.84	0.112	No (marginal)

**Table 9.** Paired  $t$ -test results for EM-LR model under All-Features and Feature-Selected configurations.

Fold	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
1	2.67	5.88	2.83	2.67	4.18	5.34	4.02	2.60
2	4.23	5.97	3.92	3.80	4.33	5.71	4.15	3.76
3	6.48	9.96	8.55	8.37	4.54	6.01	4.49	5.97
4	4.69	7.84	5.99	5.97	4.36	6.32	4.82	4.11
5	3.25	4.23	3.30	3.38	3.99	5.96	3.95	3.33
Variance	<b>2.16</b>	<b>4.80</b>	<b>5.57</b>	<b>5.43</b>	<b>1.92</b>	<b>3.77</b>	<b>2.61</b>	<b>1.59</b>

**Table 10.** Variance of RMSE across different models for the Agra Dataset.

Fold	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
1	4.68	4.28	3.94	3.88	5.55	11.45	6.02	3.89
2	5.56	6.41	5.55	5.41	5.71	11.68	6.25	5.20
3	7.40	6.56	7.91	7.16	5.76	11.91	6.42	6.52
4	5.50	5.41	6.37	5.95	5.62	12.02	6.46	5.39
5	4.52	4.76	4.56	4.55	5.82	12.22	6.25	4.33
Variance	<b>1.31</b>	<b>1.00</b>	<b>2.44</b>	<b>1.61</b>	<b>0.68</b>	<b>1.14</b>	<b>1.02</b>	<b>1.04</b>

**Table 11.** Variance of RMSE across different models for the Gorakhpur dataset.

Fold	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
1	3.11	4.10	3.23	3.34	4.18	6.91	4.41	3.02
2	5.54	6.88	5.29	5.00	4.28	7.22	4.52	4.65
3	5.10	5.16	5.88	5.63	4.33	7.35	4.61	4.78
4	5.30	5.60	6.76	7.16	4.41	7.52	4.79	4.89
5	3.73	3.98	3.87	4.15	4.09	7.39	4.40	3.71
Variance	<b>1.14</b>	<b>1.41</b>	<b>2.08</b>	<b>2.13</b>	<b>0.48</b>	<b>0.72</b>	<b>0.55</b>	<b>0.66</b>

**Table 12.** Variance of RMSE across different models for the Kanpur Nagar dataset.

Fold	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
1	3.96	4.67	3.81	3.80	6.72	8.81	6.32	4.08
2	9.31	8.11	8.10	8.22	6.94	9.05	6.58	7.66
3	8.56	9.64	10.15	10.49	6.89	9.38	6.66	7.31
4	8.81	8.20	7.25	6.07	7.05	9.52	6.83	6.10
5	5.73	5.58	5.59	5.61	6.92	9.38	6.37	4.91
Variance	<b>5.39</b>	<b>4.20</b>	<b>5.85</b>	<b>6.64</b>	<b>2.71</b>	<b>3.04</b>	<b>2.47</b>	<b>2.34</b>

**Table 13.** Variance of RMSE across different models for the Lucknow dataset.

Fold	MLP	SVR	RF	XGB	AHELM	TBLSSVR	MHKLDMR	EM-LR
1	3.01	3.62	2.81	2.66	3.59	4.62	3.52	2.85
2	4.26	5.45	3.76	3.69	3.77	4.79	3.68	3.64
3	4.11	5.81	4.60	4.30	3.93	5.09	3.81	4.03
4	3.73	5.41	4.14	4.02	3.88	5.22	3.75	3.70
5	2.89	3.22	2.88	2.95	3.64	4.48	3.59	2.83
Variance	<b>0.39</b>	<b>1.41</b>	<b>0.61</b>	<b>0.49</b>	<b>0.34</b>	<b>0.56</b>	<b>0.31</b>	<b>0.30</b>

**Table 14.** Variance of RMSE across different models for the Varanasi dataset.

## Data availability

Data sets analyzed during the current study are available from the corresponding author on reasonable request.

Received: 12 May 2025; Accepted: 5 December 2025

Published online: 09 December 2025

## References

1. Peleg, K. & Pliskin, J. S. A geographic information system simulation model of EMS: Reducing ambulance response time. *Am. J. Emerg. Med.* **22**(3), 164–170 (2004).
2. Peters, J. & Hall, G. B. Assessment of ambulance response performance using a geographic information system. *Soc. Sci. Med.* **49**(11), 1551–1566 (1999).
3. Fitch, J. Response times: Myths, measurement & management. *J. Emerg. Med. Serv.* **30**(9), 47–56 (2005).
4. Henriksen, F. L., Schorling, P., Hansen, B., Schakow, H. & Larsen, M. L. First AED emergency dispatch, global positioning of first responders with distinct roles—A solution to reduce the response times and ensuring early defibrillation in the rural area Langeland. In *Safe and Secure Cities* (Communications in Computer and Information Science), New York, NY (Springer, USA, 2014).
5. Henderson, S. G. & Mason, A. J. Ambulance service planning: simulation and data visualisation. *Int. Ser. Oper. Res. Manag. Sci.* **70**, 77–102 (2004).
6. Geroliminis, N., Karlaftis, M. G. & Skabardonis, A. A spatial queuing model for the emergency vehicle districting and location problem. *Trans. Res. Part B, Methodol.* **43**(7), 798–811 (2009).
7. Lim, C. S., Mamat, R. & Braunl, T. Impact of ambulance dispatch policies on performance of emergency medical services. *IEEE Trans. Intell. Trans. Syst.* **12**(2), 624–632 (2011).
8. Rajagopalan, H. K. Ambulance deployment and shift scheduling: An integrated approach. *J. Serv. Sci. Manag.* **04**(01), 66–78 (2011).
9. Yaseen, F. A., Alkhalidi, N. A. & Al-Raweshidy, H. S. SHE Networks: Security, health, and emergency networks traffic priority management based on ML and SDN. *IEEE Access* **10**, 92249–92258 (2022).
10. Liu, K., Li, Q. & Zhang, Z.-H. Distributionally robust optimization of an emergency medical service station location and sizing problem with joint chance constraints. *Transp. Res. Part B: Methodol.* **119**, 79–101 (2019).
11. Amorim, M., Antunes, F., Ferreira, S. & Couto, A. An integrated approach for strategic and tactical decisions for the emergency medical service: Exploring optimization and metamodel-based simulation for vehicle location. *Comput. Ind. Eng.* **137**, 106057 (2019).
12. Xia, T. et al. Measuring spatio-temporal accessibility to emergency medical services through big GPS data. *Health Place* **56**, 53–62 (2019).

13. Kvalseth, T. O. Regression models of emergency medical service demand for different types of emergencies. *IEEE Trans. Syst. Man Cybern.* **9**(1), 10–17 (1979).
14. Nicoletta, V., Guglielmi, A., Ruiz, A., Bélanger, V. & Lanzarone, E. Bayesian spatio-temporal modelling and prediction of areal demands for ambulance services. *IMA J. Manag. Math.* **33**(1), 101–121 (2022).
15. Aldrich, C. A., Hisserich, J. C. & Lave, L. B. An analysis of the demand for emergency ambulance service in an urban area. *Am. J. Public Health* **61**(6), 1156–1169 (1971).
16. Siler, K. F. Predicting demand for publicly dispatched ambulances in a metropolitan area. *Health Serv. Res.* **10**(3), 254 (1975).
17. Kvalseth, T. O. & Deems, J. M. Statistical models of the demand for emergency medical services in an urban area. *Am. J. Public Health* **69**(3), 250–255 (1979).
18. Wong, H. T. & Lai, P. C. Weather inference and daily demand for emergency ambulance services. *Emerg. Med. J.* **29**(1), 60–64 (2012).
19. Lowthian, J. A. et al. The challenges of population ageing: Accelerating demand for emergency ambulance services by older patients, 1995–2015. *Med. J. Aust.* **194**(11), 574–578 (2011).
20. Steins, K., Matinrad, N. & Granberg, T. Forecasting the demand for emergency medical services (2019).
21. Ibrahim, R. et al. Modeling and forecasting call center arrivals: A literature survey and a case study. *Int. J. Forecast.* **32**(3), 865–874 (2016).
22. Channouf, N. et al. The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health Care Manag. Sci.* **10**, 25–45 (2007).
23. Vile, J. L. et al. Predicting ambulance demand using singular spectrum analysis. *J. Oper. Res. Soc.* **63**(11), 1556–1565 (2012).
24. Wong, H.-T. & Lai, P.-C. Weather factors in the short-term forecasting of daily ambulance calls. *Int. J. Biometeorol.* **58**, 669–678 (2014).
25. Villani, M. et al. Time series modelling to forecast prehospital EMS demand for diabetic emergencies. *BMC Health Serv. Res.* **17**, 1–9 (2017).
26. Jin, R., Xia, T., Liu, X., Murata, T. & Kim, K.-S. Predicting emergency medical service demand with bipartite graph convolutional networks. *Ieee Access* **9**, 9903–9915 (2021).
27. Xia, T. et al. Measuring spatio-temporal accessibility to emergency medical services through big GPS data. *Health Place* **56**, 53–62 (2019).
28. Grekousis, G. & Liu, Y. Where will the next emergency event occur? Predicting ambulance demand in emergency medical services using artificial intelligence. *Comput. Environ. Urban Syst.* **76**, 110–122 (2019).
29. Abreu, P., Santos, D. & Barbosa-Povoa, A. Data-driven forecasting for operational planning of emergency medical services. *Socio-Econ. Plan. Sci.* **86**, 101492 (2023).
30. Martin, R. J., Mousavi, R. & Saydam, C. Predicting emergency medical service call demand: A modern spatiotemporal machine learning approach. *Oper. Res. Health Care* **28**, 100285 (2021).
31. Hermansen, A. H. & Mengshoel, O. J. Forecasting ambulance demand using machine learning: A case study from Oslo, Norway. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2021, 01–10 (2021).
32. Khatri, K. L. & Tamil, L. S. Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks. *IEEE J. Biomed. Health Inform.* **22**(1), 285–290 (2017).
33. Wong, H. T. & Lai, P. C. Weather factors in the short-term forecasting of daily ambulance calls. *Int. J. Biometeorol.* **58**, 669–678 (2014).
34. Chen, A. Y., Lu, T.-Y., Ma, M.H.-M. & Sun, W.-Z. Demand forecast using data analytics for the preallocation of ambulances. *IEEE J. Biomed. Health Inform.* **20**(4), 1178–1187 (2015).
35. Lundberg, S. & Lee, S.-I. SHAP: A Unified Approach to Interpreting Model Predictions. *Advances in neural information processing systems*, 1–10 (2017).
36. Ramgopal, S., Westling, T., Siripong, N., Salcido, D. D. & Martin-Gill, C. Use of a metalearner to predict emergency medical services demand in an urban setting. *Comput. Methods Programs Biomed.* **207**, 106201 (2021).
37. Megouo, T. G. P. & Pierre, S. A stacking ensemble machine learning model for emergency call forecasting. *IEEE Access* **12**, 115820–115837 (2024).
38. Kumari, Pratima & Toshniwal, Durga. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *J. Clean. Prod.* **279**, 123285 (2021).
39. Gupta, D., Hazarika, B. B. & Berlin, M. Robust regularized extreme learning machine with asymmetric Huber loss function. *Neural Comput. Appl.* **32**(16), 12971–12998 (2020).
40. Chen, R., Liu, M., & Ma, J. Twin Bounded Least Squares Support Vector Regression. In *Proceedings of the International Conference on Intelligence Science*, 40–54 (Springer, 2024).
41. Gupta, D., Hazarika, B. B. & Berlin, M. Wavelet kernel large margin distribution machine-based regression for modelling the river suspended sediment load. *Comput. Electr. Eng.* **120**, 109783 (2024).
42. Ranstam, Jonas & Cook, Jonathan A. LASSO Regression. *J. Br. Surg.* **105**(10), 1348–1348 (2018).
43. Awad, M. & Khanna, R. Support Vector Regression. In *Efficient Learning Machines: Theories Concepts, and Applications for Engineers and System Designers* 67–80 (Springer, 2015).
44. Taud, H. & Mas, J.-F. Multilayer Perceptron (MLP). In *Geomatic Approaches for Modeling Land Change Scenarios* 451–455 (Springer, 2017).
45. Chen, T. et al. XGBoost: Extreme Gradient Boosting. *R Package Version 0.4-2* **1**(4), 1–4 (2015).
46. Liu, J. & Yubo, X. T-Friedman test: A new statistical test for multiple comparison with an adjustable conservativeness measure. *Int. J. Comput. Intell. Syst.* **15**(1), 29 (2022).
47. Xu, M. et al. The differences and similarities between two-sample t-test and paired t-test. *Shanghai Arch. Psychiatry* **29**(3), 184 (2017).

## Author contributions

T.G. conceived and conducted the experiments. T.G., D.T., and M.P. analyzed the data and results. T.G. wrote the paper. T.G., D.T., and M.P. reviewed the manuscript.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

Correspondence and requests for materials should be addressed to T.G. or D.T.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025