IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 July 2024, accepted 13 August 2024, date of publication 19 August 2024, date of current version 29 August 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3445591

RESEARCH ARTICLE

A Stacking Ensemble Machine Learning Model for Emergency Call Forecasting

TALOTSING GAELLE PATRICIA MEGOUO AND SAMUEL PIERRE, (Senior Member, IEEE)
Mobile Computing and Networking Research Laboratory (LARIM), Department of Computer and Software Engineering, Polytechnique Montréal, Montreal, QC H3T 1J4, Canada

Corresponding author: Talotsing Gaelle Patricia Megouo (talotsing.gaelle-patricia-megouo@polymtl.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), Prompt, Flex Group, and Gestion Immobilière Isame (ISAME).

ABSTRACT One of the greatest challenges of Emergency medical services providers is to handle the large number of Emergency Medical Service (EMS) calls coming from the population. An accurate forecast of EMS calls is involved in ambulance fleet dispatching and routing to minimize response times to emergency calls and enhance the efficacy of assistance. Yet, the demand for emergency services exhibits significant variability, posing a challenge in accurately predicting the future occurrence of emergency calls and their spatial-temporal distribution. Here, we propose a stacking ensemble machine learning model to forecast EMS calls, combining different base learners to enhance the overall performance of generalization. Additionally, we conducted experiments using Boruta, Lasso, RFFI and SHAP feature selection methods to identify the most informative attributes from the EMS dataset. The proposed ensemble model integrates a base layer and a meta layer. In the base layer, we applied four base learners: Decision Tree, Gradient Boosting Regression Tree, Light Gradient Boosting Machine and Random Forest. In the meta layer, we used an optimized Random Forest model to integrate the outputs of base learners. We evaluate the performance of our proposed model using the R^2-score and four different error metrics. Based on a real data set including spatial, temporal and weather features, the findings of this study demonstrated that the proposed stacking-based ensemble model showed a better score and the minimum errors compared to the traditional single algorithms, online machine learning methods and voting ensemble methods. We achieved a higher score of 0.9954, mse of 0.8938, rmse of 0.9454, mae of 0.2923 and mape of 0.0724 compared to state-of-the-art models. This work is an aid for emergency managers in making well-informed decisions, improving outcomes for ambulance dispatch and routing, and enhancing ambulance response time.

INDEX TERMS Ambulance demand forecasting, artificial intelligence, EMS call forecasting, ensemble machine learning, feature selection, offline/online machine learning.

I. INTRODUCTION

Emergency medical services (EMS) play a crucial role in providing citizens with a higher probability of survival by ensuring short response times to emergency calls, especially urgent ones, and faster ambulance arrivals at their destination [1], [2]. EMS managers tackle this task by studying the distribution of incoming call requests and formulating resource deployment plans, specifying the number of ambulances and emergency response personnel required for current and future periods [3].

In smart cities, uncertain demand patterns pose challenges for EMS centers as emergency call volumes fluctuate significantly throughout the day. EMS providers face also significant challenges due to geographical disparities, and temporal fluctuations. Accurate forecasting is crucial for optimizing ambulance deployment, improving response times, and enhancing patient care [2]. Accurate and real-time prediction of ambulance service demand is complex, requiring routing algorithms to dynamically adapt to changing demand patterns while improving response times. Developing precise

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Ni.

algorithms that account for factors such as time, location, and weather is essential to effectively address these varying demand patterns and ensure optimal resource allocation. While several studies address EMS demand forecasting, identifying the best algorithm with minimal error and shortest execution time remains challenges. Such challenges can impact ambulance planning and allocation. Given the significance of call time and location, implementing a precise algorithm that considers these parameters is essential.

Geographic information system (GIS) emerges as a powerful tool for collecting, managing, analyzing, and representing geo-referenced health data and identifying gaps in health systems [4], [5], [6], [7], [8]. GIS enables researchers to integrate spatial (e.g., health service locations, patient addresses, and ambulance dispatch centers) and non-spatial (e.g., descriptive information about geographic features, work hours, waiting lists) data into a unified framework, facilitating better-informed decision-making [6]. Additionally, the temporal aspect must be considered, since ambulance demand can vary depending on the time of day [9]. An accurate demand forecast can aid in improved ambulance management planning. The effective management of emergencies is expected to serve as a fundamental service in modern smart cities, exerting a direct influence on urban safety and the perceived quality of life while dealing with the impacts of climatic changes [10].

The main objective of this paper is to use an ensemble machine learning model for the analysis and forecasting of EMS calls to enhance the management of ambulance fleets in time and by location. This paper introduces a stacking forecasting method described as an ensemble, incorporating spatial, temporal, and climatological parameters to support tactical decisions for ambulance deployment and planning. The workflow is summarized in four significant ways: Firstly, we started with a data collection of EMS calls by time and location and integrated the weather data to form our final dataset. Secondly, we aggregate the temporal and spatial call numbers per zone since ambulances are deployed per zone. Thirdly, we design a stacking ensemble machine learning model that predicts the number of calls considering spatial, temporal, and weather parameters. Lastly, we add online algorithms [11], as baseline to approximate real-time, stream EMS calls, and encompass temporal, spatial, and weather parameters for accurate time and location prediction. We explore various Machine learning (ML) and Deep learning (DL) algorithms both offline and online to achieve the objectives of this research work. We conduct a comparison between the best single offline, online models/algorithms and voting ensemble with our proposed ensemble model to predict EMS calls.

In this paper, we propose an ensemble EMS demand prediction method that takes into account time, location and weather parameters. The implementation of effective Emergency medical services (EMS) systems hinges on the accurate forecasting of ambulance demand, a critical aspect that significantly influences emergency response strategies and overall healthcare outcomes. The need for robust ambulance demand forecasting arises from the dynamic nature of emergency incidents, making it essential to anticipate and allocate resources efficiently. Accurate forecasts empower EMS providers to optimize ambulance deployment, strategically position medical resources, and enhance response times, ultimately leading to improved patient care and outcomes. By leveraging advanced forecasting models, EMS systems can proactively address the challenges of varying demand patterns, geographical disparities, and temporal fluctuations, fostering a more resilient and responsive emergency healthcare infrastructure.

Furthermore, the integration of computational time considerations in ambulance demand forecasting models becomes imperative. Efficient computational processes ensure timely and real-time decision-making, allowing EMS systems to dynamically adapt to evolving scenarios. By leveraging advanced forecasting models with optimized computational efficiency, EMS systems can proactively address the challenges of varying demand patterns, geographical disparities, and temporal fluctuations, fostering a more resilient and responsive emergency healthcare infrastructure. To the best of our knowledge, our system is the first to propose a stacking ensemble model for EMS call forecasting, considering spatial, temporal and weather parameters, with the performance analysis based on score, and errors. Using a real dataset, our primary objective is to identify the most accurate model for predicting the next EMS calls based on historical EMS call data. We evaluate the forecast results of the proposed model against the following ML algorithms for regression problems: Gradient Boosting Regression Tree (GBRT), Light Gradient Boosting Machine (LGBM), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), Long Short-Term Memory (LSTM), Online-Linear Regression, and Online-ANN.

The originality of this paper lies in its integration of different machine learning techniques into a stacking model that simultaneously considers time parameters, location parameters, and climatological parameters for EMS demand forecasting. Single machine learning algorithms can make different types of mistakes when predicting outcomes from data. Some may have more bias, while others may have more variance. To handle these issues and improve the score, ensemble methods combine results from multiple algorithms. This helps reduce errors overall. Stacking is one such approach that uses a meta-learner, to blend predictions from various basic machine learning models and enhance score by addressing bias and variance problems.

The contributions of the paper are as follows:

- We propose a comparative study of the feature selection method as an explanatory data analysis through the potential features that contribute towards EMS calls and ambulance demand.
- We compare the most used single offline and online machine learning models and DL models for spatial-temporal forecasting of EMS calls: GBRT,

LGBM, ANN, RF, DT, ANN and LSTM, aiming to determine the best-performing approach and identify models that offer minimum prediction error and adaptability to EMS calls patterns, which were then integrated as base learners in the first layer of our proposed model.
- We propose a novel stacking ensemble method to forecast EMS calls while incorporating spatial, temporal, and climatological parameters. The proposed model consists of two layers: the base layer and the meta layer. In the base layer, DT, GBRT and LGBM models consist of our best base learners. In the meta layer, we employ an optimized Random Forest model as our meta-learner.
- We prove that our proposed stacking strategy outperforms state-of-the-art models when applied to the real datasets. It demonstrates reduced prediction errors, ensuring reliability and robustness in capturing underlying EMS calls data patterns. Its enhanced adaptability and interpretability make it a valuable and versatile tool for practical applications in a higher resolution.

The rest of this paper is organized as follows. Section II presents the related work and provides an overview of the proposed prediction approaches in ambulance demand forecasting. Section III provides a detailed explanation of our research methodology and describes the implementation of the proposed workflow. Section IV outlines the experimental tool and performance evaluation. Finally, Section V concludes the paper.

II. RELATED WORK

In this section, we present approaches and algorithms for EMS call prediction, present in the literature review. We divided in four different groups: empirical estimation models, classical time series models, probabilistic models and learning based models.

A. EMPIRICAL ESTIMATION MODELS

Empirical estimation models are among the first classic models widely used in industry. They are characterized by predefined statistical calculations, such as averaging over observations at different time intervals, such as seasons, months, or weeks. The most well-known empirical models are the Naive Predictive (NP), the High Availability (HA), and the MEDIC [12]. The NP provides a cost-effective forecast using the last observed demand value. The demand at time t, knowing the demand at the time before $t-1$, is assumed to be proportional to the demand at time t. HA averages all available historical observations of the corresponding distribution region over the previous year to produce a forecast. MEDIC is a common industry practice deployed in cities such as Toronto and Charlotte [12], [13]. It involves averaging the last Z same-hour observations from several previous months of the forecast. The MEDIC method can be considered as a combination of the HA method and the naive method, taking into account the daily behavior of the demand. For example, in [14], the authors set $Z = 20$, in line with the practice in 2021. Similarly, the work of [12] averaged the corresponding demand densities in the previous 4 weeks for any 2-hour period in March 2007. These empirical models are static and do not account for changes in EMS demand over time and other factors that may influence the demand patterns. As a result, their predictions may not be as accurate as models that consider additional variables and spatio-temporal dynamics in EMS demand forecasting.

B. CLASSICAL TIME SERIES MODELS

Time series models are statistical methods that describe the linear autocorrelations in a variable that evolves over time. Some of the commonly used time series models include Auto-Regression (AR), Vector Auto-Regression (VAR), and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) [15], [16], [17], [18], [19]. The VAR model is particularly useful for multivariate time series as it captures the linear interdependencies between different variables, which is relevant for representing spatial correlations in our case. In [14], Wang employed AR, VAR, and SARIMA as baseline models for EMS demand forecasting. Vile [20] proposed Singular Spectral Analysis (SSA) to generate accurate demand forecasts for the Service Medical d'Urgence (SMU). SSA is a non-parametric technique used for the analysis of time series. This method allows for a multimodal decomposition of EMS demand into periodic, trend, and noise components, enhancing the understanding of the underlying patterns. Considering the time-of-day and day-of-week effects, Cheng [9] investigated the use of the SARIMA with external regressor (SARIMAX) model to forecast the hourly occupancy of the Emergency Department (ED) up to 4 hours ahead. The SARIMAX method utilizes readily available data in most emergency departments to generate prediction intervals, making it a promising technique for real-time forecasting of emergency department occupancy. While these time series models consider the temporal distribution of EMS demand, they do not explicitly account for the spatial distribution of EMS calls and the potential effects of other events on this variation. To create more comprehensive EMS demand forecasting models, it is important to incorporate both temporal and spatial aspects, along with other relevant factors that may influence EMS demand.

C. PROBABILISTIC MODELS

Probabilistic models are based on statistical inferences. There are several probabilistic models used to forecast EMS demand, such as Gaussian models, kernel density estimation, and Bayesian models. Zhou et al. [12] proposed a Gaussian Mixture Model (GMM) to estimate the distribution of ambulance demand in Toronto. This model utilizes the data distribution to fix the distributions of mixture components over all time periods, addressing data sparsity and accurately describing the spatial structure of Toronto. The GMM captures complex spatio-temporal dynamics through time-varying mixture weights, which include weekly seasonality and a conditional autoregressive priority on the mixture weights of each component. Xu et al. [21] proposed a locally adaptive Space-Time Kernel Density Estimate (ST-KDE) to model EMS queries as an InHomogeneous Poisson Process (IHPP). ST-KDE is a non-parametric method for estimating probability density functions in statistics. This method weighs the spatial kernels by functions based on the corresponding time dependence in each community area, enabling the incorporation of complex spatio-temporal variations in EMS applications. Steins et al. [22] proposed a Zero-Inflated Poisson (ZIP) regression to forecast ambulance calls in Swedish counties. ZIP models account for zero-inflation and Poisson processes, enabling more reliable predictions for count data with excess zeros. Their study estimated EMS calls per hour and geographical zone, involving historical data analysis and spatial-temporal features for improved accuracy. Standard Poisson regression assumptions may be violated in real-world scenarios with excessive zeros, posing challenges for selecting the most appropriate forecasting model for the EMS context, considering various models' strengths and weaknesses. Nicoletta et al. [23] presented a Bayesian model using MCMC for posterior inferences. The method predicted present-day requests based on past probabilities, showing effectiveness. However, the city division aspect has limitations, categorizing areas into four traffic levels. Emergency centers consider multiple factors for division. Enhancing predictions could involve considering short, medium, and long-term variations, public holidays, and other relevant factors. To improve applicability, the model should consider various characteristics beyond location-based predictions.

D. LEARNING BASED MODELS

The learning-based model integrates Machine Learning (ML) and Deep Learning (DL) methods. This model involves collecting large number of examples to identify underlying patterns and use them for predicting new ones. Hermansen et al. [26] presented Multi-Layer Perceptron (MLP) and LSTM methods for medical service request prediction in Oslo. They tested split and complete approaches with weather data in 1 km radius at 1-hour intervals. A high resolution led to sparse data, challenging predictions. MLP performed better, considering temperature and precipitation's impact. Peak days, population density, and probabilistic predictions should be explored for enhancement, along with user mobility and weather variations. Lin et al. [25] compared six machine learning methods (Regional Moving Average (RMA), LR, Support Vector Regression (SVR), MLP, Radial Basis Function Network (RBFN), and LGBM. LGBM performed best for both 7-day and 30-day predictions based on EMS demand and social aspects of populations. Response time and additional features influencing EMS demands should be considered for more comprehensive comparisons and real-life applicability. Wang et al. [14] used daily human mobility data to improve spatial correlations' representation. They introduced a Heterogeneous Multi-Graph Convolution Network (HMGCN) and a Spatio-Temporal Interlacing Attention Module (STIAM) to predict EMS demand, outperforming nine other models by incorporating dynamic human mobility. Validating the approach with small-resolution patterns (e.g., weekend mobility, daily periods, holidays) is necessary. Applying the method to more spatio-temporal prediction tasks would further validate its effectiveness. Nakai et al. [29] developed a machine learning model for predicting the number of heat stroke victims in Kobe City using past weather observation data and emergency dispatch records. The Partial Least Squares Regression (PLSR) method was employed for medium-term (4-7 days) predictions using past weather forecast data. However, they only used weekly weather forecast elements as explanatory variables, leaving room for exploration of other explanatory variables. Jin et al. [30] focused on three main factors influencing EMS demand: population density, socioeconomic factors of the study area, and hospital conditions. They proposed a bipartite computational graph neural network (BiGCN) to exploit these features and achieved promising results compared to other methods. Rautenstrauss et al. [28] introduced a Convolutional Neural Network (CNN) for ambulance demand prediction, transforming time series information into heatmaps. While the CNN method performed well in terms of prediction error, its known drawback is the time-consuming execution. In summary, the learning-based models hold promise for EMS demand forecasting, and considering various factors, exploring different resolutions, and optimizing model execution time are areas for improvement and future research.

The literature encompasses a wide range of models, which contribute significantly to the field. Accurately predicting ambulance demand for emergency medical services is of utmost importance, as it directly impacts the allocation of ambulances to emergency calls [31]. By achieving a more precise and diverse estimate of demand behavior, we can enhance the reallocation and routing of ambulances across different areas. This optimization leads to minimized ambulance response time [32], thus increasing the likelihood of timely patient care and follow-up. Table 1 shows a summary of previous studies in ambulance demand prediction, their proposed methods and the metrics used. While several studies address EMS demand forecasting, identifying the best algorithm with minimal error and shortest execution time remains challenging. Existing methodologies, such as traditional statistical models and basic machine learning algorithms, often struggle with these dynamic and complex patterns.

In the domain of EMS forecasting, there is a growing need to harness the power of artificial intelligence (AI) tools, particularly machine learning (ML) models. ML, a subset of AI, is instrumental in crafting accurate predictive models for EMS operations. However, individual ML models may exhibit weaknesses and limitations. To address these challenges effectively, the concept of ensemble learning (EL) emerges. EL involves combining multiple ML models to create a stronger, more robust forecasting framework tailored for EMS operations. Given the complex and nonlinear nature of EMS data, EL methods have garnered

TABLE 1. Methods and metrics used in other related studies.

Authors, Year	Feature names	FS Method(s)	Methods ML	Metrics
Zhou et al. [24] 2015	03 time periods 21 location cells	No FS	MEDIC, naiveKDE, GMM, **stKDE**.	Average log score
Chen et al. [4] 2015	Year, season, month, day, day of week, time bucket, weekend, rush hour, past state of EMS demand and rainfall	Data Analytics (DA)	SVR, Sinusoidal Regression, MA, ANN	RMSE, MAPE
Lin et al. [25] 2020	Spatial, temporal, and demographic	No FS	RMA Linear Regression, SVR, MLP **LightGBM**	WAPE MAE MSE
Hermansen et al. [26] 2021	Hour, day, day of week, month, precipitation, temperature	No FS	MEDIC, ANN, LSTM	MSE, MAE, CCE
Martin et al. [3] 2021	Call time, call location coordinates, responding ambulance identifier, assigned call priority, patient problem description, call response outcome the time in route and arrival time, and a primary incident number.	K-means, Boruta	ARIMA, Holts–Winters (HW), MEDIC Hourly Forecasting (MHF), MLP	Mean absolute deviation (MAD), (MAPE)
Nicoletta et al. [23] 2022	time, type of zone precipitation, temperature	Posterior inference	Markov chain Monte Carlo	MAE, Empirical coverage
Van et al. [27] 2023	Hour, day, day of week, month, precipitation, temperature	Statistical decomposition: (STL, SSA) - No FS	Simple moving average (SMA) MEDICis, MLP Naive forecast (NF) GA-ANN	MSE, MAE
Rausten et al. [28] 2023	Temperature, wind speed, humidity, dew point, sea level pressure, precipitation public holidays, school holidays, and events	Shap	CNNs, Multilayer perceptrons (MLP), Decision trees (DT), Random forests (RF), MEDIC	MSE

considerable attention. By integrating diverse ML models, EL not only mitigates the limitations of individual models but exploits their varied perspectives to enhance prediction accuracy. Furthermore, EL contributes to error reduction, faster computation, and improved generalization of EMS forecasts. By stacking different learning approaches, we can improve predictions, reduce errors, leverage computations, and create more generalizable forecasts for EMS situations.

III. METHODOLOGY

In this section, we propose an effective stacking ensemble learning model by taking advantage of ensemble learning properties. The proposed model is used for EMS call forecasting. We begin by formulating the problem of EMS call forecasting. We introduce the geo-grid division strategy along with K-means and DBSCAN for data spatial aggregation. We present both single offline and online machine learning algorithms that aim to predict EMS call. Figure 1 illustrates the workflow of our forecasting model. It encompasses the definition of the data collection, the data aggregation and feature selection, the introduction of our proposed ensemble forecasting model and the metrics for performance analysis. This figure represents the main steps we take to achieve the objectives of the analysis and prediction of EMS calls for better ambulance allocation and dispatching.

A. PROBLEM FORMULATION

Emergency medical systems provide first aid and vital medical assistance, including transportation and victim transfer [33]. Calls are received via designated emergency numbers or alarm systems, and the urgency and location of each call are assessed to dispatch ambulances promptly. The aims of this research work are to forecast the ambulance demand $y(t, z) = d_z^t$, representing the number of incidents (demand d) at time step t in zone z for $z \in Z$ (with $z \in \mathbf{N}^*$). Z represents a set of spatial clusters or zones, and $\mathbf{N}^*$ represents the set of positive integers. For a given geographical area, the larger the number of zones in Z, the higher the spatio-temporal resolution of the EMS demand forecasts. With a higher resolution and shorter time steps, more information is available to strategically dispatch ambulances. However, using a high spatio-temporal resolution also presents challenges, as the data becomes sparser and more stochastic, making forecasting more difficult.

B. DATA PREPROCESSING

Data preprocessing ensures that the data is free from inconsistencies, errors, and missing values, leading to more accurate and reliable results in predictive modeling and decision-making processes. Moreover, it involves collection, cleaning, transforming, and organizing data to improve its quality and make it suitable for forecasting. After removing

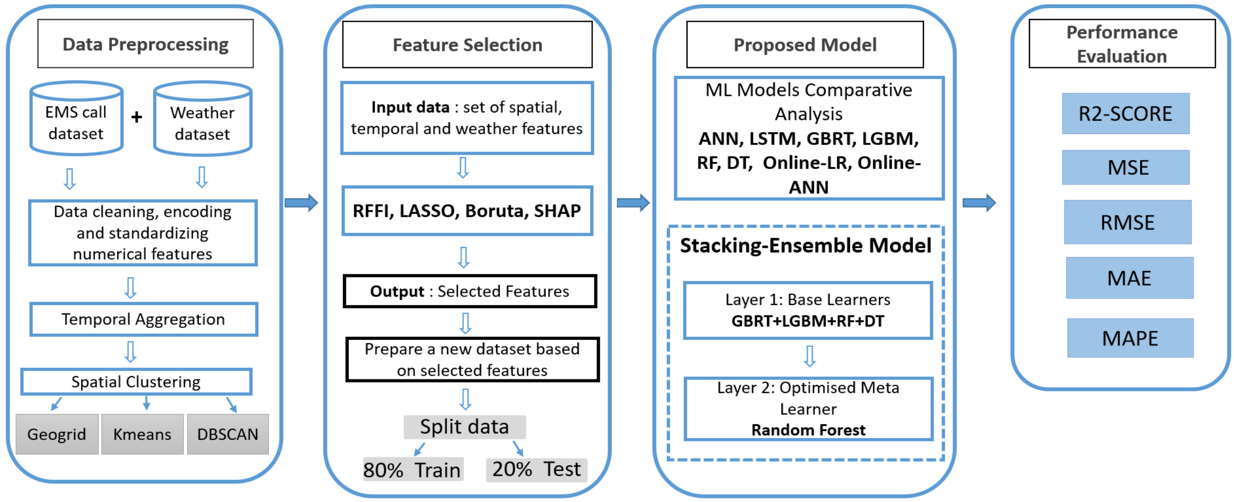


FIGURE 1. **The workflow of the proposed methodology for EMS call forecasting.**

missing and meaningless values in the dataset, we encoded categorical variables into a numerical format that can be easily understood and processed by machine learning algorithms using ordinal encoding and one-hot encoding [34]. In ordinal encoding, each unique category value is assigned an integer value. One-hot encoding creates binary vectors for each category in the categorical variable. These techniques are commonly employed when dealing with categorical features in regression tasks, and its help to enhance machine learning model compatibility and interpretability while ensuring the effective utilization of diverse types of data. After that, we evaluate the feature importance. To accurately estimate the target variable $y(t, z) = d_z^t$, we carefully assess and analyze the importance of spatial, temporal, and climatological features. Notably, the spatial and temporal features heavily rely on clustering techniques, enhancing our understanding of the data's spatial distribution and patterns.

C. SPATIAL CLUSTERING AND FEATURE SELECTION

1) SPATIAL CLUSTERING

To allow our results for being used for ambulance dispatching and routing, we divided the area of study in zones/cluster based on the location of the ambulance demand and the time. The spatial cluster represents the base stations where the ambulance is stationed. An effective aggregation in space will ensure a better coverage by EMS providers and reduce the waiting time. We aggregate data in small clusters. Inspired by [35], the comparison will focus on evaluating the effectiveness and performance of three methods into creating divisions that represent the service area efficiently.

- The geo-grid division, is a rectangular clustering based on latitude and longitude locations. The study area is divided into small rectangular zones. The number of zones is the number of splits, calculated as describe in Equation (1):

$$Nb_{zones} = Lat_{grid} \times Lng_{grid}, \tag{1}$$

  where Lat_{grid} and Lng_{grid} are integers obtained by dividing the range of latitude ($Max_{Latitude} - Min_{Latitude}$) and longitude ($Max_{Longitude} - Min_{Longitude}$) into the desired number of splits.
- K-means clustering is an unsupervised clustering algorithm within machine learning that dynamically assigns data points into K distinct, non-overlapping clusters [35], [36].
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a clustering algorithm for data sets with varying density. It groups points based on density within a specified radius [37]. DBSCAN excels in processing large databases and has been applied in real case studies [35], [38], [39], [40].

After the aggregation of data, we selected the most important feature from spatial, temporal and weather data.

2) FEATURE SELECTION

We compared recent feature selection methods for data regression using machine learning algorithms. Boruta [3], LASSO Regression (L1 Regularization), Random Forest Feature Importance (RFFI) and SHAP (SHapley Additive exPlanations) [28], used in [41], [42], and [43]. Boruta iteratively compares the importance of original features to shadow features and selects those with importance above the threshold. LASSO introduces a penalty term to the regression equation, driving some feature coefficients to exactly zero and, identifying critical features. In RFFI, feature importance is measured by the average decrease in impurity across all trees, highlighting critical features for predicting the target variable. SHAP is a unified framework for explaining the output of machine learning models by assigning a value to each feature's contribution in to a prediction [44], [45]. These feature selection methods could reduce dataset dimensionality and improve model interpretability and performance in our forecasting task.

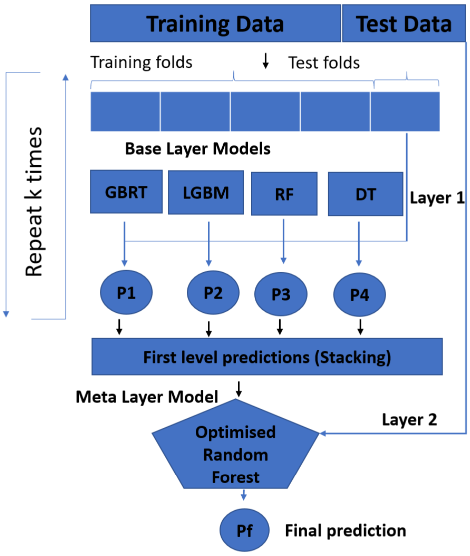


FIGURE 2. Stacking ensemble learning algorithm.

D. PROPOSED STACKING ENSEMBLE MODEL

The proposed ensemble model is designed based on the concept of stacking models, with the intention of leveraging the distinct components that can be identified in a spatio-temporal series, such as seasonality, trend, inertia, and spatial relations. By stacking different architectures, each focused on modeling a specific component, we aim to prevent redundant information from flowing through the model and to capitalize on the strengths of each approach [46]. Consequently, employing multiple layers of interpretability enables us to gain deeper insights into the problem being modeled and verify that our model is performing as expected. Figure 2 illustrates the stacking ensemble learning method. For a given input of k spatial, temporal and climatological features $x_i \in \{x_1, \ldots, x_k\}$, and the target variable $y(t, z) = d_z^t \in \mathbb{R}_+^Z$ representing the ambulance demand in zone z at time step t, we utilize a collection of different selected forecasting models denoted as $M = \{M_1, \ldots, M_m\}$. The individual model predictions are represented by $P = \{P_1, \ldots, P_m\}$.

The steps of the ensemble technique that combine information from multiple predictive models and use them as features to generate a new model can be described in Algorithm 1. The K-fold helps to find the optimal values of hyperparameters that give the best performance for each model in different subsets of data. The K-fold is a valuable tool when working with ensemble techniques, as it helps in both the evaluation and training phases, resulting in more robust and generalizable models. Stacking is a method that uses a special learning technique to figure out how to combine predictions from different machine learning models. It works by using predictions made by these models on new data, which they haven't seen before. These predictions, along with the actual outcomes, are then used to teach another model, called the meta-model. This meta-model learns how to best combine the predictions from the base models. In regression, the base models predict actual values.

Algorithm 1 Ensemble Stacking Model With Multiple Base Models and K-Fold Cross-Validation

1: **Input:**
k spatial, temporal and climatological features $x_i \in \{x_1, \ldots, x_k\}$
2: **Target:** variable $y(t, z) = d_z^t \in \mathbb{R}_+^Z$
3: **Step 1:** Select a K-fold split of the dataset
4: **Step 2:** Select m base models
Base Models: Define $M = \{M_1, \ldots, M_f\}$
GBRT, LGBM, DT, RF as describe in Table 6.
5: **Step 3:** For each base model, evaluate using K-fold cross-validation, store all out-of-fold predictions and fit the model on the full training dataset and store it.
6: **Step 4:** Fit the optimized metamodel Random Forest on the out-of-fold predictions from the base models.
7: **Step 5:** Evaluate the meta-model on the test set.
8: **Output:** The evaluation metrics

The stacking ensemble method is powerful because it can leverage the strengths of different models to produce a more accurate prediction. In our proposed model, denoted as GBRT+LGBM+DT+RF, we integrate the stacking of GBRT, LGBM, DT, and RF comparing with others single models. Our meta-model is an optimized Random Forest. The models included in our stacking ensemble were selected after many comparisons with machine learning models like GBRT, LGBM, LSTM, RF, DT, ANN, Online-LR, and Online-ANN, based on their proven effectiveness in EMS call forecasting. The chosen base learners GBRT, LGBM, DT, and RF perform well in regression tasks and capture complex patterns in EMS data. Our ensemble approach integrates its advantages, resulting in superior predictive accuracy and overall performance across multiple metrics, including R-squared (R^2), MSE, RMSE, MAE, MAPE, and computational time.

E. SELECTED SINGLE BASELINE ML ALGORITHMS

This subsection presents the selected and proposed Artificial Intelligence (AI) based algorithms for EMS call prediction. The following prominent learning-based models are considered for the assessment in this study.

1) OFFLINE FORECASTING METHODS

Offline or batch learning refers to traditional learning over all the observations in a dataset at once. We investigate six offline models for forecasting EMS call demand at different hours of the day: GBRT, LGBM, DT, RF, ANN and LSTM. These models represent the baseline models proposed by some authors [14], [25], [26], [29], [30], [47].

- Gradient Boosting Regression Tree (GBRT) is a boosting ensemble method that iteratively fits a new regression decision tree to the forecasting errors at each step. Figure 3 represents the diagram of the GBRT algorithm.
  Let $x_i, y_i(t, z)$ indicate the sample data, where $x_i \in \{x_1, \ldots, x_k\}$ represents the spatial, temporal and

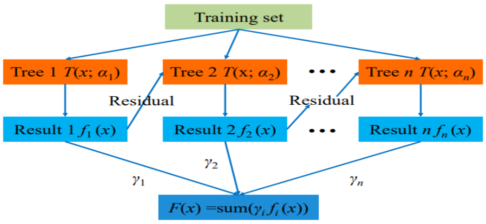


FIGURE 3. **Diagram of the GBRT algorithm [48].**

climatological features, and $y_i(t, z) = d_z^t$ denotes the target, described as the ambulance demand at time t in zone z. The specific steps of GBRT are as follows [48], [49]:

- **Step 1:** The initial constant value γ is obtained as:

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, \gamma), \tag{2}$$

where $L(y_i, \gamma)$ is the loss function.

- **Step 2:** The residual along the gradient direction is denoted by:

$$\hat{y}_i = \left\{ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right\}_{f(x)=f_{n-1}(x)}, \tag{3}$$

where n indicates the number of iterations and $n = 1, 2, \ldots, N$.

- **Step 3:** The initial model $T(x_i; \alpha_n)$ is obtained by fitting the sample data, and the parameter α_n is calculated based on the least squares method (4).

$$\alpha_n = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{N} (\hat{y}_i - \beta T(x_i; \alpha_n))^2 \tag{4}$$

- **Step 4:** By minimizing the loss function, the weight of the current model is expressed as:

$$\gamma_n = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, F_{n-1}(x) + \gamma T(x_i; \alpha_n)) \tag{5}$$

- **Step 5:** The model is updated by:

$$F_n(x) = F_{n-1}(x) + \gamma_n T(x_i; \alpha_n) \tag{6}$$

This loop is performed until the specified number of iterations or the convergence conditions are met. The GBDT has two key advantages. Firstly, it effectively captures complex nonlinear interactions between variables and the response without requiring a direct physical model [50], [51]. Additionally, GBDT demonstrates minimal overfitting issues, resulting in superior performance during the training phase compared to the test phase [26], [51], [52], [53].

- Light Gradient Boosting Machine (LGBM) belongs to the gradient boosting framework and is specifically optimized for large datasets and high-dimensional feature spaces. LGBM utilizes a tree-based ensemble approach, constructing an ensemble of decision trees sequentially, with each tree correcting the errors of its predecessors [54]. What sets LGBM apart is its ability to handle categorical features efficiently and its use of a histogram-based learning process, which speeds up training by discretizing continuous features. This algorithm is known for its fast training speed, reduced memory usage, and excellent predictive performance [55]. The pseudocode used for LGBM regression is represented in Algorithm 2.

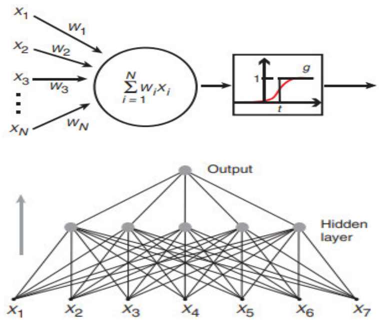


FIGURE 4. **Artificial neural networks.**

Algorithm 2 Pseudocode for LGBM Regression

Input: Data: EMS calls

1: Import packages and dataset (features and target).
2: Describe the features) and the target.
3: Train the model: Define a LGBMRegressor() function and fit the model using x_{train} and y_{train}.
4: Test the model with the testing set.

Output: R^2, MAE, MSE, RMSE, Execution Time.

- Artificial neural networks (ANN) are inspired by early models of sensory processing in neurons and brains [56], [57]. These networks can be simulated on a computer, replicating the behavior of model neurons. Through algorithms mimicking real neuron processes, ANNs can learn to solve various problems, including EMS call forecasting [13], [26], [58], [59]. The model neuron, known as a threshold unit, receives inputs from other units or external sources, weighs each input, and sums them up.
As shown in (7), the goal is to approximate some function f_i of the input $x = (x_1, \ldots, x_d)$ weighted by a vector of connection weights $w_i = (w_{i,1}, \ldots, w_{i,d})$ completed by a neuron bias b_i and associated with an activation function ϕ, which is denoted by

$$y_i = f_i(x) = \phi_i(\langle w_i, x \rangle + b_i). \tag{7}$$

The predicted output is compared with the actual output to compute the error/loss in each observation. The loss function L is the sum of differences between the observed output $\hat{y}$ and the true output y. The goal is to reduce the error values as much as possible to approximate the function in (7). To achieve this, the backpropagation algorithm is used. It compares the desired output of the neural network with the network's output, computes errors, and adjusts weights and biases to get closer to the desired output after each iteration. The weights w and biases b are updated through backpropagation during the training process, using the gradient descent algorithm to solve the optimization problem [60]. The weight and bias update between layers k and $k+1$ is performed using gradient descent, which is denoted by:

$$\begin{cases} w^{k+1} = w^k - \eta \left(\dfrac{\partial L}{\partial w^k} \right) \\ b^{k+1} = b^k - \eta \left(\dfrac{\partial L}{\partial b^k} \right), \end{cases} \tag{8}$$

where the learning rate $\eta > 0$ controls the step size towards convergence in gradient descent. A small η value ensures a careful convergence, while a high value may lead to divergence. Thus, η influences the convergence of gradient descent towards the local minimum. In this paper, we use a particular type of ANN with more than three layers called MLP (Multi-Layer Perceptron). MLP is a fundamental architecture and one of the earliest and most widely used neural network models. MLP regressor trains using backpropagation with no activation function in the output layer.

- Decision Tree (DT): A decision tree is a simple tree-like model used for both classification and regression tasks. It splits the data into different segments based on the input features, creating a hierarchical structure of decision nodes. Each leaf node represents a specific class or a numerical value for regression.
- Random Forest (RF): Random Forest is an ensemble learning method based on decision trees. It builds multiple decision trees during training and combines their predictions through voting (for classification) or averaging (for regression). Each tree in the forest is trained on a random subset of the data and features, making the model more robust and reducing overfitting [30].
- Long-Short-Term Memory (LSTM) is a kind of Recurrent Neural Network (RNN) with the ability to remember values from earlier stages for future use. An RNN is a special case of a neural network where the objective is to predict the next step in the sequence of observations concerning the previous steps observed in the sequence [26], [61].

2) ONLINE FORECASTING METHODS

Online machine learning is a type of machine learning in which data are acquired sequentially and used to update the best predictor of future data at each step. We used the online versions of two offline forecasting methods to make the models dynamic and get the most out of the available data. We adapted the online forecasting algorithm from [26]. But instead of using their hybrid approach based on using offline for training and online learning on validation /test, we used the online learning for the whole process of training and testing. Online machine learning for EMS call forecasting is described in the pseudocode Algorithm 3. The difference between online machine learning and more traditional batch machine learning is that an online model is dynamic and learns on the fly. Online learning solves a lot of pain points in real-world environments, mostly because it does not require retraining models from scratch every time new data arrives, and will be more useful in eHealth [11].

Algorithm 3 Online Prediction Algorithm for EMS Calls

Input: offline model and data calls
Initialization: *predictions* $\leftarrow$ ()
F: a number representing the frequency of data
to receive data and split in inputs and targets
$x \leftarrow inputs[t : t + F]$
$y \leftarrow targets[t : t + F]$
$z \leftarrow model.predict(x)$
predictions.append(z)
update the model (model.train(x,y))
Output: predictions

F. EVALUATION METRICS

Once the model is trained, it is important to evaluate its performance on a separate test set of data to ensure that it can accurately predict EMS calls in the region. As used differently in [3], [4], [23], [25], [26], [27], and [28], we used five different metrics to evaluate ML models, including [62]: the $R^2 - SCORE$, the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean absolute error (MAE), and the Mean Absolute Percentage Error (MAPE). Considering m as the number of observations, y_i as the observed value, $\hat{y}_i$ as the predicted value and $\bar{y}_i$ the scores of all outputs are averaged with uniform weight:

- $R^2 - SCORE$: represents the proportion of the variance of the dependent variable that is predictable from the independent variable(s). It indicates how well the model's predictions approximate the real data points, with a value closer to 1 indicating better performance.

$$R^2 - SCORE = 1 - \frac{\sum_{i=1}^{m} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{m} (y_i - \bar{y})^2} \tag{9}$$

- Mean Square Error (MSE): represents the average error between the observed values and the predicted values. MSE emphasizes larger errors due to squaring, making it useful for identifying models that make significant errors.

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

- Root Mean Square Error (RMSE): represents the square of the MSE. RMSE can show a more accurate error rate by squared MSE metric. RMSE provides a more interpretable measure of error by bringing the units back to the original scale, and it penalizes larger errors more than smaller ones.

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2} \tag{11}$$

- Mean Absolute Error (MAE): measures the average magnitude of errors in a set of forecasts, regardless of their direction. Lower values of MAE indicate better model performance.

$$MAE(y,\hat{y}) = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i| \tag{12}$$

- Mean Absolute Percentage Error (MAPE): is the average error over time as a percentage of the actual values. MAPE measures the percentage error of the forecast about the actual values. For example, a MAPE value of $p\%$ means that the average difference between the forecasted value and the actual value is $p\%$. The lowest MAPE indicates the best performance.

$$MAPE = \sum_{i=1}^{m}|\frac{\hat{y}_i - y_i}{y_i}| \tag{13}$$

Based on the evaluation results, the model can be refined by adjusting its parameters or by adding new features to improve its accuracy. Based on these evaluation metrics, we can decide which algorithm performs better on EMS call prediction. Once the model has been refined and its accuracy has been validated, it can be deployed to predict the occurrence of EMS calls in a certain zone at a certain time. This can involve integrating the model into a larger EMS dispatch system to provide more accurate and timely response to emergency calls.

IV. RESULTS AND EVALUATION

This section presents the evaluation results obtained from the offline and online machine learning algorithms and the proposed ensemble-based method for EMS call prediction. We discuss the impact of the number of clusters on RMSE, the evaluation of the machine learning algorithms in terms of scores, and their forecasting errors. Then, we compare different our stacking approach with the voting one.

A. DATA COLLECTION AND PREPROCESSING

For our experiments, we collected real historical emergency data from [63] and weather data per hour from [64]. The first dataset comprises EMS call records from the 911 EMS number in Montgomery Territory, Pennsylvania (USA). Montgomery spans an area of 1,260 km^2 with an approximate population density of 685 persons per km^2. The data covers the period from December 2015 to July 2020 and includes three types of emergencies: fire, traffic, and illness. The dataset contains 663,522 rows and 10 columns, representing various characteristics, such as latitude, longitude, emergency description, title, date and time of call, municipality (twp: township), address, index, and call type.

The frequency of EMS calls is also influenced by weather conditions [4], [23], [26], [27], [28]. Inspired by these previous works, we incorporated the weather dataset into the historical emergency dataset by adding meteorological parameters to each entry. We collect weather data through the weather API, made available by Meteomatics online shop [64]. These weather-based features serve as additional inputs for our models. In total, we include 10 weather-based features, which are concatenated with the existing spatial and temporal features. These extra features contain parameters, such as the amount of precipitation, dew point temperature, fresh snow, relative humidity, surface pressure, temperature (mean, min, or max over the selected time interval), total cloud cover, visibility, wind speed, and effective cloud cover.

We preprocess the whole data by handling missing values and unnecessary columns through imputation. We encode categorical variables and standardize numerical features. Moreover, we apply several decompositions and transformations to make the data more suitable for analysis, leading to the creation of new features. Based on date time we extracted season, month, day of the week, day, weekend, and part of the day. Using the geo-grid division in Section III-C, we obtained lat_grid and lng_grid and aggregated our target accordingly. For each occurrence in our final dataset, we are limited to 18 spatial, temporal and climatological features, consisting of 8 categorical variables, and 10 continuous variables. Each record in the final dataset includes information about a call for an ambulance, as illustrated in Table 2. This table presents an overview of the three distinct feature sets used in the paper: spatial, temporal, and weather. The feature names and their corresponding descriptions are organized clearly and concisely, facilitating an easy understanding of the data characteristics used for the forecasting model. The summary of the final dataset after preprocessing is described in Table 3, with 16 0242 rows. Moreover, for EMS call forecasting, we use the number of calls as the target and the outer parameters described above as features. Then, we split the data into 80% for training and 20% test.

B. IMPLEMENTATION TOOLS AND DETAILS

Python 3.8.10 [65] was used in making both the baseline models and our proposed stack ensemble model. We chose the Jupyter Notebook environment because its interactive and user-friendly features made iterative development and debugging much easier [66]. In the implementation process, we used various libraries, Mlens [67] is what we relied on for ensemble machine learning methods, while Scikit-learn [68] served us well with single machine learning functions and clustering methods. To design online ML algorithms, we use one extension library of Python called River [69]. As far as construction, training LSTM networks and TensorFlow were involved. Regarding data manipulation and analysis, we used Pandas, whereas Numpy handles numerical computations. The implementation takes place on the Google Colab [70] platform, while the GPU has

TABLE 2. Feature description of EMS call final dataset.

Feature Sets	Feature Names	Feature Description	Type
Spatial	Latitude	[39.00, 41.00]	Continuous
	Longitude	[-77, -74]	Continuous
	Lat_{grid}	Integer values representing the split of the latitude.	Categorical
	Lng_{grid}	Integer values representing the split of the longitude.	Categorical
Temporal	Season	Integer values (1 to 4) corresponding to the seasons (winter, spring, summer, autumn).	Categorical
	Month	Integer values (1 to 12) representing all months of the year (January to December).	Categorical
	Day of the week	Integer values (1 to 7) representing all days of the week (Monday to Sunday).	Categorical
	Day	Integer values (1 to 30 or 1 to 31) indicating the day of the month.	Categorical
	Weekend	Binary values (0 or 1), where 1 indicates a weekend and 0 otherwise.	Categorical
	Part of the day (day_part)	Three parts of the day: "8 am - 4 pm," "4 pm - 12 am," "12 am - 8 am."	Categorical
Weather	Amount of precipitation	Precipitation measured in millimeters per hour.	Continuous
	Dew point temperature	Dew point temperature at a height of 2 meters in Celsius.	Continuous
	Fresh snow	Fresh snowfall measured in centimeters per hour.	Continuous
	Relative humidity	Relative humidity at a level of 2 meters in percentage.	Continuous
	Surface pressure	Surface pressure measured in pascals.	Continuous
	Temperature	Mean, minimum, or maximum temperature at a height of 2 meters.	Continuous
	Total cloud cover	Cloud cover measurement in octas.	Continuous
	Visibility	Visibility measurement in feet.	Continuous
	Wind speed	Wind speed at a height of 10 meters in kilometers per hour.	Continuous
	Effective cloud cover	Effective cloud cover measurement in octas.	Continuous

TABLE 3. Summary statistics of input dataset.

	count	mean	std	min	25%	50%	75%	max
month	160242.0	6.46	3.76	1.0	3.0	7.0	10.0	12.00
week	160242.0	26.50	16.55	1.0	11.0	27.0	41.0	53.00
dayofweek	160242.0	2.89	1.95	0.0	1.0	3.0	5.0	6.00
day	160242.0	16.00	8.67	1.0	9.0	16.0	23.0	31.00
day_part	160242.0	1.21	0.71	0.0	1.0	1.0	2.0	2.00
lat_grid	160242.0	1.44	0.83	0.0	1.0	1.0	2.0	4.00
lng_grid	160242.0	4.25	1.12	0.0	4.0	4.0	5.0	6.00
category	160242.0	0.86	0.91	0.0	0.0	1.0	2.0	2.00
precip_1h:mm	160242.0	0.18	0.80	0.0	0.0	0.0	0.0	21.06
dew_point_2m:C	160242.0	5.69	10.28	-23.4	-1.9	5.6	14.6	25.70
fresh_snow_1h:cm	160242.0	0.01	0.10	0.0	0.0	0.0	0.0	3.30
relative_humidity_2m:p	160242.0	67.02	20.83	18.0	50.1	65.0	86.4	100.00
sfc_pressure:Pa	160242.0	100655.62	760.13	98019.0	100205.0	100691.0	101137.0	103131.00
t_min_2m_1h:C	160242.0	11.84	10.43	-15.3	3.4	11.4	20.6	34.20
total_cloud_cover:octas	160242.0	4.76	3.16	0.0	2.0	6.0	8.0	8.00
visibility:ft	160242.0	94271.56	27758.64	299.5	86993.8	104424.8	112049.9	120257.70
wind_speed_10m:kmh	160242.0	7.53	6.45	0.0	1.7	6.7	11.7	36.80
effective_cloud_cover:octas	160242.0	4.02	2.99	0.0	1.0	4.0	7.0	8.00
count	160242.0	18.30	13.71	1.0	8.0	15.0	25.0	97.00

been used as a hardware accelerator to improve processing. To implement all the methods, we did a search grid between different configurations presented in Table 6 and Table 5.

C. IMPACT OF THE NUMBER OF CLUSTERS RMSE

This section investigates the effect of increasing the number of clusters on the prediction error of machine learning algorithms. We consider different numbers of cluster labels: 2*3, 3*4, 4*5, 5*7, 7*10, 10*15 and 20*25 (Figure 5). The results demonstrate that as the number of clusters increases, the average error decreases. Furthermore, we observe that the decrease in average error for the DBSCAN and K-means algorithms is slight compared to the other algorithms, showing their weaknesses for spatial clustering of the ambulance call. However, with the geo-grid algorithm, increasing the number of clusters drops the average error dramatically; it starts at 8.005 when the number of classes is 2*3, then ends at 1.596 when the number of classes is 20*25. This indicates that geo-grid, which depends on the rectangular

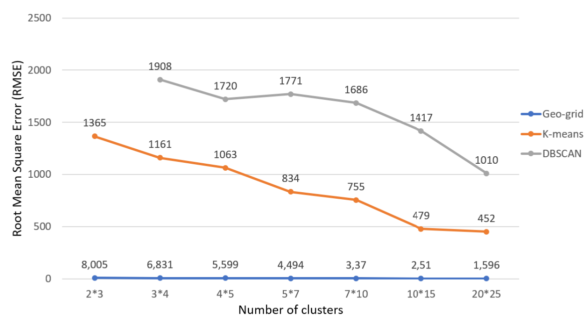


FIGURE 5. RMSE vs Numbers of zones (clusters).

TABLE 4. Spatial clustering Methods and parameters settings.

Method	Hyperparameters
Geo-grid division	2×3, 3×4, 4×5, 5×7, 7×10, 10×15, 20×25
K-means	Number of clusters = 2×3, 3×4, 4×5, 5×7, 7×10, 10×15, 20×25
DBSCAN	epsilon = 0.1, min_samples = 800

TABLE 5. Feature Selection Methods and parameters settings.

FS Method	Hyperparameters
RFFI	n_estimators = 100, random_state = 42
SHAP	No hyperparameters
Boruta	n_estimators = 'auto', verbose = 0, max_depth = 5
LASSO	alpha = 0.1

zones to separate data, is highly impacted by increasing the number of clusters. Based on these observations and the recommendation in [35], we conclude that the geo-grid division yields better performance in terms of average error compared to other algorithms when varying the number of clusters. For the remaining implementation, we assume the around the middle value of 5*7 clusters, which is close to half of 68 counties present in Montgomery County. The number of clusters has been more investigated for ambulance positioning in [71]. The clusters represent the station area for the coverage of ambulances as presented in [72]. Figure 6 provides a visualization of each division strategy.

D. FEATURE SELECTION RESULT

In this section, we present the results of the investigation of feature importance and selection. We analyzed the relevance of spatial, temporal and climatological features on the occurrence of EMS call. We considered the whole dataset and used four recent and well-known feature selection techniques: RFFI [43], LASSO [42], SHAP [28] and Boruta [3]. The results, presented in Figure 7, reveal that all variables show importance in predicting the target when using RFFI and SHAP. The RFFI values provide an indication of the relative importance of each feature in the dataset. Features such as "lat_grid", "lng_grid", and "category" appear to have relatively higher importance compared to others, while features like "fresh_snow_1h:cm" and "precip_1h:mm" have lower importance. SHAP values provide insights into the impact of each feature on model predictions. Features like "dayofweek", "day_part", and "effective_cloud_cover:octas" exhibit relatively higher SHAP values, indicating their strong influence on predictions. These values align with the RFFI analyses, highlighting the importance of all the collected features for EMS call forecasting.

However, LASSO excludes two features, namely 'fresh_snow_1h:cm' and 'total_cloud_cover:octas'. Meanwhile, Boruta recommends excluding only the feature

TABLE 6. The hyperparameter values of the ML methods for grid-search.

Methods	Hyperparameters/Configuration
LR	solver = lbfgs max_iter = 2000
GBRT, LGBM, DT, RF	n_estimators = [50, 100, 200, 300, 400, 500, 600], learning_rate = [0.2, 0.1, 0.01, 0.001, 0.0001], max_depths = [1, 2, 3, 4, 5, 10, 15, 20, 25, 30] max_depth=5, random_state=0, loss='absolute_error', warm_start = True
ANN/MLP	max iter = 200, 500, 1000 activation = relu, tanh random_state=1
LSTM	Dropout (Layer 3) Rate: 0.25 LSTM (Layer 4) Units: 50 Dropout (Layer 4) Rate: 0.25 Dense (Output) Units: 1 Compilation Optimizer: Adam, Loss: MSE Training Epochs: 50, Batch Size: 32
Online-LR	optim.SGD, ('lr': [.1, .01, .005]), optim.Adam, ('beta_1': [.01, .001], 'lr': [0.1, .01, .001]), optim.Adam, ('beta_1': [0.1], 'lr': [.001]),
Online-ANN	05 layers Activations (Sigmoid, ReLU, ReLU, Identity), optimizer=optim.SGD(1e-3), seed=42

category. These outcomes from the three methods collectively suggest that incorporating spatial, temporal, and climatological features is crucial for accurately forecasting EMS calls. LASSO values indicate the coefficients assigned to each feature by the LASSO regression model. Some features have large positive or negative coefficients, indicating their significant impact on the target variable. Features such as "dew_point_2m:C" and "sfc_pressure:Pa" have notably large coefficients, suggesting their importance in the model. Boruta feature selection technique identifies features that are deemed important for model prediction. All features are labeled as "True" by Boruta, suggesting that none of the features are redundant or can be safely removed from the dataset.

In conclusion, the analysis using different feature selection techniques provides complementary insights into the importance and impact of each feature on model predictions. While certain features consistently emerge as important across multiple techniques, the results also highlight the nuanced nature of feature importance and the need for considering multiple perspectives when selecting features for predictive modeling. This comprehensive approach enhances the understanding of the dataset and aids in building more robust and accurate predictive models. For the rest of the implementation, we consider three different scenarios of feature selection: the inclusion of all features as suggested by the output of RFFI and SHAP method, and the consideration of 17 features obtained using Boruta and 16 features resulting using LASSO. The findings emphasize the significance of integrating various types of features to achieve improved EMS call predictions.

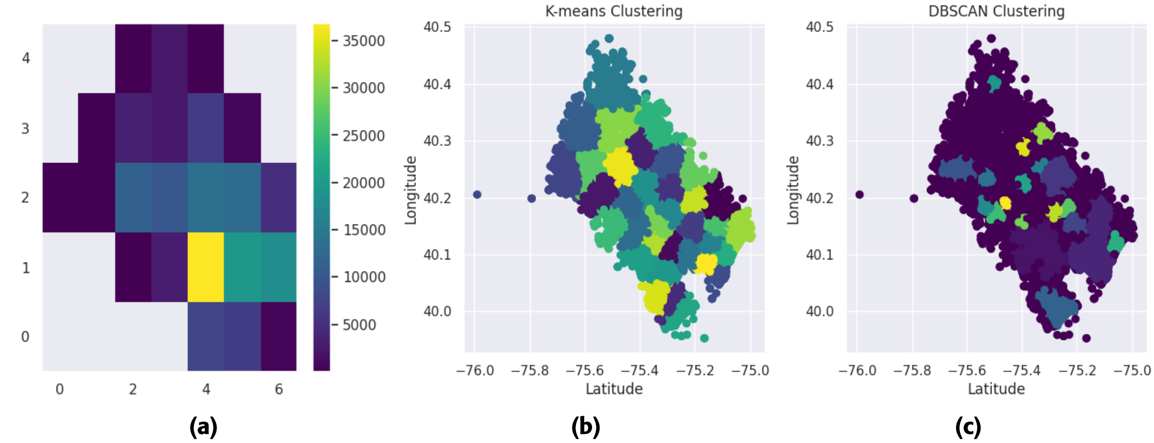


FIGURE 6. Spatial division with 03 different methods (a) with Geo-grid division. (b) with K-means. (c) with DBSCAN.

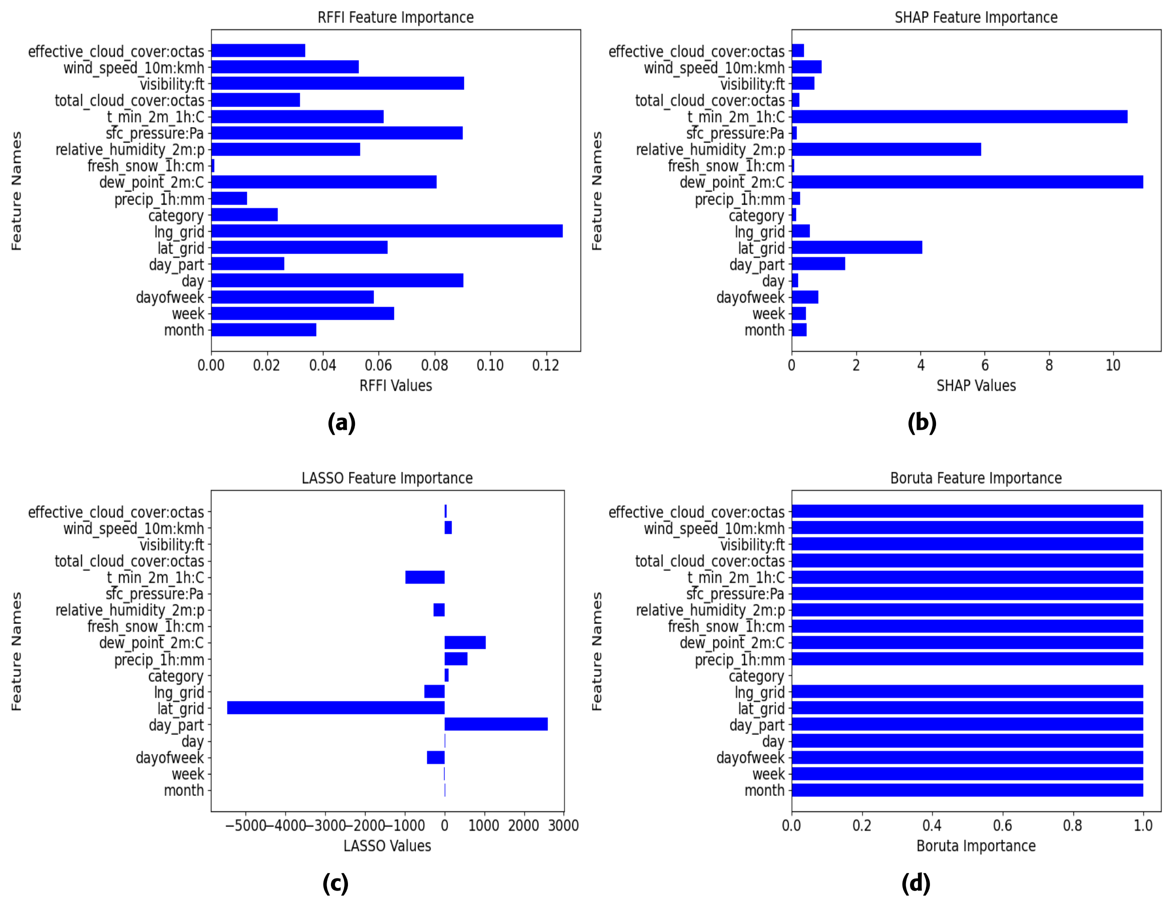


FIGURE 7. Feature importance result. (a) with RFFI. (b) with SHAP. (c) with LASSO (d) Boruta.

E. EVALUATION OF OUR PROPOSED MODEL

We compare the performance of offline and online machine learning algorithms based on the score and four different errors metrics with the aim of EMS calls forecasting. Table 7, Table 8 and Table 9 present performance analysis of various forecasting models, including Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), Light Gradient Boosting Machine (LGBM), Gradient Boosting Regression Trees (GBRT), Random Forest (RF), Decision Trees (DT), Online Linear Regression (Online-LR), Online Artificial Neural Networks (Online-ANN), and the proposed model. We conduct the implementation based on the three scenarios from the feature selection process. Each model is evaluated based on various performance metrics, including R-squared (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

1) EVALUATION BASED ON ALL FEATURES (RFFI AND SHAP)

Table 7 summarizes the performance of various baseline forecasting models equipped with spatial, temporal, and climatological features, evaluated using different metrics. The models assessed include offline models (i.e., GBRT, LGBM, LSTM, RF, DT, and ANN) [14], [25], [26], [28], [29], [30] and online learning models (i.e., Online - LR and Online - ANN) [11], [26].

TABLE 7. Performance of the baseline forecasting models with all spatial, temporal and climatological features (RFFI and Shap).

Models	R^2	MSE	RMSE	MAE	MAPE
ANN [3], [27]	0.1935	156.5079	12.5103	8.9668	0.8586
LSTM [26]	0.8681	25.580	5.0577	3.7538	0.3531
LGBM [25]	0.8874	21.834	4.6727	3.5481	0.3275
GBRT	0.8359	31.841	5.6428	3.8395	0.3093
RF	0.9933	1.2928	1.1370	0.5429	0.1103
DT	0.9883	2.2658	1.5052	0.3618	0.0884
Online-LR	0.1584	156.91	9.3519	12.5263	1.0728
Online-ANN	0.6383	67.435	8.2118	5.9029	0.4005
Proposed	0.9950	0.9705	0.9851	0.3151	0.0775

Among the models, LSTM and LGBM demonstrate notable performance with high R-squared values of 0.8681 and 0.8874, respectively. These models also exhibit relatively low MSE, RMSE, MAE, and MAPE, indicating their effectiveness in capturing the underlying patterns in the data. In contrast, the performance of some other models such as ANN and Online-LR appears to be comparatively poorer, with lower R-squared values and higher error metrics. Online models struggle to accurately predict the target variable based on all features. Interestingly, it appears that tree-based models, such as RF and DT, outperform in capturing the patterns and explaining the variance in the EMS call data, resulting in significantly lower errors and higher R-squared values. Overall, the proposed model stands out with an impressive R-squared value of 0.9950 and minimal error metrics, including a remarkably low MSE of 0.9705. These results highlight the efficacy of our proposed model in forecasting ambulance demand, showcasing its potential for tactical decisions in EMS management.

2) EVALUATION BASED ON 17 FEARURES (BORUTA)

Table 8 presents a comprehensive performance analysis of the different models using the Boruta feature selection technique. Compared with Table 7. We can easily observe that Boruta feature selection results improve the result of EMS forecasting using LGBM, GBRT, RF and DT. The results confirm the effectiveness of the choice of these models as base learners in layer 1 of our proposed model. These models also demonstrate minimal error metrics, with low MSE, RMSE, MAE, and MAPE, indicating their effectiveness in accurately predicting the number of EMS calls at a time, in a specific zone and based on weather.

TABLE 8. Performance analysis of the ML forecasting models with Boruta.

Methods	R^2	MSE	RMSE	MAE	MAPE
ANN [27]	0.1598	163.04	12.7688	9.2380	0.9857
LSTM [26]	0.8562	27.90	5.282	3.9029	0.3722
LGBM [25]	0.8897	21.3976	4.6257	3.5227	0.3282
GBRT	0.8346	32.0869	5.6645	3.8211	0.3084
RF	0.9938	1.1955	1.0934	0.4999	0.1048
DT	0.9890	2.1310	1.4598	0.3331	0.0841
Online-LR	0.1592	156.75	9.3466	12.5202	1.0721
Online-ANN	0.6466	65.8797	5.9745	8.1166	0.5398
Proposed	0.9954	0.8938	0.9454	0.2923	0.0724

Similarly, the proposed model stands out as a top-performing model, with an exceptional R-squared value of 0.9954 and minimal error metrics. The proposed model's superior performance underscores its capability to capture the underlying patterns in the spatial, temporal and weather data and make accurate EMS call forecasts, highlighting its potential for practical applications in forecasting ambulance demand. On the other hand, models such as ANN and Online-LR exhibit relatively poorer performance, with lower R-squared values and higher error metrics. These models struggle to effectively capture the complex relationships within the data and make accurate predictions.

3) EVALUATION BASED ON 16 FEARURES (LASSO)

With the aim to analyze the behavior of the base model in scale and the impact of feature selection on EMS call forecasting, Table 9 expands the analysis to the exclusion of two features, proposed by LASSO. The same evaluation metrics are provided to compare the models' performance. We can observe that, except GBRT, all the performances of other models are significantly decreasing. This table reaffirms that our ensemble model maintains its superior predictive score, even with the exclusion of two climatological features: 'fresh_snow_1h:cm' and 'total_cloud_cover:octas'. However, the performance of the ANN model significantly deteriorates, suggesting that the exclusion of the two climatological features may not be suitable for the EMS call model.

TABLE 9. Performance analysis of the ML forecasting models with Lasso.

Methods	R^2	MSE	RMSE	MAE	MAPE
ANN [27]	0.1697	161.122	12.6934	9.8007	1.1501
LSTM [26]	0.8484	29.4016	5.4223	4,000	0.3767
LGBM [25]	0.8890	21.5404	4.6411	3.5326	0.3287
GBRT	0.8349	32.0372	5.6601	3.8296	0.3085
RF	0.9936	1.2304	1.1092	0.5040	0.1052
DT	0.9885	2.2179	1.4892	0.3394	0.0861
Online-LR	0.1575	157.08	9.3589	12.5335	1.0737
Online-ANN	0.6347	68.103	5.9474	8.2524	0.6834
Proposed	0.9949	0.9823	0.9911	0.3199	0.0786

Therefore, Table 7, Table 8 and Table 9 underscore the importance of selecting appropriate features and models to achieve accurate and efficient forecasting results. The result underscores our model's ability to effectively capture underlying data patterns and produce highly accurate forecasts, positioning it as a promising solution for real-world EMS call forecasting applications and ambulance dispatch and routing. Overall, understanding the trade-offs between score, errors and computational time is critical in selecting the most suitable model for the given task.

F. COMPARATIVE ANALYSIS

1) OUR PROPOSED STACKING VS BAGGING AND VOTING

For further investigations, we compared our proposed stacking ensemble machine learning with voting and bagging strategies. Voting consists of taking only the best model in layer 1 to pursue the second level of prediction, while bagging is using the best model n times as base learners. Table 10 and Table 11 present a comparative analysis of our proposed stacking ensemble model with voting and bagging ensemble model between the best base models: GBRT, LGBM, DR, and RF. We consider the best scenario of feature selection obtained with Boruta: 17 spatial, temporal, and climatological. The performance metrics are the same.

TABLE 10. Performance analysis of 5 ML forecasting models with Boruta.

Methods	R^2	MSE	RMSE	MAE	MAPE
Bagging	0.9890	2.1253	1.4578	0.3318	0.0838
Voting	0.8897	21.3976	4.6257	3.5227	0.3282
Proposed	0.9954	0.8938	0.9454	0.2923	0.0724

TABLE 11. Performance analysis of 2 ML forecasting models with Boruta.

Methods	R^2	MSE	RMSE	MAE	MAPE
Bagging	0.9890	2.1253	1.45785	0.3318	0.0838
Voting	0.8897	21.3976	4.6257	3.5227	0.3282
Proposed	0.9954	0.8938	0.9454	0.2923	0.0724

Bagging and Voting achieved respectively scores of 0.9890 and 0.8897 using the same base learners. Furthermore, the proposed stacking maintains its superiority, achieving an outstanding R-squared value of 0.9954 and minimal error metrics. By combining the strengths of the different best models in the base layer and using the RF model in the meta layer, we proposed an effective and accurate predictive model.

2) OUR PROPOSED MODEL VS RELATED WORKS

The comparison presented in Table 12 offers valuable insights into various EMS call forecasting approaches and their respective performance metrics. For instance, Chen et al. [4] employed data analysis techniques and ANN to forecast ambulance demand, achieving notable results with a R^2 value of - and a RMSE of 0.26. Similarly, Lin et al. [25] utilized LGBM alongside other methods, obtaining a R^2 score of - and an RMSE of 10.2. Van et al. [27] leveraged data analysis and a genetic algorithm-MLP (GA-MLP) approach, yielding an R^2 of - and an RMSE of 21.68. The CNN architecture in [28] outperforms MLP, Medic, RF, and DT by 9.83%, 9.98%, 11.26%, and 14.84%, correspondingly. In comparison, the proposed method, employing feature selection techniques such as Shap and Boruta in conjunction with a stacking ensemble model, demonstrated superior forecasting performance, achieving an R^2 value of 0.9954 and an RMSE of 0.8938 for the Boruta-based approach. These findings highlight the effectiveness of our proposed EMS call forecasting model in accurately predicting ambulance demand, showcasing its potential to enhance emergency response systems and optimize resource allocation strategies.

3) COMPUTATIONAL TIME

In addition to showcasing the performance metrics of various EMS call forecasting approaches, it's crucial to underscore the significance of employing a stacking ensemble model, as demonstrated in our proposed method. Stacking leverages the collective contribution of multiple machine learning models, each with its unique strengths and weaknesses, to deliver more robust and accurate predictions. By comparing diverse models such as GBRT, LGBM, RF, DT, MLP, and LSTM, our approach harnesses the complementary capabilities of these algorithms. This amalgamation not only enhances prediction accuracy but also improves the model's resilience to uncertainties and variations in EMS call data. Thus, the adoption of a stacking ensemble model represents a strategic approach to EMS call forecasting, enabling more reliable and effective decision-making in emergency response operations.

Figure 8 presents the computational time of all the models and our proposed model. From the given data, we observe a wide variation in the time taken by different models to complete the task. The fastest models are LGBM and DT, with execution times of 1.5797 seconds and 1.6916 seconds, respectively. On the other hand, ANN takes the longest time to execute, with a time of 4934.43 seconds, followed by LSTM with 1771.80 seconds. The disparity in execution times could be attributed to various factors, including the complexity of the model, the volume of data processed, and the computational resources available. For instance, models

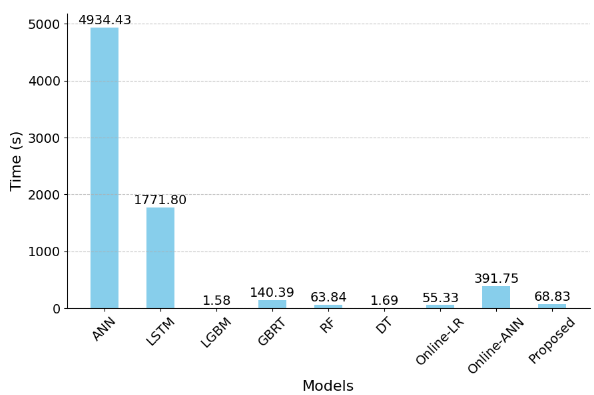


FIGURE 8. Computational Time of different models using Boruta.

TABLE 12. Comparison of Ambulance demand forecasting approaches.

Author	Feature selection	Methods	Comparison	R^2	MSE	RMSE	MAE	MAPE
Chen et al. [4]	Data Analysis	ANN	SVR, SR, MA	-	-	0.26	-	51.92
Lin et al. [25]	-	LGBM	RMA, LR, SVR, MLP	-	10.2	-	2.09	-
Van et al. [27]	Data Analysis	GA-MLP	SMA, MEDIC, MLP, NF	-	21.68	-	3.64	-
Martin et al. [3]	Boruta	MLP	ARIMA, HW, MEDIC	-	-	-	-	35.56
Rausten et al. [28]	Shap	CNN	MLPs, DTs, RF, MEDIC	14.66	-	-	-	-
Proposed	Shap	Stacking	GBRT, LGBM, RF, DT, MLP, LSTM	0.9950	0.9705	0.9851	0.3151	0.0775
Proposed	Boruta	Stacking	GBRT, LGBM, RF, DT, MLP, LSTM	0.9954	0.8938	0.9454	0.2923	0.0724

like LGBM and DT, being ensemble methods based on decision trees, tend to have faster execution times compared to more complex models like ANN and LSTM, which involve iterative optimization processes and handling sequential data. With 68.83 seconds, our proposed ensemble model gives a good trade-off between score, error minimization and execution time than the traditional and individual machine learning methods.

G. LIMITATIONS

Although our proposed stacking ensemble machine learning model demonstrates promising results, it is important to mention its limitations to ensure a well-rounded interpretation of our findings. Firstly, the effectiveness of our model is influenced by the quality and diversity of the features in the dataset. While we employed a comprehensive dataset and feature selection, having access to larger and more varied datasets could confirm the model's ability to generalize. Secondly, although our model exhibits the highest score and the minimum errors, there remain opportunities for enhancement, especially in scenarios involving real-time management. Further optimization of the model could reduce its computational time in real-world environments. Lastly, the deployment of this work could be limited by the considerations given to data privacy used in emergency intervention, especially when handling sensitive information related to emergency situations and patient details.

Despite these constraints, our stacking ensemble model for EMS call forecasting represents a notable advancement in the field, providing valuable insights for optimized ambulance dispatch and routing [72]. An effective forecasting model should accurately predict the number of incidents (volume) based on time, location (distribution) and weather. Our accurate results improve ambulance response times, enhance patient outcomes, and maximize the efficiency of emergency medical services. Additionally, this is practical because it helps decide how many staff members are needed for a shift and where resources should be located to respond quickly. The forecasting capability presented here is important because it enables informed resource and ambulance demand and is applicable across hospitals and general medical facilities.

V. CONCLUSION

In this paper, we proposed a stacking ensemble model for EMS call forecasting to facilitate the ambulance dispatch and routing planning. A precise distribution forecast plays a crucial role in strategically positioning ambulances to minimize response times. This paper provides a comprehensive literature review, emphasizing the importance of understanding and addressing the challenges posed by EMS call forecasting for ambulance dispatching and routing. We compared the most common ML models used for EMS call forecasting, including GBRT, LGBM, ANN, RF, DT, and LSTM. We used the best ones as base learners in our proposed model. In addition, we used the voting ensemble model and the online ML, an approach that embraces change and adaptability for comparative study with our proposed model.We used a real-dataset to evaluate the effectiveness of the models. During the evaluation process, we consider different metrics: the R^2-score, the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE). Moreover, we ensure the accessibility and reproducibility or our research findings since we used real and public dataset. The findings affirmed the effectiveness of our proposed stacking ensemble model to enhance the overall performance in all metrics errors and good score by including GBRT, LGBM, DT and RF. In addition, Our stacking model outperforms all the traditional single, online, bagging and voting ensemble in the three different scenario of feature selection. Our proposed model is the most accurate to predict ambulance demand in different areas and times, allowing for proactive deployment of resources. For future work, We will explore the impact of others features like demographics, sociologics and special events in the occurrence of EMS call. Integrating additional data type could contribute to the refinement of our model. Future research can access the scalability of modeling handling larger datasets and meeting growing computational demands in eHealth.We plan to evaluate the obtained results for ambulance allocation and routing in smart cities, aiming to enhance emergency response systems.

ACKNOWLEDGMENT

The authors would like to thank Dr. Franjieh El Khoury for her valuable comments and proofreading this article.

REFERENCES

[1] R. Aringhieri, M. E. Bruni, S. Khodaparasti, and J. T. van Essen, "Emergency medical services and beyond: Addressing new challenges through a wide literature review," *Comput. Operations Res.*, vol. 78, pp. 349–368, Feb. 2017.

[2] Z. Hao, Y. Wang, and X. Yang, "Every second counts: A comprehensive review of route optimization and priority control for urban emergency vehicles," *Sustainability*, vol. 16, no. 7, p. 2917, Mar. 2024.

[3] R. J. Martin, R. Mousavi, and C. Saydam, "Predicting emergency medical service call demand: A modern spatiotemporal machine learning approach," *Operations Res. Health Care*, vol. 28, Mar. 2021, Art. no. 100285.

[4] A. Y. Chen, T.-Y. Lu, M. H. Ma, and W.-Z. Sun, "Demand forecast using data analytics for the preallocation of ambulances," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 4, pp. 1178–1187, Jul. 2016.

[5] I. Franch-Pardo, B. M. Napoletano, F. Rosete-Verges, and L. Billa, "Spatial analysis and GIS in the study of COVID-19. A review," *Sci. Total Environ.*, vol. 739, Oct. 2020, Art. no. 140033.

[6] A. Azimi, N. Bagheri, S. M. Mostafavi, M. A. Furst, S. Hashtarkhani, F. H. Amin, S. Eslami, F. Kiani, R. VafaeiNezhad, T. Akbari, A. Golabpour, and B. Kiani, "Spatial-time analysis of cardiovascular emergency medical requests: Enlightening policy and practice," *BMC Public Health*, vol. 21, no. 1, pp. 1–12, Dec. 2021.

[7] A. Ali, S. A. Changezi, H. M. A. Rizwan, K. Shehzad, M. U. Nazir, and M. I. Naz, "A GIS architecture for medical disaster management to support modern healthcare management system," in *Proc. 2nd Int. Conf. Artif. Intell. (ICAI)*, Mar. 2022, pp. 13–18.

[8] M. Wang, Z. Zhang, J. Qin, Y. Chen, and B. Zhai, "Exploration on automatic management of GIS using TL-CNN and IoT," *IEEE Access*, vol. 10, pp. 40932–40944, 2022.

[9] Q. Cheng, N. T. Argon, C. S. Evans, Y. Liu, T. F. Platts-Mills, and S. Ziya, "Forecasting emergency department hourly occupancy using time series analysis," *Amer. J. Emergency Med.*, vol. 48, pp. 177–182, Oct. 2021.

[10] D. G. Costa, J. P. J. Peixoto, T. C. Jesus, P. Portugal, F. Vasques, E. Rangel, and M. Peixoto, "A survey of emergencies management systems in smart cities," *IEEE Access*, vol. 10, pp. 61843–61872, 2022.

[11] Ó. Fontenla-Romero, B. Guijarro-Berdiñas, D. Martinez-Rego, B. Pérez-Sánchez, and D. Peteiro-Barral, "Online machine learning," in *Efficiency and Scalability Methods for Computational Intellect*. Hershey, PA, USA: IGI global, 2013, pp. 27–54.

[12] Z. Zhou, D. S. Matteson, D. B. Woodard, S. G. Henderson, and A. C. Micheas, "A spatio-temporal point process model for ambulance demand," *J. Amer. Stat. Assoc.*, vol. 110, no. 509, pp. 6–15, Jan. 2015.

[13] H. Setzler, C. Saydam, and S. Park, "EMS call volume predictions: A comparative study," *Comput. Oper. Res.*, vol. 36, no. 6, pp. 1843–1851, Jun. 2009.

[14] Z. Wang, T. Xia, R. Jiang, X. Liu, K.-S. Kim, X. Song, and R. Shibasaki, "Forecasting ambulance demand with profiled human mobility via heterogeneous multi-graph neural networks," in *Proc. IEEE 37th Int. Conf. Data Eng. (ICDE)*, Apr. 2021, pp. 1751–1762.

[15] S. S. Jones, R. S. Evans, T. L. Allen, A. Thomas, P. J. Haug, S. J. Welch, and G. L. Snow, "A multivariate time series approach to modeling and forecasting demand in the emergency department," *J. Biomed. Informat.*, vol. 42, no. 1, pp. 123–139, Feb. 2009.

[16] H. J. Kam, J. O. Sung, and R. W. Park, "Prediction of daily patient numbers for a regional emergency medical center using time series analysis," *Healthcare Informat. Res.*, vol. 16, no. 3, p. 158, 2010.

[17] F. Kadri, F. Harrou, S. Chaabane, and C. Tahon, "Time series modelling and forecasting of emergency department overcrowding," *J. Med. Syst.*, vol. 38, no. 9, pp. 1–20, Sep. 2014.

[18] M. A. C. Vollmer, B. Glampson, T. Mellan, S. Mishra, L. Mercuri, C. Costello, R. Klaber, G. Cooke, S. Flaxman, and S. Bhatt, "A unified machine learning approach to time series forecasting applied to demand at emergency departments," *BMC Emergency Med.*, vol. 21, no. 1, pp. 1–14, Dec. 2021.

[19] J. Wang, X. Peng, J. Wu, Y. Ding, B. Ali, Y. Luo, Y. Hu, and K. Zhang, "Singular spectrum analysis (SSA) based hybrid models for emergency ambulance demand (EAD) time series forecasting," *IMA J. Manage. Math.*, vol. 35, no. 1, pp. 45–64, Dec. 2023.

[20] J. L. Vile, J. W. Gillard, P. R. Harper, and V. A. Knight, "Predicting ambulance demand using singular spectrum analysis," *J. Oper. Res. Soc.*, vol. 63, no. 11, pp. 1556–1565, Nov. 2012.

[21] L. Xu, M.-P. Kwan, S. McLafferty, and S. Wang, "Predicting demand for 311 non-emergency municipal services: An adaptive space-time kernel approach," *Appl. Geography*, vol. 89, pp. 133–141, Dec. 2017.

[22] K. Steins, N. Matinrad, and T. Granberg, "Forecasting the demand for emergency medical services," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2019, pp. 1–26.

[23] V. Nicoletta, A. Guglielmi, A. Ruiz, V. Bélanger, and E. Lanzarone, "Bayesian spatio-temporal modelling and prediction of areal demands for ambulance services," *IMA J. Manage. Math.*, vol. 33, no. 1, pp. 101–121, Aug. 2021.

[24] Z. Zhou and D. S. Matteson, "Predicting ambulance demand: A spatio-temporal kernel approach," in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 2297–2303.

[25] A. X. Lin, A. F. W. Ho, K. H. Cheong, Z. Li, W. Cai, M. L. Chee, Y. Y. Ng, X. Xiao, and M. E. H. Ong, "Leveraging machine learning techniques and engineering of multi-nature features for national daily regional ambulance demand prediction," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, p. 4179, Jun. 2020.

[26] A. Haugsbø Hermansen and O. J. Mengshoel, "Forecasting ambulance demand using machine learning: A case study from Oslo, Norway," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2021, pp. 1–10.

[27] E. Van De Weijer, O. A. Owren, and O. J. Mengshoel, "Forecasting hourly ambulance demand for Oslo, Norway: A neuro-symbolic method," *Norsk IKT-Konferanse Forskning Utdanning*, vol. 1, no. 1, pp. 1–14, 2023.

[28] M. Rautenstrauß and M. Schiffer, "Ambulance demand prediction via convolutional neural networks," 2023, *arXiv:2306.04994*.

[29] T. Nakai, S. Saiki, and M. Nakamura, "Medium-term prediction for ambulance demand of heat stroke using weekly weather forecast," in *Proc. 8th Int. Conf. Internet Things, Syst., Manage. Secur. (IOTSMS)*, Dec. 2021, pp. 1–8.

[30] R. Jin, T. Xia, X. Liu, T. Murata, and K.-S. Kim, "Predicting emergency medical service demand with bipartite graph convolutional networks," *IEEE Access*, vol. 9, pp. 9903–9915, 2021.

[31] Y.-C. Lee, Y.-S. Chen, and A. Y. Chen, "Lagrangian dual decomposition for the ambulance relocation and routing considering stochastic demand with the truncated Poisson," *Transp. Res. Part B: Methodol.*, vol. 157, pp. 1–23, Mar. 2022.

[32] I. Pétursson and H. Oxenholt, "Improving estimation of ambulance travel time by pre-processing and analyzing historic transports and weather data for machine learning models," Dept. Comput. Sci. Media Technol., Fac. Technol. Soc., Malmö Univ., 2024.

[33] D. Neira-Rodado, J. W. Escobar-Velasquez, and S. McClean, "Ambulances deployment problems: Categorization, evolution and dynamic problems review," *ISPRS Int. J. Geo-Information*, vol. 11, no. 2, p. 109, Feb. 2022.

[34] V. Raina, S. Krishnamurthy, V. Raina, and S. Krishnamurthy, "Data preparation," in *Building an Effective Data Science Practice: A Framework to Bootstrap and Manage a Successful Data Science Practice*. Berkeley, CA, USA: Apress, 2022, pp. 173–185.

[35] E. Kerakos, O. Lindgren, and V. Tolstoy, "Machine learning for ambulance demand prediction in Stockholm county: Towards efficient and equitable dynamic deployment systems," School Elect. Eng. Comput. Sci., Roy. Inst. Technol., Stockholm, Sweden, 2020.

[36] M. Abdollahi and Z. Boujarnezhad, "Intelligent transportation system based on the whale algorithm in Internet of Things," *J. Elect. Comput. Eng. Innov. (JECEI)*, vol. 10, no. 2, pp. 351–362, 2022.

[37] D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial–temporal data," *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, Jan. 2007.

[38] N. A. Houacine, L. S. Bendimerad, and H. Drias, "Heterogeneous DBSCAN for emergency call management: A case study of COVID-19 calls based on hospitals distribution in Saudi Arabia," in *Proc. Int. Conf. Innov. Bio-Inspired Comput. Appl.*, 2021, pp. 402–411.

[39] S. Nandita, G. Zampani, G. S. Krishnan, G. Ramadurai, and B. Ravindran, "Automated incident location identification for EMS from ambulance geospatial data," in *Proc. 5th Joint Int. Conf. Data Sci. Manage. Data*, Jan. 2022, pp. 162–168.

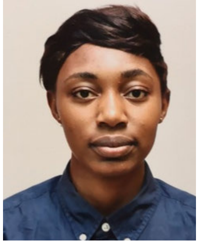
[40] A. Dubey, "Optimal placement of ambulances to best serve emergency calls," Ph.D. dissertation, School Comput., Nat. College Ireland, Dublin, Ireland, 2023.

[41] A. Altmann, L. Tolosi, O. Sander, and T. Lengauer, "Permutation importance: A corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, May 2010.

[42] V. Fonti and E. Belitser, "Feature selection using lasso," *VU Amsterdam Res. paper Bus. analytics*, vol. 30, pp. 1–25, Mar. 2017.

[43] M. M. Hosseini, S. T. M. Hosseini, K. Qayumi, S. Ahmady, and H. R. Koohestani, "The aspects of running artificial intelligence in emergency care; a scoping review," *Arch. Academic Emergency Med.*, vol. 11, no. 1, p. e38, 2023.

[44] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowl. Inf. Syst.*, vol. 41, no. 3, pp. 647–665, Dec. 2014.
[45] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–22.
[46] B. Johnston and I. Mathur, *Applied Supervised Learning With Python: Use Scikit-Learn to Build Predictive Models From Real-World Datasets and Prepare Yourself for the Future of Machine Learning*. Birmingham, U.K.: Packt Publishing Ltd, 2019.
[47] S. Ramgopal, T. Westling, N. Siripong, D. D. Salcido, and C. Martin-Gill, "Use of a metalearner to predict emergency medical services demand in an urban setting," *Comput. Methods Programs Biomed.*, vol. 207, Aug. 2021, Art. no. 106201.
[48] W. Liang, S. Luo, G. Zhao, and H. Wu, "Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms," *Mathematics*, vol. 8, no. 5, p. 765, May 2020.
[49] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput.*, vol. 74, pp. 634–642, Jan. 2019.
[50] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[51] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers Neurorobotics*, vol. 7, p. 21, 2013.
[52] L. Ma, G. Zhang, and E. Lu, "Using the gradient boosting decision tree to improve the delineation of hourly rain areas during the summer from advanced Himawari imager data," *J. Hydrometeorology*, vol. 19, no. 5, pp. 761–776, May 2018.
[53] S. Elsayed, D. Thyssens, A. Rashed, H. S. Jomaa, and L. Schmidt-Thieme, "Do we really need deep learning models for time series forecasting?" 2021, *arXiv:2101.02118*.
[54] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–26.
[55] P. S. C. Sa and G. Divyab, "A light gradient boosting machine regression model for prediction of agriculture insurance cost over linear regression," *Adv. Parallel Comput. Algorithms, Tools Paradigms*, vol. 41, p. 200, Aug. 2022.
[56] A. Krogh, "What are artificial neural networks?" *Nature Biotechnol.*, vol. 26, no. 2, pp. 195–197, Feb. 2008.
[57] B. Yegnanarayana, *Artificial Neural Networks*. Birmingham, U.K.: PHI Learning Pvt. Ltd., 2009.
[58] H. Huang, M. Jiang, Z. Ding, and M. Zhou, "Forecasting emergency calls with a Poisson neural network-based assemble model," *IEEE Access*, vol. 7, pp. 18061–18069, 2019.
[59] P. Abreu, D. Santos, and A. Barbosa-Povoa, "Data-driven forecasting for operational planning of emergency medical services," *Socio-Econ. Planning Sci.*, vol. 86, Apr. 2023, Art. no. 101492.
[60] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
[61] J. Patterson and A. Gibson, *Deep Learning: A Practitioner's Approach*. Sebastopol, CA, USA: O'Reilly Media, Inc, 2017.
[62] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021.
[63] M. Chirico. (2020). *Emergency-911 Calls*. [Online]. Available: https://www.kaggle.com/dsv/1381403
[64] A. J. Longden, "Meteomatics," in *Proc. 102nd American Meteorological Society Annual Meeting*, 2022, pp. 1–24.
[65] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA, USA: CreateSpace, 5555.
[66] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing, "Jupyter notebooks—A publishing format for reproducible computational workflows," in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Amsterdam, The Netherlands: IOS Press, 2016, pp. 87–90.
[67] S. Flennerhag, "Ml-Ensemble," School Ind. Eng. Manag. (ITM), Prod. Eng., KTH, Stockholm, Sweden, 2018.
[68] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jul. 2011.
[69] J. Montiel, M. Halford, S. M. Mastelini, G. Bolmier, R. Sourty, R. Vaysse, A. Zouitine, H. M. Gomes, J. Read, T. Abdessalem, and A. Bifet, "River: Machine learning for streaming data in Python," *J. Mach. Learn. Res.*, vol. 22, no. 1, pp. 4945–4952, 2021.
[70] *Google Colab*. Accessed: Apr. 18, 2023. [Online]. Available: https://colab.research.google.com/
[71] D. D. Desai, J. Dey, S. K. Satapathy, S. Mishra, S. N. Mohanty, P. Mishra, and S. K. Panda, "Optimal ambulance positioning for road accidents with deep embedded clustering," *IEEE Access*, vol. 11, pp. 59917–59934, 2023.
[72] R. Manaa, M. B. Talha, L. Aboueljinane, A. Benghabrit, and Y. Frichi, "Application of machine learning techniques for ambulance coverage prediction," in *Proc. IEEE 15th Int. Colloq. Logistics Supply Chain Manage. (LOGISTIQUA)*, May 2024, pp. 1–8.



TALOTSING GAELLE PATRICIA MEGOUO received the B.Eng. and M.Eng. degrees in networks and telecommunications from the Ecole Nationale Superiéure Polytechnique de Douala (ENSP, ex-FGI), Cameroon, in 2017, and the M.Sc. degree in applied mathematics from AIMS, Limbe, Cameroon, in 2020. She is currently pursuing the Ph.D. degree in computer engineering with Polytechnique Montréal, QC, Canada. She worked on several projects combining machine learning, telecommunications, and advanced networks (SDN and NFV). In 2018, she received a MasterCard scholarship as a student with African Institute for Mathematical Science (AIMS), Cameroon. In 2019, she received the IVADO Excellence Scholarship, giving her the opportunity to intern with Canada Excellence Research Chair in Real-Time Decision-Making (CERC), Montreal, Canada. Her research interests include ambulance demand prediction, ambulance allocation, and routing using deep learning and machine learning algorithms.

SAMUEL PIERRE (Senior Member, IEEE) is currently a Professor with the Department of Computer and Software Engineering, Polytechnique Montréal, Canada, and the Director of the Mobile Computing and Networking Research Laboratory (LARIM). He has authored or co-authored more than 600 technical publications, including articles in refereed archival journals, textbooks, patents, and book chapters. His research interests include wired and wireless communications, mobile computing and networking, cloud computing, and e-learning. He was a recipient of several awards, including the Prix Poly 1873 for Excellence in Teaching and Training (2001 and 2005) and the Knight of the National Order of Quebec, in 2009. He is a fellow of the Engineering Institute of Canada (2003) and Canadian Academy of Engineering (2008). In December 2011, he was appointed as a member of the Order of Canada. In May 2014, he was a recipient of an Honorary Doctorate from University of Quebec at Trois-Rivieres (UQTR) and another from the University of Quebec in Outaouais (UQO), in November 2016. In 2017, he received the El Fassi Prize from the Agence Universitaire de la Francophonie (AUF) to highlight the action of a person who has exerted a significant influence through the quality of his expertise and the innovative nature of his achievements at the international level in the fields of research, training, development and international cooperation, governance, and/or transfer of knowledge or skills. He received the Grand Prize for Professional Excellence from the Order of Engineers of Quebec in 2020, the Gold Medal from Engineers Canada, in 2021, the promotion of Officer of the National Order of Quebec, in 2022, the JM Ham Award from IEEE Canada, in 2023 for "Outstanding contributions to training the next generation of scientists and engineers in the areas of mobile computing, networking, and e-learning" and the Julian C. Smith Medal of the Engineering Institute of Canada "For distinguished achievements in the development of Canada," in 2024.