Bayesian Optimization Enhanced ARIMA Modeling for Accurate Forecasting in Emergency Medical Services

Hanaa Ghareib Hendi [1,*], Masoud E. Shaheen [2], Mohamed Hassan Ibrahim [1], and Mohamed Hassan Farag [1]

[1] Department of Information Systems, Faculty of Computers and Artificial Intelligence, Fayoum University, Cairo, Egypt
[2] Department of Computer Science, Faculty of Computers and Artificial Intelligence, Fayoum University, Cairo, Egypt
Email: hanaa_ghareib@fayoum.edu.eg (H.G.H.); mem00@fayoum.edu.eg (M.E.S.); mhi11@fayoum.edu.eg (M.H.I.); mohamed.farrag@fayoum.edu.eg (M.H.F.)
*Corresponding author

***Abstract*—Emergency Medical Services (EMS), play a vital role for community well-being, provides lifesaving assistance during emergencies. Accurately forecasting emergency call demand is crucial for optimizing resource allocation and improving response times. In this study, we analyzed an EMS dataset containing emergency call details from four U.S. states to develop a predictive model. We utilized the Autoregressive Integrated Moving Average (ARIMA) model, a widely adopted method for analyzing and forecasting stationary time series data. To fine-tune the ARIMA model's hyperparameters, we implemented three methods: Auto-ARIMA, grid search, and Bayesian Optimization (BO). Although Auto-ARIMA and grid search generated reasonable predictions, BO yielded superior accuracy with more precise forecasts. This finding underscores the superiority of BO for time series prediction tasks. The finding of this study could help EMS organizations in effectively predicting demand, leading to better resource allocation, enhanced operational efficiency, and faster response times. Additionally, these precise predictions can strengthen EMS systems' capacity to handle emergencies and improve overall health infrastructure.**

***Keywords*—Emergency Medical Services (EMS), Bayesian Optimization (BO), Autoregressive Integrated Moving Average (ARIMA) model**

I. Introduction

Emergency Medical Services (EMS) represent an important pre-hospital component of the health care system, extending beyond the familiar association with ambulances. Their primary goal is to reduce injury and mortality by providing prompt and effective treatment in an emergency. Reducing response times is of utmost importance, especially for high-priority calls involving critically ill patients [1]. Additionally, EMS plays an important role in today's health care systems by providing medical intervention and transportation from the scene of an emergency call to provide comprehensive care, often in a hospital setting, ensuring the rapid and safe transfer of patients from disaster areas to appropriate hospitals [2]. Effective EMS responses to emergencies are likely to improve patient outcomes and recovery [3]. Planning and building EMS facilities are essential to ensure rapid and timely response. Accurately predicting ambulance demand has significant benefits for various stakeholders in the emergency health care system [4]. This includes ambulance services, receiving hospitals, and citizens who rely on critical response time to ensure rapid access to critical care, so a comprehensive understanding of ambulance requirements is essential to provide informed health care efficiency and improved patient outcomes [5].

Time series forecasting is emerging as a powerful tool that EMS can use to increase response time, optimize resource allocation, and enhance patient care [6]. Time series analysis helps predict future needs for emergency services, enabling EMS to deploy resources strategically and ensure availability when and where needed [7]. Many time series forecasting models exist to improve accuracy and efficiency by reducing errors. Time series prediction focuses on various methods, including statistical methods [8] and machine learning methods [9] .Statistical methods, including Autoregressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), and Autoregressive Moving Average (ARMA) models, use statistical inference to identify underlying patterns in data sets. These methods effectively rely on detailed analysis data analysis to identify underlying patterns. The ARIMA model is a popular statistical prediction method known for its accuracy and efficiency. ARIMA model is a statistical framework specially developed for the analysis and prediction of static time series data [10]. Furthermore, in the core of the ARIMA model, there are two keys, AR and MA objects. Polynomials represent these mathematically and together form a detailed model for time series analysis. Time series predictions, especially using the ARIMA model, are valuable for optimizing EMS resource allocation and response time. By using and utilizing historical data, ARIMA can forecast ambulance needs, enabling initial

Manuscript received January 9, 2025; revised February 11, 2025; accepted March 3, 2025; published July 15, 2025.

logistics planning and strategic deployment of ambulances with time faster response times, improved patient outcomes and better decision-making for EMS leader [11, 12].

Bayesian Optimization (BO) uses a surrogate model to optimally evaluate only the most promising hyperparameter settings [13, 14]. BO offers a robust alternative, using a surrogate model to prioritize and evaluate the most promising hyperparameter schemes. This surrogate model using Bayes' theorem accounts for the posterior distribution of the objective function (e.g. prediction error). By focusing on these promising regions of hyperparameter space, BO reduces the number of experiments which is more important compared to general search methods which search for min (or max).

In this work, we recommend an investigation into the effectiveness of the ARIMA model to predict the number of incoming emergency calls, intending to improve resource utilization and response time. Although ARIMA provides a solid foundation for time series analysis due to its sensitivity to insect value parameter selection, we will further optimize its performance using a network search method using grid search and BO. This combined approach allows us to explore multiple hyperparameter settings for the ARIMA model, which can lead to more accurate and efficient prediction solutions for EMS call rates by comparing the performance of the different optimized methods used with the ARIMA model. We can imagine the potential benefits of the strategy.

The remainder of this paper is organized as follows: Section II reviews related work, Section III describes the dataset and the ARIMA model with Grid Search and Bayesian Optimization. Section IV presents the discussion, analyzing the results and comparing them with existing studies. Finally, Section V concludes the paper, summarizing the key findings and suggesting directions for future research.

II. Literature Review

Accurate forecasting of emergency needs is critical to EMS success and efficiency. It enables improved site planning and streamlines transportation, significantly increasing response times, prehospital care, and survival rates [15]. Recently, many advanced studies have been carried out to predict EMS call volume [15–17]. Prediction models are categorized into three main classes, namely: traditional techniques, artificial techniques, and hybrid techniques. Time series techniques are considered one of the widely used traditional methods for EMS forecasting [18–20]. Traditional statistical methods such as the ARIMA model have been widely used in their interpretation and efficiency. The ARIMA model is a popular and versatile method for time series forecasting, as it includes autoregressive and moving average components to capture the underlying dynamics of the data [21]. However, Hajirahimi and Khashei [22] indicate that hybrid models combining statistical and machine learning approaches yield better accuracy. The ARIMA method has been utilized in [23, 24]. According to Ref. [25], the ARIMA model has been successfully applied to various EMS-related time scales, such as ambulance dispatch, response time, and resource utilization. For instance, Al-Azzani *et al*. [26] compares four forecasting methods that use data from the Welsh Ambulance Service to predict the number of calls required. ARIMA performs well in weekly and monthly forecasts, while Single Spectrum Analysis (SSA) performs well in long-term forecasting. Also, Asghar *et al*. [27] measured rates of immobile disease over time, compare ambulatory services, and look for prognostic factors. The study analyzed monthly national health worker illness, comparing 10 regional ambulance services in England from 2009 to 2018. The study used ARIMA and Seasonal ARIMA models to predict that results showed there was a significant difference in annual sickness absence rates between ambulance services and the 10-year study period in England. Hybrid models offer several advantages over single methods. The idea behind the hybrid model is to combine complex features from a collection of simple models. Several of these investigations are described as follows: Ong *et al*. [28] used a Genetic Algorithm (GA) to define models for ARIMA and Seasonal Autoregressive Integrated Moving Average (SARIMA) to solve the issue of local optimal value. While Ervural *et al*. [29] here proposed an integrated forecasting method that incorporates a GA and ARMA model using both methods. Consequently, the results show that the proposed model outperforms traditional ARMA in cost function value reduction. Zhang *et al*. [30] used a hybrid ARIMA-supply vector machine method to predict daily radiology emergency patient flow. This combines ARIMA and Support Vector Regression (SVR) models to capture both linear and nonlinear patterns in the data. The hybrid model outperforms the single model, with Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) values of 7.02%, 19.20, and 14.97, respectively. The results show that the hybrid ARIMA-SVR method is a promising alternative for patient flow prediction. Also, several studies have applied ARIMA models for infectious disease forecasting, often optimizing their parameters using a grid search. Nichols and Abolmaali [31] conducted a comparative analysis of ARIMA and SARIMA models for COVID-19 case prediction, employing grid search to determine optimal parameters. Their findings highlight the importance of parameter tuning in improving ARIMA's forecasting accuracy, reinforcing the need for advanced optimization techniques in time-series modeling.

Studies that adopted EMS time prediction using Bayesian optimization are limited. Recent developments in time series forecasting have introduced the Bayesian Optimization-Based Dynamic Ensemble (BODE) approach [32], introduced a new cluster forecasting method for time series data. It uses Bayesian optimization to dynamically weight a combination of statistical, machine learning, and deep neural network models based on their recent performance This method is further improved by hyperparameter tuning through Bayesian optimization. Validation is achieved by testing data with different characteristics (hourly, daily, weekly, monthly).

The results show that BODE significantly outperforms the analytical methods. An ablation study highlights the importance of each factor in BODE, including data enhancement techniques, hyperparameter tuning, and various candidate models, all of which contribute to the robust performance of the method. Also, Sultana *et al*. [33] examined the statistical machine learning methods to predict hourly electricity demand in Ontario was investigated. It introduces new methods for identifying influential factors, optimizing model parameters, and testing model performance. Both models, Bayesian Optimization Algorithm-Nonlinear autoregressive with external input (BOA-NARX) and Bayesian Optimization Algorithm-Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (BOA-SARIMAX), are compared, with BOA-NARX showing consistent and superior performance in predicting daily lighting loads. Finally, Yang *et al*. [34] investigated the application of ARIMA, SARIMA, and Dynamic Bayesian Network (DBN) models for forecasting maritime EMS cases in China. Their study analyzed patient data from 2016 to 2021, optimizing ARIMA and SARIMA parameters using statistical techniques such as the Akaike Information Criterion (AIC). The results showed that SARIMA outperformed ARIMA and DBN in predictive accuracy, effectively capturing seasonal patterns in EMS demand. While ARIMA demonstrated strong short-term forecasting capabilities, its performance relied heavily on parameter tuning. The study highlights the importance of selecting appropriate forecasting models for EMS planning and suggests that advanced optimization techniques could further enhance predictive accuracy.

III. MATERIALS AND METHODS

In the initial investigation, we analyze data sets on EMS calls hosted on the online data science platform Kaggle (https://www.kaggle.com/datasets/new-york-city/ny-ems-incident-dispatch-data). This data set included approximately 1 million daily emergency calls recorded over nine months. The data contained 32 unique attributes, including a unique identifier for each incident (INCIDENT_ID), the date and time the incident entered the system (INCIDENT_DATETIME), and a critical event code (SEVERITY_CODE); representative examples of dataset characteristics are shown in Table I.

TABLE I. SAMPLE DESCRIPTION OF DATASET ATTRIBUTES

Attribute Name	Description
INCIDENT_ID	An incident unique identifier
INCIDENT_DATETIME	The date and time the incident was created in the dispatch system.
INITIAL_CALL_TYPE	The call type assigned at the time of incident creation.
SEVERITY_CODE	The segment (priority) assigned at the time of incident creation.
FINAL_CALL_TYPE	The call type at the time the incident closes.
ZIPCODE	The zip code of the incident.

Fig. 1 illustrates the structured workflow of the forecasting of volumes of emergency calls using ARIMA modeling and some optimization techniques. The input of data is done first; afterward, some preprocessing on it has been done, handling missing values, and stabilization transforms. Stationarity checking is done because ARIMA requires a time series to be stationary.

Once stationarity is established, the ARIMA model is applied at the outset. To further improve the performance of the model, its hyperparameters are tuned using grid search and Bayesian optimization to improve model efficiency. The performance of the final model is evaluated using different metrics like RMSE, MAE, and MAP. This workflow will ensure the selection of the best-fit model for reliable forecasting of emergency call volumes.

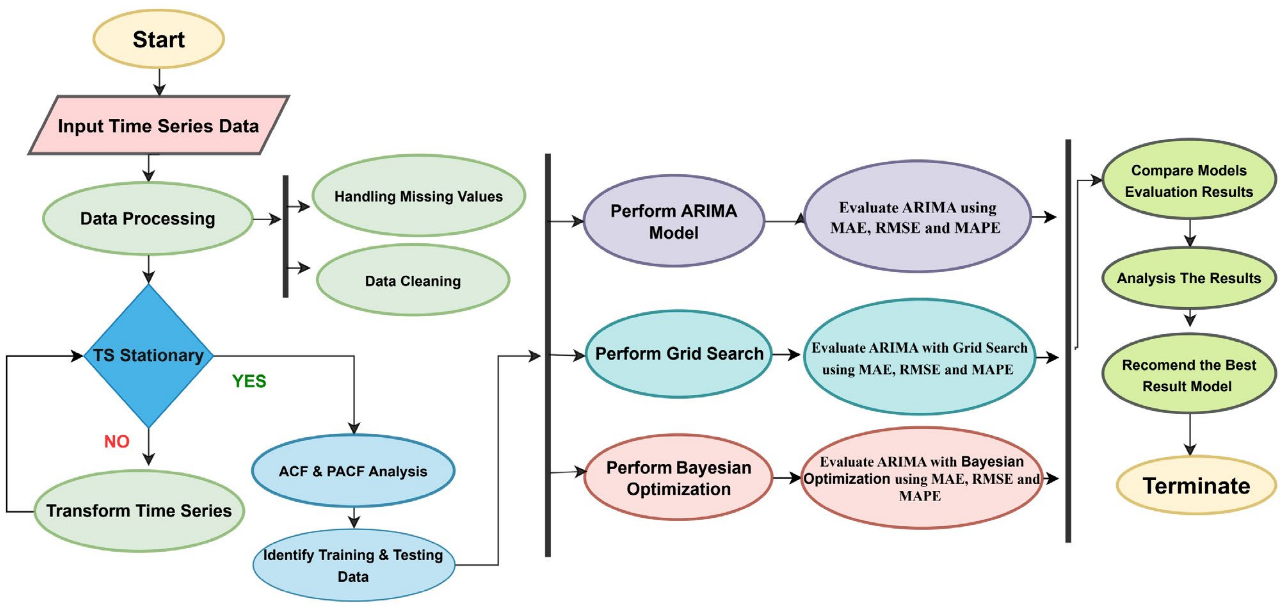

Fig. 1. Structured workflow of the forecasting of volumes of emergency calls using ARIMA modeling and some optimization techniques.

A. Data Statistical

The success of many forecasting methods depends on the stability of the underlying time series. Stability refers to the stochastic phenomenon in which statistical properties, such as mean, variance, and covariance remain constant over time [35]. Simply, a fixed timeline does not imply any trends or seasonal changes. It is important to know the stability of the series, as forecasting techniques depend on this assumption for accurate forecasting. Consequently, there are several approaches to check the stability of time series data including.

1) Visualization (Plot the series)

Analysis of time series constants is an important step in many statistical analyses. While initial analysis of time series data can be informative, it is an inherently subjective concept. Visual analyzes can overlook subtle and variable attributes, especially trends or seasonal patterns. Fig. 2 provides a graphic representation of high intensity emergency calls daily. Although the plot may provide preliminary insights, more rigorous statistical tests are needed to enable a comprehensive analysis of stability in this time series.

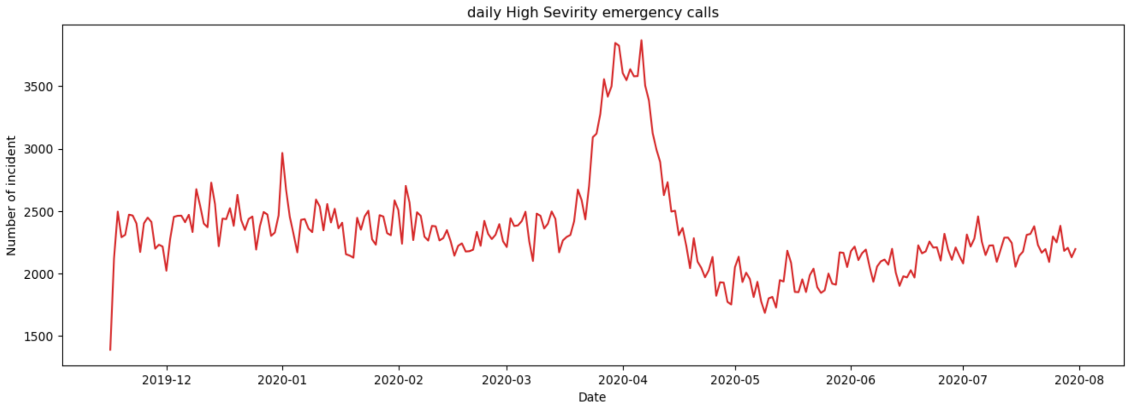

Fig. 2. Time series visualization.

2) Data exploration

To analyze the data, we divide the time series into two or three segments and compare statistical features such as the mean differences of each group. If the shape of the partition statistics is very different, we can see that the time series is not stable. We begin by constructing a histogram of time series values, shown in Fig. 3.

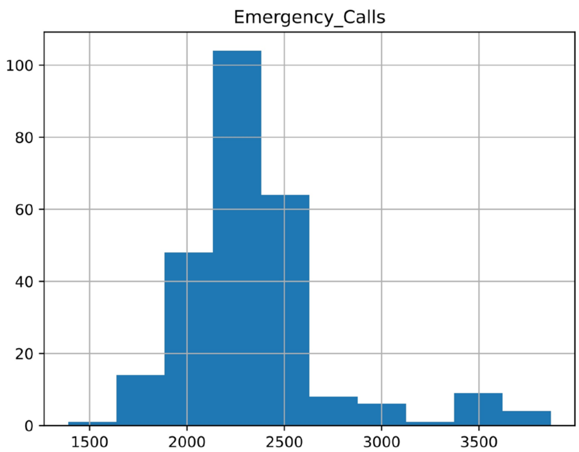

Fig. 3. Histogram of daily high severity emergency calls.

In the next step, we will divide the data into three equal parts and calculate the mean and variance of each part. As can be seen in Table II, the results are different, but not significant. In addition, Table III shows the normalized form of the log. Based on the data presented in Table II and Table III we can conclude that the time series is a stationary series.

TABLE II. SUMMARY STATISTICS OF DAILLY EMERGENCY CALLS

Criteria	Part 1	Part 2	Part 3
Mean	2391.1	2526.7	2099.0
Variance	347,009	289,348.1	26,521.9

TABLE III. SUMMARY STATISTICS OF LOG DAILLY EMERGENCY CALLS

Criteria	Part 1	Part 2	Part 3
Mean	7.78	7.81	7.65
Variance	0.0073	0.039	0.0064

3) Statistical test

In this section, statistical tests were applied to determine whether the data met or deviated from the criterion for robustness. To determine whether the time series is stable or unstable due to the unit root, we relied on the Augmented Dickey-Fuller (ADF) test [30] which developed Eq. (1).

$$\mathrm{ADF}(x_t) = \alpha + p x_{t-j} + \varepsilon_t \tag{1}$$

where α is the approximate constant value of the time series data, p is the hypothesis based on time series stability $p = 1$ or $p < 1$, t time $\{1...., j\}$, and ε is the white noise of time series data supply. As shown in Fig. 4, ADF was tested on the time series as formulated in Eq. (1), we found that the statistical value was lower than any other significant values, and the p-value was lower at 0.05. The worse this statistic is, the more likely it is that there will be nonstationary data.

```
========================
ADF Statistic: -3.086143
========================
p-value: 0.027596
Critical Values:
    1%: -3.457
    5%: -2.873
    10%: -2.573
Reject Hypothesis : Time Series is Stationary
```

Fig. 4. The Augmented Dickey Fuller (ADF) test.

B. The Correlation

The objective of time series forecasting is to estimate the value of a variable at some future point in time $(t + h)$ using the same variable information at the current point in time (t) but with the univariate characteristics of time series data (i.e. they consist of only one variable). The relationship between the values of a variable must be used. This is done by estimating the relationship between one's own timeline and subsequent versions. This type of correlation analysis helps us use past observations of a variable to predict its future behavior. To further investigate the relationship between past and future values of the time series, a lag plot was constructed. This plot shows the relationship between observations at time $t-1$ (on the x-axis) and time $t+1$ (on the y-axis). Fig. 5 provides a latency plot for daily emergency calls. The skewed pattern of the plot shows a positive correlation between past and future call volumes. This suggests that the time series exhibits a degree of reliability, where past observations can be informative for predicting future values. The Autocorrelation Function (ACF) of daily emergency call data [36] in Fig. 6, the plot reveals a significant positive relationship between the time series and its lagged values, especially at lags from $t = 1$ to $t \approx 9$. This suggests that the number of emergency calls per day is positively correlated with it, and this correlation is weaker for delays larger than about nine days (Table IV).

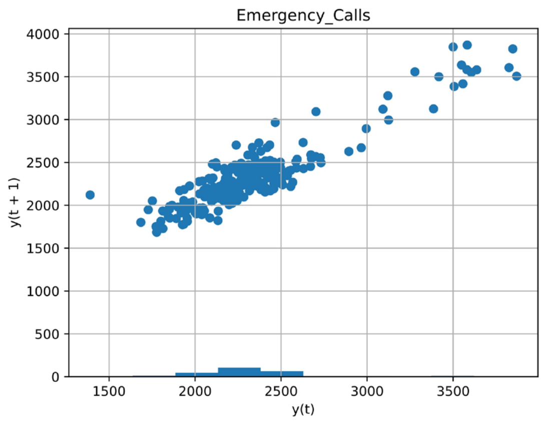

Fig. 5. The autocorrelation of daily emergency calls.

We use Partial Autocorrelation Function (PACF) in conjunction with the ACF analysis to gain a more detailed understanding of the lagged relationships in the time series data. The PACF extracts the unique association between the variable and its delay, and pays close attention to the effect of the intervening delay on the correlation coefficient. This characteristic enables PACF to identify delays that have a direct impact on the current price independently of any relationship with previous delays [38]. In simple terms, PACF identify delays that are highly correlated with the current price has, even after accounting for the correlation of the previous delay. It is extended for a period of 30 (approximately one month). ACF and PACF integrated visualization provides a comprehensive understanding of lagging relationships in the data as seen in Fig. 7.

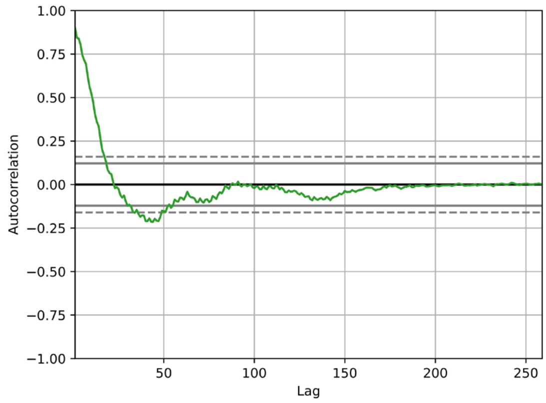

Fig. 6. The correlation by data plotting.

TABLE IV. THE PEARSON CORRELATION COEFFICIENT (PPC) CORRELATION TEST

Time	$t-1$	$t+1$
$t-1$	1.000000	0.910714
$t+1$	0.910714	1.000000

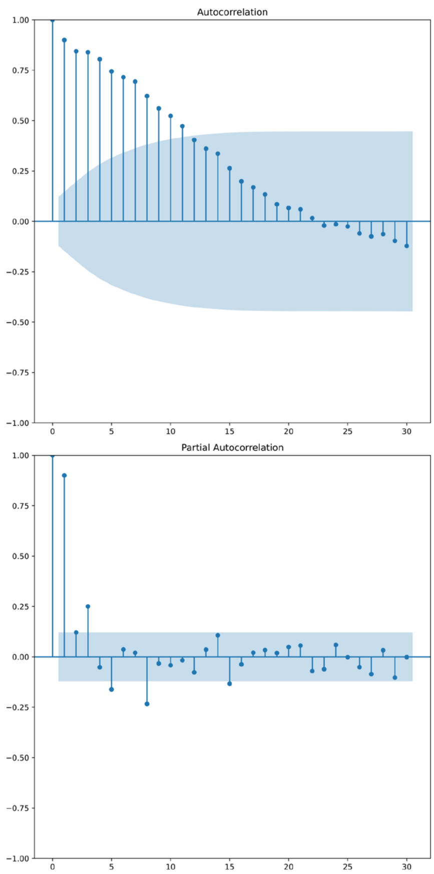

Fig. 7. The ACF and PACF graph for daily emergency calls.

C. Performance Evaluation

To evaluate the accuracy of the proposed model, its predictions compare with real-world data using established regression metrics. Three common statistical units are used in this study: MAE [39], RMSE [40], and MAPE [41] defined in Eqs. (2)–(4), respectively. MAE refers to the absolute difference between predicted and actual values and exhibits less sensitivity to noisy data compared to RMSE. Although RMSE provides a measure of the magnitude of squared forecast errors, caution is needed regarding the presence of outliers with large, squared errors, especially in the presence of noisy data [42]. This involved comparing predicted future results produced by each algorithm with actual observations.

The calculation of three statistical parameters provided a quantitative measure of the difference between predicted and observed parameters, and allowed accuracy to accuracy comparisons among different optimization algorithms as in Eqs. (2)–(4).

$$MAE = \frac{\sum_{i=1}^{n} |p_i - A_i|}{n} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_i - A_i)^2} \tag{3}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} (\frac{|p_i - A_i|}{p_i}) \times 100 \tag{4}$$

where n = the total number of days in the forecasting period, p_i = the predicted call volume on day i, and A_i = the actual call volume on day i.

D. AutoRegression Integrated Moving Average (ARIMA) Model

The ARIMA model, also known as the Box-Jenkins methodology [43], is a popular choice for time series forecasting. Several researchers have explored its applications [38, 44]. ARIMA's strength lies in its ability to capture trends and periodicities within the data, making it suitable for short-term forecasting [2, 45]. AR model Eq. (5) concerned with remembering the past; how past values of the data itself (y_{t-i}) influence the current prediction (y_t).

$$AR(p): y_t = \sum_{i=1}^{p} (\emptyset_i y_{t-i}) + \omega_t \tag{5}$$

where: y_t: The predicted value of the time series at time t, $\emptyset$: an appropriate coefficient of AR. These coefficients indicate the weight given to each lagged value (y_{t-i}), y_{t-i}: The lagged values of the time series (i = 1, 2, ..., p), and ω_t: Current error term.

On the other hand, MA Eq. (6) learning from mistakes; MA considers how much past forecasting errors (ε_{t-i}) affect the current prediction (y_t).

$$MA(q): y_t = \sum_{i=1}^{q} (\theta_i \varepsilon_{t-i}) + \omega_t \tag{6}$$

where: y_t: The predicted value of the time series at time t, θ: is a finite coefficient of MA. These coefficients indicate the weight given to each past error term, ε_{t-i}: The past error terms (residuals) from past forecasts (i = 1, 2, ..., q), εt: Current model residual (the prediction error), and ω_t : The white noise, q is a lag order.

The Autoregressive Moving Average (ARMA) model combines the strengths of the AR(p) and MA(q) processes for time series forecasting. ARMA (p, q) uses both the current and the past values of the remaining periods (errors from previous forecasts). Eq. (7) represents this model mathematically.

$$y_t = \sum_{i=1}^{p} (\emptyset_i y_{t-i}) + \sum_{i=1}^{q} (\theta_i \varepsilon_{t-i}) + \omega_t \tag{7}$$

where: y_t: Predicted value at time t, $\emptyset$: AR coefficients (weights for past values), y_{t-i}: Lagged values of the series (i = 1 to p), θ: MA coefficients (weights for past errors), ε_{t-i} : Past error terms (i = 1 to q), and ω_t: Current error term The ARIMA (p, d, q) models address the limitations of ARMA by adding an additional differentiating term to Eqs. (8–10) which is a preconditioning method for energy generation, where with the time delay between successive samples (dth order difference) calculated [46]. So far. This iterative subtraction process effectively removes inherent trends and seasonality in the data, making the model more suitable for further analysis using ARIMA.

$$D_1 y_t = y_t - y_{t-1} \textit{ (The first difference d=1)} \tag{8}$$

$$D_2 y_t = D_1 y_t - D_1 y_{t-1} \textit{ (The second difference d=2)} \tag{9}$$

$$D_d y_t = D_{d-1} y_t - D_{d-1} y_{t-1} \textit{ (the } d^{th} \textit{ difference)} \tag{10}$$

Eq. (11) depicts the mathematical formulation of the ARIMA model incorporating the differencing step (d) [47].

$$D^{\wedge}d y_t = \sum_{i=1}^{p} (\emptyset_i D^{\wedge}d y_{t-i}) + \sum_{i=1}^{q} (\theta_i \varepsilon_{t-i}) + \omega_t \tag{11}$$

where: $D^{\wedge}dy_t$: The differenced series at level d (d^{th} difference of the original series y_t).

Power is provided by differences, although in this case the data appear to be stable without differences, so the integration process (d) is set to 0. The main challenge associated with the use of ARIMA ho and AR (p), MA (q) and discrimination (d) coefficients [48]. Auto ARIMA makes it easier to determine the optimal parameters for ARIMA models. It uses the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test to determine the difference pattern (d) and subsequently applies Akaike Information Criterion (AIC) reduction [49] to select autoregressive (p) and moving average (q) terms. This automatic setting facilitates model selection based on statistical criteria, as shown in Fig. 8.

The Auto ARIMA model used obtained an ARIMA (3, 0, 2) structure as the best model low error Auto ARIMA model achieved RMSE of 180, MAPE of 150, and MEA of 120, as shown in Fig. 9. These precise metrics demonstrate the performance of the model. The corresponding time series forecast diagram is shown in Fig. 10.

```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(0,0,0)[0] intercept   : AIC=2712.170, Time=0.20 sec
 ARIMA(0,0,0)(0,0,0)[0] intercept   : AIC=3090.610, Time=0.02 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=2720.550, Time=0.08 sec
 ARIMA(0,0,1)(0,0,0)[0] intercept   : AIC=2910.454, Time=0.13 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=3813.984, Time=0.01 sec
 ARIMA(1,0,2)(0,0,0)[0] intercept   : AIC=2712.462, Time=0.29 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=inf, Time=0.71 sec
 ARIMA(3,0,2)(0,0,0)[0] intercept   : AIC=2685.118, Time=0.77 sec
 ARIMA(3,0,1)(0,0,0)[0] intercept   : AIC=2706.047, Time=0.29 sec
 ARIMA(4,0,2)(0,0,0)[0] intercept   : AIC=2687.015, Time=0.87 sec
 ARIMA(3,0,3)(0,0,0)[0] intercept   : AIC=2684.956, Time=0.64 sec
 ARIMA(2,0,3)(0,0,0)[0] intercept   : AIC=2704.879, Time=1.25 sec
 ARIMA(4,0,3)(0,0,0)[0] intercept   : AIC=2691.039, Time=1.70 sec
 ARIMA(3,0,4)(0,0,0)[0] intercept   : AIC=2687.352, Time=1.60 sec
 ARIMA(2,0,4)(0,0,0)[0] intercept   : AIC=2706.678, Time=0.38 sec
 ARIMA(4,0,4)(0,0,0)[0] intercept   : AIC=2689.620, Time=0.81 sec
 ARIMA(3,0,3)(0,0,0)[0]             : AIC=inf, Time=0.57 sec

Best model:  ARIMA(3,0,3)(0,0,0)[0] intercept
```

Fig. 8. The intercept of the auto_arima model.

```
Mean Square Error(RMSE):                || 180.284
Mean Absolute Percentage Error(MAPE):|| 0.068
Mean Absolute Error(MAE):               || 152.844
```

Fig. 9. Auto ARIMA evaluation performance metrics (RMSE, MAPE, MEA).

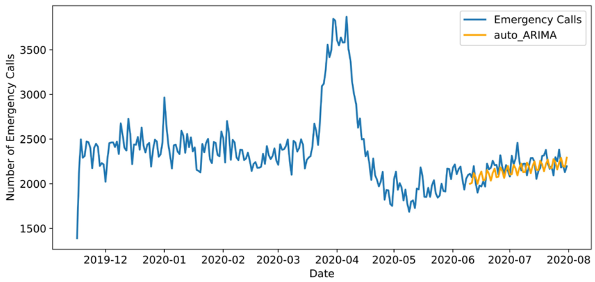

Fig. 10. Time series forecast visualization Auto ARIMA (3,0,3) model.

E. Grid Search

Grid search provides a robust method of target selection with the automated ARIMA hyperparameter, which systematically searches the model on a user-defined grid of parameter values (*p*, *d*, *q*) [50]. This guarantees that the entire parameter space will be searched in the chosen range. The analysis enables the identification of the best (*p*, *d*, *q*) combination that minimizes RMSE to increase efficiency; to strike a balance between thorough search and computational effort, we use Auto ARIMA parameter selection as a starting point. These initial choices provided basic performance and served as a valuable starting point for more sophisticated web searches. The Auto ARIMA initial choice of (*p*, *d*, *q*) = (3, 0, 3) is scientific, and the grid search focuses on narrower parameter values around the Auto ARIMA choice, especially p and *q* We searched no range from 1 to 5 for the two parameters, that Find promising nearby sites ensured that Auto ARIMA results were included. The discrimination criterion (*d*) was set to 0, in order to avoid analysis from Auto ARIMA and to avoid unnecessary data searching. The grid search identified the optimal hyperparameter combination (*p*, *d*, *q*) = (2, 0, 3) for the ARIMA model. This design reduced the number of analysis matrices; obtained an RMSE of 161.027, a MAPE of 0.064, and an MAE of 138.162, as shown in Fig. 11. The corresponding time series prediction diagram is shown in Fig 12. These results demonstrate the effectiveness of the network search method in selecting the appropriate hyperparameters for the ARIMA model.

F. Bayesian Optimization

BO helps to choose the best policy for the next attempt, making finding the best path more efficient [51]. BO is a method that uses Bayes' theorem to determine the search for a scalar objective function f(x) to be minimized for x in a bounded domain. Bayes theorem, widely used in data analysis; Time series forecast [52] and General estimation and prediction [53]. Bayesian statistics offer an alternative method for predicting single time series data. It uses a sampling to generate a probability distribution of future values, taking into account the level of uncertainty. This goes beyond detailed predictions and incorporates prior knowledge, creating a rich picture of possible future outcomes. This is valuable for businesses that need to understand different possibilities [54]. Theoretically, the Bayes theorem in Eq. (12) computes the posterior probability (P(A|B)), knowing that event Y has already happened and the probability that event A. This clarifies that it is not an absolute probability, but a conditional probability. The initial computation is based on knowing the prior probability that A (P(A)), the probability that B is true given A (P(B|A)), and the proof a there exists all for the B (P(B)). With these values, Bayes' theorem provides a mathematical way to predict the update probability of A after considering new information (event B) [55].

$$P\,(A|B) = \frac{P(B|A) \times P(A)}{P(B)}\,, P(B) > 0 \tag{12}$$

BO represents an improved method of fine-tuning the ARIMA model parameters in time series forecasting, which affects the ability of the model to better capture data patterns These parameters include autoregressive order (*p*), discriminant order (*d*), and moving average order (*q*).

```
Grid Search Best ARIMA parameters (p, d, q): (2, 0, 3)
======================================================
Mean Square Error(RMSE):                || 161.027
------------------------------------------------------
Mean Absolute Percentage Error(MAPE):|| 0.064
------------------------------------------------------
Mean Absolute Error(MAE):               || 138.162
======================================================
```

Fig. 11. Grid search evaluation performance metrics (RMSE, MAPE, MEA).

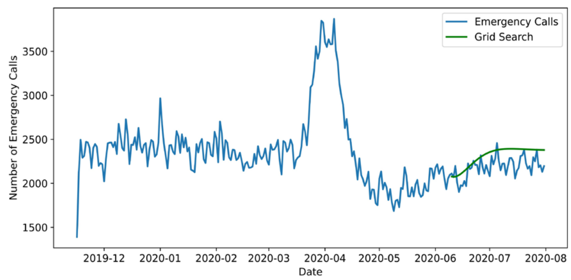

Fig. 12. Time series forecast visualization ARIMA (2,0,3) based on grid search selection.

Bayesian optimization aims to explore and systematically exploit the parameter space to minimize RMSE, a widely accepted metric for assessing prediction accuracy We define the parameter space for ARIMA (*p*, *d*, *q*) parameters. Each refinement by training the

ARIMA model on historical EMS data and calculating RMSE against the test set, BO adjusts its probability model iteratively based on observed RMSE values. This ARIMA guides BO in parameter space is analyzed, thereby reducing forecast error also ensuring convergence towards configurations that improve accuracy This iterative process assures robust parameter selection for improved ARIMA model performance in various forecasting applications. The main advantage of the Bayesian approach is its ability to incorporate prior knowledge of the parameters. This sets it apart from traditional methods based solely on data. This capability is particularly valuable for ARIMA models, where it can provide a more nuanced description of uncertainty in forecasts.

We define the Bayesian search space as the follows: the range of p and q (0–5) was chosen based on empirical evidence from previous EMS forecasting studies, where a small number of lags effectively captures temporal dependencies while avoiding overfitting, while the differencing term $d = 0$ was chosen after stationarity tests (ADF test results confirm that the time series is already stationary). The values used for ARIMA parameter selection in the EMS time series on the optimal structure with $(p, d, q) = (3, 0, 0)$ were obtained. This parameter choice resulted in the lowest RMSE of 107.776, MAPE of 0.041, and MAE of 87.055, as in Fig. 13. This indicated if the model correctly captured the underlying observations in the EMS data fewer errors, which may improve future forecasts appeal. The corresponding time-series prediction model based on the selected Bayesian parameters is shown in Fig. 14.

```
Bayesian Optimization parameters (p, d, q): (3,0,0)
=====================================================
Mean Square Error(RMSE):                    || 107.776
-----------------------------------------------------
Mean Absolute Percentage Error(MAPE):       || 0.041
-----------------------------------------------------
Mean Absolute Error(MAE):                   || 87.055
=====================================================
```

Fig. 13. Bayesian selection ARIMA (3, 0, 0) performance metrics (RMSE, MAPE, MEA).

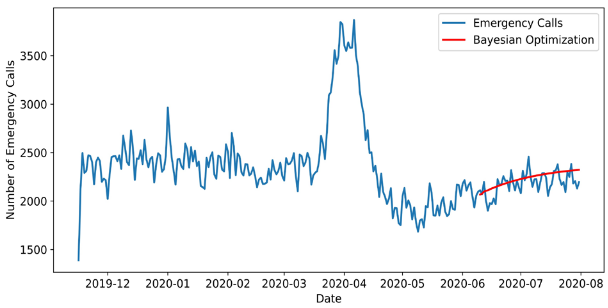

Fig. 14. ARIMA (3,0,0) based on Bayesian selection.

G. Implementation

The implementation of ARIMA and hyperparameter tuning methods was conducted using Python 3.9. The following libraries were used:

- Auto-ARIMA: pmdarima v1.8.5.
- Grid search: scikit-learn v1.0.2.
- Bayesian optimization: hyperopt v0.2.7.
- Time series analysis and visualization: statsmodels v0.13.2, matplotlib v3.5.1.

All experiments were conducted on a system with an Intel Core i9-11800H CPU (2.3 GHz, 8 cores), 16 GB RAM, and executed using Google Colab. Training the Bayesian optimization model required approximately 1.5 min, while grid search took 3.7 min on average.

IV. Discussion

In this study, we use a combination of Auto ARIMA, grid search, and Bayesian optimization to increase the prediction accuracy of EMS time series data. Each optimization method has its own trade-offs. Auto-ARIMA automates hyperparameter selection but often converges to suboptimal solutions due to its reliance on Akaike information criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC) are both penalized-likelihood information, i.e., AIC/BIC criteria. Grid search, while exhaustive, is computationally expensive, requiring multiple iterations across all parameter combinations. In contrast, Bayesian optimization balances exploration and exploitation, dynamically adjusting search based on prior evaluations, leading to a 40% reduction in RMSE compared to grid search [56]. This efficiency makes Bayesian optimization the preferred approach for time-sensitive EMS forecasting applications. The Auto ARIMA provided an initial RMSE of 180.284, indicating that it can automatically select the ARIMA parameters without manual intervention. However, seeing the possibility of further improvement, we employed a grid search to systematically explore the parameter space (p, d, q) and found an optimized algorithm by Bayesian optimization which had further improvement with RMSE decreasing to 161.027 decreasing to 107.776. Its repetitive study successfully demonstrated optimal prediction patterns. Table V presents the performance metrics of Auto-ARIMA, grid search, and Bayesian optimization. Bayesian optimization demonstrated the largest accuracy improvement, reducing RMSE by 40% compared to Auto-ARIMA and 33% compared to grid search. MAPE improved from 6.8% (Auto-ARIMA) to 4.1% (Bayesian optimization), reflecting a 39.7% increase in predictive accuracy. Similarly, MAE decreased by 43%, indicating a significant reduction in absolute forecasting errors.

TABLE V. Performance Metrecis Comparison of Parameter Selection for Each Auto Arima, Grid Search, and Bayesian Optimization

Improve % / Model	RMSE	RMSE Improvement	MAPE	MAPE Improvement	MAE	MAE Improvement
Auto ARIMA **(3, 0, 3)**	180.2	-	0.068	-	152.8	-
Grid search (2, 0, 3)	161.0	+10.7%	0.064	+5.8%	138	+9.6%
Bayesian opt. (3, 0, 0)	107.7	+40.0%	0.041	+39.7%	87.0	+43.1%

Fig. 15 presents graphical plots of the EMS time series predictions with hyperparameters selected by Auto ARIMA (3, 0, 3), grid search (3, 0, 2), and Bayesian optimization (3, 0, 0).

These findings highlight the effectiveness of Bayesian optimization as a powerful tool for enhancing the accuracy of time-series forecasting in EMS demand estimation by optimally selecting ARIMA parameters. The improved forecasting accuracy significantly enhances the reliability of EMS call volume predictions, enabling dispatch centers to proactively allocate ambulances based on anticipated demand. This leads to faster response times and, ultimately, improved patient survival rates. Accurate demand forecasting thus serves as a critical tool for optimizing EMS operations and enhancing public safety.

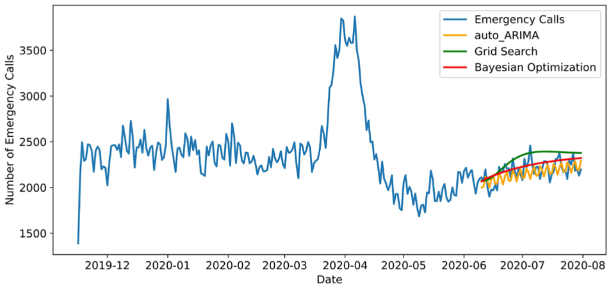

Fig. 15. EMS time series forecasting visualization using hyperparameters selected by Auto ARIMA (3, 0, 3), grid search (3, 0, 2), and bayesian optimization (3, 0, 0).

V. Conclusion

In this paper, we explored the use of the ARIMA model for time series forecasting by using techniques to improve predictive accuracy through hyperparameter tuning. The objective is to create a reliable predictive framework to support emergency service management in optimizing resource allocation, improving response times, and ultimately, enhancing public safety. To improve the predictive accuracy of the ARIMA model, we used both grid search and Bayesian optimization to find the best combination of hyperparameters. Grid search searched through parameter combinations systematically but was quite costly in terms of computation. On the other hand, Bayesian optimization gave a very efficient alternative by exploiting the knowledge it obtained to perform an iterative search through the parameter space. A comparison of the two methods showed that Bayesian optimization achieved better results with much-reduced computational resources. Performance metrics, including MAE, RMSE MAPE. Compared to Auto-ARIMA and grid search, Bayesian optimization reduces RMSE by 40%, improving accuracy while maintaining computational efficiency. These findings suggest that Bayesian optimization is a highly effective tool for optimizing EMS resource allocation, reducing response times, and improving patient outcomes. This could then be extrapolated to other mission-critical service sectors, such as health or public transportation where demand forecasting has a critical influence on operational efficiency.

This study utilized a publicly available EMS dataset to ensure a well-structured and comprehensive analysis. While the findings are based on this specific dataset, the methodology, including Bayesian optimization and ARIMA modeling, can be applied to other EMS datasets with similar characteristics. Future studies could test the model on datasets from different regions to further evaluate its adaptability and robustness.

Future research will examine multivariate time series forecasting methods. This approach includes other factors that can affect demand, which can lead to more detailed forecasts. Additionally, hybrid models combining Bayesian optimization with deep learning architectures (e.g., Long Short-Term Memory Network (LSTM)) could further improve EMS forecasting capabilities.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

MHF conducted the research and designed the study; HGH analyzed the data and performed statistical modeling; MES contributed to data curation and preprocessing; MHI wrote the initial draft of the manuscript; all authors participated in reviewing and editing the manuscript, provided critical feedback; all authors had approved the final version.

Acknowledgment

I would like to acknowledge MEAPAL Data Science @CORELLA (https://www.corelia.ai/) for their wonderful collaboration and support, and for all of the opportunities given to further research.

References

[1] S. A. Simonsen, M. Andresen, L. Michelsen *et al.*, "Evaluation of pre-hospital transport time of stroke patients to thrombolytic treatment," *Scand J Trauma Resusc Emerg Med.*, vol. 22, no. 1, pp. 1–5, 2014. doi: 10.1186/s13049-014-0065-z

[2] Ľ. Jánošíková, M. Kvet, P. Jankovič *et al.*, "An optimization and simulation approach to emergency stations relocation," *Cent. Eur. J. Oper. Res.*, vol. 27, no. 3, pp. 737–758, 2019. doi: 10.1007/s10100-019-00612-5

[3] V. Bélanger, A. Ruiz, and P. Soriano, "Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles," *Eur. J. Oper. Res.*, vol. 272, no. 1, pp. 1–23, 2019. https://doi.org/10.1016/j.ejor.2018.02.055

[4] Y. Tsai, K.-W. Chang, G.-T. Yiang *et al.*, "Demand forecast and multi-objective ambulance allocation," *Intern. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 07, 1859011, 2018.

[5] M. Swalehe and S. G. Aktas, "Dynamic ambulance deployment to reduce ambulance response times using geographic information systems: A case study of Odunpazari district of Eskisehir province, Turkey," *Procedia Environ Sci*, vol. 36, pp. 199–206, 2016.

[6] D. Tandberg, J. Tibbetts, and D. P. Sklar, "Time series forecasts of ambulance run volume," *Am J Emerg Med.*, vol. 16, no. 3, pp. 232–237, 1998. https://doi.org/10.1016/S0735-6757(98)90090-0

[7] H. Setzler, C. Saydam, and S. Park, "EMS call volume predictions: A comparative study," *Comput Oper Res*, vol. 36, no. 6, pp. 1843–1851, 2009. doi: https://doi.org/10.1016/j.cor.2008.05.010

[8] J. L. Torres, A. García, M. De Blas *et al.*, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)," *Solar Energy*, vol. 79, no. 1, pp. 65–77, 2005. https://doi.org/10.1016/j.solener.2004.09.013

[9] T. Leung and T. Zhao, "Financial time series analysis and forecasting with Hilbert–Huang transform feature generation and machine learning," *Appl Stoch Models Bus Ind*, vol. 37, no. 6, pp. 993–1016, 2021.

[10] S. W. Astuti, "Forecasting Surabaya–Jakarta train passengers with SARIMA model," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2018, 012105. https://doi.org/10.1088/1757-899x/407/ 1/012105

[11] S. D. Mohamed A. K. Al-Azzani and T. J. England, "An empirical investigation of forecasting methods for ambulance calls—A case study," *Health Systems*, vol. 10, no. 4, pp. 268–285, 2021.

[12] A. X. Lin, A. F. W. Ho, K. H. Cheong *et al*., "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, pp. 1–15, 2020.

[13] A. Blanchard and T. Sapsis, "Bayesian optimization with output-weighted optimal sampling," *J. Comput. Phys*., vol. 425, 109901, 2021. https://doi.org/10.1016/j.jcp.2020.109901

[14] A. H. Victoria and G. Maragatham, "Automatic tuning of hyperparameters using Bayesian optimization," *Evolving Systems*, vol. 12, no. 1, pp. 217–223, 2021.

[15] R. Aringhieri, G. Carello, and D. Morale, "Supporting decision making to improve the performance of an Italian emergency medical service," *Ann. Oper. Res.*, vol. 236, pp. 131–148, 2016.

[16] C. L. Watkins, S. P. Jones, M. A. Hurley *et al*., "Predictors of recognition of out of hospital cardiac arrest by emergency medical services call handlers in England: A mixed methods diagnostic accuracy study," *Scand. J. Trauma. Resusc. Emerg. Med.*, vol. 29, pp. 1–9, 2021.

[17] R. J. Martin, R. Mousavi, and C. Saydam, "Predicting emergency medical service call demand: A modern spatiotemporal machine learning approach," *Oper. Res. Health Care.*, vol. 28, 100285, 2021.

[18] M. Villani, A. Earnest, N. Nanayakkara *et al*., "Time series modelling to forecast prehospital EMS demand for diabetic emergencies," *BMC Health Serv. Res.*, vol. 17, pp. 1–9, 2017.

[19] B. Rostami-Tabar and R. J. Hyndman, "Hierarchical time series forecasting in emergency medical services," *J. Serv. Res*., 2024.

[20] P. Abreu, D. Santos, and A. Barbosa-Povoa, "Data-driven forecasting for operational planning of emergency medical services," *Socioecon. Plann. Sci.*, vol. 86, 101492, 2023.

[21] W. R. Bell, "An introduction to forecasting with time series models," *Insur. Math. Econ*., vol. 3, no. 4, pp. 241–255, 1984.

[22] Z. Hajirahimi and M. Khashei, "Hybridization of hybrid structures for time series forecasting: A review," *Artif. Intell. Rev*., vol. 56, no. 2, pp. 1201–1261, 2023.

[23] K. Zarzycki and M. Ławryńczuk, "Advanced predictive control for GRU and LSTM networks," *Inf. Sci. (N. Y.)*., vol. 616, pp. 229–254, 2022.

[24] S. Garg, S. Kumar, and P. K. Muhuri, "A novel approach for COVID-19 infection forecasting based on multi-source deep transfer learning," *Comput. Biol. Med.*, vol. 149, 105915, 2022.

[25] Q. Z. Ullah, S. Hassan, and G. M. Khan, "Adaptive resource utilization prediction system for infrastructure as a service cloud," *Comput. Intell. Neurosci.*, vol. 2017, no. 1, 4873459, 2017.

[26] M. A. K. Al-Azzani, S. Davari, and T. J. England, "An empirical investigation of forecasting methods for ambulance calls—A case study," *Health Systems*, vol. 10, no. 4, pp. 268–285, 2021.

[27] Z. B. Asghar, P. Wankhade, F. Bell *et al*., "Trends, variations and prediction of staff sickness absence rates among NHS ambulance services in England: A time series study," *BMJ Open*, vol. 11, 2021.

[28] C.-S. Ong, J.-J. Huang, and G.-H. Tzeng, "Model identification of ARIMA family using genetic algorithms," *Appl. Math. Comput.*, vol. 164, no. 3, pp. 885–912, 2005.

[29] B. C. Ervural, O. F. Beyca, and S. Zaim, "Model estimation of ARMA using genetic algorithms: A case study of forecasting natural gas consumption," *Procedia Soc. Behav. Sci.*, vol. 235, pp. 537–545, 2016. https://doi.org/10.1016/j.sbspro.2016.11.066

[30] Y. Zhang, L. Luo, J. Yang *et al*., "A hybrid ARIMA-SVR approach for forecasting emergency patient flow," *J. Ambient. Intell. Humaniz. Comput*., vol. 10, pp. 3315–3323, 2019.

[31] S. Nichols and S. Abolmaali, "Enhancing COVID-19 case forecasting in the United States: A comparative analysis of ARIMA, SARIMA, and RNN models with grid search optimization," *medRxiv*, pp. 2023–2024, 2024.

[32] L. Du, R. Gao, P. N. Suganthan *et al*., "Bayesian optimization based dynamic ensemble for time series forecasting," *Inf. Sci. (N. Y.)*., vol. 591, pp. 155–175, 2022.

[33] N. Sultana, S. M. Z. Hossain, S. H. Almuhaini *et al*., "Bayesian optimization algorithm-based statistical and machine learning approaches for forecasting short-term electricity demand," *Energies*, vol. 15, no. 9, 3425, 2022.

[34] P. Yang, P. Cheng, N. Zhang *et al*., "Statistical machine learning models for prediction of China's maritime emergency patients in dynamic: ARIMA model, SARIMA model, and dynamic Bayesian network model," *Front Public Health*, vol. 12, 1401161, 2024.

[35] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt *et al*., "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *J. Econom.*, vol. 54, no. 1–3, pp. 159–178, 1992.

[36] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods. SPRINGER Science & Business Media*, 1991.

[37] N. W. Gillham, *A Life of Sir Francis Galton: From African Exploration to the Birth of Eugenics*, Oxford University Press, 2001.

[38] G. E. P. Box, G. M. Jenkins, G. C. Reinsel *et al*., *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, 2015.

[39] C. and W. G. I. Sammut, "Mean absolute error," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2010. doi: 10.1007/978-0-387-30164-8_525

[40] A. G. Barnston, "Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score," *Weather Forecast*, vol. 7, no. 4, pp. 699–709, 1992. https://doi.org/10.1175/1520-0434(1992)007

[41] C. D. Coleman and D. A. Swanson, "On MAPE-R as a measure of cross-sectional estimation and forecast accuracy," *J. Econ. Soc. Meas.*, vol. 32, no. 4, pp. 219–233, 2007. doi: 10.3233/JEM-2007-0290

[42] S. I. Ahmed, P. Ranganathan, and H. Salehfar, "Forecasting of mid- and long-term wind power using machine learning and regression models," in *Proc. 2021 IEEE Kansas Power and Energy Conference (KPEC)*, USA: IEEE, 2021, pp. 1–6.

[43] G. E. P. Box, G. M. Jenkins, G. C. Reinsel *et al*., *Time Series Analysis: Forecasting and Control*, New Jersey: John Willey and Sons, 2016.

[44] V. V. Vandaele, "Applied time series and box-jenkins models academic press," *Nevv. York*, 1983.

[45] A. Earnest, M. I. Chen, D. Ng *et al*., "Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore," *BMC Health Serv Res.*, vol. 5, no. 1, pp. 1–8, 2005. doi: 10.1186/1472-6963-5-36

[46] F. Dama and C. Sinoquet, "Time series analysis and modeling to forecast: A survey," arXiv preprint, arXiv:2104.00164, 2021.

[47] T. T. H. Phan and X. H. Nguyen, "Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red river," *Adv. Water Resour*., vol. 142, 103656, 2020. doi: 10.1016/J.ADVWATRES.2020.103656

[48] M. Farsi, D. Hosahalli, B. R. Manjunatha *et al*., "Parallel genetic algorithms for optimizing the SARIMA model for better forecasting of the NCDC weather data," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 1299–1316, 2021. doi: 10.1016/J.AEJ.2020.10.052

[49] H. Akaike, "Likelihood of a model and information criteria," *J. Econom*., vol. 16, no. 1, pp. 3–14, 1981.

[50] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, 2018.

[51] M. I. Sameen, B. Pradhan, and S. Lee, "Self-learning random forests model for mapping groundwater yield in data-scarce areas," *Natural Resources Research*, vol. 28, pp. 757–775, 2019.

[52] Z. Zeng and M. Li, "Bayesian median autoregression for robust time series forecasting," *Int J Forecast*., vol. 37, no. 2, pp. 1000–1010, 2021. https://doi.org/10.1016/j.ijforecast.2020.11.002

[53] A. Rodriguez and G. Puggioni, "Mixed frequency models: Bayesian approaches to estimation and prediction," *Int. J. Forecast*., vol. 26, no. 2, pp. 293–311, 2010.

[54] C. Berninger, A. Stöcker, and D. Rügamer, "A Bayesian time-varying autoregressive model for improved short-term and long-term prediction," *J. Forecast*., vol. 41, no. 1, pp. 181–200, 2022.

[55] S. Edition, *Bayesian Data Analysis*, CRC Press, 2013.

[56] N. Sultana, S. M. Z. Hossain, S. H. Almuhaini *et al*., “Bayesian optimization algorithm-based statistical and machine learning approaches for forecasting short-term electricity demand,” *Energies (Basel)*, vol. 15, no. 9, 3425, 2022.